INTERACTION OF THE HIV-1 NUCLEOCAPSID PROTEIN WITH 2-MERCAPTOBENZAMIDE THIOESTERS AND ΨRNA

by

LISA M. MILLER JENKINS

(Under the Direction of Pascale Legault)

ABSTRACT

The HIV-1 nucleocapsid protein (NCp7) is a protein that plays several crucial roles throughout the retroviral lifecycle. As a nucleic acid-binding protein, NCp7 is essential for reverse transcription, viral integration, and viral assembly. The most important function of NCp7, though, is the dimerization and packaging of the genomic RNA into a new virion. Dimerization and packaging both involve the interaction of NCp7 with the stem-loop sequences in the Ψ site, located at the 5′ region of the genomic RNA. NCp7 contains two highly-conserved structural zinc-binding domains, in which three cysteine and one histidine residues are used to coordinate zinc. These zinc-binding domains are required for sequence-specific RNA-binding by NCp7. The highly conserved nature of NCp7 and its importance at various stages of the HIV-1 lifecycle have made it a target for the development of new antiviral therapies. Compounds have been developed that function by ejecting the coordinated zinc from NCp7, such as the 2-mercaptobenzamide thioester compounds. UV/Visible spectroscopy, NMR spectroscopy, mass spectrometry, and gel mobility shift assays have been used to investigate the mechanism of action of the 2-mercaptobenzamide thioesters. We have determined that these thioester compounds specifically target the Cys_{39} in the carboxyl-terminal zinc-binding domain of NCp7,

leading to the ejection of zinc and the loss of nucleic acid binding. When NCp7 is already bound to RNA, however, the 2-mercaptobenzamide thioester compounds are unable to interact with NCp7. The interaction of these compounds with several cellular zinc-binding domains has also been studied. It was found that the 2-mercaptobenzamide thioesters were able to interact with other $Cys_2HisCys$ zinc-binding domains, as well as with proteins containing Cys_4 zinc-binding domains. Proteins containing classical Cys_2His_2 and interleaved RING finger-like zinc-binding domains did not interact with the thioester compounds. In addition, NMR spectroscopy and gel mobility shift assay were used to investigate the interaction of NCp7 with the internal loop of the first stem-loop from the Ψ site. Binding of NCp7 to this RNA destabilized two base pairs around the internal loop. The internal loop could function as a nucleation point for the destabilization of the stem during dimerization. NCp7 could facilitate this destabilization as part of its dimerization function.

INDEX WORDS: Human immunodeficiency virus, HIV-1, nucleocapsid, NCp7, 2-mercaptobenzamide thioesters, RNA dimerization

INTERACTION OF THE HIV-1 NUCLEOCAPSID PROTEIN WITH 2-MERCAPTOBENZAMIDE THIOESTERS AND ΨRNA

by

LISA M. MILLER JENKINS

B.S., Winthrop University, 2000

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2005

INTERACTION OF THE HIV-1 NUCLEOCAPSID PROTEIN WITH 2-MERCAPTOBENZAMIDE THIOESTERS AND ΨRNA

by

LISA M. MILLER JENKINS

Major Professor: Pascale Legault

Committee: James Prestegard
Walter Schmidt
Robert Scott
E. Will Taylor

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2005

DEDICATION

Thanks Mom for first introducing me to science. It has become my passion.

ACKNOWLEDGEMENTS

I want to thank my family for always supporting me: my parents for their love, words of encouragement, and never-ending support … my brothers Kevin and Brian for reminding me to laugh and not take myself too seriously … Matt, my husband, I can never thank enough for standing by me through this adventure, for making me laugh when all I wanted to do was cry, for listening to my talks over and over and reading my papers, even when I know you did not understand.

I would like to thank several of my professors from Winthrop, Dr. Julian Smith III, Dr. Pat Owens, and Dr. Patricia Bossart-Whitaker. These professors encouraged me early-on to do research. They were among the first to open the door and pointed me to the path I chose to take. Without their guidance, I would never have chosen to pursue this degree.

I also have to thank my professors, Dr. James Omichinski and Dr. Pascale Legault. They have taught me so much in the years that I have worked under them and molded me into the scientist I am today. They let me explore new ideas and re-work old ones. I have been supported in their labs and shown that I can do more than I ever thought possible. Past and present members of the Legault/Omichinski lab have been invaluable along the way. Dr. Bao Nguyen, Dr. Paola DiLello, and Dr. Dean Campbell have all helped me with NMR experiments and data analysis. They have been patient with me and never laughed at my questions. Dr. Karen Abbott taught me so much in the lab and was a great source of inspiration and helpful discussion.

There are so many friends that have made this experience wonderful … thank you Jessi O'Brien for being my sounding-block and my shoulder … thank you Renu Kadirvelraj and Anita Kishore for listening and keeping me grounded in reality … thank you Wided Missoui and Teresa Ross for always smiling.

Thank you all.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

Human Immunodeficiency Virus Type-1

Acquired immunodeficiency syndrome (AIDS) is a devastating illness that impacts people worldwide. It is0 estimated that 40 million people are currently living with human immunodeficiency virus type-1 (HIV-1), the virus that causes AIDS. Each day, another 14,000 people are infected[1]. The virus itself is quite diverse: there are currently 24 genetic variants of the main group of HIV-1, 11 of which are subtypes and 13 of which are recombinant forms of those subtypes[2]. The viral genome is rapidly evolving, and current drug therapies are failing. The only way to better treat and possibly cure HIV-1 infection is to understand as much as possible about the virus itself.

HIV-1 is a member of the lentivirus family of retroviruses. It consists of a 9.5 kilobase RNA genome that codes for fifteen proteins through nine open reading frames (Figure 1.1). Three of the proteins are enzymes, the reverse transcriptase (RT), the protease (PR), and the integrase (IN) proteins, that are all encoded as part of the pro-pol polyprotein. The three structural proteins, matrix (MA), capsid (CA), and nucleocapsid (NCp7) are encoded by the gag polyprotein. The pro-pol polyprotein is always transcribed as a gag-pro-pol fusion protein. The e*nv* open reading frame codes for the viral recognition proteins, which are the surface (gp41) and transmembrane (gp120) proteins. The remaining six proteins (Tat, Vpr, Vif, Vpu, Rev, and Nef) play accessory roles in the viral lifecycle.

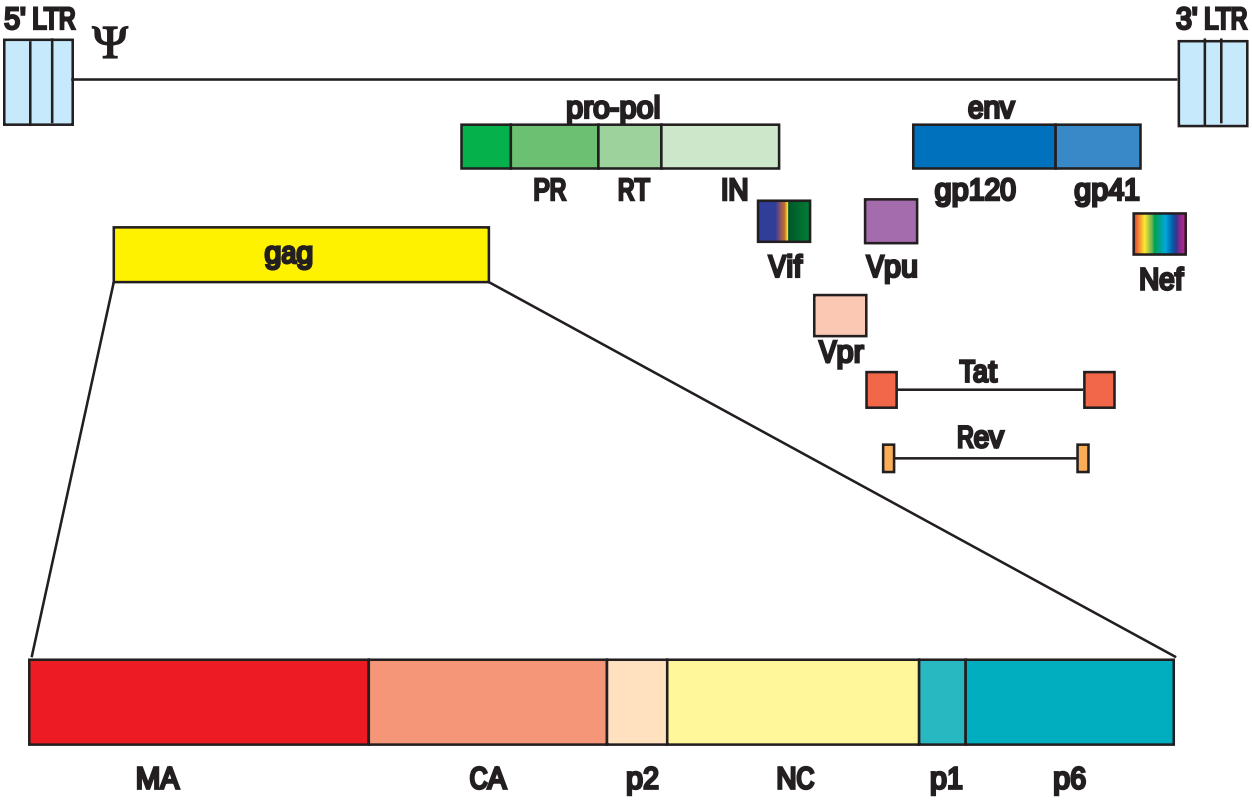


Figure 1.1: A schematic depiction of the genomic structure of HIV-1. The open reading frames of all fifteen proteins of HIV-1 are noted (MA, CA, NCp7, p6, PR, RT, IN, gp120, gp41, Vpr, Vif, Vpu, Tat, Rev, and Nef), including the cleavage products of polyproteins. In addition, the location of the Ψ-site is marked at the 5' end of the RNA.

The mature, infectious virion of most retroviruses is thought to be made up of layers of *gag*-derived proteins encasing the RNA. In the HIV-1 virion, MA is associated with the outer membrane of the virion, derived from the host cell, through amino-terminal myristylation (Figure 1.2). The viral proteins gp41 and gp120 are displayed on the outside of the virion for host cell recognition. The core of the virion is composed of ~2000 copies of CA that encase two copies of genomic RNA coated with ~2000 molecules of NCp7[3]. Several copies of RT and IN are also associated with the genomic RNA in the inner core of the virion. In HIV-1, the core is conical, but the shape varies in other retroviruses[4]. In general, the virion is constructed such that the genomic RNA is protected by several layers of structural proteins.

Cellular infection with HIV-1 begins when a mature virion attaches to a host cell through the interaction of gp41 and gp120 with cellular receptors such as CD4 (Figure 1.3). The virion membrane fuses with the cell membrane, and the core of the virion enters the cell. The RNA genome is uncoated and reverse transcribed into DNA by viral RT. Upon entering the nucleus, the DNA is integrated randomly into the cellular chromosomal DNA. Transcription and translation result in a cytoplasmic build-up of the uncleaved polyproteins and viral RNA, both spliced and unspliced. The copies of the env polyprotein are shuttled to the plasma membrane through the endoplasmic reticulum and stacks of the Golgi apparatus, and then they are exposed to the extracellular environment as part of the virion budding mechanism. Lateral movement along the membrane takes the env polyproteins to the site of budding. By contrast, the copies of the gag and gag-pro-pol polyproteins move through the cytosol to the membrane. Two copies of unspliced RNA are selectively dimerized and packaged into an immature virion that buds off of the cell. Following this release, PR cleaves the *gag* polyprotein into its constituate proteins, and

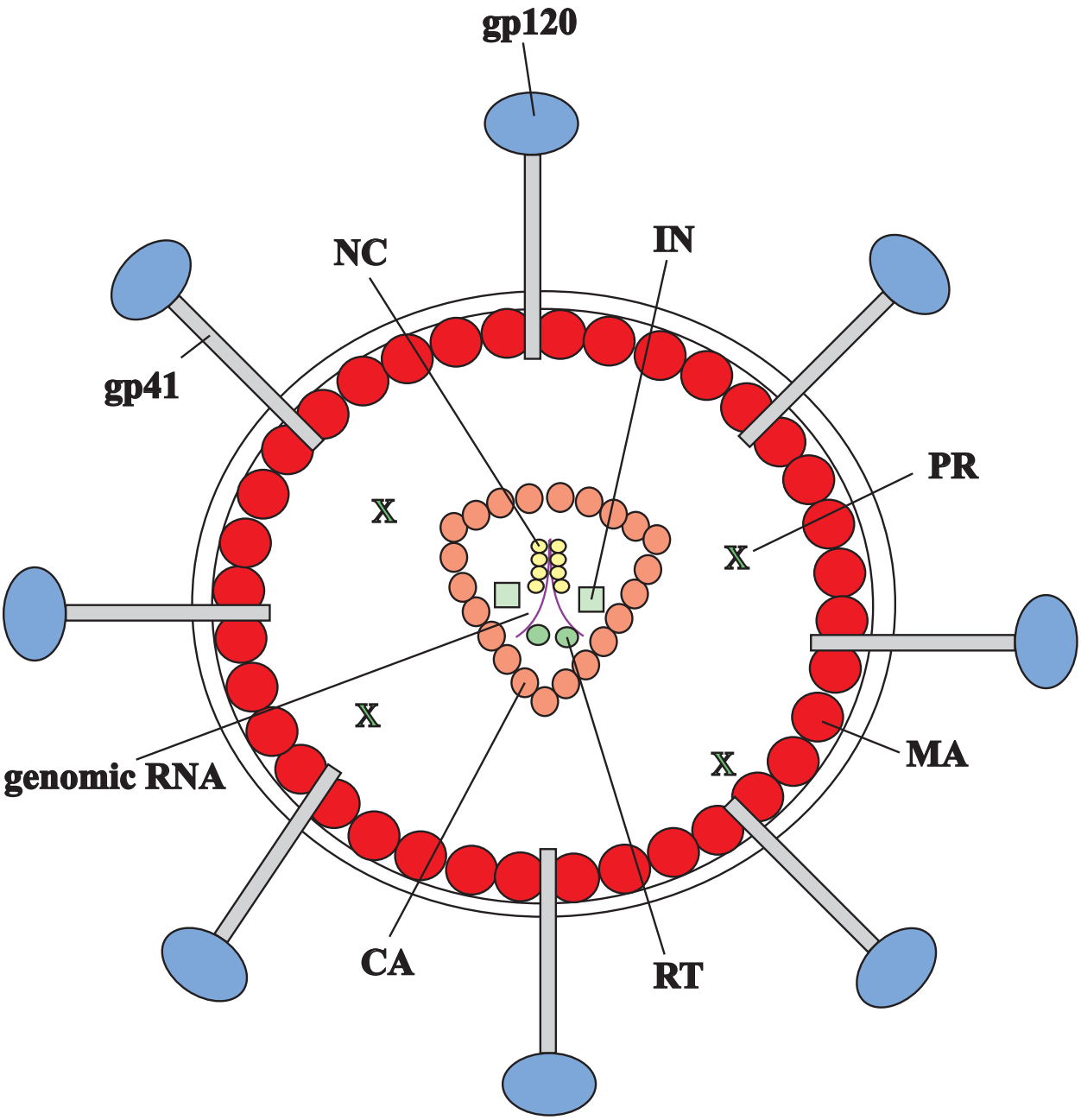


Figure 1.2: Diagram of a typical mature virion of HIV-1. The viral proteins that make up the virion are labeled.

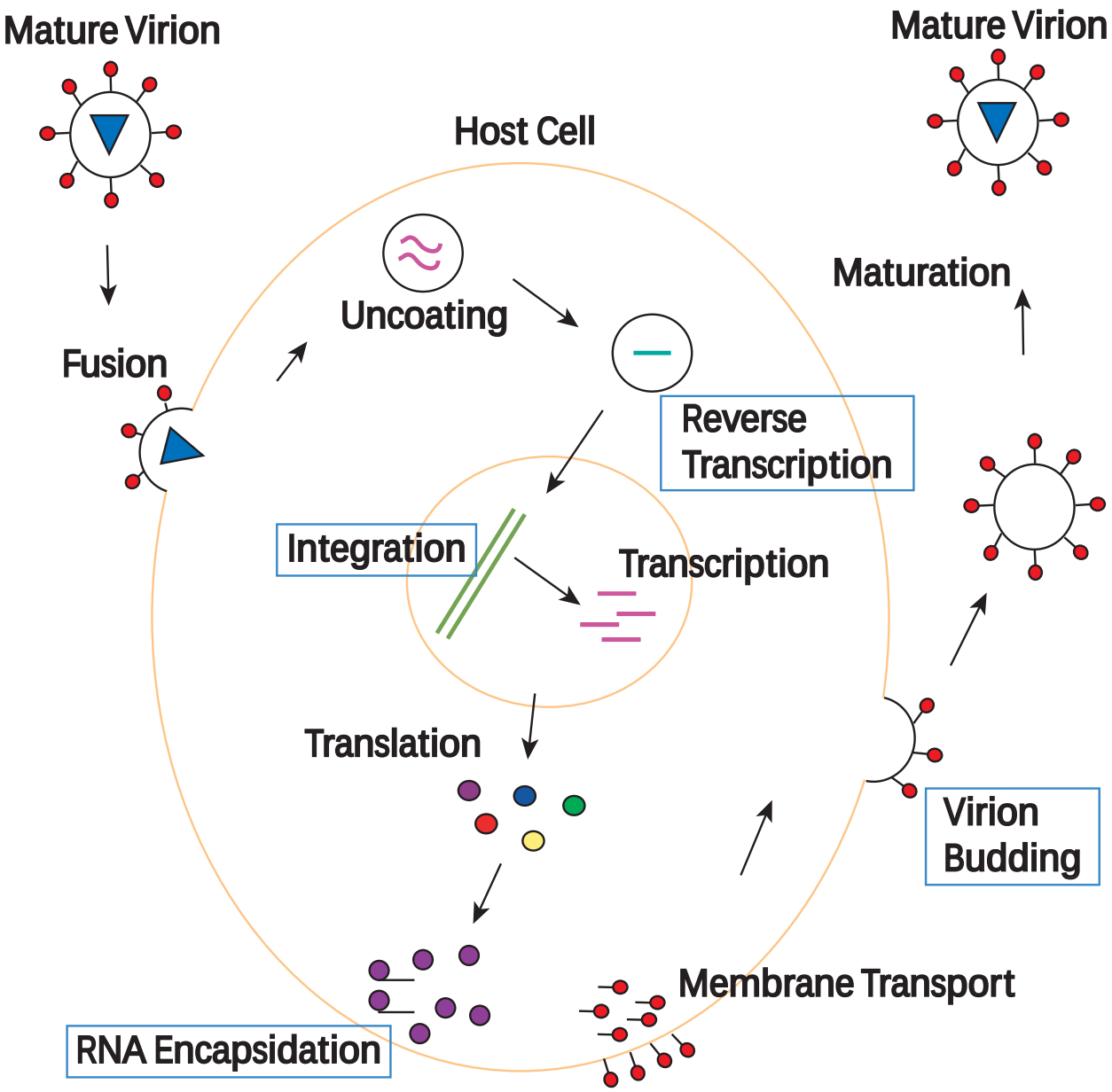


Figure 1.3: The HIV-1 lifecycle is illustrated, beginning with the fusion of a mature virion with a new cell and ending with the production of a new infectious virion. Stages in which NCp7 plays a major role are boxed in blue.

the virion undergoes a maturation event that is relatively poorly understood[4-6]. The reverse transcription, integration, virion assembly, and genome dimerization stages are described in more detail below.

The Nucleocapsid Protein (NCp7)

Sequence and Structure of NCp7

The nucleocapsid protein of all retroviruses, except for the spumaviruses, contains at least one zinc-binding domain in which one histidine and three cysteine residues chelate one atom of zinc. HIV-1 NCp7 contains two such motifs connected by a seven amino acid linker rich in basic residues (Figure 1.4A). There are several residues that are highly conserved in the sequence of zinc-binding domains of nucleocapsid proteins in addition to the chelating residues (Figure 1.4B). For example, an aromatic residue is commonly observed at position *n*+1 (where the first cysteine is position *n*), a basic or polar residue at position *n*+2, a glycine at positions *n*+4 and *n*+7, a lysine at position n+5, a basic residue at position *n*+11, and an acidic or polar residue at position *n*+12. The most distantly related retroviruses, the gammaretroviruses, show the least sequence homology.

NCp7 has been shown to bind zinc ions quite tightly, with K_D from 10^{-10} to 10^{-12} M [7]. The amino-terminal zinc-binding domain (ZD1) has been shown to coordinate zinc with an affinity three times greater than the carboxyl-terminal zinc-binding domain (ZD2)[8]. Studies of the protonation states of the zinc-coordinating cysteines of ZD2 demonstrated that the three cysteines have distinct side chain pKa values in the order: $Cys_{36} < Cys_{39} < Cys_{49}$[9]. This suggests that zinc coordination could occur through an ordered, stepwise mechanism beginning with

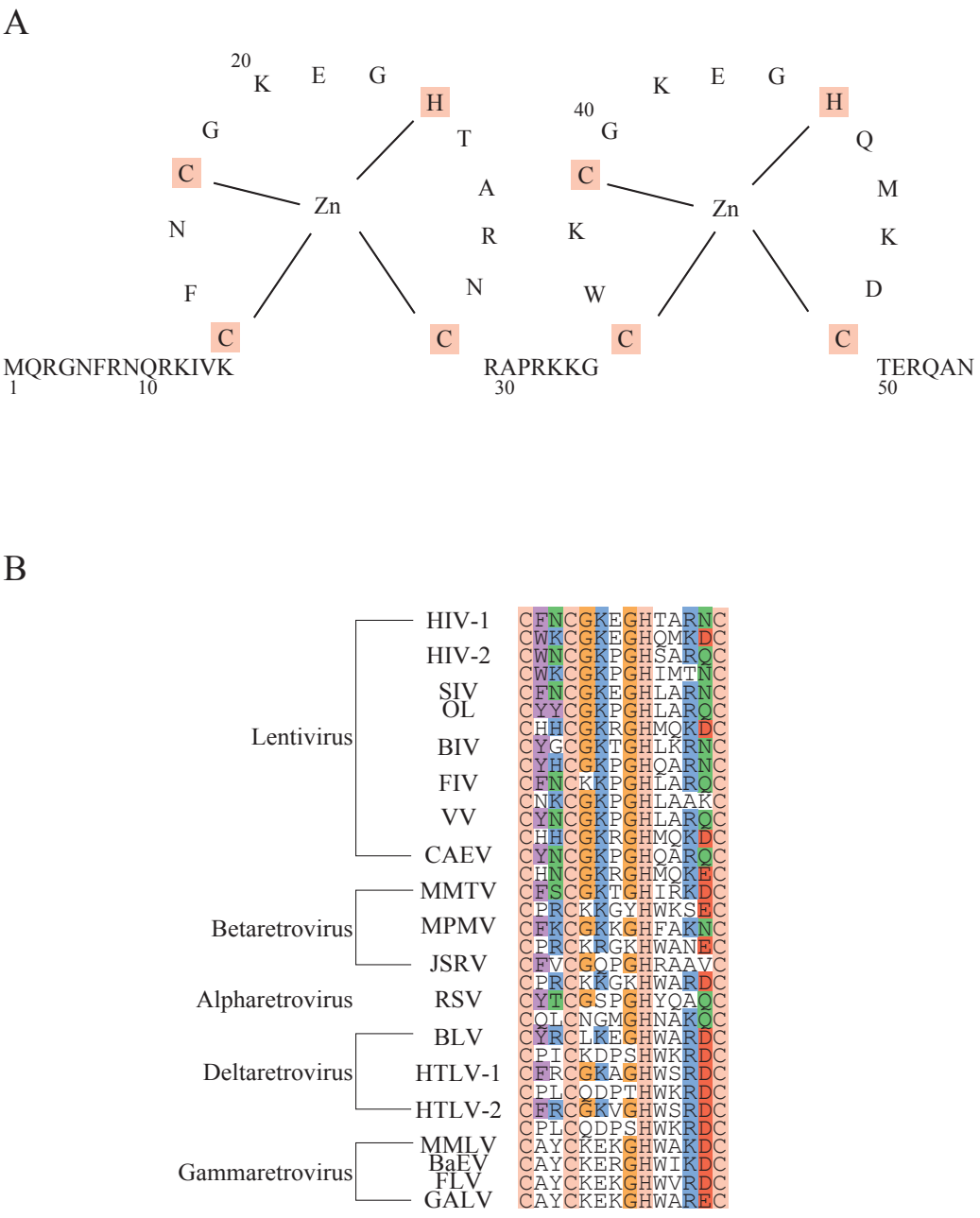


Figure 1.4: A) Amino acid sequence of NCp7 used in this study, showing the zinc- coordinating residues highlighted in orange. B) Sequence alignment of the zinc-binding domains of several retroviral nucleocapsid proteins. Aromatic residues are shown in purple, basic residues in blue, acidic residues in red, polar uncharged residues in green, and glycine residues in orange.

Cys_{36}, the first cysteine of the ZD2 sequence. A similar study has not been performed for the zinc-coordinating residues of ZD1.

NMR solution studies of NCp7 show that in the absence of zinc, the protein forms a random coil devoid of structure[10]. The metal-coordinated protein forms a somewhat ordered structure in which the two zinc-binding domains form small globular domains joined by a flexible linker (Figure 1.5)[10]. The amino and carboxyl termini of the protein remain disordered[11,12]. The conformation of each zinc-binding domain is very similar to the other, consisting of NH-S tight turns, comparable to the iron-binding domain observed in rubredoxin[13,14]. The two domains transiently interact with one another, as shown by the presence of weak NOEs[8,11,15,16].

Infectivity studies have shown that the two zinc-coordinating domains are not fully interchangeable; when the position of the zinc-binding domains are swapped, the virus is only able to infect cells at a very low frequency[17]. In addition, the exact nature of the zinc-coordinating motif is important for the function of the protein. When His_{23} was mutated to cysteine, creating a Cys_4 zinc-coordinating motif, the virus became noninfectious[18]. Structural studies revealed that this mutation induced a distortion of the amide backbone between residues 23 and 28 that led to formation of new hydrogen-bonding interactions and altered the spatial orientation of several residues in the domain[15,19]. Furthermore, mutation of Cys_{28} to histidine, creating a Cys_2His_2 zinc finger, also resulted in loss of infectivity, possibly due to structural rearrangement of the core of ZD1[20]. The functional consequences of mutations to the zinc-coordinating residues of NCp7 will be discussed below.

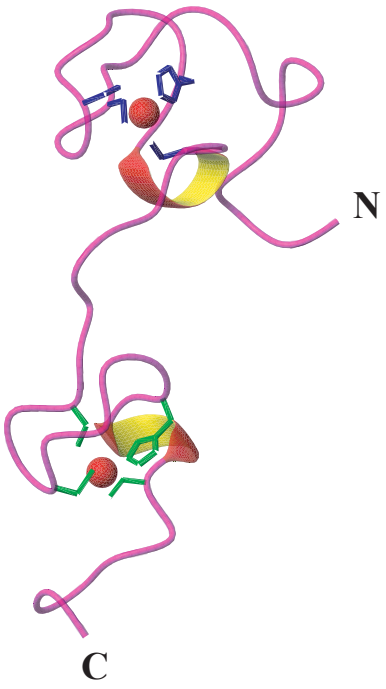


Figure 1.5: Ribbon diagram of the NMR solution structure of NCp7 showing the cysteine and histidine side chains chelating the zinc cations[16]. The zinc cations are shown as red spheres, and the side chains of the zinc-coordinating residues from ZD1 and ZD2 are shown in blue and green, respectively. This figure is adapted from PDB 1MFS using MOLMOL[21].

Functions of NCp7 Early in the Formation of a New Virion

NCp7 plays multiple roles as a nucleic acid-binding protein in the HIV-1 lifecycle. The various functions of NCp7 are found throughout the virion lifecycle: reverse transcription; integration; dimerization and packaging of genomic RNA, the most important function; and viral assembly. Early in the HIV-1 lifecycle, NCp7 is crucial for the many reactions that are part of reverse transcription. In this stage alone, NCp7 has four different roles that facilitate the generation of double-stranded DNA from the viral genomic RNA.

Reverse transcription is a complex process initiated by the annealing of cellular $tRNA_3^{Lys}$ primer to the primer binding site (PBS) of the genomic RNA (Figure 1.6). DNA synthesis begins from the $tRNA_3^{Lys}$ and continues to the 5'-end of the RNA strand, resulting in production of the minus-strand strong-stop DNA (-ssDNA) (Figure 1.6, step A). The nascent –ssDNA is transferred to the 3'-end of the genomic RNA through binding of complementary sequences repeated in the 5' and 3' untranslated regions (Figure 1.6, step B). As there are two molecules of genomic RNA per virion, the –ssDNA can be transferred to either the same strand of RNA or to the other packaged RNA molecule. Intermolecular –ssDNA transfer has the potential to increase the genetic diversity as the exact sequence of the two strands of genomic RNA do not have to be identical. After the 5'-3' transfer, minus-strand DNA synthesis continues up to the PBS (Figure 1.6, step C). Meanwhile, the template, genomic RNA is degraded by the RNase H function of RT, except for a specific purine-rich oligomer, known as the polypurine tract (PPT). The two PPT sequences anneal, and the RNA PPT is used as a primer to initiate plus-strand DNA synthesis. Plus-strand synthesis proceeds until the PBS, generating the plus-strand strong-stop DNA (+ssDNA) (Figure 1.6, step D). Completion of plus-strand DNA synthesis requires a second transfer event that involves annealing of the +ssDNA to the and minus-strand PBS

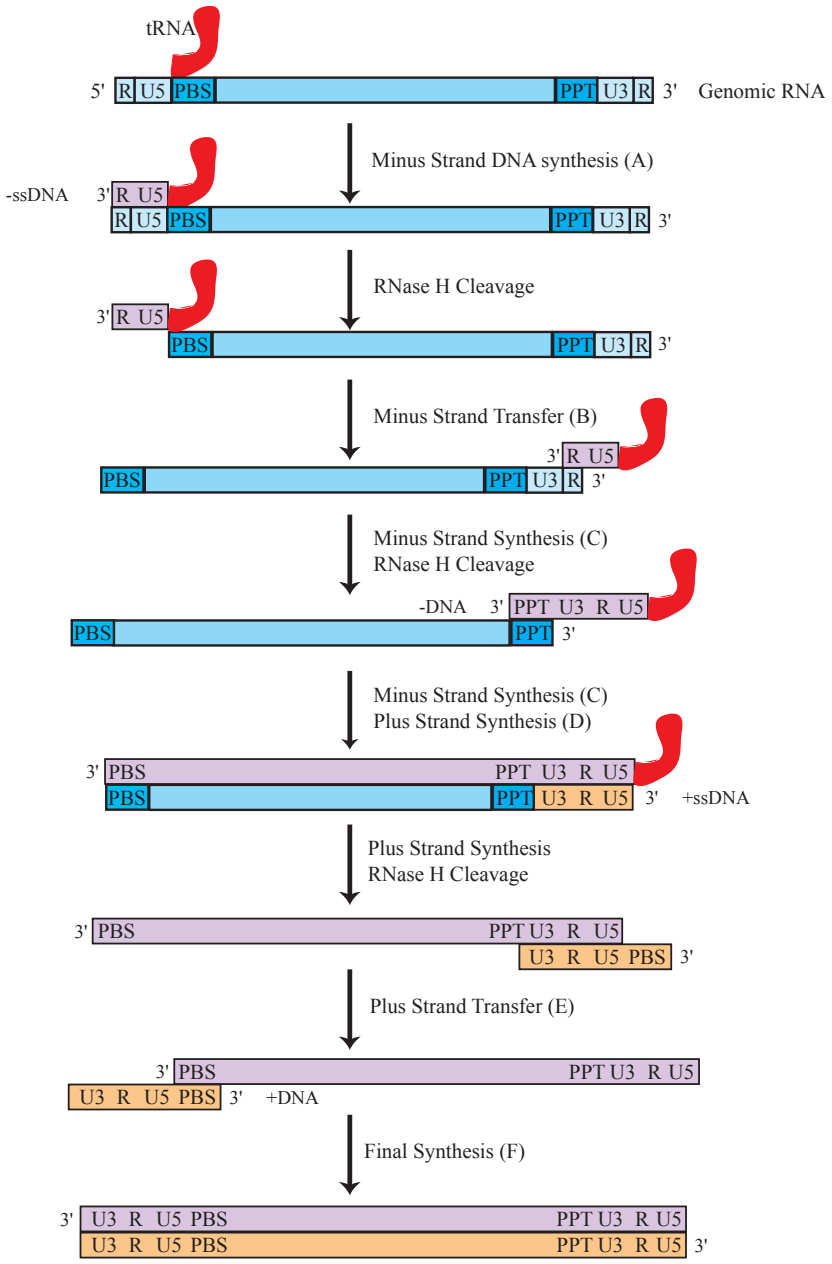


Figure 1.6: Illustration of the steps of reverse transcription in HIV-1. The genomic RNA is shown in blue, the minus-strand DNA in purple, and the plus-strand DNA in orange. In addition, $tRNA_3^{Lys}$ is drawn in red. The various steps referred to in the text (A-F) are indicated in parenthesis.

sequences (Figure 1.6, step E). Following the second DNA transfer, the full-length plus-strand DNA is generated using the minus-strand DNA as a template (Figure 1.6, step F)[4].

During reverse transcription, NCp7 has at least four distinct functions: i) it facilitates $tRNA_3^{Lys}$ annealing to the primer binding site (PBS); ii) it stimulates specific DNA synthesis by blocking self-priming reactions; iii) it enhances RT processivity; and iv) it promotes the first template switch to yield full-length minus-strand DNA. All of these functions relate to the ability of NCp7 to bind nucleic acids.

In vitro, $tRNA_3^{Lys}$ does not bind to the template RNA by itself under physiological temperatures; either high temperatures or NCp7 are required to anneal the tRNA to the PBS[22]. Gel mobility shift assays, circular dichroism spectroscopy, RNase and metal cleavage, fluorescence resonance energy transfer, and NMR spectroscopy have all been used to show that NCp7 is able to directly bind $tRNA_3^{Lys}$ [23-33]. NCp7 contacts $tRNA_3^{Lys}$ on the inside of the L-shape of the tRNA, on the acceptor and D arms as well as the TΨC loop[23,30-32]. However, when it binds $tRNA_3^{Lys}$, NCp7 does not unwind any of the helices in the tRNA or induce any other structural changes[24,27,31]. In the presence of the template genomic RNA, though, global unwinding of the $tRNA_3^{Lys}$ by NCp7 is observed, leading to complete displacement of the 5'-acceptor arm and binding of the tRNA to the PBS[27]. It has been suggested that NCp7 plays two roles in the annealing process: i) it facilitates strand exchange, thus promoting hybridization of the acceptor stem with the PBS and ii) it induces the complete destabilization of the TΨC loop that otherwise is highly stable and resistant to melting[34].

Several mutational studies have been conducted to define the residues of NCp7 required for its interaction with $tRNA_3^{Lys}$. Early reports were difficult to reconcile, as some indicated that the zinc-binding domains were required for $tRNA_3^{Lys}$ annealing, whereas other studies showed

that they were dispensable[25,26]. Mutation of His_{23} to a cysteine, therein converting the Cys_3His zinc-binding domain into a Cys_4 motif, led to loss of protein binding to the acceptor stem of the $tRNA_3^{Lys}$ and failure of primer binding[25,33]. More recent studies have clarified the seeming discrepancy using more specific mutations and more sensitive assays. NCp7 has been shown to interact with the acceptor stem and TΨC loop in a specific manner that requires the zinc-binding domains and the specific Cys_3His zinc-coordination motif. It also has been shown that NCp7 binds the D loop non-specifically, relying purely on electrostatic interactions[22,30,33,34]. Thus, NCp7 binds to the $tRNA_3^{Lys}$ through both specific and non-specific interactions.

NCp7 has been shown in many studies to promote both the minus-strand and plus-strand transfer events that are crucial for complete reverse transcription (Figure 1.6)[35-45]. It is thought that NCp7 enhances strand transfer by destabilizing secondary structural elements in the individual nucleic acid strands, such as the transactivation response element (TAR) and PBS stem loops, and by favoring formation of new, longer double-stranded molecules. NMR spectroscopy studies of the interaction between NCp7 and the PBS showed that NCp7 induces transient unpairing, "fraying", of the base pairs in the RNA stem[42]. Fluorescence studies have also shown that NCp7 can partially melt the stem base pairs of the stem-loop structure of the DNA complement to the TAR RNA (cTAR)[46-50]. Even the transient loss of some base pairs provides the opportunity for new base pairs to form. Mutational analysis revealed that the zinc-binding domains of NCp7 are required for efficient strand transfer[41,43]. In addition, conversion of the Cys_3His motif to a Cys_2His_2 motif moderately decreased strand-transfer rates, whereas conversion to a Cys_4 motif failed to stimulate strand transfer[44]. Thus, the zinc-binding domains of NCp7 are essential for optimal strand transfer.

The ability of NCp7 to unfold TAR is important for more than just maximal strand transfer. It is also crucial to inhibit self-priming reactions that could interfere with reverse transcription. The loop of the stem-loop of TAR is able to interact with cTAR, which can then act as the primer and compete for RT binding (self-priming) to form incomplete DNA sequences[51]. Only the loop/loop TAR/cTAR interaction, and not the fully annealed complex, leads to self-priming. Thus, it is important for efficient reverse transcription, and specifically strand transfer, that the loop/loop interaction between TAR and cTAR be blocked. It has been proposed that NCp7 destabilizes the cTAR stem-loop to weaken the loop/loop interaction and allow complete annealing of the RNA and DNA strands through the TAR and cTAR regions[46,47,49].

Mutational studies with NCp7 have shown that the zinc-binding domains are required for cTAR stem-loop melting and subsequent annealing between the DNA and RNA strands[43,44,52]. An NCp7 mutant in which the zinc-coordinating residues have been converted to serines was unable to block self-priming reactions[43]. In addition, when the Cys_3His motif was changed to a Cys_4 zinc-binding motif, the protein did not bind to the TAR RNA as efficiently and no annealing was observed between the TAR DNA and RNA sequences[52]. Indeed, even the order of the zinc-binding domains is important; when ZD2 was replaced with a second copy of ZD1, 50% of self-priming inhibition was lost and when ZD1 was replaced with a second copy of ZD2, there was a complete loss of inhibition[44]. Thus, the zinc-binding domains are crucial not just for placement of the $tRNA_3^{Lys}$ primer and enhancing the two strand-transfer reactions, but also for inhibiting the self-priming reactions that compete with strand-transfer. NCp7 is essential to several processes related to reverse transcription early in the HIV-1 lifecycle, and mutations of

the zinc-binding domains have been shown to disrupt all of these processes likely due to loss of the ability to bind nucleic acids.

Functions of NCp7 in the Middle of the Formation of a New Virion

NCp7 is not only important early in the HIV-1 lifecycle, it is also required for steps in the middle of the viral lifecycle. The most important function of NCp7 in this phase of the lifecycle is to facilitate the integration reaction, in which viral DNA is incorporated into the chromosomal DNA. As will be discussed below, this role is thought to involve the ability of NCp7 to bind to nucleic acids.

Once reverse transcription has been completed, the newly reverse transcribed double-stranded viral DNA (dsDNA) is transported into the nucleus and integrated into the host chromosomal DNA. While the dsDNA is still in the cytosol, integrase (IN) cleaves the 3'-terminus of each strand of the viral DNA (Figure 1.7)[4]. The cleavage removes the final two bases from the 3'-end. A viral nucleoprotein complex, termed the pre-integration complex (PIC), is able to enter the nucleus. The PIC is composed in part of the dsDNA, IN, MA, Vpr, and NCp7[4]. The PIC of HIV-1 is able to enter the nucleus both during mitosis when the nuclear membrane is disassembled, as well as during interphase via active transport through the nuclear pore[4]. Upon nuclear entry, the PIC binds to the chromosomal DNA. Highly bent DNA sites, such as those found at specific positions on nucleosomes, are highly preferred for integration, most likely because the phosphodiester backbone of the chromosomal DNA is most accesible in these types of sites[4]. There seems to be no specific sequence requirement for the integration reaction. IN then uses the 3'-hydroxyl groups at the two extremities of the dsDNA to catalyze two concerted attacks on the phosphodiester bonds of the chromosomal DNA[4] (Figure 4.7). The

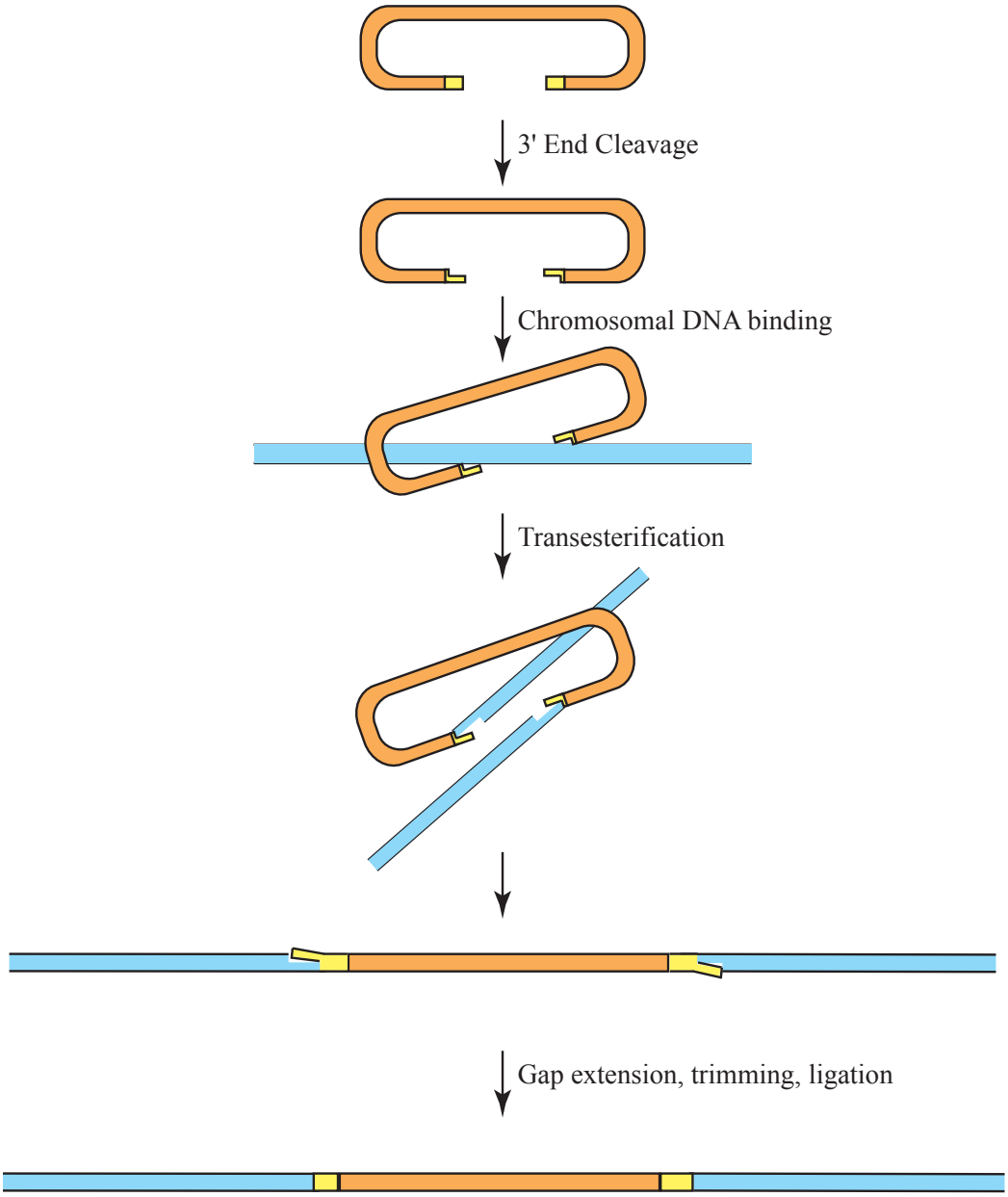


Figure 1.7: Schematic representation of the steps of HIV-1 integration. The viral DNA is drawn in orange with yellow ends and the cellular DNA is shown in blue.

two integration sites are on different DNA strands and separated by four to six bases pairs. This separation means that the two sites are on the same face of the DNA double helix, but are located on opposite sides of the major groove[4]. The transesterification reaction leaves the viral dsDNA attached to the chromosomal DNA by only one strand at each end. The 3'-end of the chromosomal DNA is extended to fill the gaps that flank the viral dsDNA, and the 5'-end of the viral dsDNA is trimmed (Figure 1.7)[4]. The resultant 3' and 5' ends can be ligated together, after which the PIC is disassembled[4].

NCp7 has been shown to greatly stimulate integration reactions both *in vitro* and *in vivo*[53-56]. Indeed, studies performed with purified IN could not produce any detectable activity under physiological conditions, yet isolated PIC could promote efficient integration *in vitro*[53]. The stimulation of integration is specific for NCp7, as other small DNA-binding proteins could not promote *in vitro* integration under physiological conditions[53]. The zinc-binding domains have been shown to be required for the 3' end cleavage activity early in integration[55]. Mutation of His_{23} to cysteine almost completely blocked 3' end cleavage by IN[55]. The zinc-binding domains also play an important role in the integration strand-transfer reaction, as mutants in which all cysteine residues were replaced with serine retained only half of the wild-type transfer activity[56]. Additionally, mutations that convert the Cys_3His zinc-binding motif of NCp7 into other zinc-binding motifs (*e.g.,* Cys_4 or Cys_2His_2) show reduced ability to stimulate integration[54]. It is thought that NCp7 enhances integration by coating the dsDNA, blocking non-specific interactions between IN and the dsDNA, and enhancing the local concentration of IN around the dsDNA end[54-56]. However, this model does not completely explain the necessity of NCp7 for end-processing. It has also been suggested that, in the presence of magnesium ions, NCp7 may

promote and stabilize specific interactions between IN and the dsDNA, possibly by favoring the correct oligomerization state of IN[53].

Function of NCp7 in Retroviral Assembly Late in the Formation of a New Virion

Once the viral DNA has been integrated into the chromosomal DNA, transcription of the integrated viral DNA and translation follow. Transcription and translation can occur minutes after integration, or the integrated viral DNA can be dormant for years[4]. The gag and gag-pro-pol polyproteins are transcribed by cytosolic ribosomes. Myristylation of the matrix domain within the polyproteins helps direct them to the cytosolic side of the plasma membrane. It is thought that the gag and gag-pro-pol polyproteins are transported to the plasma membrane either by interaction with the cytoskeleton or through interaction with cellular chaperone proteins. Aggregation of the polyproteins at the plasma membrane leads to their multimerization. The assembled gag polyproteins cause the plasma membrane to bud, eventually separating completely to form an immature virion. The exact mechanism that leads to budding is unknown. It has been suggested that budding virions sit on narrow stalks that protrude from the plasma membrane. This explanation is based upon the observation of emerging spheres on stalks prior to viral release, and mutational studies support this model[4].

Assembly of gag polyproteins at the plasma membrane is due in part to the nucleocapsid portion of the non-cleaved precursor and RNA[4]. The NCp7 domain of the polyprotein has been shown to contain two functionally-defined interaction domains that are responsible for gag-gag self-association[4]. One of the two interaction domains is located in the first seven amino acids of NCp7, with Arg_3 and Arg_7 being the most important[57]. The second interaction domain is located around ZD1[58]. Though the zinc-binding domains have been shown to be dispensable for gag

self-association, the basic residues of the amino-terminus of NCp7 seem to be crucial for this activity[57-63]. The NCp7 domain of the gag polyproteins are thought to bind to long strands of RNA, like beads on a string[57,61,64-66]. This scaffold of proteins would increase the local concentration of gag and align the proteins to facilitate optimal gag-gag interactions[61].

The most important function of NCp7 is to package two copies of the viral genomic RNA into the new virion. Full-length, genomic RNA is specifically selected out of the pool of viral and cellular mRNA transcripts where it makes up less than 1% of the total RNA[4]. The dimerization and packaging domain of the RNA is located in the Ψ site, between the 5' LTR and the gag initiation sequence (Figure 1.1). Based upon nuclease and chemical accessibility studies and energy-minimization studies, the Ψ site was predicted to be made up of four stem-loop motifs (Figure 1.8)[67]. NCp7 selectively binds to the full Ψ-site RNA[68]. Each of the four stem loops (SL1-SL4) have been shown to bind NCp7, each to a different degree. The second and third stem loops bind NCp7 with the highest affinity (K_D of 23 and 28 nM, respectively), whereas the fourth stem-loop has the weakest affinity for NCp7 (K_D = 320 nM)[69]. Binding to the first stem-loop is more complex, as it promotes RNA dimerization and packaging.

NCp7 has many critical functions throughout the HIV-1 lifecycle. At the early phase of the lifecycle, it is important for $tRNA_3^{Lys}$ primer annealing, strand transfer reactions, RT processivity, and blocking self-priming reactions in reverse transcription. In the middle phase of the lifecycle, the zinc-binding domains of NCp7 are crucial for 3' end cleavage of viral DNA by IN and stimulation of the integration reaction. Additionally, in the late phase of the lifecycle, the amino-terminus of NCp7 contains the two interaction domains that are required for gag-gag interaction and viral assembly. NCp7 is also required for dimerization and packaging of the viral

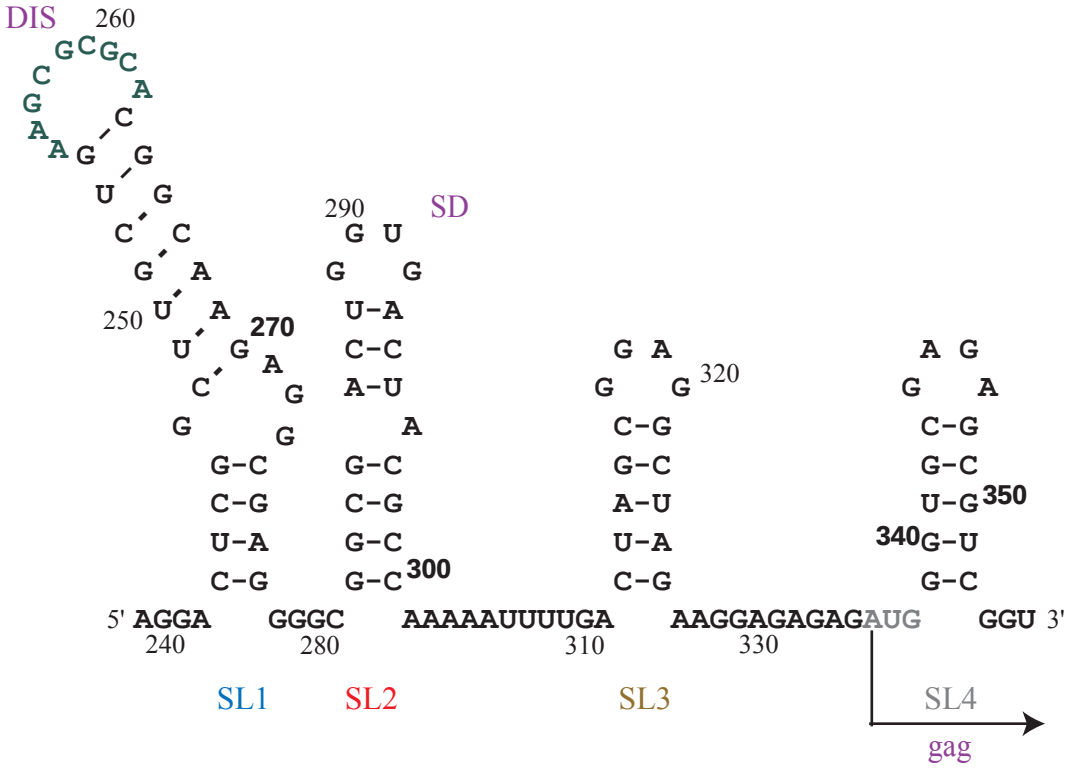


Figure 1.8: Nucleotide sequence and secondary structure of the HIV-1_{Lai} Ψ site. The dimerization initiation site (DIS), splice donor site (SD), and gag initiation codon (AUG) are highlighted. The numbering is relative to the RNA cap site (+1).

genomic RNA. Mutational studies have highlighted the necessity of NCp7 in all of these stages. Thus, NCp7 is crucial for the propagation of HIV-1 at all stages of the virus lifecycle.

NCp7 Binds to the Ψ Site RNA

Structures of the Ψ Site Stem-loops

The first structure of a Ψ RNA stem-loop to be reported was the NMR solution structure of the third stem-loop domain, SL3[70]. SL3 is the only domain specifically required for packaging of genomic RNA, though all of the other stem-loops of the Ψ site are important to produce normal virions[71]. The base-paired stem of SL3 was found to form a standard A-form helix, whereas the loop nucleotides appear flexible and adopt several conformations (Figure 1.9A)[70]. In most calculated structures, G_{318} and G_{320} are unstacked, and G_{317} and A_{319} extend into the minor groove side of the RNA stem. The Watson-Crick face of A_{319} is angled away from G_{317}, and the Hoogsteen edge of A_{319} is somewhat hidden from the solvent[70]. The structure of the SL3/NCp7 complex was also determined by NMR spectroscopy[72]. Upon binding of NCp7, the loop of the SL3 RNA adopts a more ordered conformation[72]. A_{319} moves out of its protected conformation to project away from the stem (Figure 1.9B)[72]. G_{318} also turns out of the loop to interact with NCp7[72]. G_{317} and G_{320}, which were far apart in the free structure, are joined by a hydrogen bond between the O6 of G_{317} and the NH_2 of G_{320}. The zinc-binding domains of NCp7 interact with the free, exposed bases in the loop (G_{318}, A_{319}, and G_{320}), while the amino-terminal residues of NCp7 (residues 3-10) form a 3_{10} helix that fits into the major groove of the stem and contacts the phosphodiester backbone of the RNA. The side chains of Val_{13}, Phe_{16}, Ile_{24}, and Ala_{25} in ZD1 create a hydrophobic binding pocket that interacts specifically with G_{320} in the loop[72]. Similarly, G_{318} binds in a hydrophobic pocket formed by the side chains of ZD2

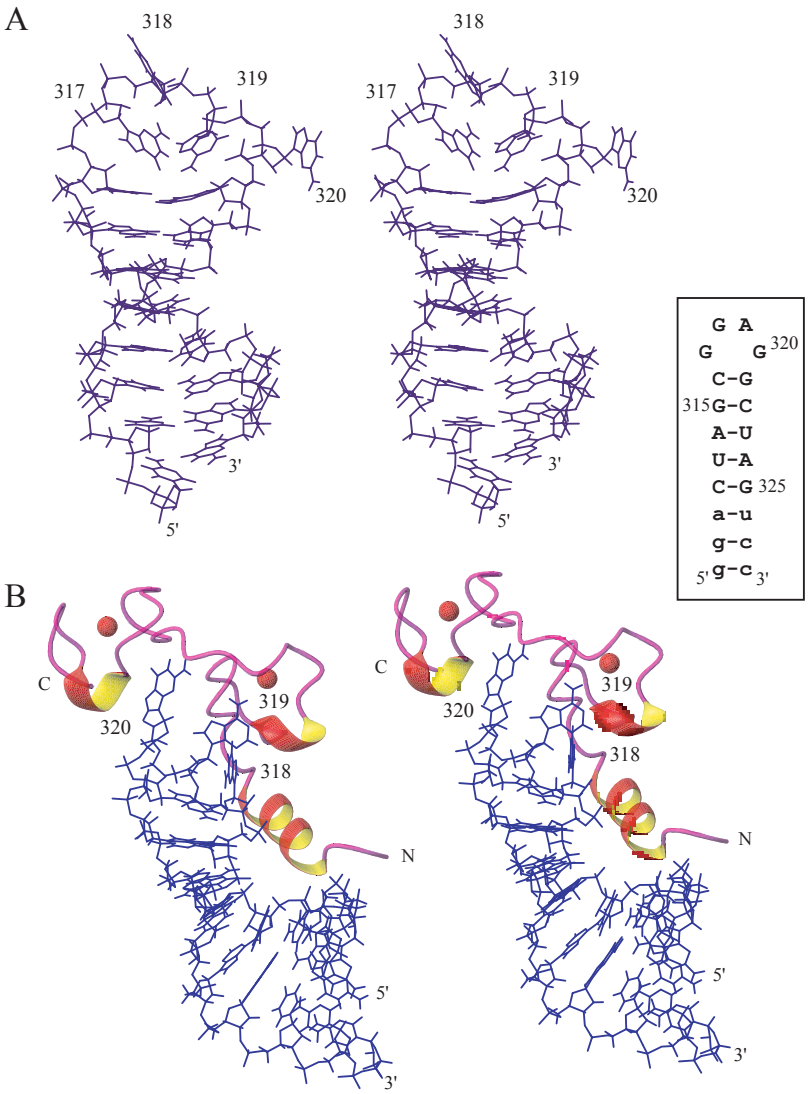


Figure 1.9: NMR structures of free SL3[70] (A) and of the SL3/NCp7 complex (B)[72]. Shown in the box is the sequence and secondary structure of SL3 used for these studies. The lower case letters represent non-natural nucleotides added to the sequence to improve transcription by T7 RNA polymerase. The figures are adapted from PDB files 1BN0 (A) and 1A1T (B) using MOLMOL[21].

residues Trp_{37}, Gln_{45}, and Met_{46}[72]. A_{319} makes hydrophobic contacts with Phe_{16}, Asn_{17}, Ala_{25}, and Arg_{32}[72]. Overall, the highly ordered SL3 structure is stabilized by interactions of loop residues with multiple side chains of NCp7.

The three-dimensional structures of the second stem-loop domain (SL2), both the free SL2 and the SL2/NCp7 complex, have been determined by NMR spectroscopy (Figure 1.10)[73,74]. The second stem-loop domain contains the highly efficient major donor splice site. It is the first 5' splice site, and it is used to produce all spliced transcripts[75]. Like free SL3, free SL2 forms an A-form helical stem and mostly disordered loop (Figure 1.10A)[74]. G_{289} is the most ordered nucleotide in the loop and is stacked on the U_{288}-A_{293} base pair that closes the loop, but is not stabilized by any internucleotide hydrogen bond[74]. The other loop bases are disordered and point out of the loop in most of the 20 lowest energy structures due to the lack of NMR evidence of internucleotide interactions. The most striking feature of the solution structure of free SL2 is an AUA base triple formed between nucleotides A_{286}, U_{295}, and A_{296} and which divides the stem into two A-form helical regions[74]. The phosphodiester backbone of the 3'-strand of the stem is bent at the U_{295}-A_{296} linkage such that the base of U^{295} is approximately coplanar with the bases of A_{286} and A_{296}. This allows A_{286} to form hydrogen bonds with the base of A_{296}, which also forms hydrogen bonds with the base of U_{295}[74].

The A-form helical stem of SL2 is not affected by the binding of NCp7; the disordered loop, however, becomes ordered due to interactions with NCp7 (Figure 1.10B)[73]. As observed in the structure of the SL3/NCp7 complex, the two exposed guanosines in the loop of SL2 bind to the hydrophobic pockets created by the two zinc-binding domains of NCp7: G_{289} binds in the pocket of ZD2 and G_{291} binds in the pocket of ZD1[73]. The primary difference between the two structures is the positioning of the 3_{10} helix formed by the amino-terminal residues of NCp7. In

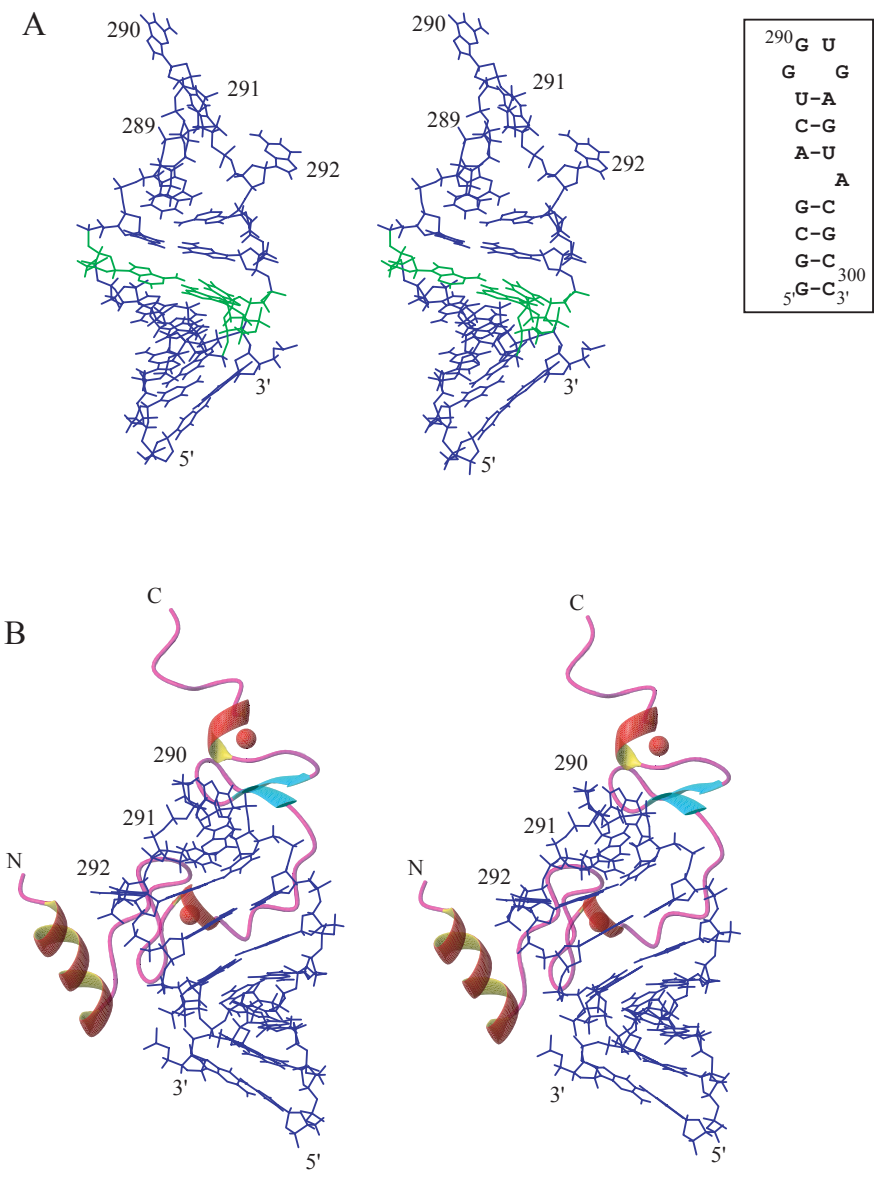


Figure 1.10: NMR structures of free SL2[74] (A) and of the SL2/NCp7 complex[73] (B). The AUA base triple is highlighted in A in green. Shown in the box is the sequence and secondary structure of SL2 used for these studies. The figures are adapted from PDB files 1ESY (A) and 1F6U (B) using MOLMOL[21].

the SL2/NCp7 complex, ZD1 packs against the A-U-A base triple in the RNA minor groove, and the 3_{10} helix packs against ZD1[73]. Thus, in the SL2/NCp7 complex, the 3_{10} helix binds along the minor groove face of the stem, as opposed to binding in the major groove as seen in the SL3/NCp7 complex.

Most recently, the NMR solution structure of the fourth stem-loop (SL4) was determined (PDB file not available)[76]. The structure of free SL4, is significantly different from that of SL2 and SL3, as the loop of SL4 is ordered, while the stem is more dynamic[76,77]. Interestingly, there is a five-base stack across the stem of SL4 that extends in the order C_{343}, G_{342}, G_{350}, U_{351}, and C_{352} (Figure 1.8)[77]. The tetraloop of SL4 has a GNRA sequence (where N is a any nucleotide and R is a purine), which has been shown to form a characteristic GNRA fold[78]. In all structures of GNRA tetraloops, the G and A are coplanar and form a sheared G-A base pair and N, R, and A all stack on the 3' side of the stem. Additionally, there is a sharp turn between the G and N in the loop. Accordingly, in the structure of SL4, G_{344} and A_{347} form a sheared G-A base pair and A_{345}, G_{346}, and A_{347} stack on each other on the 3' side of the stem[76,77]. As described above, NCp7 binds very weakly to SL4 *in vitro*. It has been suggested that *in vivo* SL4 takes part in important RNA-RNA tertiary interactions with elements outside the Ψ site that stabilize its tertiary structure[76].

Kissing-Loop Mediated Dimerization of HIV-1 Genomic RNA

In the HIV-1 virion, two copies of genomic RNA are joined by the dimer linkage structure (DLS) at the 5' end of the genome[79]. The DLS is the ~300 nucleotide, double-stranded span on the genomic RNA where the two strands of RNA are bound together. The structure of the DLS is visible by electron microscopy and has been observed in many different

retroviruses[79]. SL1 contains the specific dimerization initiation site (DIS) within the DLS. During packaging, two copies of full-length viral RNA partially unfold and bind to one-another for encapsidation[80]. The loop of SL1 contains the DIS, a 6-nucleotide palindromic sequence (Figure 1.8), GCGCGC in HIV-1 subtype B and GUGCAC in HIV-1 subtype A. These palindromic sequences have been shown to be absolutely required for dimerization to occur[81]. Dimerization takes place through a kissing-loop mediated mechanism in which the palindromic loop of one molecule of RNA makes Watson-Crick base pairs with the corresponding loop of a second RNA molecule (Figure 1.11)[82-84]. The RNAs form a stable kissing-loop homodimer (K_D of 6.2 nM and 2.4 nM for subtypes A and B, respectively)[85]. Following formation of the kissing-loop interaction, the stems of each individual RNA begin to melt and make new pairings with nucleotides in the opposite molecule (Figure 1.11). Eventually, there is a maturation event in which the kissing-loop complex is fully converted into an extended duplex formed by the SL1 nucleotides (Figure 1.11)[83]. This maturation event is essential for genomic RNA packaging.

Mutational studies have shown that the central G and C nucleotides are important for efficient loop-loop interactions, and the flanking bases can be mutated as long as the palindromic nature of the DIS sequence is preserved[86]. Indeed, selection experiments have shown that only a few sequences are capable of homodimerization[85]. Conservation of C and G at each of the two central positions in the palindromic loop occurs 96% and 89% of the time, respectively, for subtype A and 61% and 83% of the time, respectively, for subtype B[85]. Furthermore, these experiments also showed that adenosine at the first and last positions of the SL1 loop are highly favored[85]. Mutations in the palindromic sequence which interfere with dimerization, have been shown to reduce viral infectivity by 99%[87]. Thus, the sequence of the palindromic loop is integral for proper formation of the kissing-loop interaction. As the kissing-loop interaction is

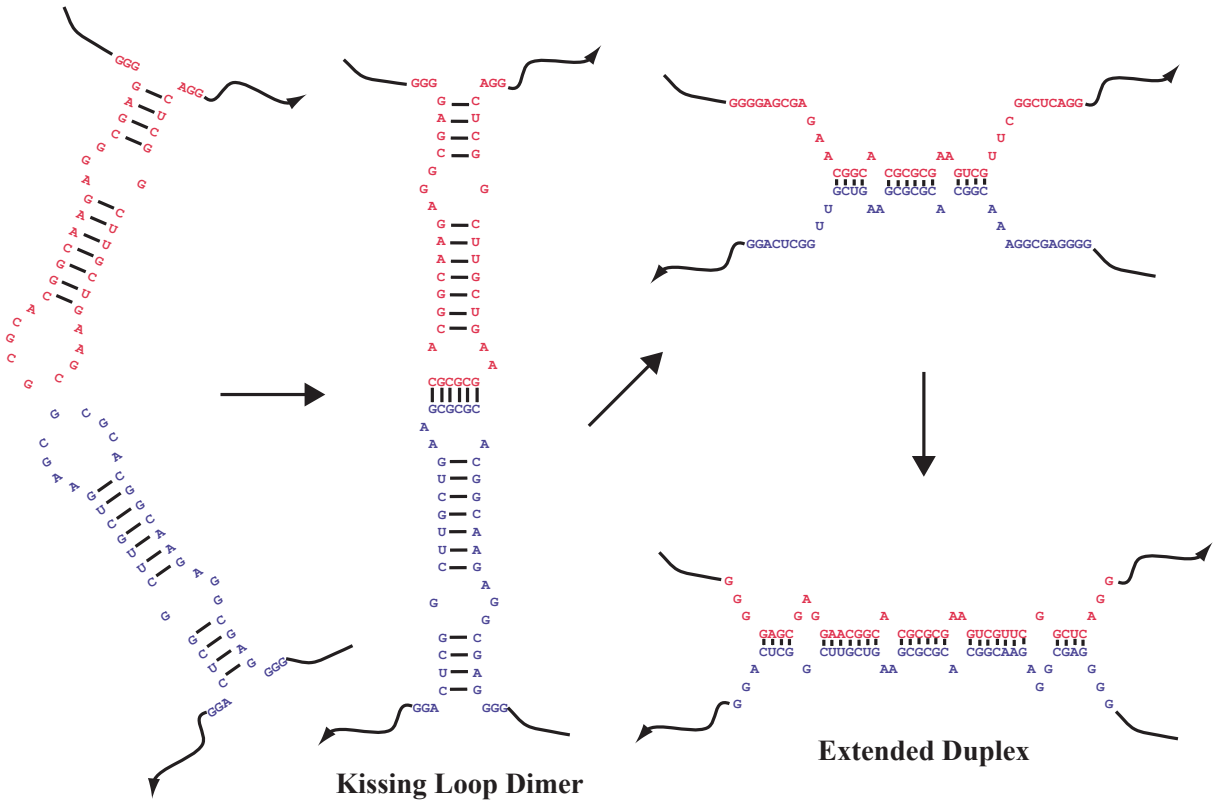


Figure 1.11: Diagram of the mechanism of HIV-1 genomic RNA dimerization. Two separate copies of full-length genomic RNA make Watson-Crick base pairs through their terminal loops, forming a kissing-loop interaction. The two RNA stems begin to melt and bind to one another. One SL1 monomer is shown in red and the other in blue. Further details on the mechanism are provided in the text.

the start of genomic RNA dimerization, anything that impairs its formation has a large impact on viral infectivity.

Structures of the SL1 Kissing-loop and Extended Duplex Conformations

There are high-resolution, three-dimensional structures available of the free SL1 monomer and of the kissing-loop and extended duplex homodimer conformations of SL1. The structure of an SL1 palindromic mutant monomer was determined by NMR spectroscopy; the subtype B palindromic sequence was mutated at position 258 to prevent dimerization[88]. The SL1 mutant monomer adopts the predicted stem and loop regions; the stem forms a typical A-form helix and the loop has some degree of structure, primarily observed in the stacking of the three cytosines in the palindrome, but is otherwise fairly dynamic[88,89].

Structures of the kissing-loop interaction have been determined by NMR spectroscopy[84,90,91] and X-ray crystallography[92] (2.76 Å and 2.6 Å for the subtype A and B structures, respectively) for both the A and B subtype sequences. The stems of each molecule of RNA form standard A-form helices, though no evidence is seen in the solution structures for formation of the G_{254}-C_{264} loop closing base pair (Figure 1.12A)[90-92]. The kissing-loop junction forms a right-handed helix, with its minor groove fully exposed[90]. In the subtype B kissing-loop, the adenosine nucleotides that flank the palindromic sequence (A_{255}, A_{256}, and A_{263}) form an intrastrand triple base stack that interacts with the stack on the other SL1 molecule (Figure 1.12A)[90,92]. In the crystal structure of the subtype B kissing-loop, however, A_{255} and A_{256} are bulged out of the loop mini-helix and stacked on one another outside of the helix[92]. It is unclear what the basis of this discrepancy is, though it could be due to crystal packing forces that alter the conformation of these nucleotides, specifically by shifting the equilibrium between a bulged-

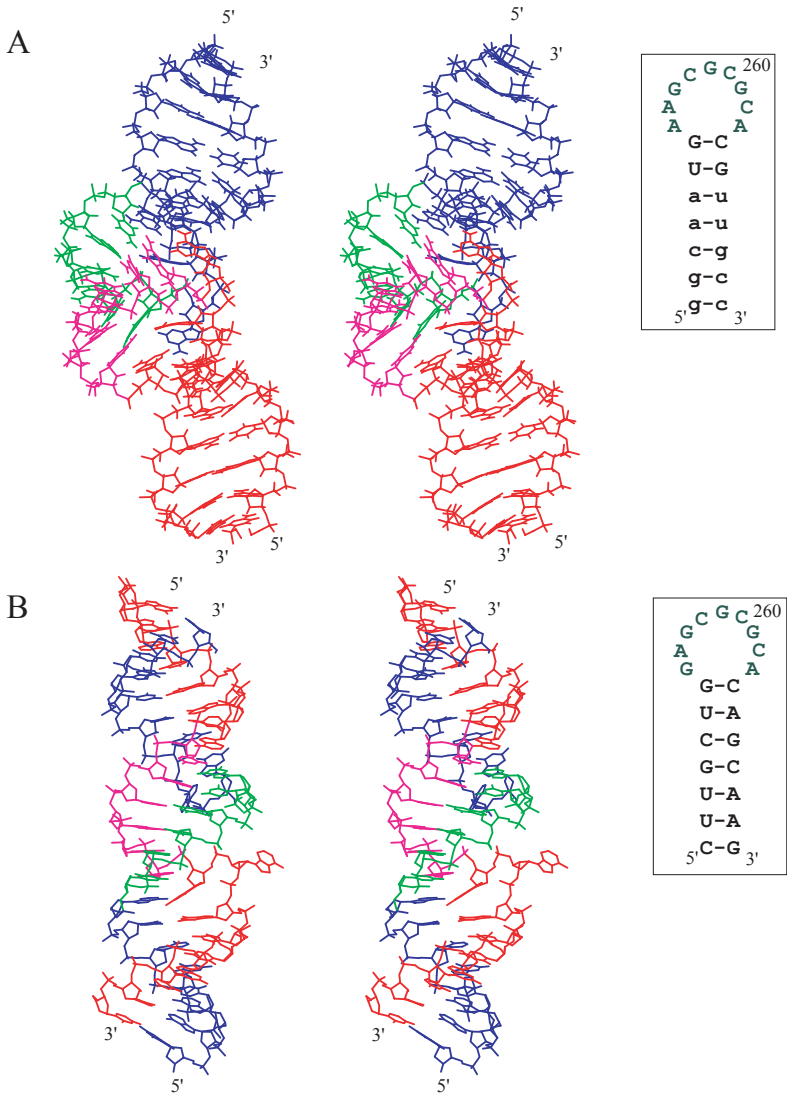


Figure 1.12: NMR solution structure of the kissing-loop[90] (A), and 2.30 Å x-ray crystal structure of the extended duplex[93] (B) forms. In both A and B, one monomer is shown in blue, with the palindromic sequence in green, and the other monomer is shown in red, with the palindromic sequence in magenta. By each structure, the sequence of the SL1 RNA used for structure determination is shown boxed. Non-natural nucleotides are shown in lower cases. The figures are adapted from PDB files 1BAU (A), and 462D (B) using MOLMOL[21].

in and bulged-out conformation. Molecular dynamics studies support this hypothesis[94]. By contrast, the flanking purines in the subtype A kissing-loop (A_{255}, G_{256}, and A_{263}) are stacked inside the loop helix, and the two adenosines (A_{255} and A_{263}) form a non-canonical A-A base pair[84,91,92].

The NMR solution structure of a subtype B extended duplex reveals the RNA to be an approximately linear, double-stranded helix with the unpaired adenine bases originating from the terminal loop bulged out[95]. The stems on either end of the molecule are similar to a standard A-form helix, while the base-paired region formed by the palindromic sequences has high roll values that distort it from A-form geometry[95]. The 2.3 Å crystal structure of a subtype A extended duplex also shows that the RNA forms a linear A-form helix (Figure 1.12A)[93]. In this structure, however, the unpaired purines originating from the terminal loop do not bulge out; rather, only one nucleotide bulges out and the other two form a non-canonical G-A base pair[93]. The difference in these two extended duplex structures is likely caused by sequence variation: in subtype A viruses, position 256 in the loop is a guanosine, whereas in subtype B it is an adenosine.

Role of NCp7 in RNA Dimerization and Packaging

The dimerized RNA is packaged into the budding virion by the NCp7 domain of the gag polyprotein. Following budding, there is a required maturation event in which PR cleaves the polyproteins, forming the individual proteins. At this point, there is an RNA maturation event as well. It is thought that the RNA is packaged as a kissing-loop dimer and during the subsequent maturation, the extended duplex is formed (Figure 1.11). Separation of RNA binding from packaging has been very difficult to accomplish. Mutagenesis studies have suggested that Lys_{11}

and Lys_{14} are crucial for RNA packaging; when mutated to alanine, the viral particles contained only 25% and 22% of the wild-type levels of RNA, respectively[96]. These residues do not seem to be important for RNA binding *in vitro*, suggesting that they may play a specific role in RNA packaging that is not related to RNA binding[96].

NCp7 has been shown to specifically enhance the conversion from the kissing-loop to the extended duplex conformation[97,98]. *In vitro*, non-physiological and high temperatures and salt concentrations are required for SL1 to form exclusively the extended duplex conformation in the absence of NCp7. In the presence of NCp7, dimerization occurred at low temperatures (37 °C) and low salt concentrations[99]. Fluorescence data showed that conversion from the kissing-loop to the extended duplex occurs most efficiently when NCp7 binds to a preformed kissing-loop[100]. There is some controversy concerning the requirement of both zinc-binding domains for genomic RNA dimerization and subsequent maturation. Mutations in either the zinc-binding residues or basic residues adjacent to the zinc-binding domains have been shown to cause deleterious defects on the specificity of encapsidation and on RNA binding[5,97,101-103]. Another study showed that a peptide containing only the basic residues flanking the amino terminus and the carboxyl terminus of ZD1 linked together by two glycine residues was sufficient for conversion of the kissing-loop to expended duplex form[101]. However, this process was significantly less efficient than observed with wild-type NCp7, requiring 10 times as much protein to convert an equal amount of RNA[101]. Mutational studies have also shown that the aromatic residues in the two zinc-binding domains are also important for dimerization[97]. It appears likely that the zinc-binding domains are responsible for the binding of specific RNA sequences and the basic residues in and around these domains interact non-specifically with the negatively-charged phosphodiester backbone of the RNA. It is also possible that the zinc-binding domains function to properly orient the basic

residues for optimal interaction with RNA. Structural studies of the complexes formed between NCp7 and both the kissing-loop and extended duplex conformations of SL1 are needed to better understand the RNA dimerization process.

Several different mechanisms have been proposed for the conversion of the kissing-loop to the extended duplex. One involves covalent modification of the RNA. More specifically, an in-line attack by the 2'-hydroxyl of one R_{256} on the 3'-phosphate of the other R_{256} would result in chain cleavage by transesterification and the formation of a 2',3' cyclic-phosphate and a 5' OH group. In the x-ray crystal structure of the kissing-loop, the two R256 residues are close in space and in an orientation with their 2'-hydroxyl groups favorably oriented for this type of reaction[92]. The two cleavage reactions could then be followed by cross-ligation of R_{256} from strand A into strand B, finally yielding the extended duplex formation without any disruption of existing base pairs[92]. It has recently been shown that in some molecules, A^{255} is protonated at physiological pH, which causes the kissing-loop interaction to be dynamic by altering the charge of the base[104]. This led the authors to propose a somewhat different mechanism of kissing-loop to extended duplex conversion: protonation of A^{255} could disrupt the kissing-loop junction and allow A_{255}, R_{256}, and A_{263} to move between multiple conformational states. Among these states would be one that is optimal for the NCp7-activated conversion to extended duplex[104]. These two mechanisms are not mutually exclusive; it is possible that the flexibility caused by the protonation state of A_{255} alters the conformation of R_{256} and makes it optimal for transesterification and cross-ligation reactions catalyzed by NCp7. The specific role of NCp7 in the conformational conversion is not known. The conformational enhancement could be due to either destabilization of the kissing-loop or stabilization of the extended duplex. Alternatively,

NCp7 could play an enzymatic role, acting as a nuclease or ligase in the transesterification and cross-ligations reactions.

Role of the Internal Loop of SL1 in RNA Dimerization

SL1 contains both a terminal palindromic loop and an asymmetric internal loop (Figure 1.8). The sequence of this internal loop is completely conserved in all HIV-1 strains, whereas the terminal palindromic loop shows some sequence variation[105]. The internal loop induces a bend in the stem of SL1 that possibly helps formation of the kissing-loop interaction *in vivo*[104]. It has also been suggested that the internal loop destabilizes the stem of SL1, facilitating the conversion from the kissing-loop to the extended duplex conformation[106,107]. Indeed, deletion of the internal loop has been shown to reduce the ability of SL1 to form an extended duplex[106,108,109]. Additionally, it has been shown that deletion or mutation of the internal loop and lower stem impairs a step of reverse transcription after production of the minus-strand DNA by 70-90%[109]. It has been suggested that this inhibition is due to improper folding of the minus-strand DNA[109].

NMR structural studies have shown that the internal loop of SL1 has a well-defined structure: the base of G^{247} stacks on the bases of the 5'-stem and that of A_{271},G_{272}, and G_{273} stack on the bases in the 3'-stem (Figure 1.13)[89,107,110]. It is this stacking of the uneven number of loop bases that causes the stem to bend by 25 °C. Additionally, G_{247} and A_{271} are coplanar and form a sheared G-A base pair[89]. Fluorescence binding studies have shown that NCp7 binds the internal loop with a relatively high affinity (K_D = 140 nM) compared to SL2 and SL3 RNAs (K_D =20 nM) [69,110]. It has been suggested that NCp7 binds to the exposed guanosine bases in the loop, but detailed structural data is needed to fully understand this interaction[110].

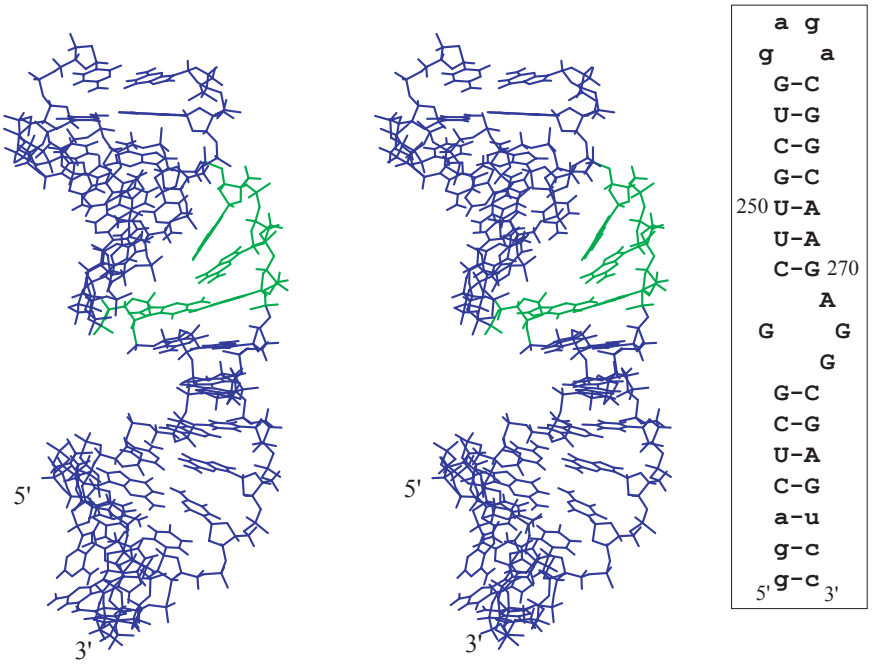


Figure 1.13: NMR solution structure of SL1 internal loop[89]. The residues of the internal loop are drawn in green. The figure is adapted from the PDB files 1N8X using MOLMOL[21].

NCp7 is a Promising Target for Antiviral Therapy

As of February 2005, there are 24 antiviral drugs that have been approved for clinical use against HIV-1 in the United States (www.thebody.com). Of these 24 drugs, eight are protease inhibitors, twelve are nucleoside analog reverse transcriptase inhibitors (NRTIs) or nucleotide analog reverse transcriptase inhibitors (NtRTIs), three are non-nucleoside reverse transcriptase inhibitors (NNRTIs), and one is a viral entry inhibitor. These four classes of drugs inhibit three different stages of the viral lifecycle, providing a broad front of attack.

Protease Inhibitors

As described above, HIV-1 protease (PR) is transcribed as part of the gag-pro-pol polyprotein that results from ribosomal frameshifting. Following an autocatalytic cleavage of gag-pro-pol, the released PR cleaves the resulting gag and pol polyproteins to yield the individual constituent proteins. These proteolytic cleavage reactions are absolutely required for viral maturation. PR is able to recognize and cleave at nine separate sites in the gag-pro-pol polyprotein, and each site has a different sequence. The cleavage efficiency varies significantly among the sequences[111]. Some generalizations have been made to describe the various cleavage sites: i) small amino acids are often found at positions two and four residues before the cleavage site; ii) glutamine, basic amino acids, or aliphatic amino acids are common three residues before the cleavage site; iii) the amino acids that flank the cleavage site are typically hydrophobic; iv) branched amino acids are never found just before the cleavage site; v) glutamine, glutamic acid, or an aliphatic residue are observed two residues after the cleavage site; and vi) aromatic residues are excluded from the third position after the cleavage site[4].

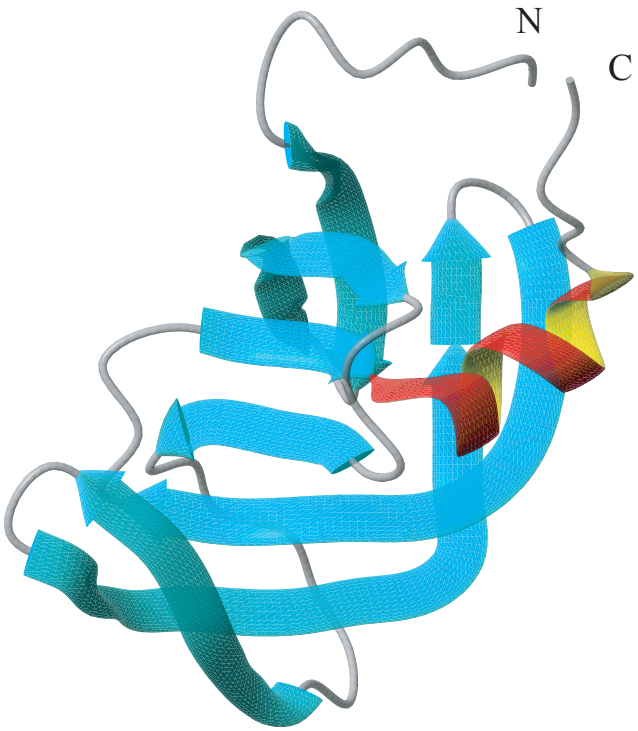


Figure 1.14: X-ray crystal structure of one monomer of the HIV-1 protease at 2.80 Å. The figure is adapted from PDB file 3HVP using MOLMOL[21].

The active PR is a symmetrical homodimer, with each monomer having a structure similar to that of the aspartyl proteases (Figure 1.14)[112]. The active site is found at the dimer interface and contains two conserved, catalytic Asp-Thr-Gly triads, one from each monomer[113]. The triads are located in loop regions and are stabilized by a network of hydrogen bonds, a feature commonly observed in aspartyl proteases[113]. The triad loops from each monomer are linked by a "fireman's grip" in which the threonine residues of two triads are joined together by hydrogen bonds[113]. The active site of the enzyme also contains well-defined hydrophobic subsites that are involved in substrate binding. Covering the active site in the free enzyme are two flexible flaps, each comprised of two antiparallel β sheets, that project over the active site and assist in binding of the substrate[114].

All eight of the currently approved protease inhibitors (e.g. ritonavir, saquinavir, amprenavir, and indinavir) act as competitive inhibitors of the substrate and contain a hydroxyethylene bond rather than a peptide bond that is not cleaved by the protease[115]. The structure of the HIV-1 PR undergoes substantial changes upon binding to an inhibitor: the flaps move towards the active site and close it off from the solvent[116]. The flap residues shift at least 1 Å in this movement, with the residues at the tip of the flaps moving as far as 7 Å[116]. Protease inhibitors have been shown to make extensive interactions with the protease active site. The crystal structure of a hexapeptide inhibitor bound to PR revealed that the inhibitor makes hydrogen bonds with the two Asp residues in the active site, forms many van der Waals contacts with the hydrophobic subsites around the active site, and connects to the bottom flap residues via a buried water molecule[116,117]. Structures of other protease inhibitors bound to PR have revealed similar binding interactions[118-120].

Reverse Transcriptase Inhibitors

As described above, RT is responsible for the conversion of genomic RNA into proviral DNA that can be integrated into the host chromosomes. HIV-1 RT contains an RNase H activity in addition to its DNA polymerase activity. The initial cleavage of RT from the gag-pro-pol polyprotein yields the p66 form of the protein, which contains both the DNA polymerase and RNase activities. *In vivo*, p66 is able to form an inactive homodimer[121]. A subsequent cleavage reaction removes the RNase H domain from one of the p66 subunits, resulting in production of p51 within the active p66/p51 heterodimer[121]. Like cellular DNA polymerases, RT uses deoxynucleoside triphosphates (dNTPs) to add deoxyribonucleoside monophosphates to the 3' OH of the primer through formation of new phosphodiester bonds, thus elongating the polynucleotide and releasing pyrophosphate. Unlike cellular DNA polymerases, though, RT does not contain a proofreading activity, making it 1,000 – 10,000 times more error-prone[121]. The RNase function of RT catalyzes the specific cleavage of the genomic RNA template in the RNA/DNA hybrid formed during the first stages of reverse transcription[121].

NMR spectroscopy and X-ray crystallography studies have been conducted with free RT[122], RT-DNA complexes[123-126], and RT-inhibitor complexes[127-132]. p51 and the DNA polymerase domain of p66 are both made up of four subdomains: the fingers, palm, and thumb subdomains, so named because of their structural resemblance to a right hand, and the connection subdomain that links the polymerase and RNase H active sites (Figure 1.15)[127]. The palm subdomain contains five β-strands that form hydrogen bonds with four β-strands at the base of the thumb subdomain, linking these two subdomains and creating a large cleft where the nucleic acid template primer complex binds (Figure 1.14)[127]. The fingers domain, comprised of a mixture of β strands and three α helices, also contributes to this catalytic cleft[124]. The

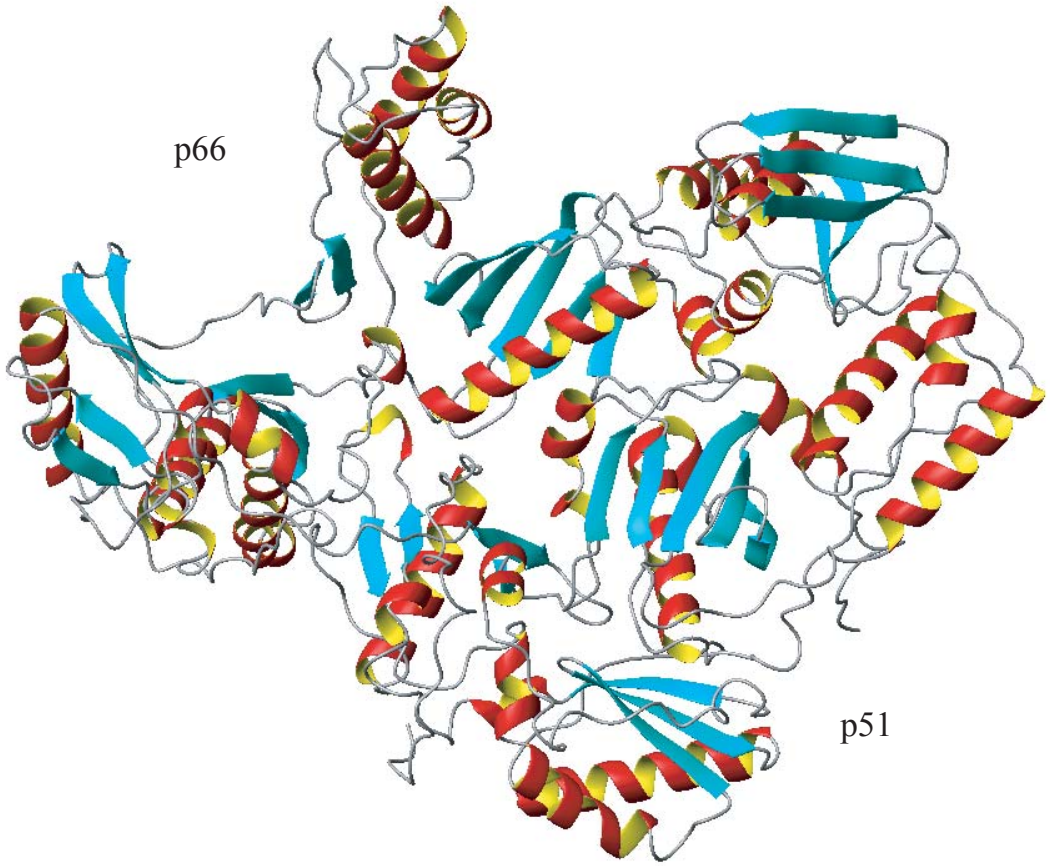


Figure 1.15: X-ray crystal structure of the HIV-1 reverse transcriptase at 2.35 Å[122]. The figure is adapted from PDB file 1RTJ using MOLMOL[21].

polymerase active site is located at the bottom of this cleft on the palm subdomain and is defined by three conserved aspartic acid residues each located on a separate antiparallel β sheet[124,125]. Binding of the nucleic acid template primer complex to RT induces the rotation of the thumb subdomain[124,133]. Further binding of a dNTP molecule to the p66 domain shifts the outer part of the fingers domain, bringing the "tip" of the finger domain closer to the palm ("closed" conformation). This shift results in a uniform distortion of the hydrophobic core of the finger domain and displacement of the outer residues of this domain by 5 Å, bringing several residues in contact with the incoming dNTP molecule[126].

The RNase H domain of p66 is folded into a five-stranded β-sheet surrounded by four α-helices (Figure 1.15)[125]. The conformation of this domain is highly dependent upon the presence of magnesium, which is required for activity[125,134]. The active site of the RNase H domain contains four highly-conserved acidic catalytic residues, located on one of the β-strands and two flanking α-helices[125]. The free and nucleic-acid bound structures of RNase H are very similar; the only significant conformational change that occurs upon nucleic acid binding is the ordering of a disordered loop that contains a highly conserved histidine[125].

Though p51 is identical in sequence to the DNA polymerase domain of p66, the tertiary fold of p51 is quite different. The fingers and connection subdomains close over the palm subdomain in such a way as to bury the catalytic aspartate residues, and the thumb subdomain is extended away from the other subdomains[127]. Together, these changes in structure render p51 inactive[114,125,127].

The nucleoside inhibitors (NRTIs and NtRTIs) are analogs of natural nucleosides or nucleotides that bind to reverse transcriptase in the active site. The NRTIs and NtRTIs function as competitive inhibitors of the dNTP substrate needed for reverse transcription. Examples of

this class of drug include didanosine, zalcitabine, stavudine, lamivudine, abacavir, and zidovudine (AZT), the last being the first drug marketed for the treatment of HIV-1. The NRTI compounds require intracellular phosphorylation by cellular kinases to yield their active, triphosphate form. They are then incorporated in the elongating chain, but since they lack a 3'-hydroxyl, they inhibit further chain elongation by RT. Once converted to their triphosphate form, all NRTIs block RT function via two mechanisms: i) they compete with dNTPs for binding of RT and ii) they cause termination of DNA synthesis. NtRTIs inhibit RT in the same manner.

Non-nucleoside inhibitors (NNRTIs) are non-competitive inhibitors of RT that bind a hydrophobic site 10 Å from the active site of the enzyme. The NNRTI-binding site lies between the β-sheets of the palm subdomain and the base of the thumb subdomain[127]. The inhibitor sits on top of a β hairpin and is held in place, in part, by interactions with the side chains of Y181 and Y188[127]. Structural studies of various NNRTIs complexed to RT show that, despite having distinct chemical structures, NNRTIs share a common binding site and mode of interaction[129,130]. It has been suggested that the molecular shape of NNRTIs resembles a butterfly with two wing sections, one being proximal and the other distal from the polymerase active site[130]. The wings make significant hydrophobic interactions with residues that line the binding site, including three conserved tyrosine residues[130]. The body of the butterfly interacts with both main chain and side chain atoms[130]. Upon NNRTI binding, the template/primer complex is repositioned on the protein, leading to the displacement of the binding groove 2 Å away from the active site and significant loss of enzyme activity[135]. Interestingly, two of the conserved tyrosines are significantly reoriented when NNRTIs bind RT in a manner that is very similar to their orientation in the inactive polymerase domain of p51[122,131]. Thus, NNRTIs distort the polymerase active site, which blocks reverse transcription.

Viral Entry Inhibitors

In March 2003, the FDA approved the first viral entry inhibitor, enfuvirtide, for clinical use[136]. Enfuvirtide is a 36-amino acid peptide that corresponds to residues 127-162 of the extracellular domain of gp41[136-139]. Binding of gp120 to both the CD4 and chemokine receptors induces the insertion of the newly exposed fusion peptide of gp120 into the host cell plasma membrane. After insertion, oligomers of gp41 fold into a 6-helix bundle consisting of 3 amino-terminal α helices (heptad repeat region-1) in the center and 3 carboxyl-terminal α helices (heptad repeat region-2) packed on the outside. Due to the formation of the 6-helix bundle, the viral and cellular membranes are brought together, enabling their fusion. Enfuvirtide mimics the sequence of heptad repeat region-2 and competitively binds to heptad repeat region-1, blocking formation of the 6-helix bundle and, therefore, membrane fusion[136-139].

Resistance to Current Antiviral Therapies

Although antiviral therapies made up of combinations of the above compounds (highly active antiviral therapy, HAART) have been relatively successful in slowing down the progression of AIDS in patients, they have also favored development of mutant viruses resistant to those very same treatments[138]. When infected individuals begin antiviral therapy, it is very likely that they already have mutant viruses that are resistant to the treatment. Those mutants exist even in patients who have never had any therapy. Thus, there is an immediate pool of resistance that has a selective replicative advantage. This resistant pool too quickly becomes the majority population and the treatment is no longer effective.

Over 30 protease inhibitor mutations have been observed in patients treated with protease inhibitor therapy[140]. Resistance mutations are seen both among residues involved in substrate

and inhibitor binding as well as at sites distant from the PR active site. When the mutated residue does contact the substrate or inhibitor, PR resistance is due to the disruption of interactions necessary for binding and positioning the inhibitor[141,142]. The mechanism of resistance for residues distal to the active sites is not entirely clear. The crystal structure of a PR mutant bound to saquinavir suggested that resistance of this mutant is caused by the formation of new, shorter van der Waals contacts between the mutant and catalytic residues that limit the flexibility of the active site such that the inhibitor cannot bind[140]. The crystal structure of the same mutant bound to indinavir supports this mechanism[142]. These mutations still allow binding of the substrate, but binding is typically reduced compared to wild type enzyme. Interestingly, secondary mutations are often observed in protease inhibitor resistant strains among residues of the various PR cleavage sites to accommodate structural changes that allow gag-pro-pol cleavage and exclude the inhibitor[143,144].

In general, only one amino acid change is required to confer resistance to a single NRTI. Many of the residues that cause NRTI resistance are localized in proximity to the dNTP binding site in the "closed" conformation of the RT. These resistance mutations affect either the binding affinity of the NRTI or its positioning on RT, relieving the inhibition while still allowing DNA synthesis. For example, when Lys_{59} is converted to arginine, a mutation that is involved in resistance to several NRTIs, the placement of the NRTI is altered, which leads to suboptimal arrangement for pairing with the coding nucleic acid strand. Since this is combined with the lack of the 3'-OH positioning interactions with the NRTI, the mutation leads to fourfold lower affinity of RT for the NRTI over the dNTP[145]. Thus, resistance is conferred by the loss of RT affinity for the NRTI.

Resistance mutations for most NNRTIs occur proximal to their binding site, resulting in steric hindrance that blocks the compounds from binding. The most frequent NNRTI resistance mutation, conversion of Lys_{103} to asparagine, changes the structure of the entrance to the NNRTI binding pocket, causes loss of hydrophobic interactions, and alters its electrostatic environment[132,146,147]. Also, a new hydrogen bond is formed by this mutation that stabilizes the unliganded RT compared to the wild-type enzyme[132]. Some NNRTI resistance mutations exert their effect via an indirect mechanism. For example, mutation of Val_{108} to isoleucine causes positional shifts of many residues in the binding pocket, including a key NNRTI-binding residue, that moves by >8 Å[148]. Therefore, this mutation leads to loss of NNRTI binding because it displaces residues that interact with the drugs.

Compounds Designed to Target NCp7

The build-up of antiviral resistant HIV-1 strains necessitates new targets for therapy. The zinc-binding domain structure and mutationally non-permissive nature of NCp7 make it a favorable possibility. Several studies have already been performed to investigate the activity of compounds that inhibit the activity of NCp7. Most molecules tested to date eject zinc from NCp7 (zinc ejectors), destroying the structure of the protein and preventing it from functioning. Initial work was done with 3-nitrosobenzamide (NOBA) compounds (Figure 1.16A). NOBA compounds were able to eject 83% of the zinc from NCp7 *in vitro* and showed inhibition of HIV-1 replication in cell culture[149,150]. However, the NOBA compounds did not show any selectivity, since they reacted with cellular zinc-binding proteins as well as with NCp7. Thus, the NOBA compounds were highly toxic.

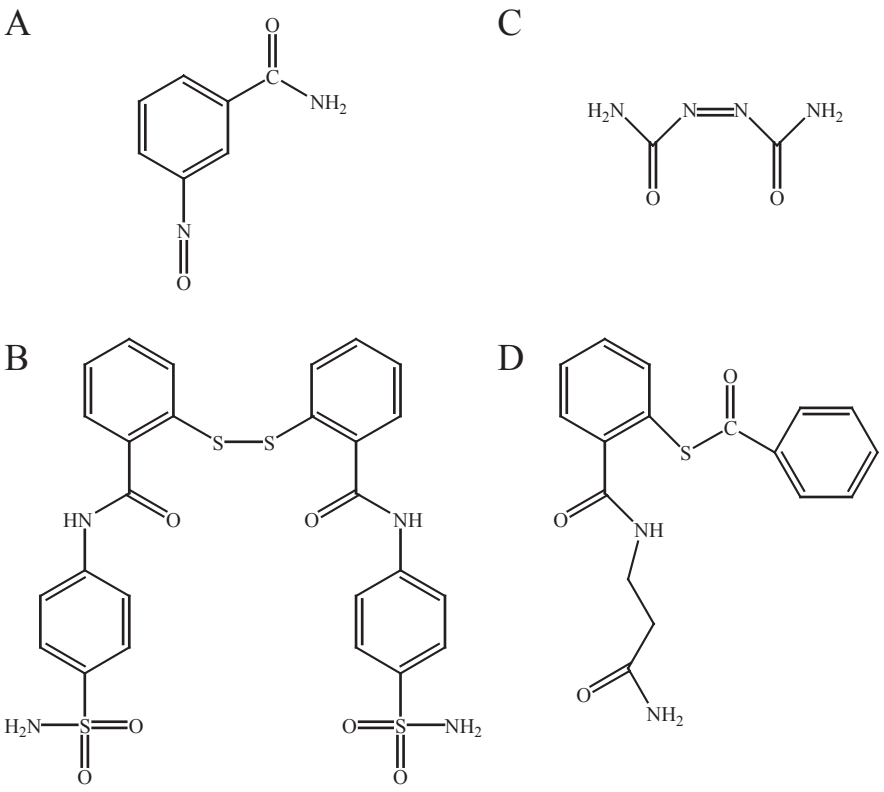


Figure 1.16: Chemical structure of representative members of the classes of zinc ejectors designed to target NCp7: A) NOBA; B) DIBA; C) ADA; D) 2-mercaptobenzamide thioester compounds.

Further studies led to the development of disulfide benzamide (DIBA) and azodicarbonamide (ADA) compounds (Figure 1.16, B and C)[151,152]. DIBA compounds are highly active against NCp7, both *in vitro* and in cell culture experiments[153-161]. They have been shown to eject zinc from NCp7 and disrupt the ability of NCp7 to bind RNA[154]. In addition, ZD2 was found to be more reactive with the DIBA compounds than ZD1[156]. The DIBA compounds have also been shown to rapidly cyclize in aqueous conditions at neutral pH, i.e. under physiological conditions, forming a benzisothiazolone and a free thiol[157]. The benzisothiazolone can eject zinc *in vitro*, and, although the free thiol does not eject zinc, it has antiviral activity in cell culture experiments[157].

Mass spectrometry studies demonstrated that the DIBA compounds covalently modify NCp7 on Cys_{36} and Cys_{49}[156]. This led to the proposal of a DIBA reaction mechanism in which the zinc-coordinated cysteine residue makes a relatively slow nucleophilic attack on a DIBA sulfur. This results in destabilization of the zinc coordination, followed by formation of new disulfide bonds and loss of zinc[156]. A similar mechanism has been proposed for ADA compounds, although in their case the nucleophilic attack of the NCp7 cysteine residue is directed on the central nitrogen double bond[162].

Neither the DIBA nor the ADA compounds inhibit the functional activity of cellular proteins containing zinc-binding domains, including poly(ADP ribose) polymerase and the Sp1 and GATA-1 transcription factors[160]. Poly(ADP ribose) polymerase has the same zinc-coordination motif as NCp7, but has many more residues between the second cysteine and the histidine resulting in longer loops than NCp7. GATA-1 contains two Cys_4 zinc-binding domains and Sp1 two Cys_2His_2 domains. In *in vitro* assays, DIBA-1 and ADA compounds did not block the poly(ADP ribose) polymerase extension activity or inhibit DNA binding and transcription by

GATA-1 and Sp1[160]. It is important to note, though, that in these assays, the protein-substrate complex was generally preformed before addition of the zinc ejector[160]. Thus, these compounds do not affect the function of poly(ADP ribose) polymerase, GATA-1, and Sp1 when the substrate is already bound.

Despite their selectivity, the DIBA compounds contain a disulfide bond that is easily reduced and this severely limits the half-life of this class of compound[151,154,163]. Thus, new antiviral compounds that target NCp7 are needed that can not be easily reduced in the cell. Among the most recent zinc-ejecting compounds to be developed are the 2-mercaptobenzamide thioesters[164]. In this class of compound, the reducible disulfide bond has been replaced with a more stable thioester linkage. These thioester compounds have high antiviral activity and low cellular toxicity. They are water soluble and have been shown to eject zinc from NCp7[164,165]. Initial studies demonstrated that silver ions had to be added to the *in vitro* reaction to "activate" the thioester compounds; no zinc ejection was observed with either silver ions alone or the thioester compounds alone[166]. Mass spectrometry revealed that these thioester compounds selectively modified NCp7 in ZD2 as was seen with the DIBA compounds[166]. These thioester compounds were also shown to have no effect on DNA binding by Sp1, suggesting that they might not target other zinc-binding domains in cellular proteins[164]. Recently, the *in vivo* antiviral activity of two 2-mercaptobenzamide thioesters, compounds **1** and **2** (Figure 1.16D) was described[167]. These compounds were shown to have antiviral activity in a murine model[167]. Orally administered compound **2** actually evoked an antiviral response similar to that of the protease inhibitor indinavir[167]. It is postulated that these new thioester compounds have antiviral activity because they eject zinc from NCp7, however the exact mechanism of zinc-ejection has

not been established. Understanding how these thioester compounds function is crucial to the design of improved compounds to target NCp7.

Research Goals

NCp7 is a small, highly basic protein that contains two highly conserved zinc-binding domains. It has many roles throughout the HIV-1 lifecycle as a nucleic acid binding protein, from early stages where it is involved in reverse transcription to late stages where it plays an essential role in genomic RNA dimerization and packaging. NCp7 is mutationally non-permissive; mutations in this protein cause a loss of viral infectivity. The zinc-binding domains of NCp7 have become new targets for antiviral therapies. As the zinc-binding domains are absolutely required for viral infectivity, it is possible that their disruption will be very difficult for the virus to circumvent. Thus, resistance mutants may be very slow to arise if they arise at all. Several 2-mercaptobenzamide thioester compounds have been developed that seem to selectively target NCp7 over other zinc-binding proteins. However, better understanding of the mechanism and specificity of these zinc ejectors are required for them to be made into viable antiviral drugs. In addition, much about the viral functions of NCp7 is still unknown. How does NCp7 facilitate the conversion of the kissing-loop dimer to the extended duplex? What is the role of NCp7 interaction with the internal loop of SL1? There is still a great deal to learn about the NCp7 before it can be effectively used to stop the devastation of HIV-1 and AIDS.

The research to be presented in the following chapters has three primary goals. First, the mechanism of action of the 2-mercaptobenzamide thioesters with NCp7 was investigated using UV/Visible spectroscopy, NMR spectroscopy, mass spectrometry, and gel mobility shift assay. These experiments were performed to determine if and how compounds **1** and **2** interact with

NCp7, and then to analyze any effect of RNA binding on these interactions. Knowledge of the mechanism of the thioester compounds is essential if these compounds are to be used to treat HIV-1. The second goal was to understand the effect of the 2-mercaptobenzamide thioesters on cellular zinc-binding proteins. UV/Visible spectroscopy, NMR spectroscopy, and gel mobility shift assay were used to analyze the effect of compounds **1** and **2** on four different types of zinc-binding domains. This study investigates the reactivity of structural zinc-binding proteins, which is important for improved antiviral drug design. Finally, the effect of NCp7 binding on the internal loop of SL1 was studied by gel mobility shift assay and NMR spectroscopy. As this loop has been shown to be important for genomic dimerization, it was important to study its interaction with NCp7. These studies have suggested a role for the internal loop in the dimerization reaction and the function of NCp7 binding to it. This report represents the first structural study of the binding of NCp7 to SL1.

Abbreviations:

ADA, azodicarbonamide; AIDS, acquired immunodeficiency syndrome; CA, capsid; DIBA, disulfide benzamide; DIS, dimerization initiation site; DLS, dimer linkage structure; dNTP, deoxynucleotide triphosphate; HIV-1, human immunodeficiency virus Type-1; IN, integrase; MA, matrix; NCp7, nucleocapsid protein; NEM, N-ethylmaleimide; NNRTI, non-nucleoside reverse transcriptase inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; NOBA, 3-nitrosobenzamide; PBS, primer binding site; PI, protease inhibitor; PIC, preintegration complex; PPT, polypurine tract; PR, protease; RT, reverse transcriptase; -ssDNA, minus-strand strong-stop DNA; TAR, transactivation response element; ZD1, amino-terminal zinc-binding domain; ZD2, carboxyl-terminal zinc-binding domain.

References:

(1) Ho, D. D., Huang, Y. The HIV-1 Vaccine Race. *Cell* **2002***, 110*, 135-138.

(2) Thomson, M. M., Perez-Alvarex, L., Najera, R. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infectious Diseases* **2002***, 2*, 461-471.

(3) Turner, B. G. a. M. F. S. Structural Biology of HIV. *Journal of Molecular Biology* **1999***, 285*, 1-32.

(4) Coffin, J. M.; Hughes, S. H.; Varmus, H. E. *Retroviruses*; Cold Spring Harbor Laboratory Press, 1997; 843.

(5) Freed, E. O. HIV-1 Gag Proteins: Diverse Functions in the Virus Life Cycle. *Virology* **1998***, 251*, 1-15.

(6) Frankel, A. D., Young, J.A.T. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **1998***, 67*, 1-25.

(7) Magyar, J. S.; Godwin, H. A. Spectropotentiometric analysis of metal binding to structural zinc-binding sites: accounting quantitatively for pH and metal ion buffering effects. *Anal. Biochem.* **2003***, 320*, 39-54.

(8) Mely, Y.; De Rocquigny, H.; Morellet, N.; Roques, B. P.; Gerard, D. Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy. *Biochemistry* **1996***, 35*, 5175-5182.

(9) Bombarda, E.; Morellet, N.; Cherradi, H.; Spiess, B.; Bouaziz, S. et al. Determination of the pK_a of the four Zn^{2+}-coordinating residues of the distal finger motif of the HIV-1

nucleocapsid protein: consequences on the binding of Zn^{2+}. *J. Mol. Biol.* **2001***, 310*, 659-672.

(10) Omichinski, J. G., Clore, G.M., Sakaguchi, K., Appella, E., Gronenborn, A.M. Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy. *FEBS Lett.* **1991***, 292*, 25-30.

(11) Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J., Roques, B.P. Determination of the structure of the nucleocapsid protein from the human immunodeficiency virus type 1 by ^{1}H NMR. *EMBO J.* **1992***, 11*, 3059-3065.

(12) South, T. L.; Summers, M. F. Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci.* **1993***, 2*, 3-19.

(13) Summers, M. F., South, T.L., Kim, B., Hare, D.R. High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* **1990***, 29*, 329-340.

(14) South, T. L.; Blake, P. R.; Hare, D. R.; Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **1991***, 30*, 6342-6349.

(15) Demene, H.; Dong, C. Z.; Ottmann, M.; Rouyez, M. C.; Jullian, N. et al. ^{1}H NMR structure and biological studies of the His23-->Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity. *Biochemistry* **1994***, 33*, 11707-11716.

(16) Lee, B. M.; De Guzman, R. N.; Turner, B. G.; Tjandra, N.; Summers, M. F. Dynamical behavior of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **1998***, 279*, 633-649.

(17) Gorelick, R. J.; Chabot, D. J.; Rein, A.; Henderson, L. E.; Arthur, L. O. The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J. Virol.* **1993***, 67*, 4027-4036.

(18) Tanchou, V.; Decimo, D.; Pechoux, C.; Lener, D.; Rogemond, 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*, 4442-4447.

(19) Jullian, N.; Demene, H.; Morellet, N.; Maigret, B.; Roques, B. P. Replacement of His23 by Cys in a zinc finger of HIV-1 NCp7 led to a change in ^{1}H NMR-derived 3D structure and to a loss of biological activity. *FEBS Lett.* **1993***, 331*, 43-48.

(20) Ramboarina, S.; Morellet, N.; Fournie-Zaluski, M. C.; Roques, B. P. Structural investigation on the requirement of CCHH zinc finger type in nucleocapsid protein of human immunodeficiency virus 1. *Biochemistry* **1999***, 38*, 9600-9607.

(21) Koradi, R.; Billeter, M.; Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **1996***, 14*, 51-55.

(22) Hargittai, M. R.; Gorelick, R. J.; Rouzina, I.; Musier-Forsyth, K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.* **2004***, 337*, 951-968.

(23) Al-Ghusein, H.; Ball, H.; Igloi, G. L.; Gbewonyo, A.; Coates, A. R. et al. Chemically synthesised human immunodeficiency virus P7 nucleocapsid protein can self-assemble into particles and binds to a specific site on the tRNA(Lys,3) primer. *Biochem. Biophys. Res. Commun.* **1996***, 224*, 191-198.

(24) Gregoire, C. J.; Gautheret, D.; Loret, E. P. No $tRNA_3^{Lys}$ unwinding in a complex with HIV NCp7. *J. Biol. Chem.* **1997***, 272*, 25143-25148.

(25) Remy, E.; de Rocquigny, H.; Petitjean, P.; Muriaux, D.; Theilleux, V. et al. The annealing of $tRNA^3_{Lys}$ to human immunodeficiency virus type 1 primer binding site is critically dependent on the NCp7 zinc fingers structure. *J. Biol. Chem.* **1998***, 273*, 4819-4822.

(26) Huang, Y.; Khorchid, A.; Gabor, J.; Wang, J.; Li, X. et al. The role of nucleocapsid and U5 stem/A-rich loop sequences in $tRNA^3_{Lys}$ genomic placement and initiation of reverse transcription in human immunodeficiency virus type 1. *J. Virol.* **1998***, 72*, 3907-3915.

(27) Chan, B.; Weidemaier, K.; Yip, W. T.; Barbara, P. F.; Musier-Forsyth, K. Intra-tRNA distance measurements for nucleocapsid protein dependent tRNA unwinding during priming of HIV reverse transcription. *Proc. Natl. Acad. Sci. USA* **1999***, 96*, 459-464.

(28) Cen, S.; Huang, Y.; Khorchid, A.; Darlix, J. L.; Wainberg, M. A. et al. The role of Pr55(gag) in the annealing of tRNA3Lys to human immunodeficiency virus type 1 genomic RNA. *J Virol* **1999***, 73*, 4485-4488.

(29) Cen, S.; Khorchid, A.; Gabor, J.; Rong, L.; Wainberg, M. A. et al. Roles of $Pr55_{gag}$ and NCp7 in $tRNA_{Lys}^3$ genomic placement and the initiation step of reverse transcription in human immunodeficiency virus type 1. *J. Virol.* **2000***, 74*, 10796-10800.

(30) Hargittai, M. R.; Mangla, A. T.; Gorelick, R. J.; Musier-Forsyth, K. HIV-1 nucleocapsid protein zinc finger structures induce $tRNA_{Lys,3}$ structural changes but are not critical for primer/template annealing. *J. Mol. Biol.* **2001***, 312*, 985-997.

(31) Tisne, C.; Roques, B. P.; Dardel, F. Heteronuclear NMR studies of the interaction of $tRNA_{Lys}^3$ with HIV-1 nucleocapsid protein. *J. Mol. Biol.* **2001***, 306*, 443-454.

(32) Brule, F.; Marquet, R.; Rong, L.; Wainberg, M. A.; Roques, B. P. et al. Structural and functional properties of the HIV-1 RNA-$tRNA_{Lys}^{3}$ primer complex annealed by the nucleocapsid protein: comparison with the heat-annealed complex. *RNA* **2002***, 8*, 8-15.

(33) Tisne, C.; Roques, B. P.; Dardel, F. Specific recognition of primer tRNA Lys 3 by HIV-1 nucleocapsid protein: involvement of the zinc fingers and the N-terminal basic extension. *Biochimie* **2003***, 85*, 557-561.

(34) Tisne, C.; Roques, B. P.; Dardel, F. The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome. *J. Biol. Chem.* **2004***, 279*, 3588-3595.

(35) You, J. C.; McHenry, C. S. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J. Biol. Chem.* **1994***, 269*, 31491-31495.

(36) Tsuchihashi, Z.; Brown, P. O. DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **1994***, 68*, 5863-5870.

(37) Rodriguez-Rodriguez, L.; Tsuchihashi, Z.; Fuentes, G. M.; Bambara, R. A.; Fay, P. J. Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro. *J. Biol. Chem.* **1995***, 270*, 15005-15011.

(38) DeStefano, J. J. Interaction of human immunodeficiency virus nucleocapsid protein with a structure mimicking a replication intermediate. Effects on stability, reverse transcriptase binding, and strand transfer. *J. Biol. Chem.* **1996***, 271*, 16350-16356.

(39) Kim, J. K.; Palaniappan, C.; Wu, W.; Fay, P. J.; Bambara, R. A. Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. *J. Biol. Chem.* **1997***, 272*, 16769-16777.

(40) Negroni, M.; Buc, H. Recombination during reverse transcription: an evaluation of the role of the nucleocapsid protein. *J. Mol. Biol.* **1999***, 286*, 15-31.

(41) Hsu, M.; Rong, L.; de Rocquigny, H.; Roques, B. P.; Wainberg, M. A. The effect of mutations in the HIV-1 nucleocapsid protein on strand transfer in cell-free reverse transcription reactions. *Nucleic Acids Res.* **2000***, 28*, 1724-1729.

(42) Johnson, P. E.; Turner, R. B.; Wu, Z. R.; Hairston, L.; Guo, J. et al. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry* **2000***, 39*, 9084-9091.

(43) Guo, J.; Wu, T.; Anderson, J.; Kane, B. F.; Johnson, D. G. et al. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.* **2000***, 74*, 8980-8988.

(44) Guo, J.; Wu, T.; Kane, B. F.; Johnson, D. G.; Henderson, L. E. et al. Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **2002***, 76*, 4370-4378.

(45) Balakrishnan, M.; Roques, B. P.; Fay, P. J.; Bambara, R. A. Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis. *J. Virol.* **2003***, 77*, 4710-4721.

(46) Bernacchi, S.; Stoylov, S.; Piemont, E.; Ficheux, D.; Roques, B. P. et al. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.* **2002***, 317*, 385-399.

(47) Hong, M. K.; Harbron, E. J.; O'Connor, D. B.; Guo, J.; Barbara, P. F. et al. Nucleic acid conformational changes essential for HIV-1 nucleocapsid protein-mediated inhibition of self-priming in minus-strand transfer. *J. Mol. Biol.* **2003***, 325*, 1-10.

(48) Azoulay, J.; Clamme, J. P.; Darlix, J. L.; Roques, B. P.; Mely, Y. Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J. Mol. Biol.* **2003***, 326*, 691-700.

(49) Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J. P.; Ficheux, D. et al. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **2003***, 328*, 95-108.

(50) Beltz, H.; Piemont, E.; Schaub, E.; Ficheux, D.; Roques, B. et al. Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **2004***, 338*, 711-723.

(51) Guo, J.; Henderson, L. E.; Bess, J.; Kane, B.; Levin, J. G. Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol.* **1997***, 71*, 5178-5188.

(52) Lee, N.; Gorelick, R. J.; Musier-Forsyth, K. Zinc finger-dependent HIV-1 nucleocapsid protein-TAR RNA interactions. *Nucleic Acids Res.* **2003***, 31*, 4847-4855.

(53) Carteau, S.; Batson, S. C.; Poljak, L.; Mouscadet, J. F.; de Rocquigny, H. et al. Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg2+-dependent DNA integration in vitro. *J. Virol.* **1997***, 71*, 6225-6229.

(54) Carteau, S.; Gorelick, R. J.; Bushman, F. D. Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J. Virol.* **1999**, *73*, 6670-6679.

(55) Buckman, J. S.; Bosche, W. J.; Gorelick, R. J. Human immunodeficiency virus type 1 nucleocapsid Zn^{2+} fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *J. Virol.* **2003**, *77*, 1469-1480.

(56) Gao, K.; Gorelick, R. J.; Johnson, D. G.; Bushman, F. Cofactors for human immunodeficiency virus type 1 cDNA integration in vitro. *J. Virol.* **2003**, *77*, 1598-1603.

(57) Sandefur, S.; Smith, R. M.; Varthakavi, V.; Spearman, P. Mapping and characterization of the N-terminal I domain of human immunodeficiency virus type 1 Pr55(Gag). *J. Virol.* **2000**, *74*, 7238-7249.

(58) Bennett, R. P.; Nelle, T. D.; Wills, J. W. Functional chimeras of the Rous Sarcoma Virus and Human Immunodeficiency Virus gag proteins. *J. Virol.* **1993**, *67*, 6487-6498.

(59) Sandefur, S.; Varthakavi, V.; Spearman, P. The I domain is required for efficient plasma membrane binding of human immunodeficiency virus type 1 Pr55Gag. *J. Virol.* **1998**, *72*, 2723-2732.

(60) Dawson, L.; Yu, X. F. The role of nucleocapsid of HIV-1 in virus assembly. *Virology* **1998**, *251*, 141-157.

(61) Cimarelli, A., Sandin, S., Hoglund, S., Luban, J. Basic residues in human immunodeficiency virus type 1 nucleocapsid protein promote virion assembly via interaction with RNA. *J. Virol.* **2000**, *74*, 3046-3057.

(62) Cimarelli, A.; Luban, J. Human immunodeficiency virus type 1 virion density is not determined by nucleocapsid basic residues. *J. Virol.* **2000***, 74*, 6734-6740.

(63) Zabransky, A.; Hunter, E.; Sakalian, M. Identification of a minimal HIV-1 gag domain sufficient for self-association. *Virology* **2002***, 294*, 141-150.

(64) Zhang, Y.; Qian, H.; Love, Z.; Barklis, E. Analysis of the assembly function of the human immunodeficiency virus type 1 gag protein nucleocapsid domain. *J. Virol.* **1998***, 72*, 1782-1789.

(65) Burniston, M. T.; Cimarelli, A.; Colgan, J.; Curtis, S. P.; Luban, J. Human immunodeficiency virus type 1 Gag polyprotein multimerization requires the nucleocapsid domain and RNA and is promoted by the capsid-dimer interface and the basic region of matrix protein. *J. Virol.* **1999***, 73*, 8527-8540.

(66) Feng, Y. X.; Li, T.; Campbell, S.; Rein, A. Reversible binding of recombinant human immunodeficiency virus type 1 gag protein to nucleic acids in virus-like particle assembly in vitro. *J. Virol.* **2002***, 76*, 11757-11762.

(67) Clever, J.; Sassetti, C.; Parslow, T. G. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J. Virol.* **1995***, 69*, 2101-2109.

(68) Berkowitz, R. D.; Luban, J.; Goff, S. P. Specific binding of human immunodeficiency virus type 1 *gag* polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J. Virol.* **1993***, 67*, 7190-7200.

(69) Shubsda, M. F.; Paoletti, A. C.; Hudson, B. S.; Borer, P. N. Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry* **2002***, 41*, 5276-5282.

(70) Pappalardo, L.; Kerwood, D. J.; Pelczer, I.; Borer, P. N. Three-dimensional folding of an RNA hairpin required for packaging HIV-1. *J. Mol. Biol.* **1998***, 282*, 801-818.

(71) Hayashi, T., Shioda, T., Iwakura, Y., Shibuta, H. RNA packaging signal of human immunodeficiency virus type 1. *Virology* **1992***, 188*, 590-599.

(72) De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N. et al. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **1998***, 279*, 384-388.

(73) Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R. et al. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* **2000***, 301*, 491-511.

(74) Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Summers, M. F. NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform. *J. Mol. Biol.* **2000***, 299*, 145-156.

(75) O'Reilly, M. M., McNally, M.T., Beemon, K.L. Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology* **1995***, 213*, 373-385.

(76) Amarasinghe, G. K.; Zhou, J.; Miskimon, M.; Chancellor, K. J.; McDonald, J. A. et al. Stem-loop SL4 of the HIV-1 Ψ RNA packaging signal exhibits weak affinity for the nucleocapsid protein. Structural studies and implications for genome recognition. *J. Mol. Biol.* **2001***, 314*, 961-970.

(77) Kerwood, D. J.; Cavaluzzi, M. J.; Borer, P. N. Structure of SL4 RNA from the HIV-1 packaging signal. *Biochemistry* **2001***, 40*, 14518-14529.

(78) Heus, H. A.; Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **1991***, 253*, 191-194.

(79) Kung, H. G.; Hu, S.; Bender, W.; Bailey, J. M.; Davidson, N. et al. RD-114, baboon, and woolly monkey viral RNA's compared in size and structure. *Cell* **1976***, 7*, 609-620.

(80) Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B., Ehresmann,C. Identification of the primary site of the HIV-1 RNA dimerization in vitro. *Proceedings of the National Academy of the USA* **1994***, 91*, 4945-4949.

(81) Paillart, J. C.; Skripkin, E.; Ehresmann, B.; Ehresmann, C.; Marquet, R. The use of chemical modification interference and inverse PCR mutagenesis to identify the dimerization initiation site of HIV-1 genomic RNA. *Pharm. Acta Helv.* **1996***, 71*, 21-28.

(82) Paillart, J. C.; Skripkin, E.; Ehresmann, B.; Ehresmann, C.; Marquet, R. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. USA* **1996***, 93*, 5572-5577.

(83) Laughrea, M.; Jette, L. 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*, 1589-1598.

(84) Paillart, J. C.; Westhof, E.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J. Mol. Biol.* **1997***, 270*, 36-49.

(85) Lodmell, J. S.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Convergence of natural and artificial evolution on an RNA loop-loop interaction: the HIV-1 dimerization initiation site. *RNA* **2000***, 6*, 1267-1276.

(86) Clever, J. L.; Wong, M. L.; Parslow, T. G. Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J. Virol.* **1996***, 70*, 5902-5908.

(87) Laughrea, M.; Jette, L.; Mak, J.; Kleiman, L.; Liang, C. 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*, 3397-3406.

(88) Kieken, F.; Arnoult, E.; Barbault, F.; Paquet, F.; Huynh-Dinh, T. et al. HIV-1 Lai genomic RNA: combined used of NMR and molecular dynamics simulation for studying the structure and internal dynamics of a mutated SL1 hairpin. *Eur. Biophys. J.* **2002***, 31*, 521-531.

(89) Lawrence, D. C.; Stover, C. C.; Noznitsky, J.; Wu, Z. R.; Summers, M. F. Structure of the intact stem and bulge of HIV-1 Ψ-RNA stem-loop SL1. *J. Mol. Biol.* **2003***, 326*, 529-542.

(90) Mujeeb, A., Clever, J.L., Billeci, T.M., James, T.L., Parslow, T.G. Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat. Struct. Biol.* **1998***, 5*, 432-436.

(91) Dardel, F.; Marquet, R.; Ehresmann, C.; Ehresmann, B.; Blanquet, S. Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res.* **1998***, 26*, 3567-3571.

(92) Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., Dumas, P Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.* **2001***, 8*, 1064-1068.

(93) Ennifar, E.; Yusupov, M.; Walter, P.; Marquet, R.; Ehresmann, B. 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*, 1439-1449.

(94) Aci, S.; Gangneux, L.; Paoletti, J.; Genest, D. On the stability of different experimental dimeric structures of the SL1 sequence from the genomic RNA of HIV-1 solution: a molecular dynamics simulation and electrophoresis study. *Biopolymers* **2004**, *74*, 177-188.

(95) Mujeeb, A., Parslow, T.G., Zarrinpar, A., Das, C., James, T.L. NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett.* **1999**, *458*, 387-392.

(96) Poon, D. T.; Wu, J.; Aldovini, A. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J. Virol.* **1996**, *70*, 6607-6616.

(97) Feng, Y. X.; Copeland, T. D.; Henderson, L. E.; Gorelick, R. J.; Bosche, W. J. et al. HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA *in vitro*. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7577-7581.

(98) Muriaux, D.; de Rocquigny, H.; Roques, B. P.; Paoletti, J. NCp7 activates HIV-1 Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J. Biol. Chem.* **1996**, *271*, 33686-33692.

(99) Takahashi, K. I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H., Kawai, G. Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA* **2000**, *6*, 96-102.

(100) Rist, M. J.; Marino, J. P. Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* **2002**, *41*, 14762-14770.

(101) Takahashi, K. I.; Baba, S.; Koyanagi, Y.; Yamamoto, N.; Takaku, H. 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*, 31274-31278.

(102) Laughrea, M.; Shen, N.; Jette, L.; Darlix, J. L.; Kleiman, L. et al. Role of distal zinc finger of nucleocapsid protein in genomic RNA dimerization of human immunodeficiency virus type 1; no role for the palindrome crowning the R-U5 hairpin. *Virology* **2001**, *281*, 109-116.

(103) Baba, S.; Takahashi, K. I.; Koyanagi, Y.; Yamamoto, N.; Takaku, H. et al. Role of the zinc fingers of HIV-1 nucleocapsid protein in maturation of genomic RNA. *J. Biochem.* **2003**, *134*, 637-639.

(104) Mihailescu, M. R.; Marino, J. P. A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1189-1194.

(105) *HIV sequence compendium*; Theoretical Biology and Biophysics, 2002.

(106) Takahashi, K. I.; Baba, S.; Chattopadhyay, P.; Koyanagi, Y.; Yamamoto, N. et al. Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA* **2000**, *6*, 96-102.

(107) Greatorex, J.; Gallego, J.; Varani, G.; Lever, A. Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal. *J. Mol. Biol.* **2002**, *322*, 543-557.

(108) Laughrea, M.; Shen, N.; Jette, L.; Wainberg, M. A. Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: role of stem-loop B in HIV replication and HIV RNA dimerization. *Biochemistry* **1999**, *38*, 226-234.

(109) Shen, N.; Jette, L.; Liang, C.; Wainberg, M. A.; Laughrea, M. Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging. *J. Virol.* **2000**, *74*, 5729-5735.

(110) Yuan, Y. Q.; Kerwood, D. J.; Paoletti, A. C.; Shubsda, M. F.; Borer, P. N. Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop. *Biochemistry* **2003**, *42*, 5259-5269.

(111) Tozser, J.; Blaha, I.; Copeland, T. D.; Wondrak, E. M.; Oroszlan, S. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Lett.* **1991**, *281*, 77-80.

(112) Navia, M. A.; Fitzgerald, P. M.; McKeever, B. M.; Leu, C. T.; Heimbach, J. C. et al. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989**, *337*, 615-620.

(113) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E. et al. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **1989**, *245*, 616-621.

(114) Turner, B. G.; Summers, M. F. Structural biology of HIV. *J. Mol. Biol.* **1999**, *285*, 1-32.

(115) De Clercq, E. New anti-HIV agents and targets. *Med. Res. Rev.* **2002**, *22*, 531-565.

(116) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R. et al. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* **1989***, 246*, 1149-1152.

(117) Miller, M.; Geller, M.; Gribskov, M.; Kent, S. B. Analysis of the structure of chemically synthesized HIV-1 protease complexed with a hexapeptide inhibitor. Part I: crystallographic refinement of 2 Å data. *Proteins* **1997***, 27*, 184-194.

(118) Fitzgerald, P. M.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C. et al. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0 Å resolution. *J. Biol. Chem.* **1990***, 265*, 14209-14219.

(119) Dohnalek, J.; Hasek, J.; Duskova, J.; Petrokova, H.; Hradilek, M. et al. A distinct binding mode of a hydroxyethylamine isostere inhibitor of HIV-1 protease. *Acta Crystallogr. D Biol. Crystallogr.* **2001***, 57*, 472-476.

(120) Dohnalek, J.; Hasek, J.; Duskova, J.; Petrokova, H.; Hradilek, M. et al. Hydroxyethylamine isostere of an HIV-1 protease inhibitor prefers its amine to the hydroxy group in binding to catalytic aspartates. A synchrotron study of HIV-1 protease in complex with a peptidomimetic inhibitor. *J. Med. Chem.* **2002***, 45*, 1432-1438.

(121) Jonckheere, H.; Anne, J.; De Clercq, E. The HIV-1 reverse transcription (RT) process as target for RT inhibitors. *Med. Res. Rev.* **2000***, 20*, 129-154.

(122) Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stammers, D. et al. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat. Struct. Biol.* **1995***, 2*, 303-308.

(123) Arnold, E.; Jacobo-Molina, A.; Nanni, R. G.; Williams, R. L.; Lu, X. et al. Structure of HIV-1 reverse transcriptase/DNA complex at 7 Å resolution showing active site locations. *Nature* **1992**, *357*, 85-89.

(124) Jacobo-Molina, A.; Ding, J.; Nanni, R. G.; Clark, A. D., Jr.; Lu, X. et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6320-6324.

(125) Ding, J.; Hughes, S. H.; Arnold, E. Protein-nucleic acid interactions and DNA conformation in a complex of human immunodeficiency virus type 1 reverse transcriptase with a double-stranded DNA template-primer. *Biopolymers* **1997**, *44*, 125-138.

(126) Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **1998**, *282*, 1669-1675.

(127) Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **1992**, *256*, 1783-1790.

(128) Smerdon, S. J.; Jager, J.; Wang, J.; Kohlstaedt, L. A.; Chirino, A. J. et al. Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3911-3915.

(129) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C. et al. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*, 293-302.

(130) Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K. et al. Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.* **1995***, 2*, 407-415.

(131) Hopkins, A. L.; Ren, J.; Esnouf, R. M.; Willcox, B. E.; Jones, E. Y. et al. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.* **1996***, 39*, 1589-1600.

(132) Ren, J.; Milton, J.; Weaver, K. L.; Short, S. A.; Stuart, D. I. et al. Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Structure Fold. Des.* **2000***, 8*, 1089-1094.

(133) Ding, J.; Das, K.; Hsiou, Y.; Sarafianos, S. G.; Clark, A. D., Jr. et al. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **1998***, 284*, 1095-1111.

(134) Pari, K.; Mueller, G. A.; DeRose, E. F.; Kirby, T. W.; London, R. E. Solution structure of the RNase H domain of the HIV-1 reverse transcriptase in the presence of magnesium. *Biochemistry* **2003***, 42*, 639-650.

(135) Balzarini, J. Suppression of resistance to drugs targeted to human immunodeficiency virus reverse transcriptase by combination therapy. *Biochem. Pharmacol.* **1999***, 58*, 1-27.

(136) Fung, H. B.; Guo, Y. Enfuvirtide: a fusion inhibitor for the treatment of HIV infection. *Clin. Ther.* **2004***, 26*, 352-378.

(137) Cervia, J. S.; Smith, M. A. Enfuvirtide (T-20): a novel human immunodeficiency virus type 1 fusion inhibitor. *Clin. Infect. Dis.* **2003***, 37*, 1102-1106.

(138) De Clercq, E. HIV-chemotherapy and -prophylaxis: new drugs, leads and approaches. *Int. J. Biochem. Cell Biol.* **2004***, 36*, 1800-1822.

(139) Hardy, H.; Skolnik, P. R. Enfuvirtide, a new fusion inhibitor for therapy of human immunodeficiency virus infection. *Pharmacotherapy* **2004***, 24*, 198-211.

(140) Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance. *Protein Sci.* **2000***, 9*, 1898-1904.

(141) Sham, H. L.; Kempf, D. J.; Molla, A.; Marsh, K. C.; Kumar, G. N. et al. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* **1998***, 42*, 3218-3224.

(142) Mahalingam, B.; Wang, Y. F.; Boross, P. I.; Tozser, J.; Louis, J. M. et al. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004***, 271*, 1516-1524.

(143) Cote, H. C.; Brumme, Z. L.; Harrigan, P. R. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir. *J. Virol.* **2001***, 75*, 589-594.

(144) Gatanaga, H.; Suzuki, Y.; Tsang, H.; Yoshimura, K.; Kavlick, M. F. et al. Amino acid substitutions in Gag protein at non-cleavage sites are indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J. Biol. Chem.* **2002***, 277*, 5952-5961.

(145) Sluis-Cremer, N.; Arion, D.; Parniak, M. A. Molecular mechanisms of HIV-1 resistance to nucleoside reverse transcriptase inhibitors (NRTIs). *Cell Mol. Life Sci.* **2000***, 57*, 1408-1422.

(146) Lindberg, J.; Sigurdsson, S.; Lowgren, S.; Andersson, H. O.; Sahlberg, C. et al. Structural basis for the inhibitory efficacy of efavirenz (DMP-266), MSC194 and PNU142721 towards the HIV-1 RT K103N mutant. *Eur. J. Biochem.* **2002***, 269*, 1670-1677.

(147) Udier-Blagovic, M.; Tirado-Rives, J.; Jorgensen, W. L. Structural and energetic analyses of the effects of the K103N mutation of HIV-1 reverse transcriptase on efavirenz analogues. *J. Med. Chem.* **2004***, 47*, 2389-2392.

(148) Ren, J.; Nichols, C. E.; Chamberlain, P. P.; Weaver, K. L.; Short, S. A. et al. Crystal structures of HIV-1 reverse transcriptases mutated at codons 100, 106 and 108 and mechanisms of resistance to non-nucleoside inhibitors. *J. Mol. Biol.* **2004***, 336*, 569-578.

(149) Rice, W. G.; Schaeffer, C. A.; Harten, B.; Villinger, F.; South, T. L. et al. Inhibition of HIV-1 infectivity by zinc-ejecting aromatic *C*-nitroso compounds. *Nature* **1993***, 361*, 473-475.

(150) Rice, W. G.; Schaeffer, C. A.; Graham, L.; Bu, M.; McDougal, J. S. et al. The site of antiviral action of 3-nitrosobenzamide on the infectivity process of human immunodeficiency virus in human lymphocytes. *Proc. Natl. Acad. Sci. USA* **1993***, 90*, 9721-9724.

(151) Rice, W. G.; Supko, J. G.; Malspeis, L.; Buckheit, R. W., Jr.; Clanton, D. et al. Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. *Science* **1995***, 270*, 1194-1197.

(152) Rice, W. G.; Turpin, J. A.; Huang, M.; Clanton, D.; Buckheit, R. W., Jr. et al. Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. *Nat. Med.* **1997***, 3*, 341-345.

(153) Ryser, H. J. J.; Levy, E. M.; Mandel, R.; DiSciullo, G. J. Inhibition of human immunodeficiency virus infection by agents that interfere with thiol-disulfide interchange upon virus-receptor interaction. *Proc. Natl. Acad. Sci. USA* **1994***, 91*, 4559-4563.

(154) Tummino, P. J.; Scholten, J. D.; Harvey, P. J.; Holler, T. P.; Maloney, L. et al. The in vitro ejection of zinc from human immunodeficiency virus (HIV) type 1 nucleocapsid protein by disulfide benzamides with cellular anti-HIV activity. *Proc. Natl. Acad. Sci. USA* **1996***, 93*, 969-973.

(155) Turpin, J. A.; Terpening, S. J.; Schaeffer, C. A.; Yu, G.; Glover, C. J. et al. Inhibitors of human immunodeficiency virus type 1 zinc fingers prevent normal processing of gag precursors and result in the release of noninfectious virus particles. *J. Virol.* **1996***, 70*, 6180-6189.

(156) Loo, J. A.; Holler, T. P.; Sanchez, J.; Gogliotti, R.; Maloney, L. et al. Biophysical characterization of zinc ejection from HIV nucleocapsid protein by anti-HIV 2,2'-dithiobis[benzamides] and benzisothiazolones. *J. Med. Chem.* **1996***, 39*, 4313-4320.

(157) Tummino, P. J.; Harvey, P. J.; McQuade, T.; Domagala, J.; Gogliotti, R. et al. The human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein zinc ejection activity of disulfide benzamides and benzisothiazolones: correlation with anti-HIV and virucidal activities. *Antimicrob. Agents Chemother.* **1997***, 41*, 394-400.

(158) Domagala, J. M.; Bader, J. P.; Gogliotti, R. D.; Sanchez, J. P.; Stier, M. A. et al. A new class of anti-HIV-1 agents targeted toward the nucleocapsid protein NCp7: the 2,2'-dithiobisbenzamides. *Bioorg. Med. Chem.* **1997***, 5*, 569-579.

(159) Domagala, J. M.; Gogliotti, R.; Sanchez, J. P.; Stier, M. A.; Musa, K. et al. 2,2'-Dithiobisbenzamides and 2-benzisothiazolones, two new classes of antiretroviral agents: SAR and mechanistic considerations. *Drug Des. Discov.* **1997***, 15*, 49-61.

(160) Huang, M., Maynard, A., Turpin, J.A., Graham, L., Janini, G.M., Covell, D.G., Rice, W.G. Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J. Med. Chem.* **1998***, 41*, 1371-1381.

(161) Prasad, J. V.; Loo, J. A.; Boyer, F. E.; Stier, M. A.; Gogliotti, R. D. et al. 2,2'-Dithiobisbenzamides derived from alpha-, beta- and gamma-amino acids possessing anti-HIV activities: synthesis and structure-activity relationship. *Bioorg. Med. Chem.* **1998***, 6*, 1707-1730.

(162) Topol, I. A.; Nemukhin, A. V.; Dobrogorskaya, Y. I.; Burt, S. K. Interactions of azodicarbonamide (ADA) species with the model zinc finger site: theoretical support of the zinc finger domain destruction in the HIV-1 nucleocapsid protein (NCp7) by ADA. *J. Phys. Chem. B* **2001***, 105*, 10.

(163) Rice, W. G.; Baker, D. C.; Schaeffer, C. A.; Graham, L.; Bu, M. et al. Inhibition of multiple phases of human immunodeficiency virus type 1 replication by a dithiane compound that attacks the conserved zinc fingers of retroviral nucleocapsid proteins. *Antimicrob. Agents Chemother.* **1997***, 41*, 419-426.

(164) Turpin, J. A.; Song, Y.; Inman, J. K.; Huang, M.; Wallqvist, A. et al. Synthesis and biological properties of novel pyridinioalkanoyl thiolesters (PATE) as anti-HIV-1 agents that target the viral nucleocapsid protein zinc fingers. *J. Med. Chem.* **1999***, 42*, 67-86.

(165) Goel, A.; Mazur, S. J.; Fattah, R. J.; Hartman, T. L.; Turpin, J. A. et al. Benzamide-based thiolcarbamates: a new class of HIV-1 NCp7 inhibitors. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 767-770.

(166) Basrur, V.; Song, Y.; Mazur, S. J.; Higashimoto, Y.; Turpin, J. A. et al. Inactivation of HIV-1 nucleocapsid protein P7 by pyridinioalkanoyl thioesters. Characterization of reaction products and proposed mechanism of action. *J. Biol. Chem.* **2000**, *275*, 14890-14897.

(167) Schito, M. L.; Goel, A.; Song, Y.; Inman, J. K.; Fattah, R. J. et al. In vivo antiviral activity of novel human immunodeficiency virus type 1 nucleocapsid p7 zinc finger inhibitors in a transgenic murine model. *AIDS Res. Hum. Retrov.* **2003**, *19*, 91-101.

CHAPTER 2

STUDIES ON THE MECHANISM OF INACTIVATION OF THE HIV-1 NUCLEOCAPSID PROTEIN NCP7 WITH 2-MERCAPTOBENZAMIDE THIOESTERS[1]

[1]Reprinted with permission from Miller Jenkins, L. M., Byrd, J. C., Hara, T., Srivastava, P., Mazur, S. J., Stahl, S. J., Inman, J. K., Appella, E., Omichinski, J. G., Legault, P. 2005 Apr 21; 48(8):2847-58. Copyright 2005 American Chemical Society.

Abstract

The HIV-1 nucleocapsid protein (NCp7) is a small basic protein with two CysCysHisCys zinc-binding domains that specifically recognizes the Ψ-site of the viral RNA. NCp7 plays a number of crucial roles in the viral lifecycle, including reverse transcription and RNA encapsidation. Several classes of potential anti-HIV compounds have been designed to inactivate NCp7 through zinc ejection, including a special class of thioester compounds. We have investigated the mechanism of action of two N-substituted-S-acyl-2 mercaptobenzamide compounds (compounds **1** and **2**) that target NCp7. UV/Visible spectroscopy studies demonstrated that both thioesters were able to eject metal from NCp7. NMR and mass spectroscopy studies showed that the thioester compounds specifically ejected zinc from the carboxyl-terminal zinc-binding domain of NCp7 by covalent modification of Cys_{39}. Exposure of NCp7 to compounds **1** and **2** destroyed its ability to specifically bind RNA, whereas NCp7 already bound to RNA was protected from zinc ejection by the thioesters. The thiol component of the thioesters (compound **3**, 2-mercaptobenzoyl-β-alaninamide) did not eject zinc from NCp7, but when compound **3** was incubated with acetyl CoA prior to incubation with NCp7, we observed extensive metal ejection. Thus, the thiol released by the reaction of compounds **1** and **2** could be re-acylated *in vivo* by acyl CoA to form a new thioester compound that is able to react with NCp7. These studies provide a better understanding of the mechanism of action of thioester compounds, which is important for future design of anti-HIV-1 compounds that target NCp7.

[1]Abbreviations: ADA, azodicarbonamide; CID, collision-induced dissociation; DIBA, disulfide benzamide; DTT, dithiothreitol; HIV-1, Human Immunodeficiency Virus Type-1; HPLC, high-performance liquid chromatography; IPTG, isopropyl-β-D-thiogalactopyranoside; NCp7, nucleocapsid protein; NEM, N-ethylmaleimide; NOBA, 3-nitrosobenzamide; TFA, trifluoroacetic acid; water-sLED, water-suppressed longitudinal encode-decode; ZD1, amino-terminal zinc-binding domain; ZD2, carboxyl-terminal zinc-binding domain.

.

Introduction

HIV-1 infection afflicts millions of people worldwide, killing three million people in the last year alone (www.unaids.org). Though much work has been done to stop the progression of HIV-1 infection to AIDS, several issues remain. Traditional combination therapies do not completely eradicate the virus. Consequently, long-term treatment is required, which is complicated by problems of patient compliance and metabolic side effects[1-4]. Over the long course of combination therapy, drug-resistant strains of HIV-1 often arise. These resistant strains necessitate the continual development of new antiviral drugs.

One potential target for antiviral therapy is the HIV-1 nucleocapsid protein (NCp7). NCp7 plays many crucial roles throughout the retrovirus lifecycle as a general nucleic acid binding protein. During reverse transcription, NCp7 facilitates binding of the $tRNA_3^{Lys}$ to the primer binding site and promotes strand annealing[5-7]. In addition, NCp7 enhances the integration of viral DNA into the host chromosome[8,9]. It is also essential for dimerization and packaging of full-length viral RNA into the new virions[10-12].

NCp7 is a small basic protein that contains two zinc-binding domains (ZD1 and ZD2). The two zinc-binding domains each chelate one zinc ion using three cysteine and one histidine residues as ligands (Figure 2.1A). NMR solution studies of NCp7 show that in the free protein, the two zinc-binding domains form small globular structures, whereas the amino and carboxyl termini of the protein are disordered[13-15]. The zinc-binding domains and several basic amino acids at the amino terminus of NCp7 are required for sequence-specific RNA binding. The structure of the zinc-binding domains and the mutationally non-permissive nature of NCp7 suggest that HIV-1 strains resistant to drugs targeted against this protein are unlikely to develop[10,16-19].

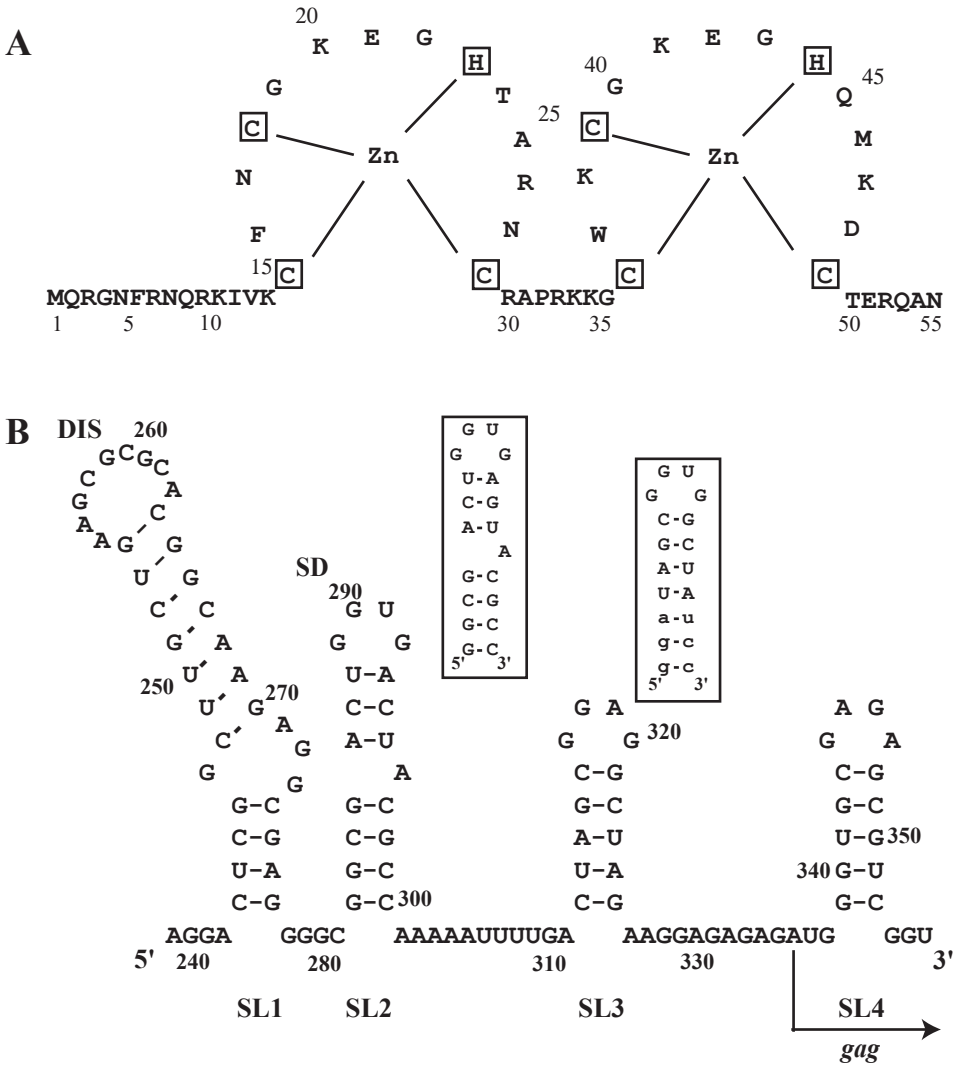


Figure 1. A) Amino acid sequence of NCp7 used in this study, showing the zinc- coordinating residues highlighted in boxes. B) Nucleotide sequence and secondary structure of the HIV-1 Ψ-site. Boxed sequences correspond to the sequences of the ΨSL2 and ΨSL3 RNAs used in this study. Nucleotides in lower case are nonnatural residues added for improved *in vitro* transcription by T7 RNA polymerase.

Over the last several years, numerous studies have been reported on compounds that were designed to inhibit NCp7. Most of the compounds eject zinc from NCp7 (zinc ejectors), destroying the structure of the protein and preventing it from functioning. The first zinc ejector was 3-nitrosobenzamide (NOBA), which showed inhibition of HIV-1 replication, but was highly reactive, and thus toxic[20]. Compounds of this type lacked selectivity, as they reacted indiscriminately with both zinc-binding domains of NCp7 as well as with other zinc-binding domains present in cellular proteins. Further studies led to the development of less toxic and more selective compounds, such as the disulfide benzamide (DIBA) and azodicarbonamide (ADA) compounds[21,22]. Both the DIBA and ADA compounds inhibit HIV-1 replication and show a degree of specificity for NCp7. For example, the carboxyl-terminal zinc-binding domain (ZD2) of NCp7 was found to be more reactive with the DIBA compounds than the amino-terminal zinc-binding domain (ZD1)[23]. In addition, the DIBA compounds did not inhibit the functional activity of cellular proteins containing zinc-binding domains, including the poly(ADP ribose) polymerase and the Sp1 and GATA-1 transcription factors[24]. Mass spectrometry demonstrated that both Cys_{36} and Cys_{49} of ZD2 were covalently modified by the DIBA compounds[23]. Despite their selectivity, the DIBA compounds had the disadvantage that they contained an easily reduced disulfide bond that severely limited the half-life of the compound[25].

Among the most recent class of zinc ejectors to be developed are the pyridinioalkanoyl thioesters[26]. These thioester compounds have high antiviral activity and low cellular toxicity. They are water soluble and have been shown to eject zinc from NCp7[26,27]. Initial studies demonstrated that silver ions had to be added to the reaction *in vitro* to "activate" the thioester compounds; no zinc ejection was observed with either silver ions alone or the thioester compounds alone[28]. Mass spectrometry revealed that these thioester compounds selectively

modified NCp7 in ZD2 as was seen with the DIBA compounds[28]. These thioester compounds were also shown to have no effect on DNA binding by Sp1, suggesting that they might not target other zinc-binding domains in cellular proteins[26]. Recently, the *in vivo* antiviral activity of two S-acyl-2-mercaptobenzamide thioesters (Figure 2.2, compounds **1** and **2**) was described[29]. These compounds were shown to have antiviral activity in a murine model[29]. Orally-administered compound **2** actually evoked an antiviral response similar to that of the protease inhibitor indinavir[29]. Though it is postulated that these new thioester compounds have antiviral activity because they eject zinc from NCp7, the exact mechanism of zinc ejection has not been established. Understanding how these thioester compounds function should be very useful for the design of improved compounds to target NCp7.

Here, we have used UV/Visible spectroscopy to evaluate the ability of the thioester compounds **1** and **2** to eject metal from purified, recombinant NCp7. NMR spectroscopy and mass spectrometry were used to study the structural interaction between the thioester compounds and NCp7, as well as to determine if these compounds covalently modify NCp7 and to identify any reaction products. Finally, we used gel mobility shift assay to study the ability of the thioester compounds to disrupt NCp7's function as an RNA-binding protein. These studies provide a better understanding of the mechanism of action of thioester compounds.

Methods

Synthesis of Compounds

The preparation of compound **1** (N-[2-(5-pyridiniovaleroylthio) benzoyl]-β-alaninamide bromide) and compound **3** (N^{β}-(2-mercaptobenzoyl)-β-alaninamide) (Figure 2) were carried out as previously described[30]. Compound **2** (N-[2-(nicotinoylthio)benzoyl]-β-alaninamide

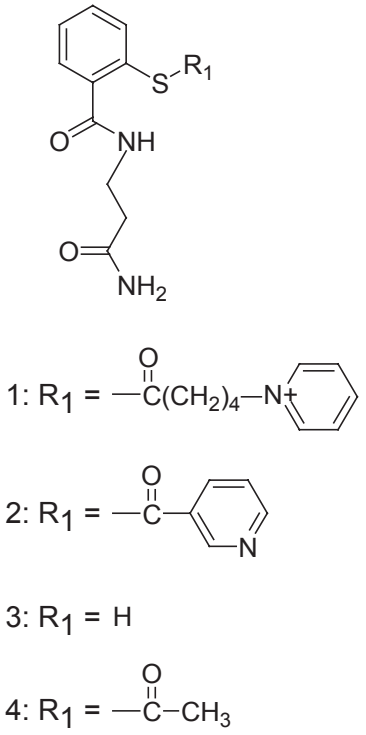


Figure 2.2. Chemical structure of thioester compounds and derivatives used in this study.

hydrochloride) was synthesized as outlined[26,30]. The uncharged compound **2** was prepared as described[31]. Compound **4** (N-[2-acetylthiobenzoyl]-β-alaninamide) (Figure 2) was synthesized as described[31]. Compound **3** and the thioester compounds were lyophilized in 500 μL aliquots at a concentration of 1 mM and stored at –20 ºC.

Sample Preparation

The coding sequence for HIV-1 NCp7 (HXB2 isolate) was cloned into the *Escherichia coli* vector pET11a and expressed in host strain Rosetta (DE3) (Novagen, WI). Uniformly ^{15}N-labeled or ^{13}C/^{15}N-labeled NCp7 was obtained by growing the cells in modified minimal media containing ^{15}N-labeled (>98%) ammonium chloride and/or ^{13}C-labeled (>99%) glucose as the sole nitrogen and carbon sources, respectively. Cells were grown overnight at 37 ºC, and protein expression was induced for 4 h with 0.66 mM or 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) for cultures grown in minimal media or Luria Broth, respectively. The cells were pelleted and resuspended in 25 mM Tris pH 8.0, 1 mM EDTA, 2 mM dithiothreitol (DTT), 6 mM benzamidine. The cells were then lysed by French press and centrifuged at 100,000 x g for 45 min. The supernatant was applied to a DEAE-Sepharose Fast Flow (Amersham Biosciences, NJ) column (300 mL bed volume), equilibrated with Buffer A (25 mM Tris pH 8.0, 1 mM EDTA, 2 mM DTT) and eluted using a gradient (0-100% B over 1500 mL) of Buffer B (25 mM Tris pH 8.0, 1 mM EDTA, 2 mM DTT, 1 M NaCl). The pooled NCp7 fractions were then applied to a SP-Sepharose Fast Flow (Amersham Biosciences, NJ) column (300 mL bed volume), equilibrated with Buffer A and eluted using a gradient (0 -100% B over 1500 mL) of Buffer B. The fractions containing NCp7 were pooled and purified on a C-8 reverse phase (Vydac) high-performance liquid chromatography (HPLC) column using a gradient of 10-30%

acetonitrile in 0.05% aqueous trifluoroacetic acid (TFA). The purified NCp7 was flash frozen and lyophilized.

Purified NCp7 was refolded using the following procedure. Lyophilized protein was first resuspended at a concentration of ~0.2 mM in 0.05% TFA. Five equivalents of $ZnCl_2$ (zinc-refolded NCp7) or $CoCl_2$ (cobalt-refolded NCp7) were then added to the NCp7 solution as 50 mM metal solutions in 0.05% TFA. The solution was slowly titrated to pH 6.0 with 0.2 M NaOH. The zinc-refolded NCp7 was flash frozen and lyophilized. Prior to use, zinc-refolded NCp7 (1 mM) was resuspended in NMR buffer A (20 mM sodium phosphate pH 7.0) in 90% H_2O/10% D_2O. The buffer of the cobalt-refolded NCp7 was adjusted to 20 mM sodium phosphate pH 7.0. The cobalt-refolded NCp7 was used immediately after refolding.

ΨSL2 and ΨSL3 RNAs (Figure 1B) were transcribed *in vitro* using T7 RNA polymerase, synthetic oligonucleotide templates, and nucleoside triphosphates. The T7 RNA polymerase was purified based upon a published procedure[32]. The RNAs were purified to single-nucleotide resolution by 20% denaturing polyacrylamide gel electrophoresis. Each RNA was further purified by DEAE-Sephacel chromatography (Amersham Biosciences, NJ). The ΨSL2 and ΨSL3 RNAs were concentrated with an Amicon Centricon-3 concentrator (Millipore, MA) and exchanged into NMR buffer A. Prior to use, the RNA was heated to 95 ºC for 1 min and snap-cooled in ice water to promote hairpin formation. For the gel mobility shift assay, the ΨSL2 RNA was first dephosphorylated on the 5'-end with calf alkaline phosphatase (Roche Molecular Biochemicals, IN) and then 5'-end-labeled with γ-(^{32}P) ATP (MP Biomedicals, CA) using T4 polynucleotide kinase (New England Biolabs, MA) according to the manufacturer's instructions.

The complex of ^{15}N-labeled NCp7 with ΨSL3 RNA was prepared by titrating ^{15}N-labeled NCp7 into ΨSL3 RNA in NMR buffer B (20 mM sodium phosphate buffer pH 7.0, 25

mM NaCl) in 90% H_2O/10% D_2O. Complex formation was monitored by 1D ^{15}N-decoupled watergate[33] and 2D ^{1}H-^{15}N HSQC[34]. Once a 1:1 complex was prepared, the sample was flash frozen and lyophilized. The sample was then resuspended to achieve a final buffer composition of 20 mM sodium phosphate pH 7.0, 100 mM NaCl in 90% H_2O/10% D_2O.

UV/Visible Spectroscopy

For UV/Visible spectroscopy studies, cobalt-refolded NCp7 (150 μM) was incubated with the appropriate thioester compound or compound **3** (150 μM unless specified otherwise) at 25 ºC. The UV/Visible spectrum was recorded from 220-800 nm every 0.5 h for 3 h using a Shimadzu UV-1601 spectrophotometer equipped with PC control via the UVPC software (Shimadzu Scientific Instruments, MD). The tetrahedrally-coordinated cobalt ions of NCp7 have absorption maxima at 642 nm and 698 nm. The absorbance at these two wavelengths was specifically monitored and compared against a control sample in which no thioester compound was added. Experiments were repeated a minimum of three times. Normalized absorbance values were calculated by subtracting the absorbance monitored by that of the control sample. The normalized absorbance at 642 nm and 698 nm was plotted versus time, and the initial rate of absorbance loss (0 – 3 h) was obtained from the slope of the linear regression analysis. In one set of experiments, compound **3** (150 μM) was incubated with acetyl CoA (Sigma-Aldrich, MO) (150 μM) for 1 h prior to addition of NCp7 (150 μM). Once NCp7 was added, the UV/Visible spectrum was monitored as described above. The rate of absorbance loss was calculated as described above using absorbance values normalized against a control sample in which NCp7 was incubated with acetyl CoA alone.

NMR Spectroscopy

The zinc-refolded ^{15}N-labeled NCp7 (1 mM) was incubated with an equimolar concentration of thioester compound at 25 ºC for 48 h in NMR buffer A and 90% H_2O/10% D_2O. The pH of the sample was checked both before and after the incubation period to ensure that any changes observed in the NMR spectra were not due to changes in pH. The sample was analyzed on Varian ^{UNITY}INOVA 500 MHz or 600 MHz spectrometers equipped with HCN triple resonance probes with actively shielded z- gradients. Three experiments were recorded at various times: a 1D ^{15}N-decoupled watergate[33], a 1D difference water-sLED[35], and a 2D ^{1}H-^{15}N HSQC[34]. 1D difference water-sLED experiments[35] were recorded to detect the resonances of the faster diffusing component(s) of the sample, the free thioester compounds and their derivatives produced by incubation with NCp7. In this 1D difference experiment, two water-sLED spectra, A and B, were recorded. The only parameter that differs between these two spectra is the strength of the encoding and decoding pulse-field gradient pulses, 7 G/cm in spectrum A and 14 G/cm in spectrum B. The duration of these pulses and the time delay between these pulses was set to 0.007 and 1.7 s, respectively. The 1D difference water-sLED spectrum was obtained by Fourier transformation of [FID(A) - alphaFID(B)], where the value of alpha was determined empirically to filter out the resonances of the slowly diffusing component of the sample, the NCp7 protein. The 2D ^{1}H-^{15}N HSQC spectra were processed with NMRPipe[36] and analyzed with PiPP[37]. The measured signal intensities of the HSQC signals were plotted versus time (from 0 – 48 h), and the initial rate of signal intensity loss from 0 – 24 hours was obtained by linear regression.

^{1}H, ^{15}N, and^{13}C backbone assignments for NCp7 in 90% H_2O/10% D_2O at pH 6.0, 25 °C were obtained with a ^{13}C/^{15}N-labeled sample from the following experiments: 2D ^{1}H-^{15}N

HSQC[34], 3D HNCO[38], 3D HNCACB[39-41], and 3D (HB)CBCA(CO)NNH[39,41]. ^{1}H and ^{15}N chemical shift assignments for NCp7 complexed with ΨSL3 RNA were generously provided by Dr. M. F. Summers (University of Maryland, Baltimore County). ^{1}H chemical shift assignments for compound **1** were obtained from the following experiments: 1D ^{1}H watergate[33], 2D ^{1}H-^{1}H DQF-COSY and 2D ^{1}H-^{1}H NOESY. Partial ^{1}H chemical shift assignments for compound **2** were obtained by comparison of the 1D ^{1}H watergate spectrum of the thioester compound with that of compound **1**.

Mass Spectrometry

A 1 mM sample of zinc-refolded NCp7 was incubated for 48 h with 1 mM compound **1** or **2** at 25 ºC. Aliquots were taken immediately following thioester addition and then after 1 h, 3 h, 5 h, 7 h, 9 h, 24 h, and 48 h. The samples were promptly flash frozen in liquid nitrogen and lyophilized to stop the thioester activity. Lyophilized samples were resuspended in 0.05% TFA and separated on a C-18 reverse phase column using a linear gradient (5-50%) of 90% acetonitrile containing 0.04% TFA. The lyophilized NCp7 was redissolved in 50% methanol and 1% acetic acid. Samples were introduced into an LCQ Classic ion trap mass spectrometer (Thermo Electron Corp) fitted with an electrospray ionization (ESI) source. Under acidic conditions, ESI produces multiple charge states of the NCp7 protein within the range 0 to 2000 m/z. Spectra were deconvoluted over the range of 5000 to 7500 Da using Xcalibur Biomass software (Thermo Electron Corp). Peak areas were calculated assuming Gaussian peak shape.

For ESI-MS studies, the lyophilized NCp7 was redissolved in a minimal volume of 20 mM Tris pH 7.8, 5 mM $CaCl_2$ and 2 mM DTT. Clostripain (Promega) was added at a ratio of

1:20 with NCp7 and incubated for 3 h at 37 ºC. Mass spectrometric data for the clostripain peptides were obtained from collision induced dissociation (CID) spectra[28].

For the identification of the site of NCp7 modification, 250 μM zinc-refolded chemically-synthesized NCp7 was incubated with equimolar concentration of compound **4** in 20 mM phosphate buffer pH 7.0 in a total volume of 250 μL for 3 h at 25 °C (chemical synthesis was performed as described[28]). The samples were purified on a C-18 reverse phase (Vydac) HPLC column using a gradient of 5-30% acetonitrile in 0.05% aqueous trifluoroacetic acid (TFA), and each peak was collected separately. These fractions were analyzed by MALDI-TOF and lyophilized. The lyophilized acetylated fraction was incubated for 1 h with N-ethylmaleimide (NEM) at a peptide:NEM molar ratio of 1:150 in 20 mM phosphate buffer pH 7.0, 0.5 mM EDTA at 30 °C. The modified peptide was then separated by HPLC separation, followed by MALDI-TOF analysis. These fractions were next subjected to clostripain digestion, HPLC separation, and MALDI-TOF analysis. The fractions of interest were further treated with chymotrypsin at 30 °C for 1 h in 10 mM Tris pH 7.0, 1.5 mM $CaCl_2$, 1.5 mM EDTA. HPLC separation, MALDI-TOF analysis, and finally ESI-MS/MS were performed.

Gel Mobility Shift Assay

A 1 μM sample of zinc-refolded NCp7 was incubated with 5.0 μM of either compound **1** or **2** in 50 mM Tris pH 7.5, 10% glycerol at 25 ºC for 0 min, 5 h, and 24 h. After each incubation period, a 5 μL aliquot was removed and incubated with an equal volume of 5'-^{32}P-labeled ΨSL2 RNA (2.5 nM) in 50 mM Tris pH 7.5, 10% glycerol, 25 μg/ml yeast tRNA, 200 mM KCl, and 40 mM $MgCl_2$ for 30 min at 25 ºC. The binding reactions were analyzed on an 8% native polyacrylamide gel in 50 mM Tris-Borate pH 8.0, and 1 mM EDTA at 200 V and 4 ºC. The 5'-

^{32}P-labeled ΨSL2 RNA was detected using a Storm PhosphorImager (Amersham Biosciences, NJ).

Results

*Thioester Compounds **1** and **2** Cause Loss of Metal Coordination by NCp7*

The ability of compounds **1** and **2** (Figure 2.2) to eject metal from NCp7 was initially studied by UV/Visible spectroscopy. In these experiments, apo-NCp7 was refolded with the spectroscopically active metal Co^{2+} rather than the spectroscopically inactive metal Zn^{2+}. Whereas zinc has a full outer shell, cobalt does not, allowing d-d transitions that absorb in the UV/Visible region of the spectrum. The cobalt-refolded NCp7 was incubated with various concentrations of each thioester compound, and the cobalt ejection was monitored. NCp7 refolded with Co^{2+} has two absorption maxima in the visible range: one at 642 nm and one at 698 nm[42]. These two absorption maxima correlate with the two different d-d transitions characteristic of tetrahedral geometry in which the ligands are three cysteines and one histidine[43]. If cobalt is ejected from NCp7 in the presence of the thioester compounds, a reduction in absorbance at 642 nm and 698 nm will be observed. Because both cobalt-bound zinc-binding domains absorb at these wavelengths, it was not possible using this assay to determine if the metal is ejected from both metal-binding domains or selectively from a single one.

Both compounds **1** and **2** cause a loss of absorbance over time at 642 nm and 698 nm (Figure 2.3), indicating that cobalt is no longer tetrahedrally coordinated by NCp7, and most likely ejected. Based on these results, it is clear that both thioester compounds are capable of ejecting metal from NCp7, and this process does not require activation with silver ions. Reduced

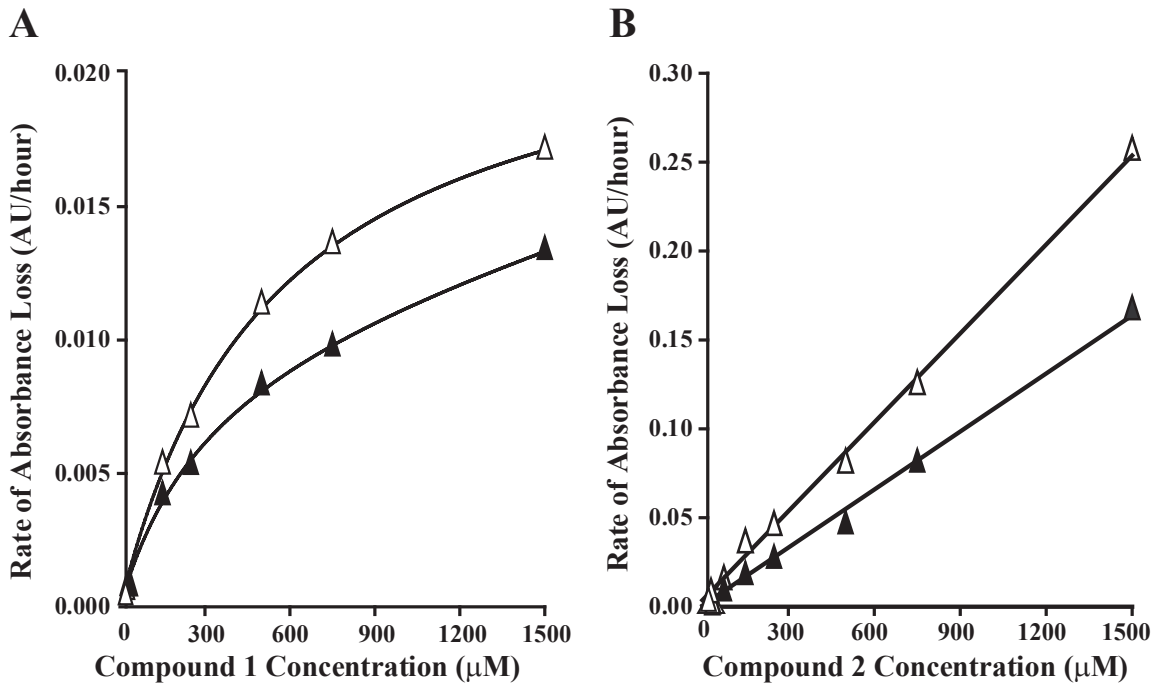


Figure 2.3. Rate of absorbance loss from the UV/Visible spectra recorded of cobalt- refolded NCp7 incubated with either compound **1** (A) or compound **2** (B). In all experiments, the cobalt-refolded NCp7 was kept at a constant concentration (150 μM), whereas the thioester concentration was varied from 0.15-1500 μM. In each figure, △ is the rate of absorbance loss at 642 nm and ▲ is the rate of absorbance loss at 698 nm. Rates were calculated as described in Materials and Methods.

absorption at 642 nm and 698 nm was tabulated at various times and used to calculate the initial rate of cobalt ejection from the zinc-binding domains. The absorbance at these two wavelengths decreased linearly over the period studied (0 – 3 h). The total absorbance loss was never more than 20% of the original absorbance. The rate of absorbance loss can therefore be considered as initial rates of metal ejection. These rates of absorbance loss were also calculated at various concentrations of compounds **1** and **2** (Figure 2.3). The observed rate of absorbance loss increases linearly with thioester concentration from 0.15 μM to 1500 μM for compound **2** and from 0.15 μM to 500 μM for compound **1** (Figure 2.3). Compound **1** and compound **2** show very different reaction rates from one another. At equimolar concentrations with NCp7 (150 μM), compound **2** reacts nine times more rapidly with cobalt-bound NCp7 than does compound **1** (Figure 2.3 and Table 2.1).

Thioester Compounds Show Specificity in the Reactivity with the Zinc-binding Domains of NCp7

From the UV/Visible spectroscopy experiments, it is evident that both compound **1** and compound **2** eject cobalt from NCp7. However, as previously mentioned, whether one or both zinc-binding domains are being affected cannot be determined using this experimental technique. Thus, NMR spectroscopy was used to gain a more detailed understanding of the mechanism of metal ejection. Equimolar amounts of zinc-refolded ^{15}N-labeled NCp7 and compound **1** were mixed, and 2D ^{1}H-^{15}N HSQC experiments were used to observe changes in the ^{1}H-^{15}N correlations of the ^{15}N-labeled NCp7 over a period of 48 h. Immediately after the addition of compound **1**, there was no change in the 2D ^{1}H-^{15}N HSQC spectrum of NCp7 (Supplementary Figure 2.1A). This demonstrates that NCp7 did not undergo a structural change when the thioester compound was added and that no stable interaction was formed between NCp7 and

Table 2.1. Rate of UV/Visible absorbance loss at 642 nm due to interaction of various compounds with NCp7.

Compound	Rate (mAU/h)	
1	3.9	± 1.3
2	36	± 3.7
3	0.4	± 0.1
4	29.6	± 2.0
3 + acetyl CoA	5.1	± 1.6

compound **1**. Indeed, chemical shift changes would have been expected if NCp7 were making a stable complex with the thioester compound. The primary change observed in the 2D ^{1}H-^{15}N HSQC spectrum over 48 hours was a loss of signal intensity for several signals (Supplementary Figure 2.1B). Loss of NCp7 signal intensity in the 2D ^{1}H-^{15}N HSQC spectrum was significantly greater for residues in ZD2 than for residues in ZD1. After 24 h, Cys_{36} and Cys_{39} in ZD2 retained only 40% of their original signal intensity, whereas Cys_{15} and Cys_{28} in ZD1 retained 85% of their original signal intensity (Figure 2.4A). When the change in signal intensity of zinc-binding residues is plotted over time, the distinction between the two zinc-binding domains is clear (Figure 2.4A). The ^{1}H-^{15}N correlations of the ZD1 zinc-coordinating residues Cys_{15}, His_{23}, and Cys_{28} show only a small overall change in signal intensity over 48 h. However, the ^{1}H-^{15}N correlations of the ZD2 zinc-coordinating residues Cys_{36}, Cys_{39}, His_{44}, and Cys_{49} lose at least 50% of their signal intensity. The initial rate of signal intensity loss is on average 10 times faster for residues in ZD2 than for residues in ZD1 (Table 2.2). The disparity in the signal intensity changes between residues in ZD1 and residues in ZD2 clearly indicates that compound **1** preferentially reacts with ZD2 over ZD1.

The signals for a few residues in ZD1 did undergo amide chemical shift changes during the incubation with compound **1** (Supplementary Figure 2.1B). We observed that both the amide ^{1}H and ^{15}N chemical shifts changed by at least 0.2 ppm for Phe_{16}, Asn_{17}, Cys_{18}, and Ala_{25} in ZD1. Similar amide chemical shift changes have been reported for Asn_{17} for NCp7 with no zinc coordinated in ZD2[23]. Over the course of the assay, the intensity in these shifted signals increased. Interestingly, we observed that the rate of chemical shift change for these signals is similar to the rate of disappearance of signals in ZD2 (data not shown). It therefore appears that unfolding ZD2 by compound **1** correlates with small chemical shift changes in residues located

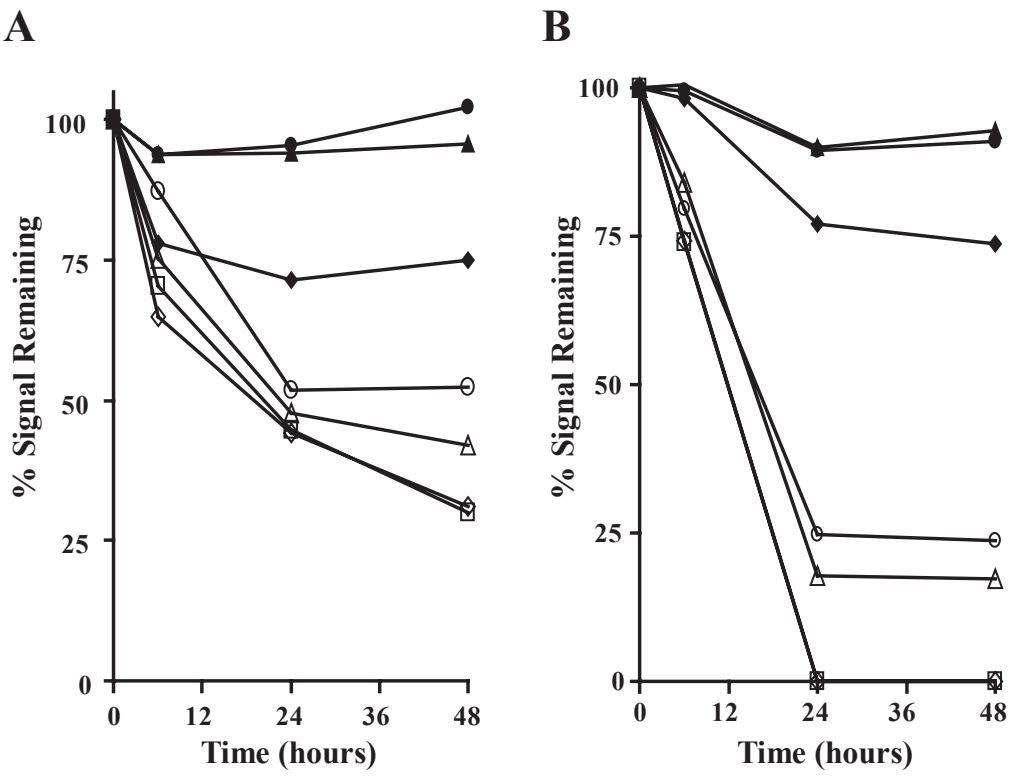


Figure 2.4. Change in 2D ^{1}H-^{15}N HSQC signal intensity of NCp7 zinc-coordinating residues after addition of compound **1** (A) and compound **2** (B). In each plot, ◆ is Cys_{15}, ▲ is His_{23}, ● is Cys_{28}, ◇ is Cys_{36}, □ is Cys_{39}, △ is His_{44}, ○ is Cys_{49}. The signal intensity of the ZD1 zinc-coordinating residue Cys_{18} was not monitored because its chemical shift changes during the course of the incubation with thioester compounds **1** and **2**.

Table 2. Average rate of loss of 2D ^{1}H-^{15}N HSQC signal intensity for zinc-coordinating residues in NCp7.

Thioester	NCp7 State	Rate (% Intensity/h) ZD1		ZD2	
1	Free	0.11	± 0.18	1.20	± 0.10
	Complex	0.27	± 0.22	0.34	± 0.21
2	Free	0.33	± 0.18	3.70	± 0.40
	Complex	0.01	± 0.05	0.21	± 0.01
3	Free	0.08	± 0.01	0.45	± 0.04

in ZD1. The fact that residues in ZD1 are affected by unfolding of ZD2 is not surprising; a dynamic interaction between the two zinc-binding domains has been reported by others[14,28,44].

The NMR results obtained when compound **2** was incubated with NCp7 were similar to those obtained with compound **1**. The 2D ^{1}H-^{15}N HSQC spectrum does not change much immediately following compound **2** addition (Supplementary Figure 2.2A), indicating that the NCp7 structure remains intact. Over time, most amide signals lose intensity rather than undergo chemical shift changes (Supplementary Figure 2.2B). After 24 h, signals that correspond to residues in ZD2 lost more intensity than those from ZD1 (Figure 2.4B). Signals for Cys_{36} and Cys_{39} disappeared within 24 h of addition of compound **2**. His_{44} and Cys_{49} signals retained less than 25% of their original intensity by 24 h. In contrast, ZD1 residues Cys_{15}, His_{23}, and Cys_{28} retained at least 80% of their original intensity even after 48 h of incubation. Like compound **1**, compound **2** shows a clear specificity for ZD2. However, the reaction of compound **2** with ZD2 is three times faster than that of compound **1** (Table 2.2). This general trend agrees with that observed in the UV/Visible experiments with cobalt-refolded NCp7. The overall rate of change was, however, faster with cobalt-refolded NCp7 than with zinc-refolded NCp7. This difference likely reflects the different affinities of NCp7 for the two metals. As zinc is coordinated by NCp7 with a 10^3 lower K_D than cobalt[45], it is expected that the ejection of zinc would take longer than that of cobalt. The specificity towards ZD2 is also more clearly observed with compound **2** than compound **1** (Figure 2.4). The more pronounced specificity of compound **2** is likely a result of the higher activity of this thioester compared to compound **1**.

Disappearance of Free Thioester Compounds Upon Incubation with NCp7

The 1D difference water-sLED spectrum was used to monitor changes in compound **1** and compound **2** when they were mixed with equimolar concentrations of ^{15}N-labeled NCp7. The difference water-sLED spectrum collected immediately after addition of compound **1** to NCp7 (Figure 2.5A) is identical to the spectrum of the free thioester compound (data not shown). Over time, all compound **1** signals decreased in intensity, and by 48 h the signals are generally indistinguishable from subtraction artifacts (Figure 2.5A). To ensure that the decrease in signal intensity in the difference water-sLED spectrum was not due to the natural instability of compound **1**, we incubated the thioester compound in NMR buffer over 48 h. After 48 h, there was no change in the 1D ^{1}H spectrum of compound **1** (data not shown). This indicates that the loss of signal intensity seen in the difference water-sLED spectra is due to either the interaction of compound **1** with NCp7 or chemical reactivity due to the presence of NCp7. In the case of compound **2**, we also did not see any change immediately after addition of the thioester compound to NCp7 (Figure 2.5B). Over time, the difference water-sLED spectra show that some peaks from compound **2** disappear, whereas others decrease in intensity, but never completely disappear (Figure 2.5B). After only 6 h, three peaks between 8.1 – 9.4 ppm have completely disappeared. These peaks correspond to protons in the nicotinoyl moiety unique to compound **2** (Figure 2.2). Peaks from the 2-mercaptobenzamide portion of compound **2** decrease in intensity, but do not disappear completely. This suggests that the nicotinoyl moiety is rapidly binding and/or reacting with NCp7, whereas the 2-mercaptobenzamide moiety undergoes a slower uncharacterized secondary binding event and/or reaction.

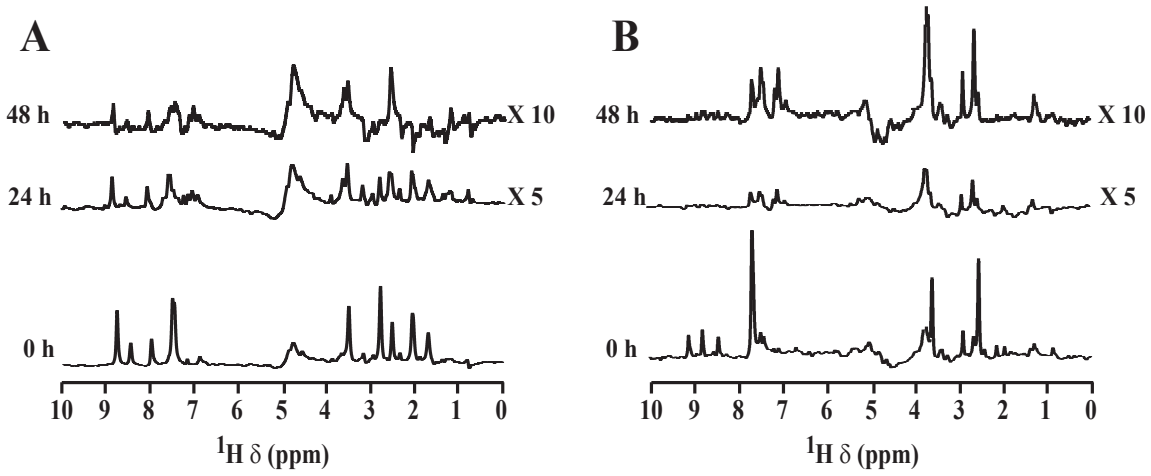


Figure 2.5. 1D difference water-sLED spectra of NCp7 incubated with compound **1** (A) or compound **2** (B). In both A and B, spectra are shown after 0 h, 24 h, and 48 h.

Thioester Compounds Covalently Modify NCp7

From the UV/Visible and NMR spectroscopy experiments, it was shown that compounds **1** and **2** specifically react with ZD2 of NCp7, eject coordinated metal from the protein, and cause unfolding of ZD2. Based on these results, we hypothesized that the thioester compounds are modifying cysteine residues in ZD2, leading to zinc ejection. To test this hypothesis, we used mass spectrometry to identify any modifications to NCp7 following addition of compound **1**. Equimolar amount of zinc-refolded NCp7 and each thioester compound were incubated over a period of 48 h. THE ESI-MS spectrum of an aliquot removed 1 h after addition of compound **1** shows an envelope of peaks corresponding to multiple charge states of unmodified NCp7 with an additional, lower-intensity envelope of higher-mass peaks (Supplementary Figure 3A). The deconvoluted mass spectrum (Supplementary Figure 3B) contained two peaks, one at 6425 ±1 Da, corresponding to intact NCp7, and one at 6587 ±1 Da, corresponding to a modified form. The net difference between the two masses results from the loss of a proton and addition of a modification with a mass of 163 Da, consistent with the pyridinioalkanoyl moiety of compound **1**. During the initial seven hours of incubation, the amount of the singly-modified species increases while the unmodified species decreases. The deconvoluted mass spectra of aliquots taken at later times contain peaks of higher mass that are consistent with NCp7 being modified with two or three similar adducts.

Reactivity of NCp7 with the Free Thiol

The mass spectral analysis raises a new question: What happens to the thiol group if only the acyl moiety is found attached to NCp7? The 1D difference water-sLED spectra show that all

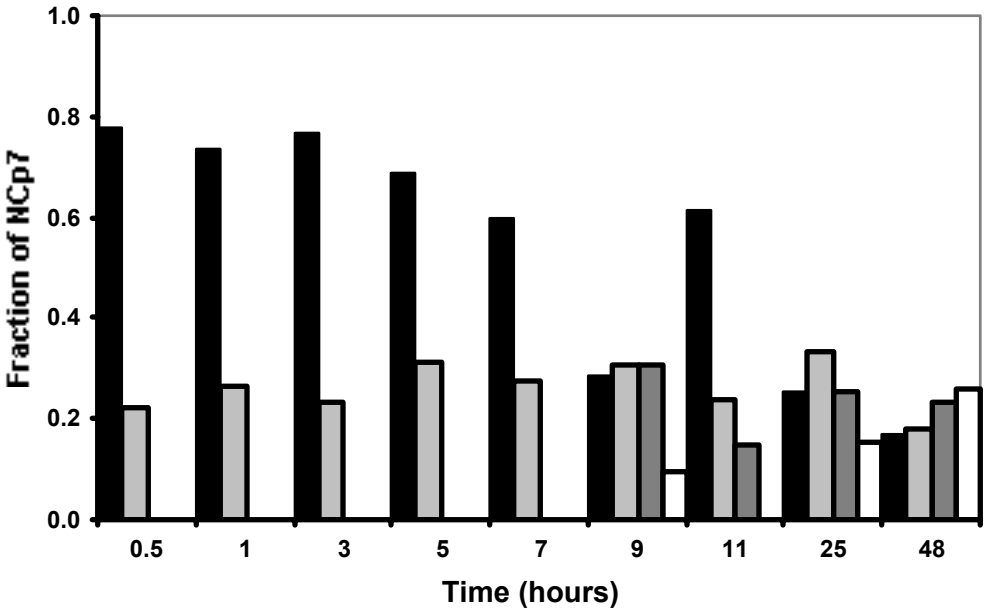


Figure 2.6. Mass spectrometry analysis of NCp7 modifications by compound **1** at various times. The relative amount of NCp7 with no modification is shown in black, one modification in light gray, two modifications in dark gray, or three modifications in white.

of the protons from compound **1** eventually disappear (Figure 2.5A). The difference water-sLED spectra from the experiments with compound **2** also show large loss of signal intensity from protons in the thiol moiety (Figure 2.5B). Thus, the NMR experiments suggest that something does happen to the thiol moiety.

In order to determine the effect of the thiol group on NCp7, we studied the interaction of NCp7 with the free thiol formed after thioester cleavage (compound **3**, Figure 2.2). The activity of compound **3** was analyzed first by UV/Visible spectroscopy experiments with cobalt-refolded NCp7 in a sealed cuvette with degassed buffer to prevent oxidation (data not shown). Over a period of three hours, there was no significant change in absorbance at 642 nm compared to a control sample (Table 2.1). Thus, in this assay, compound **3** was at least 9.8 times slower than compound **1** and 90 times slower than compound **2** at ejecting the bound cobalt from NCp7.

NMR experiments were next performed with compound **3** using degassed NMR buffer in a sealed NMR tube. The difference water-sLED spectra for compound **3** did not change over the entire incubation period (data not shown). There was a small loss of NCp7 signal intensity in the 2D ^{1}H-^{15}N HSQC spectra for residues in ZD2 when incubated with compound **3** (Figure 2.7, Table 2.2). After 48 h, the signals for zinc-coordinating residues in ZD2 retained 71-78% of their original intensity, whereas the corresponding signals in ZD1 retained 85-100% of their original intensity (Figure 2.7). The average rate of signal intensity loss in ZD2 due to compound **3** is 2.5 times lower than that seen with compound **1** (Table 2.2). Like the active thioester compounds, the activity of compound **3** seems to be specific for ZD2, but its overall activity is very low.

In the cellular environment, there are many potential reactions that a free thiol, such as compound **3**, can undergo. For example, the thiol can react with glutathione or be acylated by

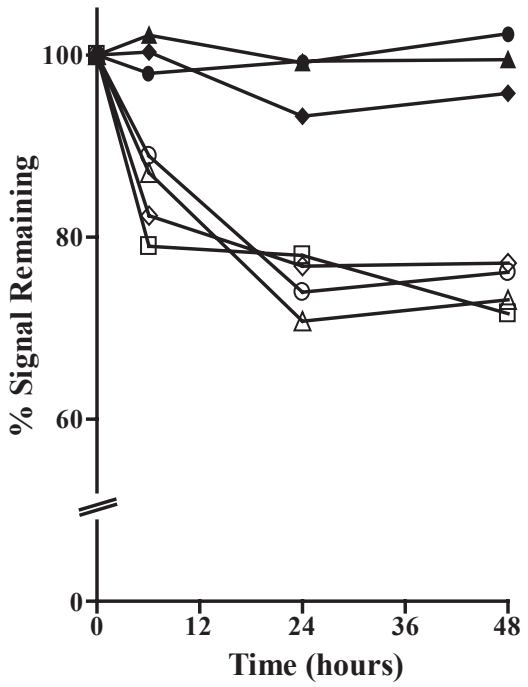


Figure 2.7. Change in 2D ^{1}H-^{15}N HSQC signal intensity of zinc-coordinating residues after compound **3** addition to NCp7. In the plot, ◆ is Cys_{15}, ▲ is His_{23}, ● is Cys_{28}, ◇ is Cys_{36}, □ is Cys_{39}, △ is His_{44}, ○ is Cys_{49}.

acyl CoA derivatives. To test the potential reactivity of one reaction product of compound **3**, we used UV/Visible spectroscopy to analyze the effect of compound **3** incubated with acetyl CoA and the effect of the acetylated form of the thiol (compound **4**). When we first incubated compound **3** with acetyl CoA and then with cobalt-refolded NCp7, we observed a loss of cobalt absorbance (Table 2.1). We also incubated cobalt-refolded NCp7 with preformed acetylated-compound **3**, compound **4** (Figure 2.2). With compound **4**, we also observed a loss of cobalt absorbance (Table 2.1). The rate of absorbance loss for compound **4** was similar to that observed for compound **2**, demonstrating high reactivity of this compound towards cobalt-refolded NCp7. However, the lower rate observed for thiol mixed with acetyl CoA suggests that the transfer of acetyl group from CoA to thiol may have been incomplete under the conditions of the experiment.

We further characterized the interaction of compound **4** with NCp7 by NMR spectroscopy. As with compounds **1** and **2**, we incubated equimolar concentrations of compound **4** with NCp7 and monitored changes in NCp7 by 2D ^{1}H-^{15}N HSQC spectra. There was no change in the 2D HSQC spectrum of NCp7 immediately following the addition of compound **4**, indicating that there was no conformational change in NCp7 upon mixing with the thioester compound (data not shown). After 48 hours of incubation with compound **4**, we observed loss of signal intensity for residues of ZD2 (data not shown). These changes are the same as those seen upon NCp7 incubation with compounds **1** and **2**. This indicates that compound **4** interacts with NCp7 in a similar manner as the other thioester compounds that we have analyzed. These results suggest that the thiols freed by reaction of the thioester compounds with NCp7, or by enzyme-promoted hydrolysis, could be re-acylated intracellularly by acyl CoA derivatives. The newly formed thioester would then be available to interact with additional NCp7 molecules.

Thioester Reaction with NCp7 Occurs First with Cys39

Mass spectrometry was next used to determine which residues of NCp7 are the first to interact with the thioester compounds. Following incubation of compound **4** with NCp7, samples were subjected to tandem mass spectral analysis of constituent peptides (see methods) to identify residues of NCp7 that had reacted. This analysis revealed that the first modification of NCp7 occurred primarily on Cys_{39}, with a minor population with modification to Cys_{49} (data not shown). As NMR spectroscopy analysis showed that compound **4** interacts with NCp7 in a similar manner as compounds **1** and **2**, it is likely that the modification on Cys_{39} is a typical feature of thioester interaction with NCp7.

Nucleic Acid Binding Both Inhibits and is Inhibited by Thioester Compound Interaction

Since NCp7 has nucleic acid binding activity *in vivo*, it is important to analyze the effect of thioester compounds not just on the free protein, but also on NCp7 complexed with nucleic acid. We made a complex between NCp7 and the third stem-loop (ΨSL3) from the Ψ-region at the 5'-end of the HIV-1 viral RNA (Figure 2.1B). NCp7 has been shown to bind ΨSL3 with a K_d = 28 nM[46], and the three-dimensional structure of the NCp7/ΨSL3 RNA complex has been determined by NMR spectroscopy[47]. We first determined that compound **1** did not have an effect on the free ΨSL3 RNA (data not shown). Then, the zinc-refolded NCp7/ΨSL3 complex was incubated with an equimolar concentration of either compound **1** or compound **2**. The effect of the thioester compound with the complex was analyzed using NMR spectroscopy over a period of 48 h.

The difference water-sLED spectrum recorded immediately after addition of compound **1** to the complex showed no changes in the thioester compound (data not shown). Likewise, the

2D ^{1}H-^{15}N HQSC spectrum recorded immediately after compound **1** addition shows no chemical shift changes in the complexed protein. Together these two observations indicate that neither the protein/RNA complex nor compound **1** undergo any change upon mixing.

Over time, the difference water-sLED spectra show only minor changes in compound **1** (data not shown). There is a small amount (<10%) of signal intensity loss after 48 h. This result is very different from what was seen with the uncomplexed protein, where all of the compound **1** signals disappeared by 48 h. This suggests that the compound **1** does not interact with the NCp7/RNA complex, as its diffusion rate does not change over 48 h. The 2D ^{1}H-^{15}N HSQC spectra of complexed NCp7 also reveal only a small loss of signal intensity in both zinc-binding domains over the period of the incubation (Figure 2.8A, Table 2.2). After 48 h, the zinc-coordinating residues of both ZD1 and ZD2 had lost 10% of their signal intensity (Figure 2.8A). The overall reactivity of compound **1** towards ΨSL3-bound NCp7 is greatly reduced compared with the reactivity towards free NCp7. The rate of signal intensity loss of ZD2 with free NCp7 is 3.5 – 4 times faster than that of ΨSL3-bound NCp7 (Table 2.2).

The effect of compound **2** on the ΨSL3-bound NCp7 is similar to that of compound **1**. The difference water-sLED spectra recorded during the incubation with compound **2** do not change over 48 h (data not shown). There is no change in signal intensity and chemical shift. As with compound **1**, there is no evidence from the 1D difference water-sLED experiments that compound **2** interacts with NCp7 bound to ΨSL3 RNA. The 2D ^{1}H-^{15}N HSQC spectra also remain the same throughout the incubation. There is essentially no change in signal intensity and chemical shift for any residues in NCp7 (Figure 2.8B, Table 2.2). Thus, interaction with the HIV-1 ΨSL3 RNA protects the zinc-coordinating residues of NCp7 from reacting with these two thioester compounds.

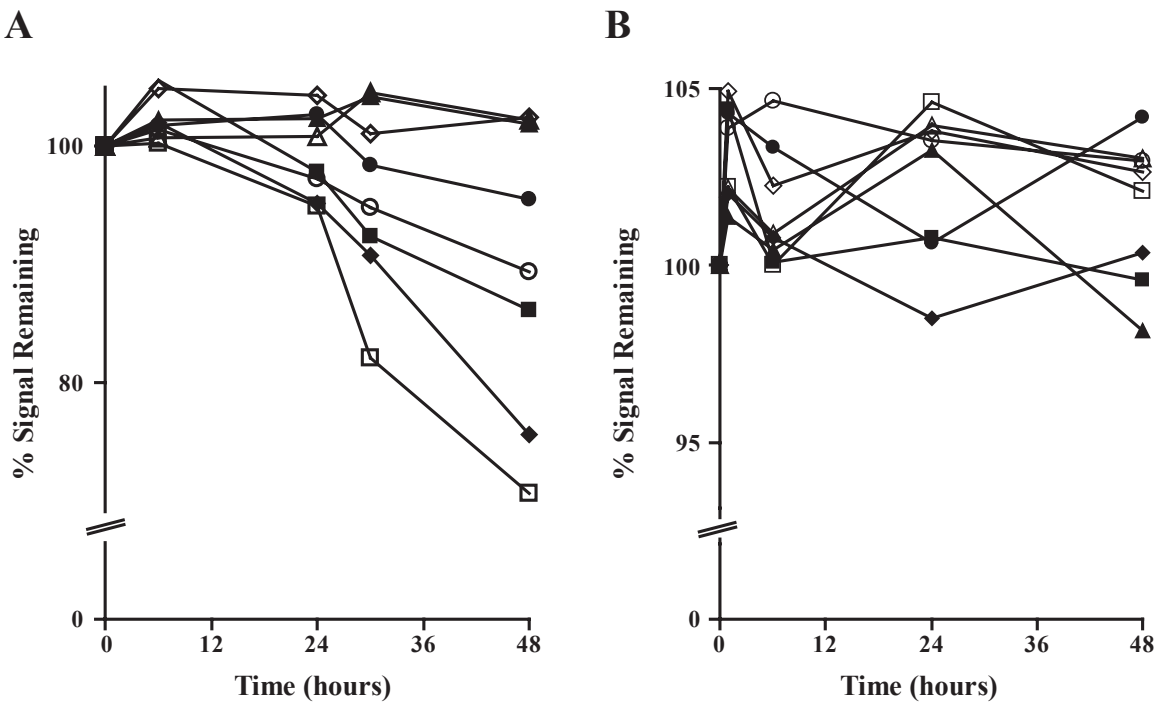


Figure 2.8. Change in 2D ^{1}H-^{15}N HSQC signal intensity of zinc-coordinating residues after addition of compound **1** (A) or compound **2** (B) to an NCp7/ΨSL3 RNA complex. In each plot ◆ is Cys_{15}, ■ is Cys_{18}, ▲ is His_{23}, ● is Cys_{28}, ◇ is Cys_{36}, □ is Cys_{39}, △ is His_{44}, ○ is Cys_{49}.

Gel mobility shift assays were used to investigate the effect of the thioester compounds on the nucleic acid binding properties of NCp7. Zinc-refolded NCp7 was incubated with either compound **1** or compound **2** and then mixed with the second stem loop (ΨSL2) from the HIV-1 Ψ site (K_d = 23 nM[46]). Addition of NCp7 to ΨSL2 RNA caused a gel mobility shift of the ^{32}P-labeled RNA (Figure 2.9, lanes 1, 2, and 5). The presence of a gel mobility shift in samples where NCp7, ΨSL2 RNA, and thioester compounds were briefly mixed indicates that neither compound **1** nor compound **2** immediately affects the RNA binding properties of NCp7 (Figure 9, lanes 3 and 4). However, after 24 h of NCp7 and thioester incubation, NCp7 no longer caused retardation of ΨSL2 RNA in the native gel (Figure 2.9, lanes 6 and 7). Since the ejection of zinc from NCp7 by compound **1** or compound **2** causes a loss of the structural integrity of ZD2, NCp7 likely loses its affinity for ΨSL2 RNA after reaction with the thioester compounds.

Discussion

Our studies demonstrate that the new thioesters (compounds **1** and **2**) are able to eject coordinated metals from the zinc-binding domains of NCp7. The new compounds appear to be more reactive than earlier thioesters, as they did not require the addition of silver ions to activate metal ejection from NCp7. Like the DIBA compounds and earlier thioester compounds, both compound **1** and compound **2** showed specificity for ZD2 of NCp7[23,25,26,28,48,49]. From NMR experiments, we observed that compound **2** completely unfolded ZD2, whereas ZD1 was essentially unaffected. Clearly, even though the two zinc-binding domains of NCp7 are structurally alike and contain similar amino acid composition[50], there is a large variation in their reactivity. The surrounding residues in each zinc-binding domain must make a significant contribution to the overall reactivity and/or accessibility of each zinc-binding domain.

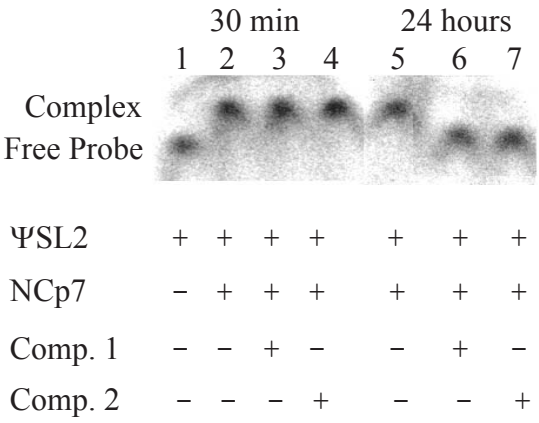


Figure 2.9. Gel mobility shift assay to study the effect of thioester compounds on binding of NCp7 to ΨSL2 RNA. NCp7 was incubated without (lanes 2 and 5) or with a thioester compound [compound **1** (lanes 3 and 6) or compound **2** (lanes 4 and 7)] for 30 sec or 24 h. Except for the control lane 1, ^{32}P-labeled ΨSL2 RNA was then added and incubated for 30 min at 25 ºC before being run on the 8% native gel.

With the combined information from the UV/Visible spectroscopy, NMR spectroscopy, and mass spectrometry, we can propose a reaction mechanism for the interaction of the thioester compounds **1** and **2** with NCp7 (Figure 2.10). The sulfhydryl group of Cys_{39} makes a nucleophilic attack on the carbonyl carbon of the thioester compound, leading to a covalent modification of the sulfhydryl group (acyl transfer). This results in a covalently modified cysteine in NCp7 and release of a free thiol from the thioester (compound **3)**. In this process, a new thioester bond is formed between the cysteine sulfur in NCp7 and the pyridinioalkanoyl moiety of compound **1** or the nicotinoyl moiety of compound **2**. Loss of zinc coordination by Cys_{39} makes the other cysteine residues in the zinc-binding domain more susceptible to subsequent reaction with the thioester compound or the free thiol. At some point, zinc is ejected from the zinc-binding domain causing a loss of the ZD2 fold. Once ZD2 is unfolded, ZD1 is likely destabilized and also undergoes a structural change. As we have shown, NCp7 is then no longer able to bind to nucleic acids. The loss of structure prevents the protein from fulfilling its numerous functions associated with its ability to specifically bind RNA.

In all of our experiments, compound **2** was more reactive than compound **1**. In the UV/Visible spectroscopy experiments, the rate of absorbance loss for compound **2** was ~10 times faster than that calculated for compound **1** (Figure 2.3, Table 2.1). Likewise, compound **2** caused 3 times faster loss of signal intensity than compound **1** for signals of residues in ZD2 in the 2D ^{1}H-^{15}N HSQC spectra. Interestingly, compound **4** ejected metal from cobalt-refolded NCp7 at approximately the same rate as compound **2** in the UV/Visible spectroscopy experiments. Compounds **1, 2** and **4** all contain the same 2-mercaptobenzamide thiol group (Figure 2). Thus, the variation in the activity of the three thioester compounds must be due to inherent differences between the pyridinioalkanoyl moiety of compound **1**, the nicotinoyl moiety of compound **2**, and

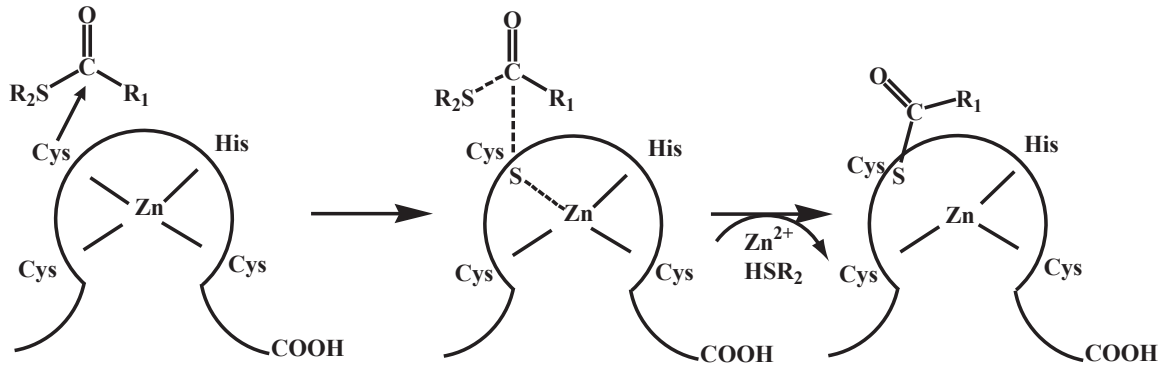


Figure 2.10. Schematic of the proposed thioester reaction mechanism. Cys_{39} in ZD2 makes a nucleophilic attack on the carbonyl carbon of the thioester compound. This results in loss of zinc coordination by Cys_{39} and destruction of the thioester bond of the thioester compound. A new thioester linkage is formed between the R_1 moiety and Cys_{39}. Once zinc coordination is destabilized, other cysteine residues in the zinc-binding domain become more susceptible to thioester reaction. Eventually, the zinc is ejected from the zinc-binding domain, which leads to loss of structure.

the acetyl moiety of compound **4**. The acetyl and nicotinoyl groups have smaller electron donating effects on the carbonyl carbon of the thioester, which could make the nucleophilic reaction with Cys_{39} proceed faster with compound **2** or **4** than with compound **1** (Table 2.1).

A similar study to that reported here was conducted with DIBA compounds, in which it was found that the compound reacted primarily with ZD2[23]. The proposed reaction mechanism of the DIBA compounds *in vitro* is dissimilar to the mechanism we propose for the thioesters: whereas thioester compounds transfer their acyl group initially to Cys_{39}, there is a thiol-disulfide interchange between the DIBA compound and NCp7 at Cys_{36} and Cys_{49}[23]. Modifications by both the thioester and DIBA compounds lead to loss of zinc coordination. Additionally, the study of NCp7 reactivity with NEM found Cys_{49} to be the first site of reaction with this alkylating agent[49]. In the 2D ^{1}H-^{15}N HSQC experiments with compounds **1** and **2**, Cys_{49} showed the lowest loss of signal intensity compared with the other zinc-coordinating residues in ZD2. After 48 h of incubation with compound **2**, the signal for Cys_{49} retained 25% of its original signal intensity, whereas the signal for Cys_{39} had completely disappeared after by 24 h. Our experiments show that Cys_{49} is not affected structurally as much as Cys_{39} by the interaction with the thioester compounds. The different sites of the first reaction between the DIBA compound and the thioester compounds studied here could reflect a distinction in the binding site of the two types of compounds. When various electrophiles were docked to NCp7, it was observed that the most productive docking occurred in a pocket adjacent to Cys_{39} and Cys_{49}[51]. The DIBA compounds may interact in an orientation with easier access to Cys_{49} and the thioester compounds may more readily approach Cys_{39}. However, stable complexes between the thioester compounds and NCp7 could not be demonstrated by NMR spectroscopy.

The functions of NCp7 in the HIV-1 lifecycle revolve around the ability of the protein to bind nucleic acids. Thus, it was important to study the effect of the thioester compounds not only on the free protein, but also on RNA-bound NCp7. We showed that NCp7 is resistant to reaction with the thioester compounds when complexed to RNA. In the bound state, the zinc-coordinating residues are not solvent-accessible, and the RNA is able to protect the protein[47,52]. Thus, NCp7 is mostly susceptible to reactions with the thioester compounds when it is not bound to RNA. Once NCp7 has reacted with the thioester compounds, however, it loses its structure and is no longer able to bind RNA. Throughout the HIV-1 lifecycle, NCp7 binds to several different nucleic acids: for example, the $tRNA_3^{Lys}$ primer, the reverse transcribed dsDNA, and the Ψ-site RNA. As these different nucleic acids show a large variation in sequence, secondary structure, and NCp7 binding affinity, there is likely also a range of susceptibilities of these NCp7/nucleic acid complexes from reaction with the thioester compounds. Additionally, it is unlikely that the conformation of NCp7 in the complex with each of these nucleic acids is exactly the same. The solution structure of NCp7 bound to ΨSL2 RNA is different from the solution structure of NCp7 bound to ΨSL3 RNA even though both RNAs form a stem-loop structure and have similar loop sequences[47,52]. A previous study with N-ethylmaleimide (NEM) showed a range of reactivity when NCp7 was complexed with short DNA molecules of different sequences[49]. Thus, it is possible that when bound to different nucleic acid sequences, NCp7 will be differentially susceptible to reaction with the thioester compounds.

We propose that compound **3** is acylated, possibly by acyl CoA's, *in vivo*, producing new thioester compounds. This means that the thiol group produced upon cleavage of the thioester bond could be re-acylated and available for a second reaction with NCp7. The ability of the reaction product to undergo subsequent reactions *in vivo* appears to make the thioester

compounds unique among zinc-ejecting compounds designed to target NCp7. For example, ADA is converted into biurea *in vivo*, which has no antiviral effect[53]. Interestingly, in aqueous conditions, many disulfide compounds rapidly cyclize to form a somewhat less reactive benzisothiazolone derivative and an aromatic thiol compound[23,54-56]. The aromatic thiol groups released from the DIBA compounds by glutathione reduction may behave like compound **3**, as they do not function as zinc ejectors *in vitro*[55]. However, it is conceivable that the DIBA-released thiols are acylated *in vivo* by acyl CoA to form an active thioester as we are predicting for compound **3**. Thus, the antiviral activity associated with the DIBA compounds might be due in part to their ability to form reactive thioester compounds following reduction.

The experiments shown here demonstrate that the thioester compounds **1** and **2** are able to eject coordinated zinc from NCp7 through covalent modification of Cys_{39}. The covalent modification of Cys_{39} following reaction with compounds **1** and **2** leads to structural changes that prevent NCp7 from binding to nucleic acids. In addition to modifying Cys_{39}, the reaction of compounds **1** and **2** with NCp7 results in the release of the 2-mercaptobenzoyl-β-alanamide, compound **3**. Although compound **3** itself has very little intrinsic reactivity toward NCp7, we have demonstrated that this thiol could be acylated *in vitro* by acetyl CoA to form a new thioester, compound **4**, which is capable of ejecting zinc from NCp7 by acylating Cys_{39}. These results suggest that the thiol formed from reaction of the thioester compounds **1** and **2** with NCp7 could be acylated *in vivo* by acetyl CoA or other acyl transfer agents to form a new thioester, which is able to undergo a second reaction with NCp7. This may explain the *in vivo* reactivity of these thioester compounds, which have the potential to be reactivated.

One feature of the thioester compounds that may be important for antiviral activity is the stability of these compounds, which is much higher in general than that of the DIBA compounds.

In a recent study, the antiviral activity and the stability of a number of thioester compounds were evaluated[31]. This study demonstrated that hydrolytic cleavage of the thioester bond in human serum varied considerably depending on the structure of the S-acyl group[31]. For example, compound **1** [$t_{1/2}$ = 156 min (unpublished data)] is more stable than compound **2** ($t_{1/2}$ = 49 min[31]) and compound **4** [$t_{1/2}$ = 19 min (unpublished data)] in 10 % human serum. The higher stability of compound **1** does not correlate with a higher *in vitro* reactivity for this compound, since we have shown here that compounds **2** and **4** reacted faster than compound **1** with NCp7. Furthermore, thorough analysis of the hundreds of compounds tested failed to establish a correlation between hydrolytic stability and antiviral potency[31]. One possible explanation is that the free thiols produced by hydrolytic degradation may be converted *in vivo* into other thioester compounds at different rates and, as a result, possess considerably different levels of antiviral activities. The same process could explain the antiviral activity of DIBA compounds, which can be reduced in solution to produce free thiols.

Through these studies, we have gained a more detailed understanding of the mechanism of action of the thioester compounds, which is important for future design of compounds that target NCp7. Interestingly, our NMR results clearly demonstrate that neither the 2-mercaptobenzoyl-β-alanamide nor the S-acyl moieties have intrinsic binding activity towards NCp7. These results suggest that it might be possible to design thioesters with improved affinity and selectivity for NCp7 by altering the framework of the thioester compounds, especially the thiol component, so that they would be capable of specifically interacting with amino acid residues in NCp7 that are in proximity to Cys_{39}. One potential amino acid target is Trp_{37}, which is close in space to the reactive cysteine and is relatively well exposed to the solvent. Therefore, it might be possible to design thioester compounds that both interact with Trp_{37} and acylate

Cys_{39}, and such compounds may possess enhanced specificity and affinity for NCp7 relative to the thioester compounds studied here. To understand the specificity of the thioesters compounds studied here, as well as those that will be designed in the future, it is crucial to investigate their effect on cellular zinc-binding proteins. Studies similar to the one presented here are currently underway to examine the mechanism of action of the thioester compounds with cellular proteins containing a variety of zinc-binding motifs.

Acknowledgements

We thank Dr. Venkatesha Basrur for his help with preliminary mass spectrometry data. This work was supported by American Cancer Society Grant RPG LBC-100183 (J.G.O.) and National Institutes of Health Grant RO1 GM60298-01 (to J.G.O and P.L.) L.M.M.J. is supported by a Presidential Graduate Fellowship from the University of Georgia. This project was funded in part by the Intramural AIDS Targeted Antiretroviral Program (IATAP).

Supporting Information Available

Supplementary Figure 1 shows superposition of 2D ^{1}H-^{15}N spectra of NCp7 free and after addition of compound **1** (at 0 h and 48 h). Supplementary Figure 2 shows superposition of 2D ^{1}H-^{15}N spectra of NCp7 free and after addition of compound **2** (at 0 h and 48 h). Supplementary Figure 3 shows electrospray ionization and deconvoluted mass spectra of NCp7 following addition of compound **1** (at 1 h). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Cohen, J. The daunting challenge of keeping HIV suppressed. *Science* **1997***, 277*, 32-33.

(2) Finzi, D.; Hermankova, M.; Pierson, T.; Carruth, L. M.; Buck, C. et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **1997***, 278*, 1295-1300.

(3) Wong, J. K.; Hezareh, M.; Gunthard, H. F.; Havlir, D. V.; Ignacio, C. C. et al. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **1997***, 278*, 1291-1295.

(4) Tozser, J. Part III clinical insight - HIV inhibitors: problems and reality. *Ann. NY Acad. Sci.* **2001***, 946*, 15.

(5) Guo, J.; Henderson, L. E.; Bess, J.; Kane, B.; Levin, J. G. Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol.* **1997***, 71*, 5178-5188.

(6) Negroni, M.; Buc, H. Mechanisms of retroviral recombination. *Annu. Rev. Genet.* **2001***, 35*, 275-302.

(7) Tisne, C.; Roques, B. P.; Dardel, F. Heteronuclear NMR studies of the interaction of $tRNA_{Lys}^{3}$ with HIV-1 nucleocapsid protein. *J. Mol. Biol.* **2001***, 306*, 443-454.

(8) Carteau, S.; Batson, S. C.; Poljak, L.; Mouscadet, J. F.; de Rocquigny, H. et al. Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg2+-dependent DNA integration in vitro. *J. Virol.* **1997***, 71*, 6225-6229.

(9) Carteau, S.; Gorelick, R. J.; Bushman, F. D. Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J. Virol.* **1999**, *73*, 6670-6679.

(10) Meric, C.; Goff, S. P. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. *J. Virol.* **1989**, *63*, 1558-1568.

(11) Darlix, J. L.; Gabus, C.; Nugeyre, M. T.; Clavel, F.; Barre-Sinoussi, F. Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **1990**, *216*, 689-699.

(12) Gorelick, R. J.; Nigida, S. M., Jr.; Bess, J. W., Jr.; Arthur, L. O.; Henderson, L. E. et al. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **1990**, *64*, 3207-3211.

(13) Omichinski, J. G., Clore, G.M., Sakaguchi, K., Appella, E., Gronenborn, A.M. Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy. *FEBS Lett.* **1991**, *292*, 25-30.

(14) Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J., Roques, B.P. Determination of the structure of the nucleocapsid protein from the human immunodeficiency virus type 1 by ^{1}H NMR. *EMBO J.* **1992**, *11*, 3059-3065.

(15) South, T. L.; Summers, M. F. Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci.* **1993**, *2*, 3-19.

(16) Dorfman, T.; Luban, J.; Goff, S. P.; Haseltine, W. A.; Gottlinger, H. G. Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **1993***, 67*, 6159-6169.

(17) Housset, V.; De Rocquigny, H.; Roques, B. P.; Darlix, J. L. Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity. *J. Virol.* **1993***, 67*, 2537-2545.

(18) Poon, D. T.; Wu, J.; Aldovini, A. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J. Virol.* **1996***, 70*, 6607-6616.

(19) Tanchou, V.; Decimo, D.; Pechoux, C.; Lener, D.; Rogemond, 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*, 4442-4447.

(20) Rice, W. G.; Schaeffer, C. A.; Graham, L.; Bu, M.; McDougal, J. S. et al. The site of antiviral action of 3-nitrosobenzamide on the infectivity process of human immunodeficiency virus in human lymphocytes. *Proc. Natl. Acad. Sci. USA* **1993***, 90*, 9721-9724.

(21) Rice, W. G.; Supko, J. G.; Malspeis, L.; Buckheit, R. W., Jr.; Clanton, D. et al. Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. *Science* **1995***, 270*, 1194-1197.

(22) Rice, W. G.; Turpin, J. A.; Huang, M.; Clanton, D.; Buckheit, R. W., Jr. et al. Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. *Nat. Med.* **1997***, 3*, 341-345.

(23) Loo, J. A.; Holler, T. P.; Sanchez, J.; Gogliotti, R.; Maloney, L. et al. Biophysical characterization of zinc ejection from HIV nucleocapsid protein by anti-HIV 2,2'-dithiobis[benzamides] and benzisothiazolones. *J. Med. Chem.* **1996**, *39*, 4313-4320.

(24) Huang, M., Maynard, A., Turpin, J.A., Graham, L., Janini, G.M., Covell, D.G., Rice, W.G. Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J. Med. Chem.* **1998**, *41*, 1371-1381.

(25) Tummino, P. J.; Scholten, J. D.; Harvey, P. J.; Holler, T. P.; Maloney, L. et al. The in vitro ejection of zinc from human immunodeficiency virus (HIV) type 1 nucleocapsid protein by disulfide benzamides with cellular anti-HIV activity. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 969-973.

(26) Turpin, J. A.; Song, Y.; Inman, J. K.; Huang, M.; Wallqvist, A. et al. Synthesis and biological properties of novel pyridinioalkanoyl thiolesters (PATE) as anti-HIV-1 agents that target the viral nucleocapsid protein zinc fingers. *J. Med. Chem.* **1999**, *42*, 67-86.

(27) Goel, A.; Mazur, S. J.; Fattah, R. J.; Hartman, T. L.; Turpin, J. A. et al. Benzamide-based thiolcarbamates: a new class of HIV-1 NCp7 inhibitors. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 767-770.

(28) Basrur, V.; Song, Y.; Mazur, S. J.; Higashimoto, Y.; Turpin, J. A. et al. Inactivation of HIV-1 nucleocapsid protein p7 by pyridinioalkanoyl thioesters. Characterization of reaction products and proposed mechanism of action. *J. Biol. Chem.* **2000**, *275*, 14890-14897.

(29) Schito, M. L.; Goel, A.; Song, Y.; Inman, J. K.; Fattah, R. J. et al. In vivo antiviral activity of novel human immunodeficiency virus type 1 nucleocapsid p7 zinc finger inhibitors in a transgenic murine model. *AIDS Res. Hum. Retrov.* **2003**, *19*, 91-101.

(30) Song, Y.; Goel, A.; Basrur, V.; Roberts, P. E.; Mikovits, J. A. et al. Synthesis and biological properties of amino acid amide ligand-based pyridinioalkanoyl thioesters as anti-HIV agents. *Bioorg. Med. Chem.* **2002***, 10*, 1263-1273.

(31) Srivastava, P.; Schito, M.; Fattah, R. J.; Hara, T.; Hartman, T. et al. Optimization of unique, uncharged thioesters as inhibitors of HIV replication. *Bioorgan. Med. Chem.* **2004***, 12*, 6437-6450.

(32) Davanloo, P.; Rosenberg, A. H.; Dunn, J. J.; Studier, F. W. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **1984***, 81*, 2035-2039.

(33) Piotto, M.; Saudek, V.; Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **1992***, 2*, 661-665.

(34) Kay, L. E.; Keifer, P.; Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **1992***, 114*, 10663.

(35) Altieri, A. S.; Hinton, D. P.; Byrd, R. A. Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J. Am. Chem. Soc.* **1995***, 117*, 7566-7567.

(36) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995***, 6*, 277-293.

(37) Garrett, D. S.; Powers, R.; Gronenborn, A. M.; Clore, G. M. A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **1991***, 95*, 214-220.

(38) Grzesiek, S.; Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **1992***, 99*, 201.

(39) Grzesiek, S.; Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **1992***, 114*, 6291.

(40) Wittekind, M.; Mueller, L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson. Ser. B* **1993***, 101*, 201-205.

(41) Muhandiram, D. R.; Kay, L. E. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Ser. B* **1994***, 103*, 203-216.

(42) Fitzgerald, D. W.; Coleman, J. E. Physicochemical properties of cloned nucleocapsid protein from HIV. Interactions with metal ions. *Biochemistry* **1991***, 30*, 5195-5201.

(43) Chen, X.; Chu, M.; Giedroc, D. P. Spectroscopic characterization of Co(II)-, Ni(II)-, and Cd(II)-substituted wild-type and non-native retroviral-type zinc finger peptides. *J. Biol. Inorg. Chem.* **2000***, 5*, 93-101.

(44) Lee, B. M.; De Guzman, R. N.; Turner, B. G.; Tjandra, N.; Summers, M. F. Dynamical behavior of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **1998***, 279*, 633-649.

(45) Magyar, J. S.; Godwin, H. A. Spectropotentiometric analysis of metal binding to structural zinc-binding sites: accounting quantitatively for pH and metal ion buffering effects. *Anal. Biochem.* **2003***, 320*, 39-54.

(46) Shubsda, M. F.; Paoletti, A. C.; Hudson, B. S.; Borer, P. N. Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry* **2002***, 41*, 5276-5282.

(47) De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N. et al. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **1998**, *279*, 384-388.

(48) Hathout, Y.; Fabris, D.; Han, M. S.; Sowder, R. C. I.; Henderson, L. E. et al. Characterization of intermediates in the oxidation of zinc fingers in human immunodeficiency virus type 1 nucleocapsid protein p7. *Drug Metab. Dispos.* **1996**, *24*, 6.

(49) Chertova, E. N.; Kane, B. P.; McGrath, C.; Johnson, D. G.; Sowder, R. C., 2nd et al. Probing the topography of HIV-1 nucleocapsid protein with the alkylating agent N-ethylmaleimide. *Biochemistry* **1998**, *37*, 17890-17897.

(50) South, T. L.; Blake, P. R.; Hare, D. R.; Summers, M. F. C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain. *Biochemistry* **1991**, *30*, 6342-6349.

(51) Maynard, A. T.; Huang, M.; Rice, W. G.; Covell, D. G. Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11578-11583.

(52) Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R. et al. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* **2000**, *301*, 491-511.

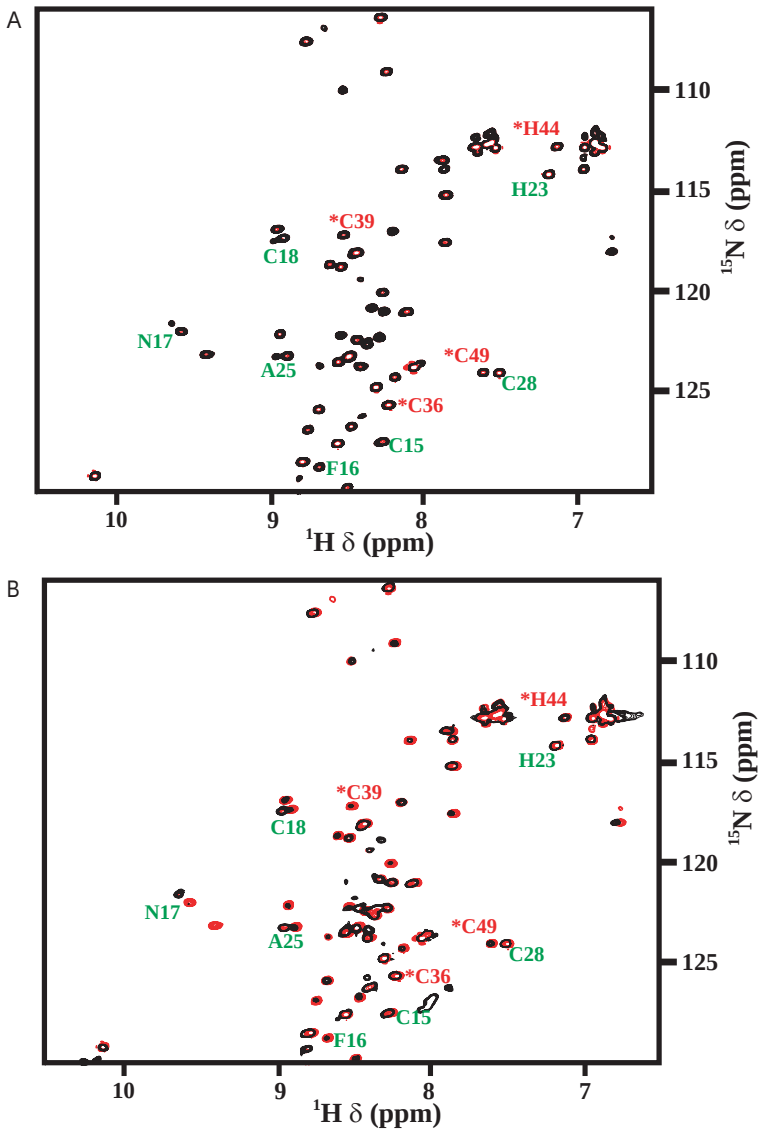
(53) Goebel, F. D.; Hemmer, R.; Schmit, J. C.; Bogner, J. R.; de Clercq, E. et al. Phase I/II dose escalation and randomized withdrawal study with add-on azodicarbonamide in patients failing on current antiretroviral therapy. *AIDS* **2001**, *15*, 33-45.

(54) Domagala, J. M.; Gogliotti, R.; Sanchez, J. P.; Stier, M. A.; Musa, K. et al. 2,2'-Dithiobisbenzamides and 2-benzisothiazolones, two new classes of antiretroviral agents: SAR and mechanistic considerations. *Drug Des. Discov.* **1997***, 15*, 49-61.

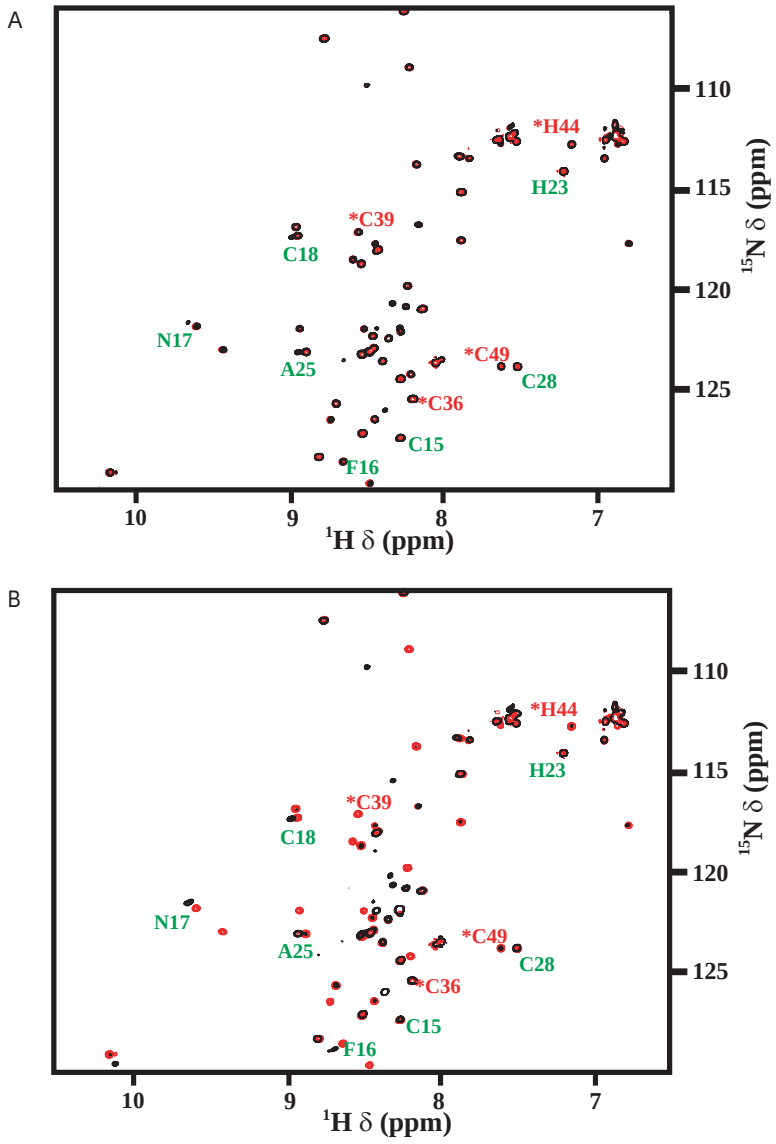
(55) Tummino, P. J.; Harvey, P. J.; McQuade, T.; Domagala, J.; Gogliotti, R. et al. The human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein zinc ejection activity of disulfide benzamides and benzisothiazolones: correlation with anti-HIV and virucidal activities. *Antimicrob. Agents Chemother.* **1997***, 41*, 394-400.

(56) Phillips, L. R.; Malspeis, L.; Tubbs, E. K.; Supko, J. G. Characterization of a novel degradation product of 2,2'-dithiobis[N-isoleucylbenzamide], an inhibitor of HIV nucleocapsid protein zinc fingers. *J. Pharm. Biomed. Anal.* **2000***, 23*, 395-402.

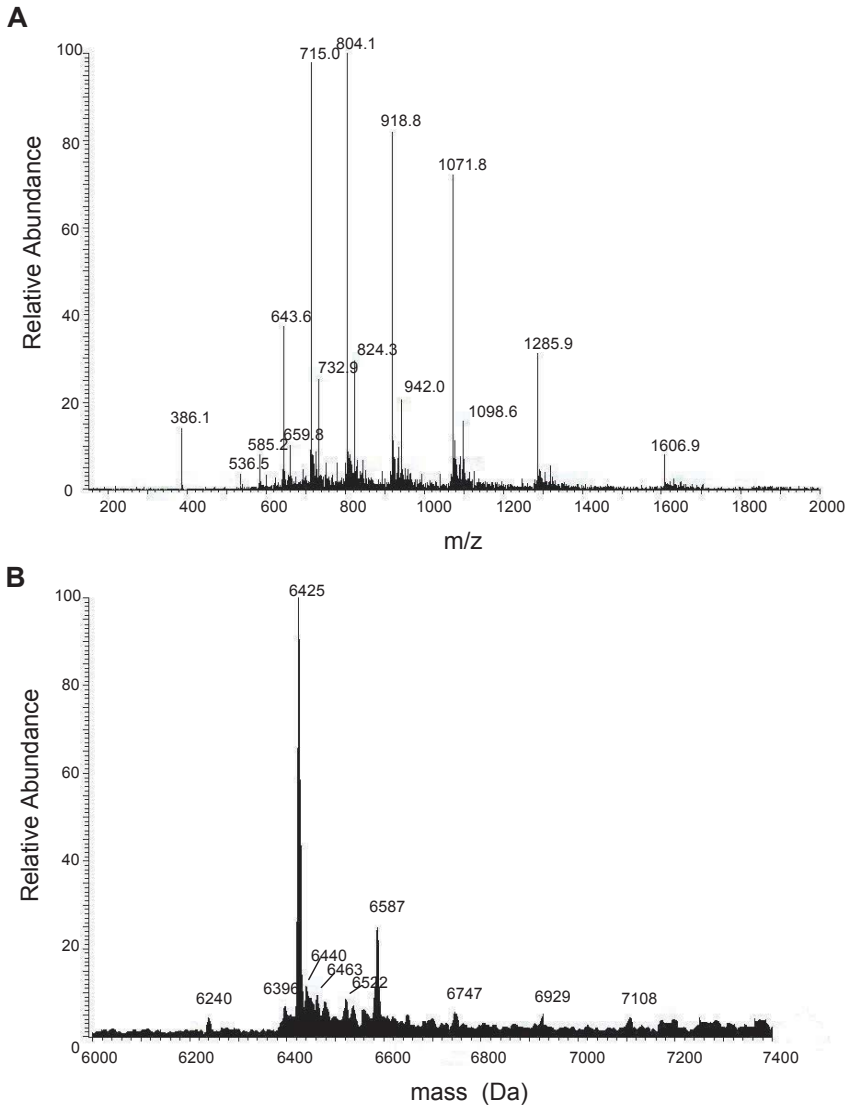
Supplementary Data



Supplementary Figure 2.1: 2D ^{1}H-^{15}N HSQC spectrum of NCp7 alone superimposed on the 2D ^{1}H-^{15}N HSQC spectrum of NCp7 immediately after compound **1** addition (A) or after 48 h of incubation (B). In both A and B, the signals from NCp7 alone are shown in red and the signals from the NCp7 with compound **1** are shown in black. The signals for zinc-coordinating residues are labeled, those from ZD1 in green and those from ZD2 in red. Also labeled in green are the signals for residues in ZD1 that shift during the incubation.



Supplementary Figure 2.2: 2D ^{1}H-^{15}N HSQC spectrum of NCp7 alone superimposed on the 2D ^{1}H-^{15}N HSQC spectrum of NCp7 immediately after compound **2** addition (A) or after 48 h of incubation (B). In both A and B, the signals from NCp7 alone are shown in red and the signals from the NCp7 with compound **2** are shown in black. The signals for zinc-coordinating residues are labeled, those from ZD1 in green and those from ZD2 in red. Also labeled in green are the signals for residues in ZD1 that shift during the incubation.



S4

Supplementary Figure 2.3: Electrospray ionization (A) and deconvoluted mass spectra (B) of NCp7 following addition of compound **1** (after 1 h). The mass of the unmodified and adducted NCp7 are 6425 and 6587 Da, respectively.

CHAPTER 3

STUDIES ON THE SPECIFICITY OF INTERACTION OF STRUCTURAL ZINC-BINDING DOMAINS WITH 2-MERCAPTOBENZAMIDE THIOESTERS[1]

[1] Miller Jenkins, L. M., Appella, E., Stahl, S. J., Inman, J. K., Legault, P., Omichinski, J. G. To be submitted to *Journal of Medicinal Chemistry*.

Abstract

The interactions of two mercaptobenzamide thioester compounds with six structural zinc-binding domains (ZBDs) have been analyzed by UV/Visible spectroscopy, NMR spectroscopy, and nucleic acid binding assays. The thioester compounds are antiviral agents that selectively modify Cys_{39} in the carboxyl-terminal ZBD ($Cys_2HisCys$) of the HIV-1 nucleocapsid protein (HIV-1 NCp7). This covalent modification results in zinc ejection and loss of protein structure and function. In our studies, the classical (Cys_2His_2) zinc finger ZBDs, the interleaved RING finger-like ZBDs of PKCδ ($Cys_2HisCys$ and $HisCys_3$) and the carboxyl-terminal ($Cys_2HisCys$) ZBD of Mouse Mammary Tumor Virus nucleocapsid protein (MMTV NCp10) were determined to be resistant to zinc ejection when exposed to the thioester compounds. In contrast, the thioester compounds were able to efficiently eject zinc from the amino-terminal ($Cys_2HisCys$) ZBD of MMTV NCp10, a $Cys_2HisCys$ ZBD from Friend of GATA-1 (FOG-1), and from both Cys_4 ZBDs of GATA-1. In all cases, zinc ejection led to a loss of protein structure as previously observed with the HIV-1 NCp7. Interestingly, the GATA-1 protein was resistant to zinc ejection by the thioester compounds when bound to its target DNA sequence, which also occurred with HIV-1 NCp7. Based on these results, it appears that both first shell and second shell interactions within the zinc-binding domain, as well as nucleic acid binding, play important roles in determining the susceptibility of various types of ZBDs to zinc ejection by the thioester compounds tested here. These studies not only provide information regarding the relative reactivity of cysteine residues within structural ZBDs, but are crucial for the design of future therapeutic agents that specifically target ZBDs such as those found in the HIV-1 NCp7 protein.

Abbreviations: ADA, azodicarbonamide; DIBA, disulfide benzamide; DTT, dithiothreitol; FOG-1, Friend of GATA-1; GST, glutathione-S-transferase; HIV-1, Human Immunodeficiency Virus Type-1; HIV-1 NCp7, HIV-1nucleocapsid protein ; HPLC, high-performance liquid chromatography; IPTG, isopropyl-β-D-thiogalactopyranoside; MBP-1, major histocompatibility binding protein-1; MMTV, Mouse Mammary Tumor Virus; MMTV NCp10, Mouse Mammary Tumor Virus nucleocapsid protein; NOBA, 3-nitrosobenzamide; PARP, poly(ADP-ribose) polymerase; PKCδ, protein kinase C δ; TFA, trifluoroacetic acid; water-sLED, water-suppressed longitudinal encode-decode; ZBD, zinc-binding domain; ZD1, amino-terminal zinc-binding domain; ZD2, carboxyl-terminal zinc-binding domain.

Introduction

Structural zinc-binding proteins are estimated to make up 3% of the human genome, and they are extremely prevalent in sequence-specific DNA-binding transcription factors[1]. The first ZBD to be described was the "classical zinc finger" from the *Xenopus* transcription factor TFIIIA[2]. TFIIIA contains nine ZBDs, and each ZBD coordinates a zinc ion using two cysteine and two histidine residues[2]. Structural zinc-binding domains (ZBDs)[1] are often similarly classified according to the distribution of cysteine and histidine residues used to coordinate the zinc ion (Cys_2His_2, Cys_3His, and Cys_4)[3]

Numerous structural studies indicate that the bound zinc ion serves to stabilize the three-dimensional fold of the ZBD by participating in the formation of local secondary structure elements. Various structural motifs have now been associated with the common classes of ZBD. Most retroviral *gag* polyproteins contain at least one copy of a $Cys_2HisCys$ ZBD that is thought to be essential for specific binding to nucleic acid target sites in the retroviral genome[4,5].

In the Human Immunodeficiency Virus Type-1 (HIV-1), there are two retroviral $Cys_2HisCys$ ZBDs in the 55-amino acid HIV-1 NCp7 protein of the gag polyprotein. The two ZBDs are connected by a seven amino acid basic linker region, and the ZBDs are essential for the nucleic acid binding properties of HIV-1 NCp7[4-7]. NMR solution studies of HIV-1 NCp7 have shown that the two ZBDs form small globular structures that participate in specific RNA base recognition[8-11]. The RNA binding properties of HIV-1 NCp7 are absolutely essential for the survival and spread of the HIV-1 virus. The unique structural and functional properties of the HIV-1 NCp7 ZBDs, and the mutationally non-permissive nature of HIV-1 NCp7, suggest that HIV-1 may not easily become resistant to drugs targeted against this protein[12-16]. For this

reason, HIV-1 NCp7 has become an important target for the development of new drug-based antiviral therapy.

Several different classes of electrophilic compounds have been reported to directly function as antiviral agents through their ability to inhibit the activities of HIV-1 NCp7[17-28]. Most of the compounds function by interacting with cysteine residues in the ZBDs and ejecting zinc from HIV-1 NCp7 (zinc ejectors)[22]. Their activity is highly dependent on the electrophilic nature of the compounds, and zinc ejection leads to a loss of both protein structure and function[19,22,29]. The initial zinc ejectors were 3-nitrosobenzamide (NOBA) compounds[18]. Although the NOBA compounds did eject zinc from HIV1 NCp7 and inhibit HIV-1 replication, they lacked specificity for the ZBDs of HIV-1 NCp7 and were highly cytotoxic. Additional studies led to the development of compounds that were more specific for the ZBDs of HIV-1 NCp7 and, thus, less cytotoxic. The new compounds included the disulfide benzamide (DIBA) and azodicarbonamide (ADA) compounds[20,21]. The DIBA and ADA compounds inhibit HIV-1 replication and appear to selectively eject zinc from the carboxyl-terminal ZBD (ZD2) following reaction with Cys_{49} of HIV-1 NCp7[22,30]. In contrast to the NOBA compounds, the DIBA compounds did not inhibit the specific nucleic acid binding activity of some cellular proteins containing structural ZBDs, namely the poly(ADP ribose) polymerase (PARP), Sp1, and GATA-1 proteins[23]. Despite their antiviral activity and selectivity, the DIBA compounds are unstable in aqueous solutions, thus limiting their potential utility in drug therapy[20,31].

Recently, thioester compounds that selectively target the HIV-1 NCp7 protein have been developed in attempts to overcome the stability problems associated with the DIBA compounds[25,27,28,32-34]. Like the DIBA compounds, the thioester compounds display potent antiviral activity and low cellular toxicity[25,32,34]. In addition, the thioester compounds were

shown to have no effect on Sp1 binding to DNA, strongly suggesting they selectively target the ZBDs of HIV-1 NCp7[25]. We have recently shown that two highly active thioester compounds, compounds **1** and **2** (Figure 3.1A), selectively eject zinc from NCp7 by covalently modifying Cys_{39} in ZD2 of HIV-1 NCp7 (Figure 3.1B)[35]. This modification leads to loss of protein structure and loss of sequence-specific RNA binding by HIV-1 NCp7. Interestingly, the loss of HIV-1 NCp7 RNA binding properties requires pre-incubation of the thioesters with unbound HIV-1 NCp7 prior to conducting the RNA-binding assays, and RNA-bound NCp7 is resistant to modification by these thioester compounds[35].

As the modification of a specific cysteine residue in HIV-1 NCp7 by thioester compounds leads to a loss of NCp7 function, it is important to determine if these compounds are also capable of modifying cysteine residues in structural ZBDs from cellular proteins (Figure 3.1B). Theoretical studies have attempted to predict the relative reactivity of different classes of ZBDs towards electrophilic compounds such as ADA, DIBA, and the thioester compounds[23,30]. Unfortunately, very few experimental studies have been conducted to either support or refute the theoretical predictions. Most of the experimental studies have examined the interactions of electrophilic compounds, such as the DIBAs and thioesters, with various classes of structural ZBDs only in the presence of nucleic acids[23,25]. In the case of the interaction of HIV-1 NCp7 with compounds **1** and **2**, we observed dramatically different rates of reactivity depending on whether the protein was in the free or RNA-bound state[35]. Additional studies comparing the reactivity of structural ZBDs with these electrophilic compounds in both the free and RNA-bound states are important to more thoroughly understand the reactivity of ZBDs with these electrophilic compounds. This information is crucial for the design of new antiviral agents with improved *in vivo* activity and decreased cytotoxicity.

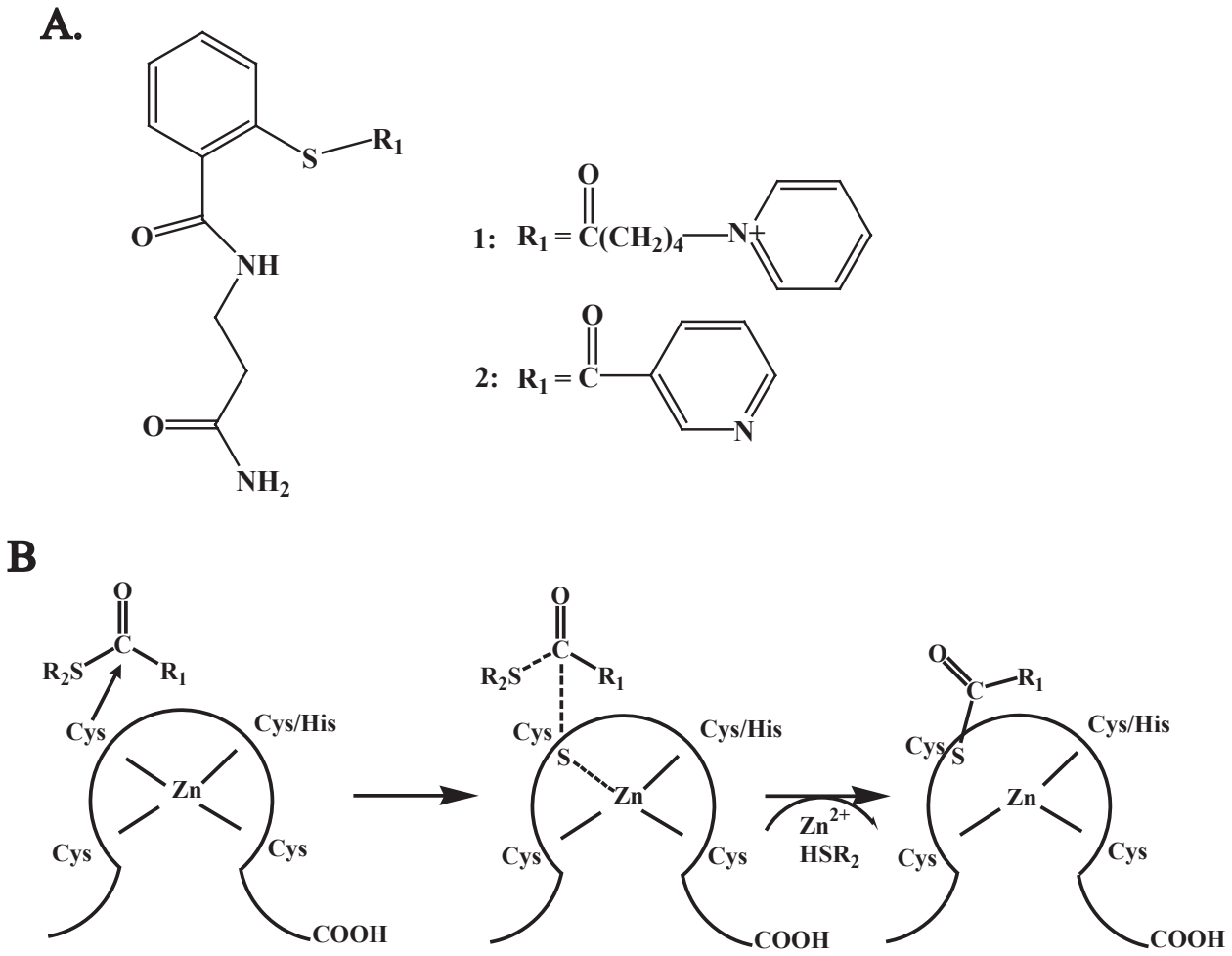


Figure 3.1: A) Chemical structure of thioester compounds used in this study. B) General mechanism for the covalent modification of a cysteine residue in ZBDs by thioester antiviral compounds **1** and **2**. Note that the identity of the first modified cysteine can vary and that the other cysteines of the ZBD can be subsequently modified.

In this manuscript, we have analyzed the reactivity of thioester compounds **1** and **2** with structural ZBDs from six different proteins. These six proteins represent three different classes (Cys_2His_2, Cys_3His, Cys_4) of ZBDs with five distinct structural motifs. We have used UV/Visible spectroscopy and NMR spectroscopy to evaluate the ability of the thioester compounds to eject metal from the various classes of ZBDs and to alter their structure. In addition, we have determined the ability of these thioester compounds to disrupt the sequence-specific nucleic acid binding properties of selected ZBDs. The results demonstrate that the thioester compounds show reactivity towards specific classes of structural ZBDs found in cellular proteins. The relative reactivity of the ZBDs with the thioester compounds is determined by both first and second shell interactions within the zinc-binding site. In addition, the interaction of the thioester compounds is greatly inhibited when the ZBDs are bound to their target nucleic acid. These results help to explain the excellent antiviral properties and the low cytotoxicity observed with thioester compounds **1** and **2** and provide important information that can be incorporated into the design of future antiviral agent specifically directed towards HIV-1 NCp7[32].

Experimental Procedures

Protein Subcloning- The coding sequences for the *Drosophila* GAGA ZBD ($GAGA_{310-372}$ - amino acids 310-372)[37], the double ZBD of human GATA-1 protein ($hGATA\text{-}1_{200-317}$ - amino acids 200-317)[38], the carboxyl-terminal ZBD of chicken GATA-1 protein ($cGATA\text{-}1_{158-223}$ - amino acids 158-223), and a glutathione-S-transferase (GST) fusion of the second cysteine-rich region of human protein kinase C delta protein ($PKC\delta_{231-280}$ - amino acids 231-280)[39] were transformed and expressed in host strain BL21(DE3) (Novagen, WI), as previously described.

The coding sequence for the double ZBD of the major histocompatibility binding protein-1 (MBP-$1_{2085\text{-}2142}$ - amino acids 2085-2142) was PCR amplified from pL-EB(F12)[40]. The amplified coding sequence was subcloned into pET3a (Novagen, WI) using the *Nde*I – *Bam*HI restriction sites and expressed in BL21(DE3)pLysS (Novagen, WI). The coding sequence for Mouse Mammary Tumor Virus NCp10 (C3H strain) was PCR amplified from the tetracycline-resistant plasmid p202 containing the partial gag-pol region[41] obtained from the American Type Culture Collection (#45005). The MMTV NCp10 sequence was subcloned into the *Bam*HI and *Nco*I sites of pET11a (Novagen, WI)[42] and expressed in *Escherichia coli* host strain Rosetta (DE3) (Novagen, WI). The coding sequence for the ninth ZBD of mouse Friend of GATA-1 protein (FOG-$1_{950\text{-}995}$ - amino acids 950-995) was subcloned into pGEX5X-1 (Amersham, NJ) and expressed as a GST fusion protein in BL21(DE3).

Protein Expression and Purification- Cells expressing $GAGA_{310\text{-}372}$[43], hGATA-$1_{200\text{-}317}$[38], cGATA-$1_{158\text{-}223}$[44] or MMTV NCp10 were grown in Luria Broth overnight at 37 °C. Uniformly (>98%) ^{15}N-labeled cGATA-$1_{158\text{-}223}$, ^{15}N-labeled hGATA-$1_{200\text{-}317}$, or ^{15}N-labeled MMTV NCp10 were obtained by growing the cells in modified minimal media containing ^{15}N-labeled ammonium chloride as the sole nitrogen source. Protein expression was induced for 4 h with 0.66 mM or 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) for cultures grown in minimal media or Luria Broth, respectively. The cells were resuspended in 25 mM Tris pH 8.0, 1 mM EDTA, 2 mM dithiothreitol (DTT), and 6 mM benzamidine. The cells were then lysed by French press and centrifuged at 100,000 x g for 45 min. The supernatant was applied to a DEAE-Sepharose Fast Flow (Amersham Biosciences, NJ) column (300 mL bed volume), equilibrated with Buffer A (25 mM Tris pH 8.0, 1 mM EDTA, and 2 mM DTT) and eluted using

a gradient (0-100% B over 1500 ml) of Buffer B (25 mM Tris pH 8.0, 1 mM EDTA, 2 mM DTT, and 1 M NaCl). The pooled protein fractions were then applied to a SP-Sepharose Fast Flow (Amersham Biosciences, NJ) column (300 ml bed volume), equilibrated with Buffer A and eluted using a gradient (0-100% B over 1500 ml) of Buffer B. The fractions containing the desired protein were pooled and subsequently purified on a C-4 (for $GAGA_{310-372}$, hGATA-$1_{200-317}$, and cGATA-$1_{158-223}$) or on a C-8 (for MMTV NCp10) reverse phase (Vydac) high performance liquid chromatography (HPLC) column. The following acetonitrile gradients in 0.05% aqueous trifluoroacetic acid (TFA) were used: 20-40% for $GAGA_{310-372}$, 20-40% for cGATA-$1_{158-223}$, 15-35% for hGATA-$1_{200-317}$, and 15-35% for MMTV NCp10. The eluted proteins were subsequently flash frozen and lyophilized.

^{15}N-labeled MBP-$1_{2085-2142}$ was obtained by growing the cells in modified minimal media containing ^{15}N-labeled (>98%) ammonium chloride as the sole nitrogen source. Protein expression was induced for 4 h with 0.66 mM IPTG. The cells were resuspended in 25 mM sodium phosphate pH 7.0, 1 mM EDTA, and 2 mM DTT. The cells were then lysed by French press and centrifuged at 19,500 x g for 30 min. ^{15}N-labeled MBP-$1_{2085-2142}$ was purified from an inclusion body by resuspending the pellet in 25 mM sodium phosphate pH 7.0, 0.5% Triton X-100, 2 M urea, and 2 mM DTT and centrifuging at 19,500 x g for 30 min. The pellet was resuspended in 6 M guanidine hydrochloride and incubated at 50 ºC for 10 min. This suspension was centrifuged at 100,000 x g for 45 min and the resulting supernatant dialyzed into 5% aqueous acetic acid containing 5 mM DTT. The ^{15}N-labeled MBP-$1_{2085-2142}$ was further purified from the dialysate on a C-4 reverse phase (Vydac) HPLC column using a 20-40% gradient of acetonitrile in 0.05% aqueous TFA. The eluted protein was flash frozen and lyophililzed.

PKC$\delta_{231-280}$ and FOG-$1_{950-995}$ were prepared as GST-fusion proteins. Cells expressing GST-PKC$\delta_{231-280}$ or GST-FOG-$1_{950-995}$ were grown in Luria Broth overnight at 37°C. Protein expression was induced for 4 h with 1 mM IPTG at 30 °C. The cells were resuspended in EBC Buffer (50 mM Tris pH 8.0, 120 mM NaCl, 0.5% NP-40, and 2 mM DTT), then lysed by French press and centrifuged at 100,000 x g for 45 min. The supernatant was incubated for 1 h at 4 ºC with glutathione sepharose (Amersham Biosciences, NJ) equilibrated in EBC Buffer. The bound GST-PKC$\delta_{231-280}$ was equilibrated in Phosphate Buffered Saline and digested with 100 U thrombin (Calbiochem, CA) for 2 h at 25 ºC to cleave PKC$\delta_{231-280}$ from the GST tag. The bound GST-FOG$1_{950-995}$ was equilibrated in 50 mM Tris pH 8.0, 100 mM NaCl, and 1 mM $CaCl_2$ and digested with Factor Xa (Novagen, WI) for 16 h at 25 ºC to cleave FOG$1_{950-995}$ from the GST tag. The cleaved PKC$\delta_{231-280}$ or FOG-$1_{950-995}$ was dialyzed into 10% acetic acid and subsequently purified on a C-4 reverse phase (Vydac) HPLC column using a 25-45% or 15-35% gradient of acetonitrile in 0.05% aqueous TFA, respectively. The eluted protein was subsequently flash frozen and lyophilized.

Sample Preparation - Purified proteins were refolded using the following procedure. First, the lyophilized material was resuspended in 0.05% TFA. Five equivalents of $ZnCl_2$ (zinc-refolded protein) or $CoCl_2$ (cobalt-refolded protein) were then added to the protein solution as a 50 mM metal solution in 0.05% TFA. The pH of the solution was titrated to pH 6.0 with 0.2 M NaOH. The zinc-refolded MMTV NCp10 and MBP-$1_{2085-2142}$ were then flash frozen and lyophilized. Prior to use, zinc-refolded MMTV NCp10 or MBP-$1_{2085-2142}$ (1 mM) were resuspended in NMR Buffer A (20 mM sodium phosphate pH 7.0) in 90% H_2O/10% D_2O. The zinc-refolded PKC$\delta_{231-280}$, cGATA-$1_{158-223}$, and hGATA-$1_{200-317}$ were concentrated with an Amicon Centricon-3

concentrator (Millipore, MA) and exchanged into NMR Buffer A in 90% H_2O/10% D_2O. The buffer of the cobalt-refolded proteins was adjusted to 20 mM sodium phosphate pH 7.0. The cobalt-refolded proteins were then used immediately after refolding.

The complex of ^{15}N-labeled cGATA-$1_{158-223}$ (1 mM) with AGATAA DNA (1 mM) was prepared by titrating ^{15}N-labeled $cGATA_{158-223}$ into double-stranded AGATAA DNA (Midland Certified Reagents, TX) in NMR buffer A in 90% H_2O/10% D_2O. The complex was concentrated to 500 µl using an Amicon Centricon-3 concentrator (Millipore, Bedford, MA). The sequence of the top strand of the double-stranded AGATAA DNA used for NMR studies was 5' GTTGCAGATAAACATT 3'.

Synthesis of Compounds - Compound **1** (N-[2-(5-pyridinovalerylthio) benzoyl]-β-alaninamide bromide) and compound **2** (N-[2-nicotinoylthiobenzoyl]-β-alaninamide hydrochloride) were synthesized as described[25,33]. The thioester compounds were lyophilized in aliquots such that the concentration was 1 mM in a volume of 500 µl and stored at -20 ºC.

UV/Visible Spectroscopy- For UV/Visible spectroscopy studies, cobalt-refolded proteins (150 µM) were incubated with compound **2** (150 µM) at 25 ºC. The UV/Visible spectrum was recorded from 220-800 nm every 0.5 h for 3 h using a Shimadzu UV-1601 spectrophotometer equipped with PC control via the UVPC software (Shimadzu Scientific Instruments, MD). The wavelengths of maximum absorbance (λ_{max}) of the tetrahedrally-coordinated cobalt ions of each protein are listed in Table 3.1. The absorbance at λ_{max} was specifically monitored and compared against a control sample in which no thioester compound was added. Experiments were repeated a minimum of three times. Normalized absorbance values were calculated by subtracting the

absorbance monitored by that of the control sample. The normalized absorbance at λ_{max} was plotted versus time, and the initial rate of absorbance loss (0 – 3 h) was obtained from the slope of the linear regression analysis.

NMR Spectroscopy- Zinc-refolded ^{15}N-labeled PKC$\delta_{231-280}$, ^{15}N-labeled hGATA-$1_{200-317}$, ^{15}N-labeled cGATA-$1_{158-223}$, ^{15}N-labeled cGATA-$1_{158-223}$/AGATAA DNA complex, ^{15}N-labeled MMTV NCp10, or ^{15}N-labeled MBP-$1_{2085-2142}$ (1 mM) were incubated with an equimolar amount (1 mM) of compound **1** at 25 ºC for 48 h. The samples were analyzed before and at various times after addition of compound **1** on Varian ^{UNITY}Inova 500 MHz and 600 MHz spectrometers equipped with an HCN triple resonance probe with an actively shielded z-gradient. Three experiments were recorded at each time point: a 1D ^{15}N-decoupled watergate[45], a 1D water-sLED[46], and a 2D ^{1}H-^{15}N HSQC[47]. 1D difference water-sLED spectra[46] were recorded as previously described[35]. The 2D ^{1}H-^{15}N HSQC spectra were processed with the NMRPipe package[48] and analyzed with PiPP[49]. The measured signal intensities were plotted against time (0-48 h) and the initial rate of signal intensity loss was obtained by linear regression. ^{1}H and ^{15}N chemical shift assignments for MMTV NCp10[42], hGATA-$1_{200-217}$ (JGO, unpublished data) and cGATA-$1_{158-223}$ complexed with AGATAA DNA (JGO, unpublished data) and the ^{1}H chemical shift assignments for MBP-$1_{2085-2142}$[50] were obtained from previous studies.

Gel Mobility Shift Assay- Zinc-refolded cGATA-$1_{158-223}$ (100 nM) was incubated with either compound **1** (500 nM) or **2** (500 nM) in 50 mM Tris pH 7.0, 0.125% NP-40, 2 mM EDTA, 20 μg/ml poly(dI-dC), and 10% glycerol at 25 ºC for 0 min, 4 h, and 24 h. After each incubation period, an aliquot of cGATA-$1_{158-223}$ (final cGATA-$1_{158-223}$ concentration 5 nM) was removed

and incubated with 1 nM 5'-^{32}P-labeled double-stranded AGATAA DNA (Integrated DNA Technologies, Inc, IA) for 30 min at 25 ºC. The sequence of the AGATAA DNA used for gel-binding studies was: 5'AGCTTCGGTTGCAGATAAACATTGAATTCA 3'. The top-strand of the double-stranded DNA was 5'-end-labeled with γ-(^{32}P)ATP using T4 polynucleotide kinase (New England Biolabs, MA) according to the manufacturer's instructions.

The cGATA-$1_{158-223}$/DNA binding reactions were analyzed by 8% native gel electrophoresis in 10 mM Tris, 10 mM HEPES pH 7.5, and 1 mM EDTA at 150 V and 4 ºC. The 5'-^{32}P-labeled DNA was detected using a Storm PhosphorImager (Amersham Biosciences, NJ). The density of the bands of the free and bound DNA was determined with the ImageQuant program (Molecular Dynamics). The percent free DNA was calculated as: $D_f/(D_b+D_f)$ where D_f is the normalized density of the free DNA and D_b is the normalized density of the bound DNA. Normalized densities were calculated by subtracting out the measured densities of a control binding reaction.

Results

$Cys_2HisCys$ ZBDs interact with thioester compounds - Thioester compounds **1** and **2** have been shown to selectively eject tetrahedrally-coordinated zinc from the $Cys_2HisCys$ ZD2 of NCp7 by covalently modifying Cys_{39} (Figure 3.1)[35]. Here, we analyze the effect of compounds **1** and **2** on a variety of structural ZBDs from cellular and retroviral proteins. The metal ejecting properties of compound **2** towards several types of ZBDs was initially determined by UV/Visible spectroscopy using cobalt-refolded ZBDs (Table 3.1). Unlike zinc, cobalt is a spectroscopically active metal that will give rise to absorption maxima in the visible region of the spectrum as a consequence of *d-d* ligand field transitions when it is tetrahedrally coordinated[51]. The

Table 3.1: Rate of UV/Visible absorbance loss of cobalt-refolded proteins when incubated with compound **2**.

Protein	Type of ZBD	λ max (nm)	Rate (AU/h)
HIV-1 NCp7[35]	2 CCHC	642, 698	36 ± 3.7
MMTV NCp10	2 CCHC	642, 698	12.1 ± 0.5
FOG-$1_{950\text{-}995}$	CCHC	642, 698	17.9 ± 0.5
$GAGA_{310\text{-}372}$	CCHH	634	1 ± 0.7
$PKC\delta_{231\text{-}280}$	HCCC, CCHC	665	2.8 ± 0.6
$cGATA_{158\text{-}223}$	2 CCCC	741, 705	15.2 ± 0.2

wavelengths of the absorption maxima are highly dependent upon the specific makeup of the four cobalt bound ligands, in our case a variable mixture of sulfurs (cysteine) and/or nitrogens (histidine) depending on the ZBD. In each case, the wavelength of the absorbance maxima (λ_{max}) varies with the type of ZBD that is binding the cobalt (Table 3.1).

We first analyzed the interaction of compound **2** on $Cys_2HisCys$ ZBDs from two proteins, one cellular (FOG-$1_{950-995}$) and one retroviral (MMTV NCp10). The two ZBDs use the same combination of cysteine and histidine ligands for metal binding as HIV-1 NCp7. MMTV NCp10 is the nucleocapsid protein from Mouse Mammary Tumor Virus (MMTV) and, like HIV-1 NCp7, it contains two retroviral $Cys_2HisCys$ domains. The two ZBDs in MMTV NCp10 have the same amino acid spacing between the metal binding ligands as those found in HIV-1 NCp7. The two ZBDs of MMTV NCp10 are structurally very similar, but not identical to, the two ZBDs of HIV-1 NCp7[42]. FOG-$1_{950-995}$ is a $Cys_2HisCys$ domain from the mouse Friend of GATA-1 (FOG-1) protein. The composition of the FOG-$1_{950-995}$ metal-binding ligands is the same as HIV-1 NCp7, but the amino acid spacing between the binding residues is quite different than the in HIV-1 NCp7. In addition, the $Cys_2HisCys$ domain of FOG-1 adopts a $\beta\beta\alpha$ fold in the presence of zinc ions[52]. The three-dimensional structure of this ZBD is closely related to the structures of the Cys_2His_2 classical zinc fingers and bears little resemblance to the structure of the known retroviral ZBDs.

Compound **2** was incubated with cobalt-refolded MMTV NCp10 and cobalt-refolded FOG-$1_{950-995}$. As was previously observed for cobalt-refolded HIV-1 NCp7, there was a significant loss of absorbance in the visible region of the UV/Visible spectra with both cobalt-refolded proteins over a period of 3 h (Table 3.1). The λ_{max} are consistent with previously reported values[53]. The loss of absorbance is due to ejection of the tetrahedrally-coordinated

cobalt by compound **2.** The rate of absorbance loss for MMTV NCp10 was similar to that observed for FOG-$1_{950\text{-}995}$ (Table 3.1). The rate of absorbance loss for MMTV NCp10 was ~3 times lower and that of FOG-$1_{950\text{-}995}$ was ~2 times lower than what had been observed with HIV-1 NCp7[35] (Table 3.1). These results clearly indicate that compound **2** is capable of ejecting metal from the $Cys_2HisCys$ ZBDs of both MMTV NCp10 and $FOG1_{950\text{-}995}$, although the rate of metal ejection appears slightly reduced when compared with NCp7.

Though it was clear that the thioester compounds were able to eject tetrahedrally-coordinated cobalt from both MMTV NCp10 and FOG-$1_{950\text{-}995}$, the UV/Visible spectroscopy experiments failed to provide detailed information on the mechanism of metal ejection. Therefore, NMR spectroscopy experiments were used in an attempt to obtain a more detailed understanding of the mechanism of metal ejection by the thioesters. Zinc-refolded ^{15}N-labeled MMTV NCp10 was incubated with compound **1** for 48 h at 25 ºC. The 2D ^{1}H-^{15}N HSQC spectrum collected immediately after addition of compound **1** to the protein was identical to that of free MMTV NCp10 (data not shown). Thus, there were no changes in the protein amide signals, indicating that the MMTV NCp10 did not undergo a significant structural change upon addition of compound **1**. This result also suggests that a stable complex was not formed between the protein and thioester compound. Indeed, chemical shift changes would have been expected if MMTV NCp10 and compound **1** were forming a stable complex. This result is similar to what we previously observed when HIV-1 NCp7 was mixed with either compound **1** or compound **2**[35].

During the 48 h incubation with compound **1**, no change in chemical shifts were observed in the 2D ^{1}H-^{15}N HSQC spectra of MMTV NCp10. The primary changes observed in the 2D ^{1}H-^{15}N HSQC spectra were loss of signal intensity for several amide signals (Supplementary Figure 3.1). The loss of MMTV NCp10 amide signal intensity in the 2D ^{1}H-^{15}N HSQC spectra was

seen specifically for residues in the amino-terminal zinc-binding domain (ZD1). After 24 h, the signals for Cys_{31} and His_{39} had only 75% of their original signal intensity, whereas the equivalent zinc-coordinating residues in ZD2 (Cys_{58} and His_{66}) retained 95% of their original signal intensity (Figure 3.2). In general, the zinc-coordinating residues from ZD2 (Cys_{58}, Cys_{61}, His_{66}, and Cys_{71}) showed only a small overall change in amide signal intensity over the 48 h incubation period (< 5%), whereas the zinc-coordinating residues from ZD1 (Cys_{31}, Cys_{34}, His_{39}, Cys_{44}) lost > 55% of their original amide signal intensity over the same period (Figure 3.2). The average rate of signal intensity loss for the zinc-coordinating residues of ZD1 was 7.5 times faster than that observed for ZD2 (Table 3.2). This disparity in the signal intensity changes between ZD1 and ZD2 indicates that compound **1** preferentially interacts with ZD1 over ZD2. This result is very similar to what was observed with HIV-1 NCp7 and compound **1,** except that for the HIV-1 NCp7 compound **1** preferentially interacted with ZD2 over ZD1[35].

The interleaved ZBD of PKCδ does not interact with thioester compounds – The next zinc-binding domains to be studied were the interleaved ZBDs from PKCδ. In $PKC\delta_{231-280}$, the interleaved ZBDs can actually be divided into two ZBDs formed by alternating pairs of Cys and His ligands. The metal binding ligands are arranged as His-Cys_2-Cys_2-His/Cys-Cys[39]. The first Cys_2 pair chelates a zinc ion with the His/Cys pair to form a Cys_2HisCys ZBD. The second zinc ion is bound by the first His residue, the second Cys_2 pair, and the final Cys residue to form a $HisCys_3$ ZBD[54]. The two ZBDs of $PKC\delta_{231-280}$ are thus a Cys_2HisCys domain and a $HisCys_3$ domain. These domains have an identical combination of first shell ligands as the ZBDs in HIV-1 NCp7, MMTV NCp10, and FOG-$1_{950-995}$. We first incubated cobalt-refolded $PKC\delta_{231-280}$ with compound **2** and monitored changes in the UV/Visible absorbance of the tetrahedrally-

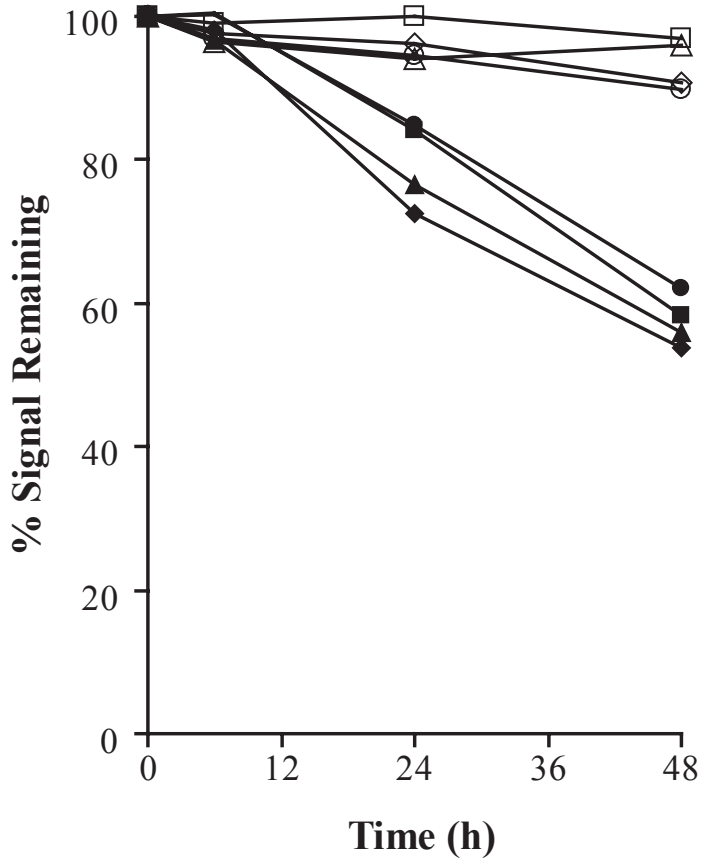


Figure 3.2: Changes in 2D ^{1}H-^{15}N HSQC signal intensity of MMTV NCp10 zinc-coordinating residues after addition of compound **1**. In the plot, ◆ is Cys_{31}, ■ is Cys_{34}, ▲ is Cys_{39}, ● is Cys_{44}, ◇ is Cys_{58}, □ is Cys_{61}, Δ is His_{66}, ○ is Cys_{71}.

Table 3.2: Average rate of loss of 2D ^{1}H-^{15}N HSQC signal intensity for zinc-coordinating residues during incubation with compound **1**.

Protein	ZBD	Protein State	Rate (IU/h) ZD1		ZD2	
HIV-1 NCp7[35]	CCHC	Free	0.11	± 0.18	1.20	± 0.10
MBP-$1_{2085\text{-}2142}$	CCHH	Free	0.04	± 0.04	0.09	± 0.04
MMTV NCp10	CCHC	Free	0.93	± 0.06	0.12	± 0.06
hGATA-$1_{200\text{-}317}$	CCCC	Free	1.90	± 0.30	2.40	± 0.30
$cGATA_{158\text{-}223}$	CCCC	Complex	N.D.	N.D.	0.10	± 0.02

coordinated cobalt. Over the course of the 3 h incubation, very small changes in absorbance were observed (Table 3.1). The rate of absorbance loss of $PKC\delta_{231-280}$ was 6.5 times lower than that observed for FOG-$1_{950-995}$ and 4.5 times lower than that observed for MMTV NCp10 (Table 3.1). This indicates that compound **2** was able to eject only a minimal amount of metal from the interleaved ZBDs of $PKC\delta_{231-280}$. This result is interesting, as the interleaved zinc-binding domains of $PKC\delta_{231-280}$ have the same distribution of Cys and His zinc-coordinating ligands as FOG-$1_{950-995}$, MMTV NCp10, and HIV-1 NCp7.

Next, NMR spectroscopy was used to examine the interactions between $PKC\delta_{231-280}$ and the thioester compounds in more detail. Zinc-refolded ^{15}N-labeled $PKC\delta_{231-280}$ was incubated with compound **1** for 48 h at 25 ºC. The 2D ^{1}H-^{15}N HSQC spectrum recorded immediately after addition of compound **1** showed no changes to the $PKC\delta_{231-280}$ chemical shifts (Figure 3.3). This indicates that the conformation of $PKC\delta_{231-280}$ remained intact after addition of compound **1** and that the thioester compound did not form a stable complex with $PKC\delta_{231-280}$.

The 2D ^{1}H-^{15}N HSQC spectra recorded throughout the 48 h incubation period with compound **1** revealed only a minor decrease in the intensity of $PKC\delta_{231-280}$ amide signal intensity (Figure 3.3). These results confirm what was observed by UV/Visible spectroscopy with the cobalt-refolded $PKC\delta_{231-280}$; the thioester compounds are unable to eject significant quantities of tetrahedrally-coordinated metal from $PKC\delta_{231-280}$ over a 48 hour time period.

Thioester compounds do not react with classical Cys_2His_2 ZBDs - In order to examine the reactivity of the thioesters with other ZBDs, we next incubated compound **2** with the cobalt-refolded ZBD (Cys_2His_2) from the DNA-binding protein GAGA ($GAGA_{310-372}$). As described for MMTV NCp10 and FOG-$1_{950-995}$, we monitored changes in the UV/Visible spectrum of

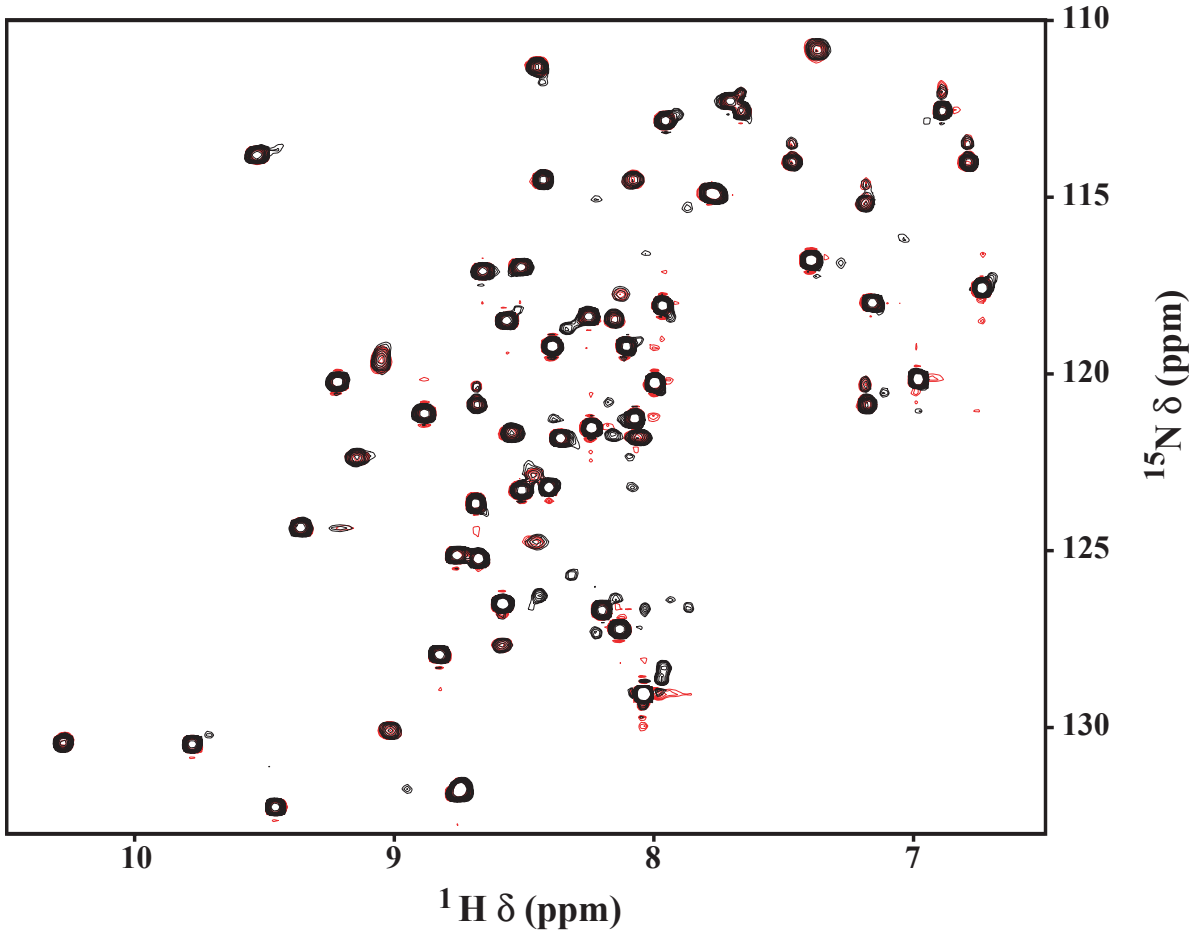


Figure 3.3: 2D ^{1}H-^{15}N HSQC spectrum of free PKCδ$_{231–280}$ (red) superimposed on the 2D ^{1}H-^{15}N HSQC spectrum of PKCδ$_{231–280}$ after 48 hours of incubation with compound **1** (black).

cobalt-refolded $GAGA_{310-372}$ over a period of 3 h. In the case of the cobalt-refolded $GAGA_{310-372}$, we failed to observe any significant changes in the absorbance at λ_{max} due to ejection of the tetrahedrally-coordinated cobalt (Table 3.1). Based on this result, we concluded that compound **2** was unable to eject coordinated cobalt from the Cys_2His_2 ZBD of $GAGA_{310-372}$ to any significant extent over the period studied.

Since compound **2** was unable to eject cobalt from $GAGA_{310-372}$, we next used NMR spectroscopy to look at interactions involving a second classical Cys_2His_2 zinc-binding protein, MBP-1, with compound **1**. The goal of these studies was to determine if other Cys_2His_2 ZBDs were also resistant to metal ejection by thioester compounds, as well as examine in more precise detail if there were any interactions between the thioester compounds and the proteins that did not involve metal ejection. $MBP\text{-}1_{2085-2142}$ contains two tandem Cys_2His_2 ZBDs separated by a seven amino acid linker[50]. Zinc-refolded ^{15}N-labeled $MBP\text{-}1_{2085-2142}$ was incubated with compound **1** for 48 h at 25 ºC. 2D ^{1}H-^{15}N HSQC experiments were used to observe changes in the ^{1}H-^{15}N correlations in the protein over time. Upon initial addition of compound **1** to zinc-refolded ^{15}N-labeled $MBP\text{-}1_{2085-2142}$, there was no change in amide signal intensity and chemical shift in the 2D ^{1}H-^{15}N HSQC spectrum (data not shown). This result indicates that there was no structural change in $MBP\text{-}1_{2085-2142}$ upon addition of compound **1**. In addition, the results also demonstrate $MBP\text{-}1_{2085-2142}$ and compound **1** do not form a stable complex.

The $MBP\text{-}1_{2085-2142}$ amide signals in the 2D ^{1}H-^{15}N HSQC spectra did not change significantly over the 48 h incubation period (Figure 3.4). If compound **1** were interacting with $MBP\text{-}1_{2085-2142}$, differences in the intensity or chemical shift of the amide signals would be expected, similar to what we observed for MMTV NCp10. Thus, the lack of change in the amide

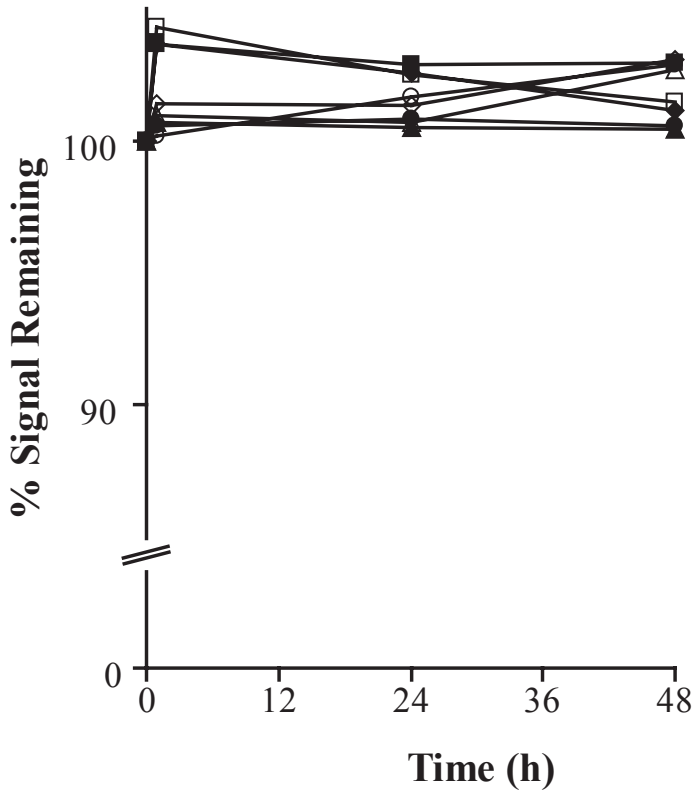


Figure 3.4: Changes in 2D ^{1}H-^{15}N HSQC signal intensity of MBP-$1_{2085\text{-}2142}$ zinc-coordinating residues after addition of compound **1**. In the plot, ◆ is Cys_{35}, ■ is Cys_{38}, ▲ is His_{51}, ● is His_{55}, ◇ is Cys_{63}, □ is Cys_{66}, Δ is His_{79}, ○ is His_{85}.

signals of MBP-$1_{2085-2142}$ demonstrates that the protein did not undergo any conformational change during the incubation period with compound **1**. It is unlikely, then, that compound **1** was able to interact with MBP-$1_{2085-2142}$ to any degree, similar to what we observed in the UV/Visible spectroscopy experiments with Cys_2His_2 ZBD of $GAGA_{310-372}$. Thus, we did not observe significant metal ejection from these ZBDs, suggesting that the classical Cys_2His_2 ZBDs are less susceptible to covalent modification by the thioester compounds then the $Cys_2HisCys$ ZBDs.

GATA Cys_4 ZBDs interact with thioester compounds - In the previous experiments, we examined the interaction of thioester compounds **1** and **2** with structural ZBDs containing either two or three cysteine residues. We next looked at the ability of compounds **1** and **2** to interact with structural ZBDs that bound metal with four cysteine residues (Cys_4). Compound **2** was incubated with cobalt-refolded $cGATA_{158-223}$, a fragment of the chicken GATA-1 transcription factor that contains one Cys_4 ZBD[44,55]. As with the other ZBDs, we monitored the change in UV/Visible absorbance over a 3 h period of time. The rate of absorbance loss observed for cGATA-$1_{158-223}$ was similar to that observed for FOG-$1_{950-995}$ and MMTV NCp10 (Table 3.1). This observed rate of absorbance loss for cGATA-$1_{158-223}$ was ~2.5 times slower than that observed for HIV-1 NCp7[35]. Compound **2** was able to eject the coordinated cobalt from the Cys_4 ZBD of cGATA-$1_{158-223}$ and the rate of cobalt metal ejection is similar to what is observed with $Cys_2HisCys$ ZBDs of FOG-$1_{950-995}$, HIV-1 NCp7 and MMTV NCp10.

To gain a more detailed understanding of the interaction between the thioester compounds and the Cys_4 ZBD of GATA-1, we again used NMR spectroscopy. Zinc-refolded hGATA-$1_{200-317}$ was incubated with compound **1** at 25 ºC for 48 h. hGATA-$1_{200-317}$ is a human GATA-1 fragment that contains two Cys_4 ZBDs separated by a 29 amino acid linker[38]. The 2D

^{1}H-^{15}N HSQC spectrum (Supplementary Figure 3.2A) recorded immediately after addition of compound **1** did not display any significant changes in the amide signals of the ^{15}N-labeled hGATA-$1_{200\text{-}317}$. The lack of change demonstrates that the structure of hGATA-$1_{200\text{-}317}$ remained intact after addition of compound **1**. These results also indicate that a stable complex was not formed between compound **1** and hGATA-$1_{200\text{-}317}$.

The 2D ^{1}H-^{15}N HSQC spectra recorded throughout the compound **1** incubation with hGATA-$1_{200\text{-}317}$ showed a loss of signal intensity as was seen with MMTV NCp10 (Figure 3.5A). By 24 h, all eight zinc-coordinating cysteine residues had lost >60% of their original signal intensity. After 48 h, many signals had almost completely disappeared. The rate of signal intensity loss is very similar for both ZBDs of hGATA-$1_{200\text{-}317}$ (Table 3.2), indicating that compound **1** was capable of covalently modifying cysteine residues in both ZBDs of GATA-1. The rate of signal intensity loss calculated for hGATA-$1_{200\text{-}317}$ was ~2.5 times greater than that observed for MMTV NCp10 (Table 3.1). Our results demonstrate that both Cys_4 ZBDs of GATA-1 are susceptible to metal ejection by the thioester compounds which results in loss of protein structure.

Next, we wanted to determine if the complexation state of the GATA-1 protein would have an effect on the interaction with the thioester compound. We first determined that the unbound carboxyl-terminal ZBD of chicken GATA-1 (cGATA-$1_{158\text{-}223}$) was able to interact with compound **1** in a manner identical to that observed by NMR spectroscopy with the two ZBDs of hGATA-$1_{200\text{-}317}$ (data not shown). Previous studies have shown that the domain comprised by cGATA-$1_{158\text{-}223}$ is the minimum domain needed for specific DNA binding to target double-stranded AGATAA DNA sequences[55] and the NMR solution structure of the cGATA-$1_{158\text{-}223}$/AGATAA DNA complex has been determined[44]. The zinc-refolded ^{15}N-labeled cGATA-

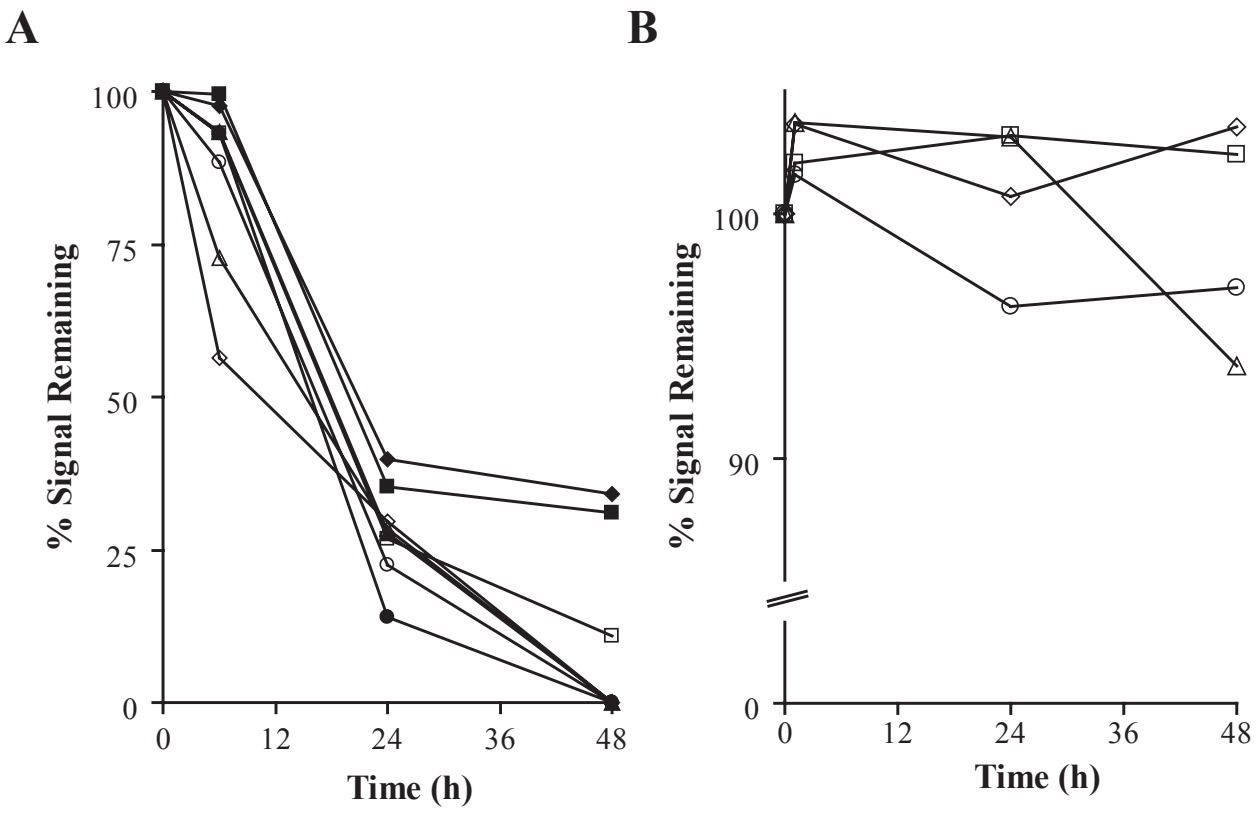


Figure 3.5: Changes in 2D ^{1}H-^{15}N HSQC signal intensity of zinc-coordinating residues of hGATA-$1_{200\text{-}317}$ (A) and AGATAA DNA-bound cGATA-$1_{158\text{-}223}$ (B) after addition of compound **1**. In A, ◆ is Cys_{205}, ■ is Cys_{208}, ▲ is Cys_{226}, ● is Cys_{229}, ◇ is Cys_{259}, □ is Cys_{262}, Δ is Cys_{280}, ○ is Cys_{283}. In B, ◇ is Cys_{164}, □ is Cys_{167}, Δ is Cys_{185}, is ○ Cys_{188}.

$1_{158-223}$/AGATAA DNA complex was incubated with compound **1** for 48 h at 25 ºC. The 2D ^{1}H-^{15}N HSQC spectrum recorded immediately after addition of compound **1** to the zinc-refolded ^{15}N-labeled cGATA-$1_{158-223}$/ AGATAA DNA complex did not undergo any changes in the amide signals of the DNA-bound cGATA-$1_{158-223}$ (data not shown). This again indicated that the structure of the complex was unchanged upon addition of compound **1** and that cGATA-$1_{158-223}$ and compound **1** did not form a stable complex. Surprisingly, the 2D ^{1}H-^{15}N HSQC spectra did not change during the entire 48 h incubation with compound **1** (Figure 3.5B). The signal intensities of the zinc-coordinating residues remained the same for the entire incubation period. By 48 h, all signals of the DNA-bound cGATA-$1_{158-223}$ retained at least 94% of their original intensity, which is dramatically different from what was observed when compound **1** was incubated with either the free cGATA-$1_{158-223}$ or the free hGATA-$1_{200-317}$. Thus, over the 48 h incubation period, compound **1** is unable to eject significant quantities of zinc from cGATA-$1_{158-223}$ when it bound to its target AGATAA DNA sequence.

We next used gel mobility shift assays to determine if preincubation of cGATA-$1_{158-223}$ with compound **1** would have an effect on the ability of cGATA-$1_{158-223}$ to bind DNA. Compound **1** or **2** were incubated with zinc-refolded cGATA-$1_{158-223}$ for 24 h at 25 ºC. After 30 sec, 4 h, and 24 h, 5′-^{32}P-labeled DNA was added to the zinc-refolded cGATA-$1_{158-223}$. The native gel ran after the 30 sec incubation with compound **1** revealed that the amount of DNA bound to cGATA-$1_{158-223}$ was not significantly changed (Figure 3.6A). After 4 h of incubation with compound **1**, there was a small decrease in the amount of DNA bound (< 5%), and after 24 h, we observed a significant decrease in the amount of DNA bound to cGATA-$1_{158-223}$ (~ 10%) (Figure 3.6B). In contrast, after a 30 sec pre-incubation with compound **2**, there was already a significant increase in the amount of DNA that was not bound to cGATA-$1_{158-223}$ (~ 10%)

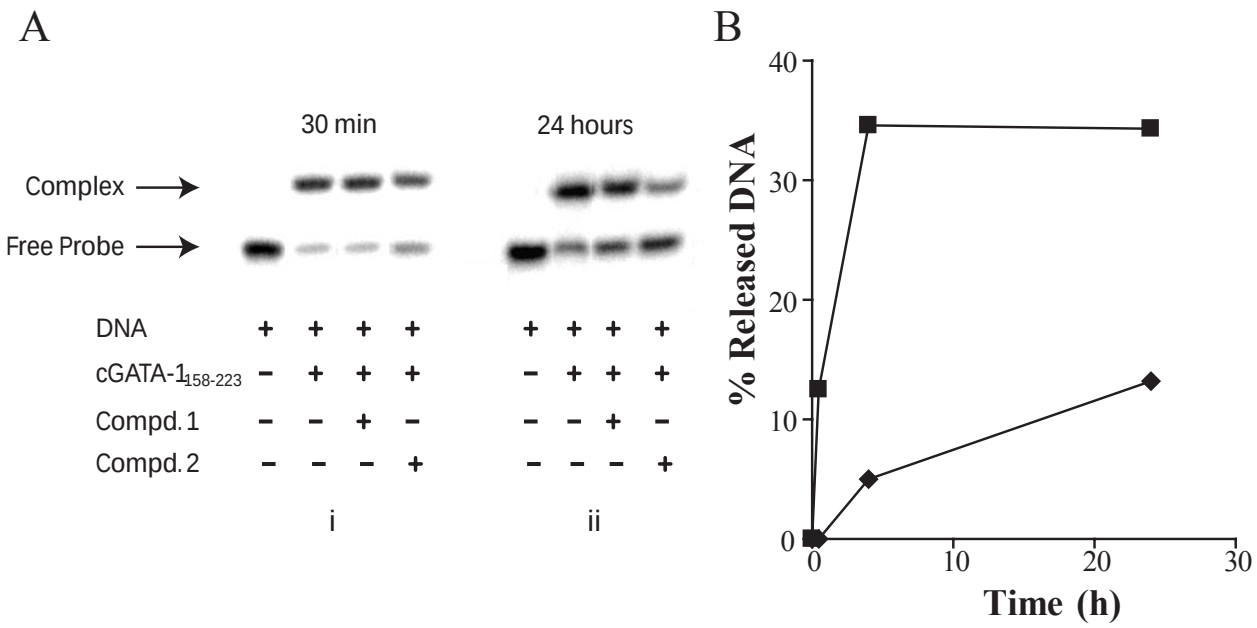


Figure 3.6: Gel mobility shift assay of binding of cGATA-$1_{158-223}$ to 5'-^{32}P-labeled AGATAA DNA upon incubation with thioester compounds. A) cGATA-$1_{158-223}$ was incubated with compound **1** or **2** for 30 seconds (i) or 24 hours (ii). AGATAA DNA was then added and incubated for 30 minutes before being run on an 8% native polyacrylamide gel. In each gel, the first lane is a control of labeled DNA prior to annealing. B) Plot of change in free DNA over thioester incubation time. The density of each band was measured and used to calculate the percent free DNA as described in Materials and Methods. In the plot, ◆ is compound **1** and ■ is compound **2**.

(Figure 3.6A). After 4 h, 65% of the DNA remained bound to cGATA-$1_{158-223}$. However, longer incubations with compound **2** (24 h) did not seem to cause much increase in unbound DNA (Figure 3.6B). These results demonstrate that after thioester-induced metal ejection, the cGATA-$1_{158-223}$ protein loses its ability to bind specifically to its target DNA sequence. This indicates that the thioester compounds likely disrupt the structure of the ZBD, preventing sequence specific DNA binding. In this assay, compound **2** seems to have a more pronounced effect than compound **1**, a result that has also been observed previously with HIV-1 NCp7[32,35].

*Compound **1** is modified upon incubation with select ZBDs* - The 1D difference water-sLED spectrum was used to monitor changes in compound **1** when mixed with equimolar concentration of ^{15}N-labeled ZBDs[46]. The 1D difference water-sLED spectrum that resulted from the spectra collected immediately after addition of compound **1** to MBP-$1_{2085-2142}$, MMTV NCp10, hGATA-$1_{200-317}$, cGATA-$1_{158-223}$/AGATAA DNA complex, or PKC$\delta_{231-280}$ were identical to the spectrum of the free thioester compound (data not shown). Therefore, compound **1** did not change upon addition to any of the proteins. Over the 48 h incubation of compound **1** with MBP-$1_{2085-2142}$, there was no change in ^{1}H chemical shifts or signal intensities attributed to compound **1** in the 1D difference water-sLED spectra at any time (data not shown). Thus, compound **1** did not undergo any alterations that would affect its diffusion during the incubation with MBP-$1_{2085-2142}$. These results further support our earlier conclusions that compound **1** fails to interact with MBP-$1_{2085-2142}$.

During the 48 h incubation of MMTV NCp10 with compound **1**, the 1D difference water-sLED spectra show that over time the signals of compound **1** lost some intensity (data not shown). After 48 h, the signals of compound **1** were almost indistinguishable from subtraction

artifacts. Changes in the 1D difference water-sLED spectra could be due to either alterations in the diffusion of the thioester compound in solution or degradation of compound **1**. Since compound **1** is completely stable in NMR buffer over the 48 h incubation period[35], the loss of signal intensity observed in the 1D difference water-sLED spectra is caused by either an interaction of compound **1** with MMTV NCp10 or covalent modification of MMTV NCp10 by compound **1**. Either of these two possibilities would lead to changes in the diffusion of the thioester compound. Based on these results, we conclude that compound **1** interacts in some fashion with the $Cys_2HisCys$ ZBD of MMTV NCp10.

In the incubation of compound **1** with hGATA-$1_{200\text{-}317}$, the 1D difference water-sLED spectra showed a progressive loss of intensity in the signals of compound **1**. This loss in signal intensity was observed for all ^1H signals of compound **1**, and by 48 h the signals had lost over half of their original signal intensity (Figure 3.7A). In contrast, during the incubation of compound **1** with the cGATA-$1_{158\text{-}223}$/AGATAA DNA complex, there was only a small change in the intensity of the signals from the thioester compound (Figure 3.7B). Thus, this data supports the fact that compound **1** interacts with the unbound GATA-1 proteins, but it is not able to interact with the DNA-bound GATA-1 protein.

After the 48 h incubation of compound **1** with PKC$\delta_{231\text{-}280}$, there were only minor changes in the 1D difference water-sLED spectra (data not shown). The signals of compound **1** lost only a small percentage of their original signal intensity. The primary change was the appearance of new signals (data not shown). The new signals were relatively weak and were only observed at the end of the incubation period, and may reflect a very weak interaction with PKC$\delta_{231\text{-}280}$. Overall, the signals of compound **1** remain intact, indicating that PKC$\delta_{231\text{-}280}$ does not interact significantly with the thioester compound.

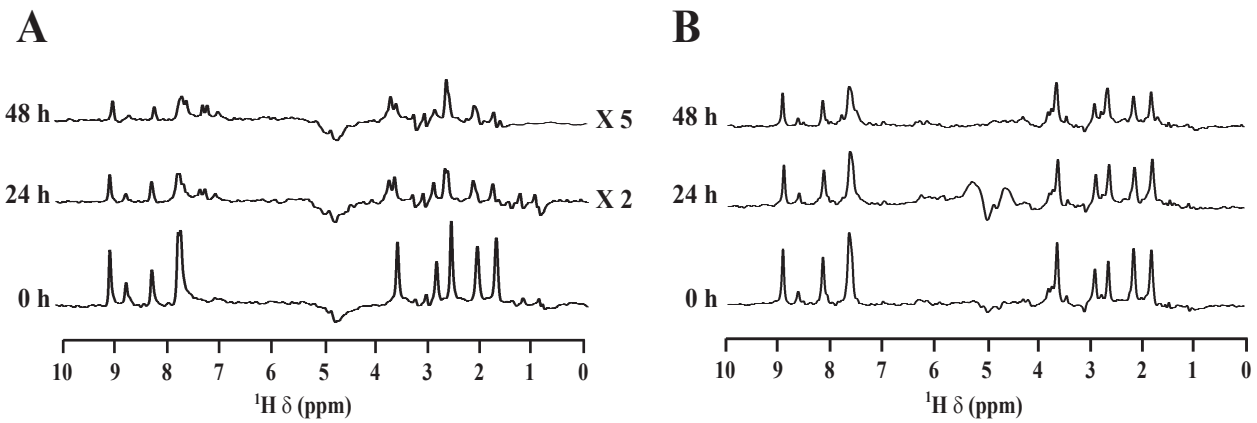


Figure 3.7: 1D difference water-sLED spectra of compound **1** incubated with hGATA-$1_{200\text{-}317}$ (A) or AGATAA DNA-bound cGATA-$1_{158\text{-}223}$ (B). In both A and B, spectra are shown after 0 hours, 24 hours, and 48 hours.

Discussion

In an attempt to develop new drug therapies to resistant strains of the HIV-1 virus, a number of different classes of small molecules have been designed to specifically target the HIV-1 NCp7 protein[17-28]. In most cases, these compounds are electrophilic and function by covalently modifying one or more cysteine thiolates in the HIV-1 NCp7 ZBDs. This ultimately leads to zinc ejection and loss of protein function since the zinc ions play critical roles in maintaining the structure of HIV-1 NCp7[22]. Given the highly reactive nature of these electrophilic compounds, it is absolutely essential to assess their specificity for the ZBDs of HIV-1 NCp7 in order to gauge their potential cytotoxicity. In particular, it is important to experimentally measure the reactivity of these compounds with ZBDs from cellular proteins in both their free and bound states.

Despite the potential importance of these compounds for controlling the spread of HIV infections, there have been surprisingly few studies addressing the reactivity of these compounds with ZBDs from cellular proteins[23,25]. Several theoretical studies suggest that the various classes of ZBDs most likely have widely varying reactivity with electrophilic reagents[30,36]. It has been predicted that the reactivity towards ZBDs for a given group of compounds is highly dependent on both first shell and second shell interactions with the zinc ion. Unfortunately, previous experimental studies have never thoroughly examined the reactivity of these compounds with a diverse group of ZBDs from cellular proteins[23,25]. The reactivity of a particular group of compounds has been examined with one or two zinc binding proteins specifically bound to nucleic acids[23]. In particular, no attempt has been made to compare the reactivity of the compounds with these zinc-binding proteins in both their nucleic acid-bound and unbound states. This lack of extensive experimental data has limited the refinement of the initial theoretical predictions. In this manuscript, we have investigated the reactivity of two

electrophilic thioester compounds with ZBDs from six different proteins. The ZBDs originate from five cellular proteins and one retroviral protein, and, combined, they possess a diverse range of structural motifs.

The two thioester compounds were tested with three classes (Cys_2His_2, Cys_3His, and Cys_4) of zinc ZBDs including the classical zinc finger proteins. Zinc fingers chelate zinc with two moles of cysteine and two moles of histidine to form a very compact $\beta\beta\alpha$ fold. This is an extremely important and abundant class of proteins in mammalian systems. Therefore, it is extremely important that this class of ZBDs show little or no reactivity with a compound if this compound is to have any potential as an antiviral agent. We examined two zinc finger Cys_2His_2 ZBDs ($GAGA_{310\text{-}372}$ and $MBP\text{-}1_{2085\text{-}2142}$) and demonstrated that both were essentially unreactive towards the thioester compounds. These results are consistent with what has previously been published with other thioester and DIBA compounds[23,25]. The only difference is that in our studies we have looked at the reactivity of the cellular zinc fingers in the free forms, whereas other studies examined their reactivity in their DNA-bound form. These results are also consistent with theoretical studies that predict that Cys_2His_2 ZBDs are the least reactive type of ZBDs[36]. Thus, the experimental and theoretical data indicate that free or DNA-bound classical zinc finger proteins do not appear to be targets for most electrophilic compounds, including the thioester compounds.

We also tested the reactivity of three ZBDs that bind zinc with three cysteine residues and one histidine residue (Cys_3His). This class of ZBDs is structurally very diverse, and the three ZBDs we selected were chosen to explore experimentally this structural diversity. Theoretical studies predict that Cys_3His ZBDs are more reactive towards electrophilic compounds than the Cys_2His_2 ZBDs due to the increase in negative charge around the zinc ion

from the three cysteine thiolates[30,36]. However, it is also predicted that Cys_3His ZBDs should have a wide range of reactivity towards electrophilic reagents[36]. The variation in reactivity within the Cys_3His ZBDs has been suggested to depend on second shell interactions about the zinc ion. As predicted, we did observe a wide range of reactivities between our thioester compounds and the three different Cys_3His ZBDs tested. We determined that $FOG_{950-995}$ and the amino-terminal ZBD of MMTV NCp10 reacted with the thioesters and their reactivities were similar to what we previously observed with the same thioesters for the Cys_3His carboxy-terminal ZBD of HIV-1 NCp7[35]. On the other hand, the interleaved Cys_3His ZBDs of $PKC\delta_{231-280}$ and the carboxyl-terminal ZBD of MMTV NCp10 possess low reactivity towards the thioester, analogous to what we observed for the amino-terminal ZBD of HIV-1 NCp7 (Table 3.2)[35].

Our results are in good agreement with theoretical predictions of the potential reactivity of the Cys_3His ZBDs of HIV-1 NCp7 and $PKC\delta_{231-280}$[36]. Theoretical predictions have not been performed to assess the potential reactivity of either $FOG_{950-995}$ or MMTV NCp10. However, the NMR structure of MMTV NCp10 clearly demonstrates that its carboxyl-terminal ZBD is considerably more shielded from interactions with the solvent than its amino-terminal ZBD[42]. This increased shielding about the carboxyl-terminal ZBD could easily explain the differences in reactivity observed with these two similar ZBDs. The shielding is provided by a series of hydrophobic and electrostatic interactions between the carboxyl-terminal ZBD and a β-hairpin that is located immediately adjacent to it. Our experimental results with these Cys_3His ZBDs strongly support theoretical predictions that the first shell ligands are not the sole determinant of reactivity within this class of ZBDs.

The Cys_4 ZBDs were the third class of ZBDs tested with the thioester compounds. Experimental model systems and theoretical predictions suggest that this class of ZBDs is the most reactive towards electrophiles by virtue of the four cysteine thiolates present in the zinc-binding domain[36]. For these studies we used the ZBDs from the GATA-1 protein as a test case. The GATA-1 protein contains two Cys_4 ZBDs separated by a 29-amino acid linker. The two ZBDs are structurally very similar but they serve distinct biological functions in hematopoietic development and appear to be structurally independent domains within the protein[38]. GATA-1 was chosen to determine whether the two ZBDs display different relative reactivities towards the thioesters, as observed with the two ZBDs of HIV-1 NCp7 and MMTV NCp10[35]. Based on the initial metal ejection experiments with protein fragments containing either one or both ZBDs from GATA-1, it was clear that both Cys_4 ZBDs in the GATA-1 protein are very susceptible to zinc ejection when incubated in the presence of the thioester compounds. Initially, it was believed that these results were in disagreement with earlier experimental studies and theoretical predictions that suggested that the carboxyl-terminal ZBD of GATA-1 would have low reactivity with electrophilic compounds[36]. However, the early experimental studies with DIBA compounds were done in the presence of DNA, and the early theoretical predictions were made with the structure of the carboxyl-terminal ZBD of GATA-1 in the DNA bound state[23,36]. Therefore, we examined the reactivity of the carboxyl-terminal ZBD of GATA-1 in both the free and the DNA bound state. Our results clearly demonstrated that the free carboxyl-terminal ZBD was susceptible to attack by the thioesters, whereas the DNA-bound form was not. In addition, we determined that the DIBA compounds are also capable of ejecting metal from the free GATA-1 ZBDs (LMJ, PL, and JGO unpublished data).

This variation in reactivity between the free carboxyl-terminal ZBD of GATA-1 and the DNA-bound form can be understood by examining the structures of the GATA-1 ZBDs. The binding of the carboxyl-terminal ZBD of GATA-1 to DNA induces a significant structural change in regions adjacent to the ZBD[55]. This change led to a significant increase in second shell interactions around the zinc ion by other residues in the ZBD[55]. In particular, there is a significant increase in the electrostatic screening around the zinc ion provided by several basic amino acids that make salt bridges to help stabilize the coordination of the zinc ion. In addition, the results with the GATA-1 ZBD are very similar to what was previously observed with HIV-1 NCp7 where binding to RNA significantly reduced the rate of zinc ejection from the ZD2 of HIV-1 NCp7[35]. In summary, the GATA-1 Cys_4 ZBDs reacted readily with the thioesters, but like HIV-1 NCp7 specific binding to its target nucleic acid significantly diminished the zinc ejecting properties of the thioesters.

In conclusion, these results provide important information regarding the reactivity of various classes of ZBDs with thioester compounds specifically designed to eject zinc from HIV-1 NCp7. The results demonstrate that the thioester compounds have considerable *in vitro* reactivity with Cys_4 GATA ZBDs and some Cys_3His ZBDs found in both cellular and other retroviral proteins. In the case of GATA-1, this reactivity with the thioester compounds appears to be inhibited by binding to its specific nucleic acid target, as observed for HIV-1 NCp7. Since HIV-1 NCp7 is expressed at high levels in HIV-1 infected cells, there is probably a significant level of HIV-1 NCp7 in the cell that is not bound to nucleic acids and, thus, would be reactive with the thioesters. Likewise, GATA-1 is highly expressed in red blood cells and, therefore, free GATA-1 in red blood cells could react with the thioesters and lead to potential toxicicity. However, studies of the effect of the thioester compounds on viral infection in murine models

indicate that the thioester compounds have low cellular toxicity at the doses required to inhibit HIV-1 infection[32]. More detailed studies in primate species are currently being conducted to more thoroughly assess the effectivity and toxicity associated with the thioesters, as our observations indicate potential toxicity risks. Our experimental results greatly improve our understanding of the reactivity of thiolates present in various classes of ZBDs. In addition, the results provide valuable information that can be incorporated in the future designs of antiviral agents with improved reactivity and specificity for the ZBDs of HIV-1 NCp7.

Acknowledgements

This work was supported by American Cancer Society Grant RPG LBC-100183 (J.G.O.) and National Institutes of Health Grant RO1 GM60298-01 (to J.G.O and P.L.) L.M.M.J. is supported by a Presidential Graduate Fellowship from the University of Georgia. This project was funded in part by the Intramural AIDS Targeted Antiretroviral Program (IATAP).

Supporting Information Available

Supplementary Figure 1 shows the overaly of the 2D ^{1}H-^{15}N spectra of ^{15}N-labeled MMTV NCp10 free and after addition of compound **1** (at 0 h and 48 h). Supplementary Figure 2 shows the overlay of the 2D ^{1}H-^{15}N spectra of ^{15}N-labeled hGATA-$1_{200\text{-}317}$ free and after addition of compound **1** (at 0 h and 48 h). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **2001***, 409*, 860-921.

(2) Miller, J.; McLachlan, A. D.; Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* **1985***, 4*, 1609-1614.

(3) Harrison, S. C. A structural taxonomy of DNA-binding domains. *Nature* **1991***, 353*, 715-719.

(4) Darlix, J. L.; Gabus, C.; Nugeyre, M. T.; Clavel, F.; Barre-Sinoussi, F. Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **1990***, 216*, 689-699.

(5) Gorelick, R. J.; Nigida, S. M., Jr.; Bess, J. W., Jr.; Arthur, L. O.; Henderson, L. E. et al. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **1990***, 64*, 3207-3211.

(6) Tisne, C.; Roques, B. P.; Dardel, F. Heteronuclear NMR studies of the interaction of $tRNA_{Lys}^{3}$ with HIV-1 nucleocapsid protein. *J. Mol. Biol.* **2001***, 306*, 443-454.

(7) Carteau, S.; Gorelick, R. J.; Bushman, F. D. Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J. Virol.* **1999***, 73*, 6670-6679.

(8) Omichinski, J. G., Clore, G.M., Sakaguchi, K., Appella, E., Gronenborn, A.M. Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy. *FEBS Lett.* **1991***, 292*, 25-30.

(9) Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J., Roques, B.P. Determination of the structure of the nucleocapsid protein from the human immunodeficiency virus type 1 by ^{1}H NMR. *EMBO J.* **1992***, 11*, 3059-3065.

(10) South, T. L.; Summers, M. F. Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci.* **1993***, 2*, 3-19.

(11) Lee, B. M.; De Guzman, R. N.; Turner, B. G.; Tjandra, N.; Summers, M. F. Dynamical behavior of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **1998***, 279*, 633-649.

(12) Meric, C.; Goff, S. P. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. *J. Virol.* **1989***, 63*, 1558-1568.

(13) Dorfman, T.; Luban, J.; Goff, S. P.; Haseltine, W. A.; Gottlinger, H. G. Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **1993***, 67*, 6159-6169.

(14) Housset, V.; De Rocquigny, H.; Roques, B. P.; Darlix, J. L. Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity. *J. Virol.* **1993***, 67*, 2537-2545.

(15) Poon, D. T.; Wu, J.; Aldovini, A. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J. Virol.* **1996***, 70*, 6607-6616.

(16) Tanchou, V.; Decimo, D.; Pechoux, C.; Lener, D.; Rogemond, 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*, 4442-4447.

(17) Rice, W. G.; Baker, D. C.; Schaeffer, C. A.; Graham, L.; Bu, M. et al. Inhibition of multiple phases of human immunodeficiency virus type 1 replication by a dithiane compound that attacks the conserved zinc fingers of retroviral nucleocapsid proteins. *Antimicrob. Agents Chemother.* **1997***, 41*, 419-426.

(18) Rice, W. G.; Schaeffer, C. A.; Graham, L.; Bu, M.; McDougal, J. S. et al. The site of antiviral action of 3-nitrosobenzamide on the infectivity process of human immunodeficiency virus in human lymphocytes. *Proc. Natl. Acad. Sci. USA* **1993***, 90*, 9721-9724.

(19) Rice, W. G.; Schaeffer, C. A.; Harten, B.; Villinger, F.; South, T. L. et al. Inhibition of HIV-1 infectivity by zinc-ejecting aromatic *C*-nitroso compounds. *Nature* **1993***, 361*, 473-475.

(20) Rice, W. G.; Supko, J. G.; Malspeis, L.; Buckheit, R. W., Jr.; Clanton, D. et al. Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. *Science* **1995***, 270*, 1194-1197.

(21) Rice, W. G.; Turpin, J. A.; Huang, M.; Clanton, D.; Buckheit, R. W., Jr. et al. Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. *Nat. Med.* **1997***, 3*, 341-345.

(22) Loo, J. A.; Holler, T. P.; Sanchez, J.; Gogliotti, R.; Maloney, L. et al. Biophysical characterization of zinc ejection from HIV nucleocapsid protein by anti-HIV 2,2'-dithiobis[benzamides] and benzisothiazolones. *J. Med. Chem.* **1996***, 39*, 4313-4320.

(23) Huang, M., Maynard, A., Turpin, J.A., Graham, L., Janini, G.M., Covell, D.G., Rice, W.G. Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J. Med. Chem.* **1998***, 41*, 1371-1381.

(24) Tummino, P. J.; Harvey, P. J.; McQuade, T.; Domagala, J.; Gogliotti, R. et al. The human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein zinc ejection activity of disulfide benzamides and benzisothiazolones: correlation with anti-HIV and virucidal activities. *Antimicrob. Agents Chemother.* **1997***, 41*, 394-400.

(25) Turpin, J. A.; Song, Y.; Inman, J. K.; Huang, M.; Wallqvist, A. et al. Synthesis and biological properties of novel pyridinioalkanoyl thiolesters (PATE) as anti-HIV-1 agents that target the viral nucleocapsid protein zinc fingers. *J. Med. Chem.* **1999***, 42*, 67-86.

(26) Turpin, J. A.; Terpening, S. J.; Schaeffer, C. A.; Yu, G.; Glover, C. J. et al. Inhibitors of human immunodeficiency virus type 1 zinc fingers prevent normal processing of gag precursors and result in the release of noninfectious virus particles. *J. Virol.* **1996***, 70*, 6180-6189.

(27) Goel, A.; Mazur, S. J.; Fattah, R. J.; Hartman, T. L.; Turpin, J. A. et al. Benzamide-based thiolcarbamates: a new class of HIV-1 NCp7 inhibitors. *Bioorg. Med. Chem. Lett.* **2002***, 12*, 767-770.

(28) Basrur, V.; Song, Y.; Mazur, S. J.; Higashimoto, Y.; Turpin, J. A. et al. Inactivation of HIV-1 nucleocapsid protein p7 by pyridinioalkanoyl thioesters. Characterization of reaction products and proposed mechanism of action. *J. Biol. Chem.* **2000***, 275*, 14890-14897.

(29) Topol, I. A.; Nemukhin, A. V.; Dobrogorskaya, Y. I.; Burt, S. K. Interactions of azodicarbonamide (ADA) species with the model zinc finger site: theoretical support of the zinc finger domain destruction in the HIV-1 nucleocapsid protein (NCp7) by ADA. *J. Phys. Chem. B* **2001***, 105*, 10.

(30) Maynard, A. T.; Huang, M.; Rice, W. G.; Covell, D. G. Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11578-11583.

(31) Tummino, P. J.; Scholten, J. D.; Harvey, P. J.; Holler, T. P.; Maloney, L. et al. The in vitro ejection of zinc from human immunodeficiency virus (HIV) type 1 nucleocapsid protein by disulfide benzamides with cellular anti-HIV activity. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 969-973.

(32) Schito, M. L.; Goel, A.; Song, Y.; Inman, J. K.; Fattah, R. J. et al. In vivo antiviral activity of novel human immunodeficiency virus type 1 nucleocapsid p7 zinc finger inhibitors in a transgenic murine model. *AIDS Res. Hum. Retrov.* **2003**, *19*, 91-101.

(33) Song, Y.; Goel, A.; Basrur, V.; Roberts, P. E.; Mikovits, J. A. et al. Synthesis and biological properties of amino acid amide ligand-based pyridinioalkanoyl thioesters as anti-HIV agents. *Bioorg. Med. Chem.* **2002**, *10*, 1263-1273.

(34) Srivastava, P.; Schito, M.; Fattah, R. J.; Hara, T.; Hartman, T. et al. Optimization of unique, uncharged thioesters as inhibitors of HIV replication. *Bioorgan. Med. Chem.* **2004**, *12*, 6437-6450.

(35) Miller Jenkins, L. M.; Byrd, J. C.; Hara, T.; Srivastava, P.; Mazur, S. et al. Studies on the inactivation of the HIV-1 nucleocapsid protein NCp7 with 2-mercaptobenzamide thioesters. *J Med Chem* **2005**, *48*, 2487-2458.

(36) Maynard, A. T., Covell, D.G. Reactivity of zinc finger cores: analysis of protein packing and electrostatic screening. *J. Am. Chem. Soc.* **2001**, *123*, 1047-1058.

(37) Omichinski, J. G.; Pedone, P. V.; Felsenfeld, G.; Gronenborn, A. M.; Clore, G. M. The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat. Struct. Biol.* **1997***, 4*, 122-132.

(38) Trainor, C. D.; Omichinski, J. G.; Vandergon, T. L.; Gronenborn, A. M.; Clore, G. M. et al. A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction. *Mol. Cell. Biol.* **1996***, 16*, 2238-2247.

(39) Kazanietz, M. G.; Barchi, J. J., Jr.; Omichinski, J. G.; Blumberg, P. M. Low affinity binding of phorbol esters to protein kinase C and its recombinant cysteine-rich region in the absence of phospholipids. *J. Biol. Chem.* **1995***, 270*, 14679-14684.

(40) Scharf, S. J.; Horn, G. T.; Erlich, H. A. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science* **1986***, 233*, 1076-1078.

(41) Majors, J. E.; Varmus, H. E. Nucleotide sequences at host-proviral junctions for mouse mammary tumour virus. *Nature* **1981***, 289*, 253-258.

(42) Klein, D. J.; Johnson, P. E.; Zollars, E. S.; De Guzman, R. N.; Summers, M. F. The NMR structure of the nucleocapsid protein from the mouse mammary tumor virus reveals unusual folding of the C-terminal zinc knuckle. *Biochemistry* **2000***, 39*, 1604-1612.

(43) Pedone, P. V.; Ghirlando, R.; Clore, G. M.; Gronenborn, A. M.; Felsenfeld, G. et al. The single Cys_2-His_2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. *Proc Natl Acad Sci U S A* **1996***, 93*, 2822-2826.

(44) Omichinski, J. G.; Clore, G. M.; Schaad, O.; Felsenfeld, G.; Trainor, C. et al. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **1993***, 261*, 438-446.

(45) Piotto, M.; Saudek, V.; Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **1992***, 2*, 661-665.

(46) Altieri, A. S.; Hinton, D. P.; Byrd, R. A. Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J. Am. Chem. Soc.* **1995***, 117*, 7566-7567.

(47) Kay, L. E.; Keifer, P.; Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **1992***, 114*, 10663.

(48) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995***, 6*, 277-293.

(49) Garrett, D. S.; Powers, R.; Gronenborn, A. M.; Clore, G. M. A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **1991***, 95*, 214-220.

(50) Omichinski, J. G.; Clore, G. M.; Robien, M.; Sakaguchi, K.; Appella, E. et al. High-resolution solution structure of the double Cys_2His_2 zinc finger from the human enhancer binding protein MBP-1. *Biochemistry* **1992***, 31*, 3907-3917.

(51) Maret, W.; Vallee, B. L. Cobalt as probe and label of proteins. *Methods Enzymol.* **1993***, 226*, 52-71.

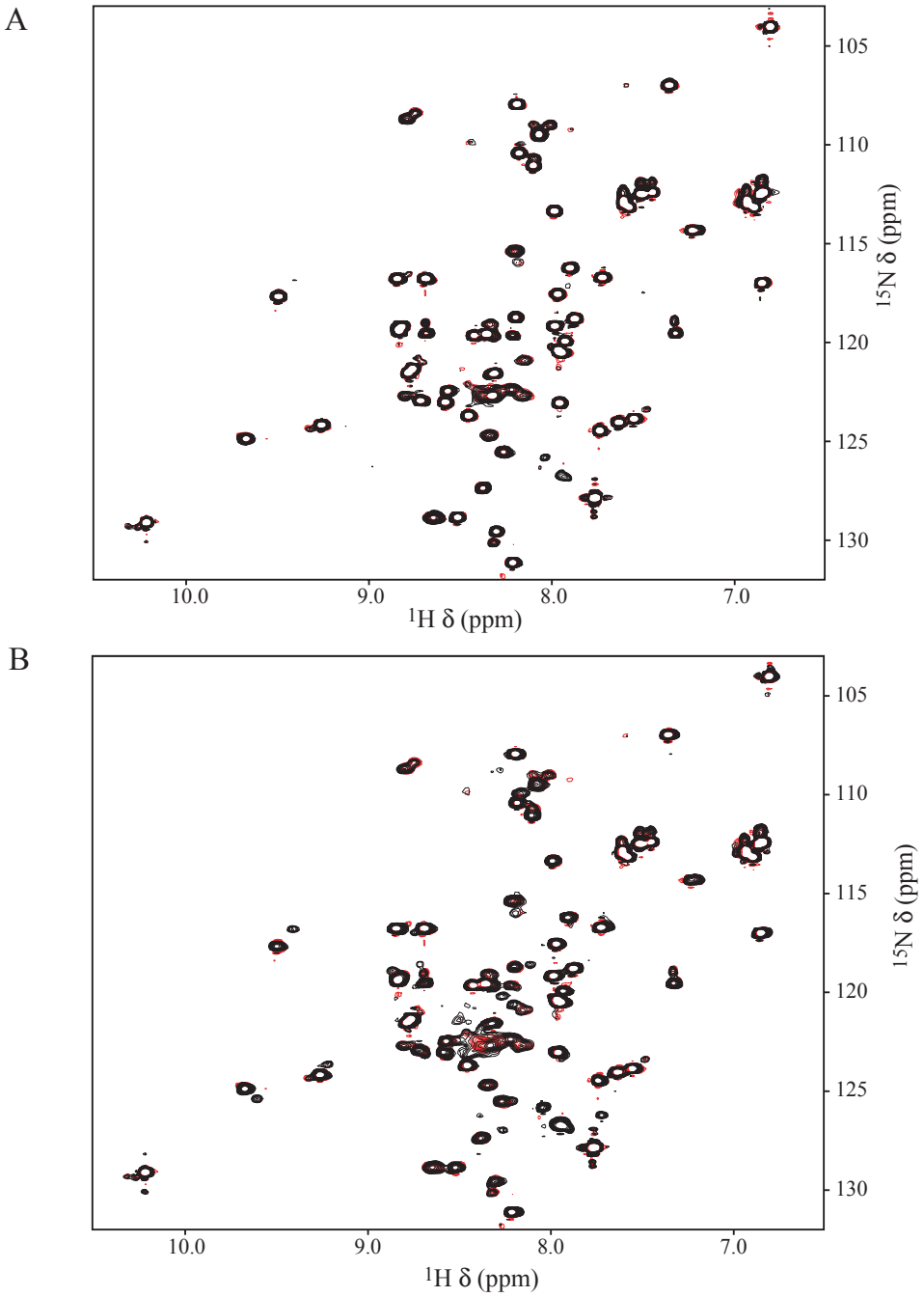
(52) Tsang, A. P.; Visvader, J. E.; Turner, C. A.; Fujiwara, Y.; Yu, C. N. et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **1997***, 90*, 109-119.

(53) Chen, X.; Chu, M.; Giedroc, D. P. Spectroscopic characterization of Co(II)-, Ni(II)-, and Cd(II)-substituted wild-type and non-native retroviral-type zinc finger peptides. *J. Biol. Inorg. Chem.* **2000***, 5*, 93-101.

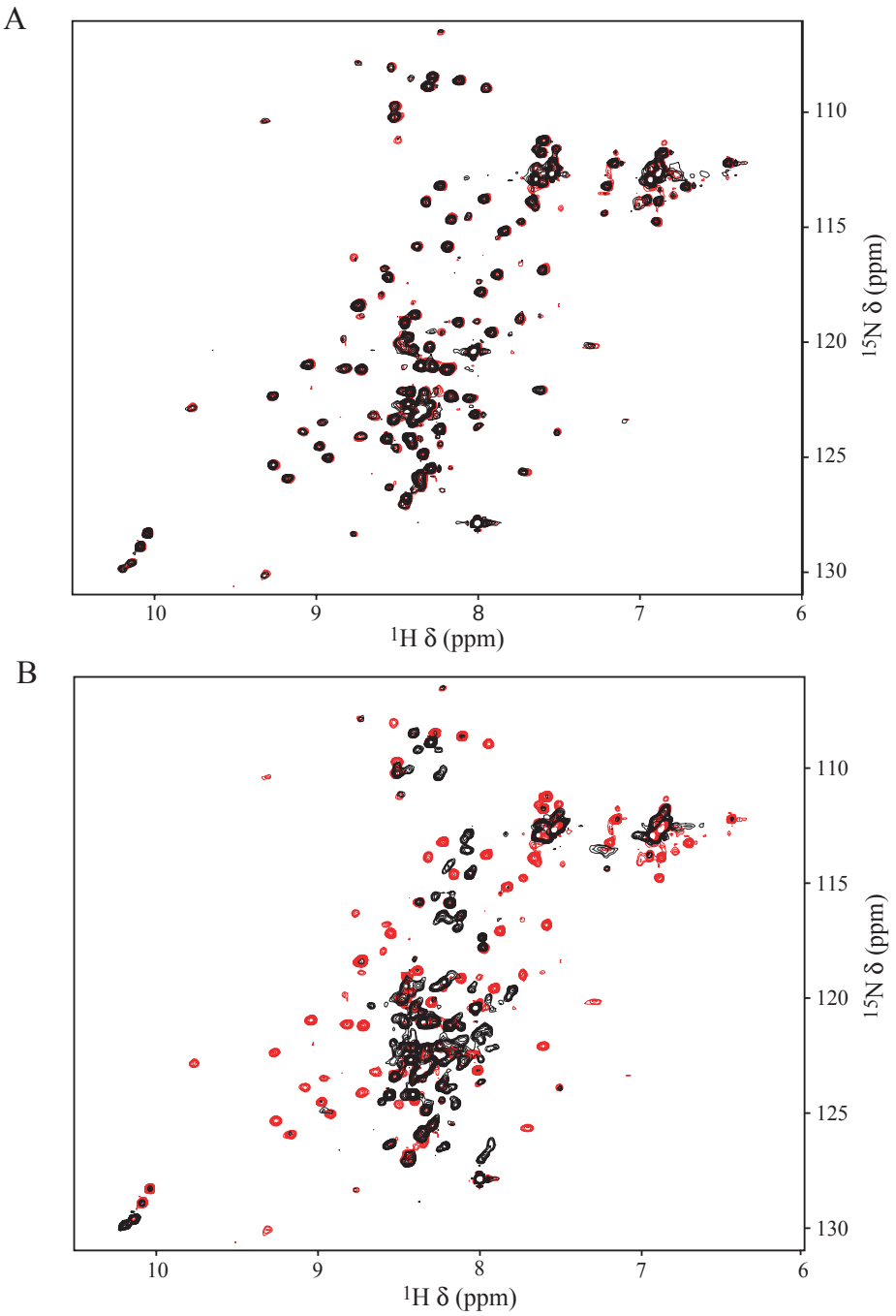
(54) Hommel, U.; Zurini, M.; Luyten, M. Solution structure of a cysteine rich domain of rat protein kinase C. *Nat. Struct. Biol.* **1994***, 1*, 383-387.

(55) Omichinski, J. G.; Trainor, C.; Evans, T.; Gronenborn, A. M.; Clore, G. M. et al. A small single-"finger" peptide from the erythroid transcription factor GATA-1 binds specifically to DNA as a zinc or iron complex. *Proc. Natl. Acad. Sci. USA* **1993***, 90*, 1676-1680.

Supplementary Data



Supplementary Figure 3.1: Overlay of the 2D ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled MMTV NCp10 in the free form (red signals) and after incubation with compound **1** (black signals) for either 0 h (A) or 48 h (B).



Supplementary Figure 3.2: Overlay of the 2D ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled hGATA-$1_{200\text{-}317}$ in the free form (red signals) and after the incubation with compound **1** (black signals) for either a 0 h (A) or 48 h (B).

CHAPTER 4

STUDIES ON THE INTERACTION OF HIV-1 NCP7 WITH THE INTERNAL LOOP OF ΨSL1[1]

[1] Miller Jenkins, L. M., Omichinski, J. G., Legault, P. To be submitted to *Journal of Molecular Biology*.

Abstract

The HIV-1 ΨSL1 stem-loop is crucial for the dimerization of HIV-1 genomic RNA. It includes two loops, one terminal loop and one internal loop. The terminal loop contains the six-nucleotide, palindromic dimerization initiation site that has been shown to mediate dimerization through formation of a kissing-loop. The HIV-1 nucleocapsid protein (NCp7) has been shown to be important for this dimerization reaction. The internal loop of ΨSL1 is invariant among HIV-1 strains and has been shown to be important for efficient dimerization, yet the function of this loop is not clear. Gel mobility shift assays and NMR spectroscopy were used to investigate the interaction between NCp7 and the internal loop of ΨSL1. NCp7 binds the internal loop and appears to destabilize the stem-loop structure, specifically two base pairs around the internal loop. However, the overall structure of the stem does not appear to change. This interaction does, though, result in large conformational changes in NCp7. We propose that the internal loop functions as the nucleation site for the melting of the ΨSL1 stem that is required for extended duplex formation, RNA dimerization, and virion infectivity.

Abbreviations

DIS, dimerization initiation site; DTT, dithiothreitol; HIV-1, human immunodeficiency virus type-1; HPLC, high-performance liquid chromatography; IPTG, isopropyl-β-D-thiogalactopyranoside; NCp7, HIV-1 nucleocapsid protein; SD, splice donor; SL1i, internal loop of stem loop 1; TAR, transactivation response element; TFA, trifluoroacetic acid; ZBD, zinc-binding domain; ZD1, amino-terminal zinc-binding domain; ZD2, carboxyl-terminal zinc-binding domain

Introduction

There are several important RNA structures in the HIV-1 virion that are required for complete viral production, such as the viral transactivation response element (TAR) and REV sequences. These structures often bind viral proteins as part of their functions. For example, tat binds to the TAR RNA to enhance the transcription of the viral DNA. Another crucial RNA element is the HIV-1 Ψ site, located at the 5' end of the HIV-1 genome. This region is made up of four stem-loop structures (ΨSL1-4), each with a different role in the viral lifecycle[1]. ΨSL2 contains the major splice donor site and ΨSL3 is the packaging signal. ΨSL4 is thought to be important for tertiary contacts required for the proper folding of the genomic RNA[2]. ΨSL1 contains the dimerization initiation site (DIS) and is essential for dimerization of the viral genomic RNA. The HIV-1 nucleocapsid protein (NCp7) has been shown to bind to these stem-loop structures to varying degrees[3], in the case of ΨSL1 and ΨSL3, to facilitate their roles in the virus lifecycle[4-6].

ΨSL1 is comprised of two loop regions – a nine nucleotide terminal loop and a four nucleotide internal loop (Figure 4.1A). The terminal loop of ΨSL1 contains the palindromic DIS sequence. During packaging, two copies of full-length viral RNA partially unfold and bind to one-another for encapsidation[7]. The dimerization reaction takes place through formation of a homodimeric kissing-loop complex. In the kissing-loop complex the palindromic loop of one monomer of genomic RNA makes Watson-Crick base pairs with the corresponding loop of a second monomer of genomic RNA[8-10]. Following formation of the dimeric kissing-loop complex, the stems of each individual RNA begin to melt and make new pairings with nucleotides in the opposite molecule.

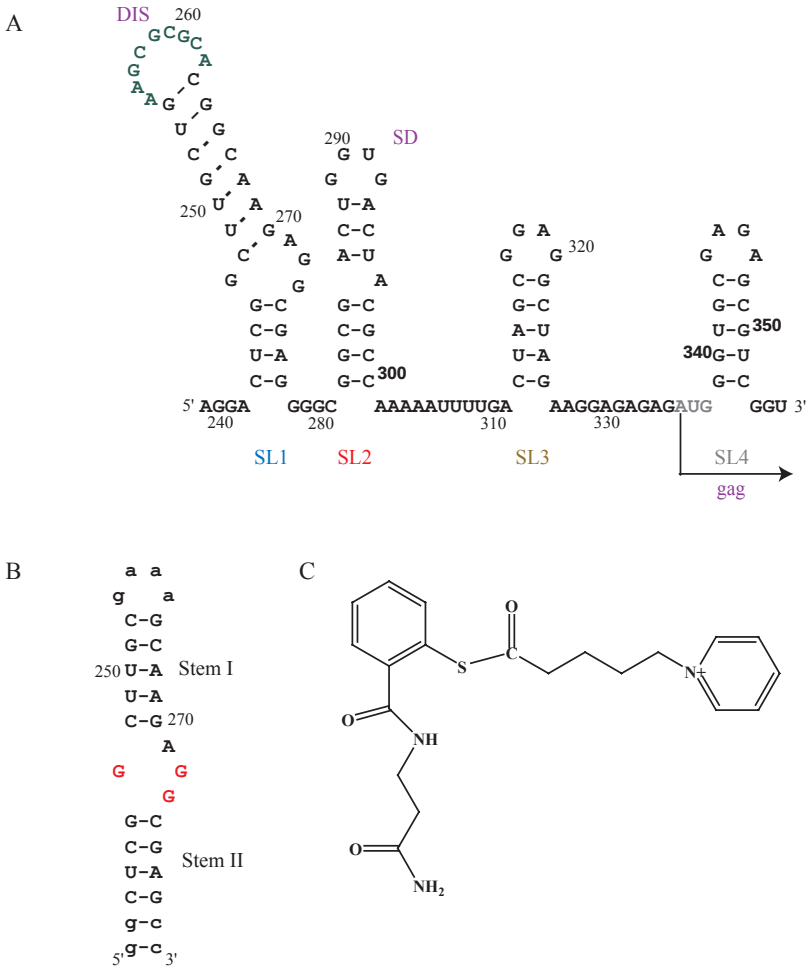


Figure 4.1: A) Schematic diagram of the sequence and secondary structure of the HIV-1 Ψ site. The dimerization initiation site (DIS), splice donor (SD), and gag initiation codon (AUG) are highlighted. The numbering is relative to the RNA cap site (+1). B) Sequence and secondary structure of the ΨSL1i construct used in this study. The three guanines in the internal loop that were mutated to adenines in $SL1i_m$ are shown in red. The lowercase letters are non-wild-type bases added to the sequence for improved transcription by T7 RNA polymerase. Also, the bases from the terminal loop were mutated from the original $HIV\text{-}1_{Lai}$ sequence to a GAAA tetraloop to prevent formation of dimers, improve the stability of the of the RNA hairpin, and facilitate NMR resonance assignment. C) Chemical structure of the 2-mercaptobenzamide thioester, compound **1**, used in this study.

Eventually, there is a maturation event in which the kissing-loop complex is fully converted into an extended duplex in which the stems of the two ΨSL1 molecules form intermolecular base pairs[9]. This maturation event is essential for HIV-1 infectivity[11]. The exact mechanism of the conversion of the kissing-loop complex to the extended duplex is unknown. Different mechanisms have been proposed for this reaction. One involves covalent modification of the RNA via an in-line attack of a nucleotide from one monomer on the other monomer. This would result in chain cleavage by transesterification and the formation of a 2',3' cyclic-phosphate and a 5' OH group[12]. The two cleavage reactions could then be followed by cross-ligation of strand A into strand B, finally yielding the extended duplex formation without any disruption of existing base pairs[12]. It has recently been shown that in some molecules, A^{255} is protonated at physiological pH, which causes the kissing-loop interaction to be dynamic by altering the charge of the base[13]. This led the authors to propose a somewhat different mechanism of kissing-loop to extended duplex conversion: protonation of a base in the terminal loop could disrupt the kissing-loop junction and allow other bases to move between multiple conformational states. Among these states would be one that is optimal for the conversion to occur[13].

NCp7 has been shown to specifically enhance the conversion from the kissing loop complex to the extended duplex[4,5]. *In vitro*, high, nonphysiological temperatures and salt concentrations are required for ΨSL1 to form exclusively the extended duplex conformation in the absence of NCp7. In the presence of NCp7, dimerization occurred at lower temperatures (37 °C) and lower salt concentrations[14]. Just as the mechanism of the conversion from kissing loop complex to extended duplex complex is unknown, the

function of NCp7 in this process is also unclear. NCp7 could destabilize the kissing loop complex or stabilize the extended duplex. Alternatively, NCp7 could play an enzymatic role, acting as a nuclease or ligase in the transesterification and cross-ligations reactions that have been proposed as the mechanism of kissing loop complex to extended duplex conversion[12].

The sequence of the internal loop of ΨSL1 is absolutely conserved amongst HIV-1 strains [11,15]. As yet, however, the function of this internal loop is unknown. It has been suggested to orient the palindromic terminal loop of ΨSL1 for optimal dimerization [16]. It is also possible that the internal loop is important for formation of the more-stable extended duplex[14]. Mutations in the internal loop have been shown to drastically reduce the infectivity of the resulting mutant virions and reduce the ability of ΨSL1 to form the extended duplex conformation [14,16,17]. The NMR solution structures of the internal loop revealed that G_{247} forms a base pair with A_{271} across the loop[18,19]. In addition, the base G_{247} stacks between G_{246} and C_{248}[18,19]. One structure showed that the bases of A_{271}, G_{272}, and G_{273} stack between G_{270} and C_{274} (Figure 4.1)[18]. A second NMR solution structure found G_{272} and G_{273} to be partially ordered, but flexible[19]. The stacking of an uneven number of loop bases causes the stem to bend by 25 °C [18,19]. This bend is thought to orient the palindromic terminal loop in a preferred orientation for dimerization[18,19].

Fluorescence studies have demonstrated that NCp7 binds to the both the terminal and internal loops of ΨSL1 with moderate affinity (K_d = 100 nM and 140 nM, respectively)[3]. However, it has been suggested that the internal loop does not serve as a high-affinity binding site for NCp7 because the guanosines in the internal loop are stacked on one-another in one of the NMR solution structures and are thus not readily

accessible for binding to the protein[18]. As the internal loop of ΨSL1 has been shown to have an effect on this dimerization process and is absolutely conserved, it is crucial to better understand the function of this region. Since NCp7 facilitates genomic RNA dimerization by binding the terminal loop of ΨSL1[20], it is possible that it also is important for the function of the internal loop. The ability of NCp7 to interact with the internal loop of ΨSL1 merits further analysis. In addition, there are no structural studies of NCp7 binding to ΨSL1 and this research was performed to begin to address this issue.

In this report, the interaction of NCp7 with the internal loop of ΨSL1 was investigated using gel mobility shift analysis and NMR spectroscopy. In addition, the effect of two 2-mercaptobenzamide thioesters on the NCp7/SL1 complex was investigated. It was found that NCp7 is able to interact with the internal loop, and the effect of this interaction on both the protein and the RNA was studied. We observed that NCp7 appears to destabilize the two base pairs about the internal loop and propose that this interaction is important for the dimerization reaction.

Methods

Protein Purification- The coding sequence for HIV-1 NCp7 (HXB2 isolate) was cloned into the *Escherichia coli* vector pET11a and expressed in host strain Rosetta (DE3) (Novagen, WI). Uniformly ^{15}N-labeled or ^{13}C/^{15}N-labeled NCp7 was obtained by growing the cells in modified minimal media containing ^{13}C-labeled (>99%) glucose and/or ^{15}N-labeled (>98%) ammonium chloride as the sole carbon and nitrogen sources, respectively. Cells were grown overnight at 37 ºC, and protein expression was induced for 4 h with 0.66 mM or 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) for

cultures grown in minimal media or Luria Broth, respectively. The cells were pelleted and resuspended in 25 mM Tris pH 8.0, 1 mM EDTA, 2 mM dithiothreitol (DTT), 6 mM benzamidine. The cells were then lysed by French press and centrifuged at 100,000 x g for 45 min. The supernatant was applied to a DEAE-Sepharose Fast Flow (Amersham Biosciences, NJ) column (300 mL bed volume), equilibrated with Buffer A (25 mM Tris pH 8.0, 1 mM EDTA, and 2 mM DTT) and eluted using a gradient (0-100% B over 1500 mL) of Buffer B (25 mM Tris pH 8.0, 1 mM EDTA, 2 mM DTT, and 1 M NaCl). The pooled NCp7 fractions were then applied to a SP-Sepharose Fast Flow (Amersham Biosciences, NJ) column (300 mL bed volume), equilibrated with Buffer A and eluted using a gradient (0 -100% B over 1500 mL) of Buffer B. The fractions containing NCp7 were pooled and purified on a C-8 reverse phase (Vydac) high-performance liquid chromatography (HPLC) column using a gradient of 10-30% acetonitrile in 0.05% aqueous trifluoroacetic acid (TFA). The purified NCp7 was flash frozen and lyophilized.

Purified NCp7 was refolded using the following procedure. Lyophilized protein was first resuspended at a concentration of ~0.2 mM in 0.05% TFA. Five equivalents of $ZnCl_2$ (zinc-refolded NCp7) were then added to the NCp7 solution as 50 mM metal solutions in 0.05% TFA. The solution was slowly titrated to pH 6.0 with 0.2 M NaOH. The zinc-refolded NCp7 was flash frozen and lyophilized. Prior to use, zinc-refolded NCp7 (1 mM) was resuspended in NMR buffer A (20 mM sodium phosphate pH 7.0) in 90% H_2O/10% D_2O.

RNA Purification- SL1i and $SL1i_m$ RNAs (Figure 4.1B) were transcribed *in vitro* using T7 RNA polymerase, double-stranded synthetic oligonucleotide templates, and

nucleoside triphosphates (NTPs). The T7 RNA polymerase was purified based upon a published procedure [21]. The ^{15}N-labeled and $^{15}N/^{13}C$-labeled SL1i was transcribed using ^{15}N-labeled and $^{15}N/^{13}C$-labeled NTPs, respectively[22].

The RNAs were purified to single-nucleotide resolution by 20% denaturing polyacrylamide gel electrophoresis. Each RNA was dephosphorylated on the 5'-end with calf alkaline phosphatase (Roche Molecular Biochemicals, IN) and then further purified by DEAE-Sephacel chromatography (Amersham Biosciences, NJ). The RNAs were concentrated with an Amicon Centricon-3 concentrator (Millipore, MA) and exchanged into water. Prior to use, the RNA was heated to 95 ºC for 1 min and snap-cooled in ice water to promote hairpin formation. For the gel mobility shift assay, the RNAs were 5'-end-labeled with γ-(^{32}P) ATP (MP Biomedicals, CA) using T4 polynucleotide kinase (New England Biolabs, MA) according to the manufacturer's instructions.

Gel Mobility Shift Assay- A sample of zinc-refolded NCp7 (0-500 nM) was incubated with an equal volume of 5'-^{32}P-labeled SL1i or $SL1i_m$ RNA (1 nM) in gel shift buffer (50 mM Tris pH 7.5, 10% glycerol, 25 μg/ml yeast tRNA, 200 mM KCl, and 40 mM $MgCl_2$) for 30 min at 25 ºC. In the cold competition experiments, unlabeled SL1i (1-10 nM) was added to the 5'-^{32}P-labeled SL1i and incubated with the NCp7 for 30 min at 25 ºC. The binding reactions were analyzed on an 8% native polyacrylamide gel in 50 mM Tris-Borate pH 8.0, and 1 mM EDTA at 200 V and 4 ºC. The 5'-^{32}P-labeled SL1i or $SL1i_m$ RNA was detected using a Storm PhosphorImager (Amersham Biosciences, NJ). Quantification of bound and unbound 5'-^{32}P-labeled SL1i were made using ImageQuant (Molecular Dynamics). The percentage of bound SL1i was calculated as: % Bound =

$D_B/(D_B+D_F)$ where D_B is the amount of the bound SL1i and D_F is the amount of the free SL1i after normalization for the background intensity. The K_D was calculated from the concentration of 50% binding on the plot of the % bound RNA versus the NCp7 concentration.

A 250 nM sample of zinc-refolded NCp7 was incubated with 2.5 μM of compound **1** in 50 mM Tris pH 7.5 and 10% glycerol at 25 ºC for 0 min, 6 h, and 24 h. A 5 μL aliquot of this incubation was removed and incubated with an equal volume of 5'-^{32}P-labeled SL1i RNA (2.5 nM) in gel shift buffer for 30 min at 25 ºC. Alternatively, the NCp7 was incubated with 5'-^{32}P-labeled SL1i RNA in gel shift buffer for 30 min at 25 ºC. Then, 2.5 μM compound **1** was incubated with this preformed complex for 0 min, 1 h, 6 h, and 24 h at 25 ºC. The binding reactions were analyzed on an 8% native polyacrylamide gel in 50 mM Tris-Borate pH 8.0, and 1 mM EDTA at 200 V and 4 ºC. The 5'-^{32}P-labeled SL1i RNA was detected using a Storm PhosphorImager (Amersham Biosciences, NJ).

Preparations of Samples for NMR Spectroscopy- NCp7/SL1i complexes were prepared by titrating NCp7 (1 mM) into SL1i (1 mM) in NMR buffer (20 mM sodium phosphate buffer pH 7.0 and 25 mM NaCl) in 90% H_2O/10% D_2O. Complexes were prepared in which one molecule was ^{15}N-labeled or ^{13}C/^{15}N-labeled and the other was not isotopically labeled. Complex formation was monitored by 1D ^{15}N-decoupled watergate[23]. Once a 1:1 1 mM complex was prepared, the sample was flash frozen and lyophilized. The sample was then resuspended to achieve a final composition of 2 mM NCp7/RNA, 40

mM sodium phosphate pH 7.0 and 200 mM NaCl in 90% H_2O/10% D_2O, unless otherwise stated.

NMR Spectroscopy- The NCp7/SL1i complexes were analyzed on Varian ^{UNITY}INOVA 500 MHz or 600 MHz spectrometers equipped with HCN triple resonance probes with actively shielded z- gradients. ^{1}H, ^{15}N, and ^{13}C assignments for the SL1i-complexed NCp7 were obtained at 35 °C from the following experiments: 2D ^{1}H-^{15}N HSQC[24], 2D ^{1}H-^{13}C CT-HSQC, 3D HNCO[25], 3D HNCACB[26-28], 3D (HB)CBCA(CO)NNH[26,28], 3D HCCCH-COSY[29-31], 3D HCCCH-TOCSY[32], 3D H(CCO)NNH-TOCSY[28], and 3D C(CO)NNH-TOCSY[28]. Distance constraints were obtained at 35 ˚C from the following experiments: 3D ^{15}N-edited NOESY-HSQC[33] and 3D ^{13}C-edited NOESY-HMQC (60 and 120 ms mixing times at 600 MHz)[34]. Assignments for NCp7 in the NCp7/ΨSL3 complex were generously provided by M. F. Summers.

Assignments of the free SL1i were obtained from 2D ^{1}H, ^{1}H flip-back watergate-NOESY[35] at 5 ˚C and 15 ˚C under two buffer conditions, low salt buffer (25 mM NaCl and 40 mM sodium phosphate pH 7.0) and high salt buffer (200 mM NaCl and 40 mM sodium phosphate pH 7.0). Assignments of the NCp7-bound SL1i were obtained at 25 ˚C and 35 ˚C from the following experiments: 2D ^{1}H-^{15}N HSQC optimized for imino or amino protons, 4D ^{13}C/^{15}N-edited NOESY[36], 2D HNN-COSY[37], and 2D CPMG-NOESY[38]. 1D ^{1}H NMR spectra of unlabeled SL1i were also collected in the high and low salt buffers at 5, 15, 25, 35, and 45 ˚C. The chemical shift was referenced against the chemical shift of the water peak, for which the exact chemical shift at various temperatures is known relative to a DSS standard.

Results

NCp7 binds to the internal loop of ΨSL1- Gel mobility shift analysis was used to study the binding of NCp7 to SL1i at various concentrations. In SL1i, the palindromic loop of ΨSL1 was replaced with a GNRA sequence that would favor hairpin conformation and prevent dimerization (Figure 1B). This loop sequence has been shown to form a highly stable tetraloop that should stabilize proper base pairing in the stem. SL1i (1 nM) was incubated with a 1-500 nM zinc-refolded NCp7 (Figure 4.2A). The presence of a shift in the mobility of SL1i upon the addition of 5 nM NCp7 (5:1 NCp7:SL1i) demonstrates that NCp7 is able to bind SL1i (Figure 4.2A, lane 2). From the binding curve, the K_D of this interaction is ~100 nM, corresponding to the point at which 50% of the RNA is bound (Figure 4.2A, lane 5). Thus, NCp7 was able to bind the internal loop of ΨSL1 with moderate affinity.

As we replaced the terminal loop of ΨSL1 with a GNRA tetraloop, it was important to ascertain that the binding observed in the gel shift assay specific to the internal loop and not to the non-natural GNRA tetraloop. Since NCp7 has been shown to prefer exposed guanine bases and has no affinity for adenine bases, we mutated the three guanine bases in the internal loop to adenines ($SL1i_m$, Figure 4.1B). Therefore, if the binding of NCp7 was specific for the the internal loop, and not the GNRA tetraloop, we should not observe any binding to $SL1i_m$. When NCp7 was incubated with ^{32}P-labeled $SL1i_m$ (1 nM), no change in the mobility of this mutant RNA was observed even at an NCp7 concentration of 1 μM (1000:1 NCp7:RNA) (Figure 4.2B, lane 4 8). Thus, NCp7 binds specifically to the guanines in the internal loop of ΨSL1 at these concentrations.

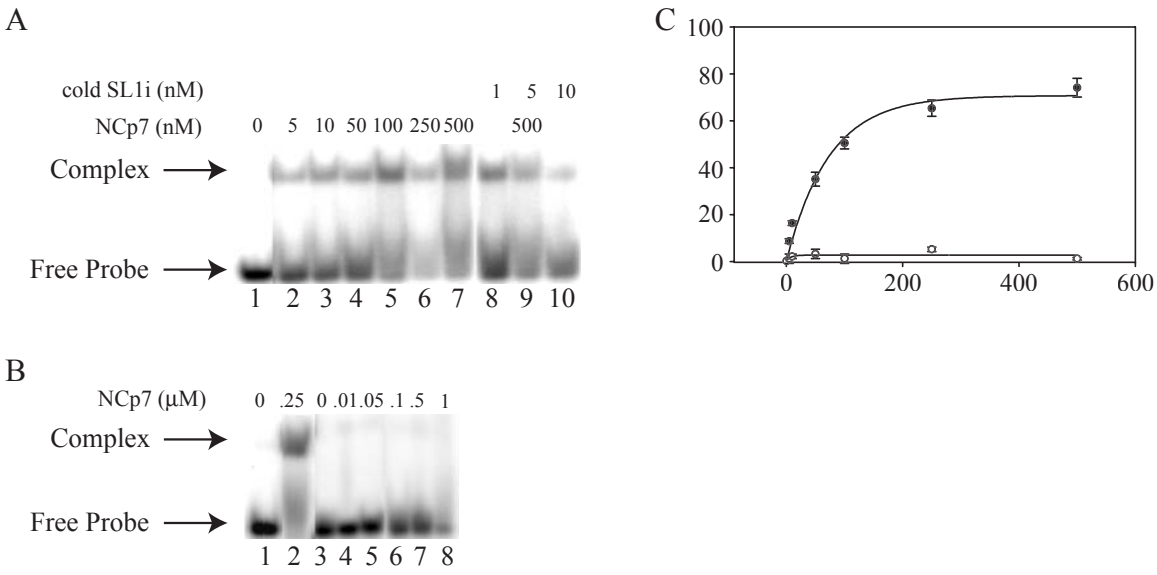


Figure 4.2: Gel mobility shift assay to study the binding of NCp7 to SL1i. A) 5'-^{32}P-labeled SL1i (1 nM) was incubated with NCp7 at a range of concentrations from 0-500 nM (lanes 1-7). In lanes 8-10, unlabeled SL1i (1-10 nM) was added to a mixture of 5' ^{32}P-labeled SL1i (1 nM) and NCp7 (500 nM). B) 5' ^{32}P-labeled SL1i$_m$ (1 nM, lanes 3-8) was incubated with NCp7 at a range of concentrations from 0-1000 nM (lanes 3-8). Lanes 1 and 2 show 5' ^{32}P-labeled SL1i without and with 250 nM NCp7, respectively. C) Plot of percentage bound SL1i (closed circles) and SL1i$_m$ (open circles) over the range of NCp7 concentrations. The amount of the free and bound SL1i in A was measured using ImageQuant, and the background intensity was subtracted out. The % Bound RNA was calculated as described in Methods.

*Effect of compound **1** on NCp7 binding to SL1i-* In our effort to probe the interaction between NCp7 and SL1i, we studied the effect of a 2-mercaptobenzamide thioester on the ability of the protein to bind to the RNA. NCp7 was incubated with compound **1** (Figure 4.1C) for 1 h, 6 h, and 24 h before being incubated with ^{32}P-labeled SL1i. The samples where then run on a native gel and the gel mobility shift was analyzed (Figure 4.3, lanes 1-5). After 1 h, there was no change in the mobility of SL1i (Figure 4.3, lane 3). After 6 h, there was a small loss in the amount of shifted RNA and an increase in the amount of free RNA (Figure 4.3A, lane 4). After 24 h, no shift was observed (Figure 4.3A, lane 5). This indicates that whereas compound **1** had no large effect on the ability of NCp7 to bind SL1i after 6 h, NCp7 was completely unable to bind SL1i after 24h. In our previous studies with compound **1**, we observed a similar loss of NCp7 binding to ΨSL2 when incubated with this compound [39]. These results indicate that long incubation with compound **1** abolish the ability of NCp7 to bind SL1i.

We next analyzed the effect of compound **1** on pre-formed NCp7/SL1i complexes. NCp7 was first mixed with ^{32}P-labeled SL1i, and then incubated with compound **1** for 1 h, 6 h, and 24 h. The samples were run on a native gel and the gel mobility shift analyzed (Figure 4.3, lanes 6-10). At all times, a shift in the mobility of the RNA was observed, indicating that NCp7 was bound to SL1i. Even after 24 h, most of the SL1i remained bound (Figure 4.3B, lane 10). Measurement of the amount of the free and bound SL1i showed that 90% of the RNA was bound (data not shown). This result suggests that binding to SL1i protected the NCp7 from interaction with compound **1**. Our earlier experiments with a preformed NCp7/SL3 complex showed that binding of NCp7 to this RNA also protected the protein from interaction with the compound **1**[39].

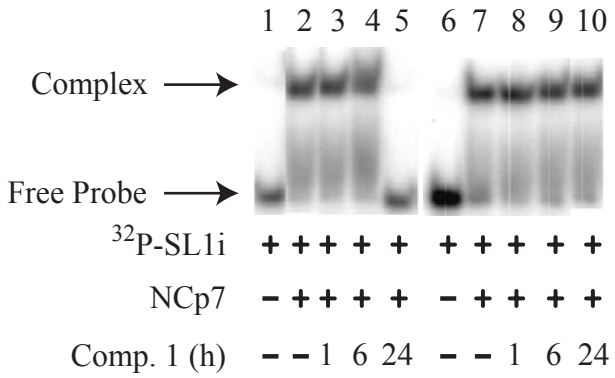


Figure 4.3: Gel mobility shift assay to study the effect of compound **1** on NCp7 binding to SL1i. NCp7 (250 nM) was incubated without (lanes 2 and 7) or with compound **1** (2.5 μM) (lanes 3-5, 7-10) for 1 h, 6h, or 24 h at 25 ºC. For lanes 3-5, 5' ^{32}P-labeled SL1i (1 nM) was added to the mixture after the incubation with compound **1**. For lanes 8-10, the NCp7/SL1i complex was preformed by incubation for 30 min at 25 ºC prior to compound **1** addition.

Therefore, the interaction of NCp7 with SL1i, or other Ψ-site stem loops, shows similar properties with regard to the reactivity of NCp7 towards the 2-mercaptobenzamide thioesters.

Resonance assignment of NCp7- The backbone resonance assignments of the free were obtained at 25 ˚C using $^{13}C/^{15}N$-labeled NCp7 as described in Materials and Methods. The 2D ^{1}H-^{15}N HSQC spectrum of the free protein is characterized by very good chemical shift dispersion in both the ^{1}H and ^{15}N dimensions (Figure 4.4). The first amino acid and several amino acids in the linker region were not observed in the 2D ^{1}H-^{15}N HSQC spectrum, likely due to rapid proton exchange. For the ^{1}H, ^{15}N, and ^{13}C backbone nuclei, nearly complete assignment was achieved: 93% of the $^{1}H^{N}$, 93% of the ^{15}N, 89% of the $^{13}C'$, and 95% of the $^{13}C^{\alpha}$ (Supplementary Table 4.1). None of the side chain resonances were assigned for free NCp7.

Resonance assignment of SL1 - Assignment of the signals of SL1i in the free state was performed using the experiments described in Materials and Methods (Supplementary Table 4.2). The exchangeable imino proton assignments were determined first according to NOEs observed between guanine and uracil imino signals in consecutive base pairs in the RNA stem using a 2D ^{1}H-^{1}H flip-back watergate-NOESY[35] (Figure 4.5). These assignments were confirmed, and further assignments were obtained, by analyzing NOEs observed between the imino protons and amino protons on cytosines and the H2 proton on the adenines (data not shown). U_4, G_{28}, U_{10}, G_{17}, G_{11}, G_{26}, and G_{13} were assigned from N-H/N-H NOEs. G_2, G_6, and U_9 were assigned using NOEs observed between the

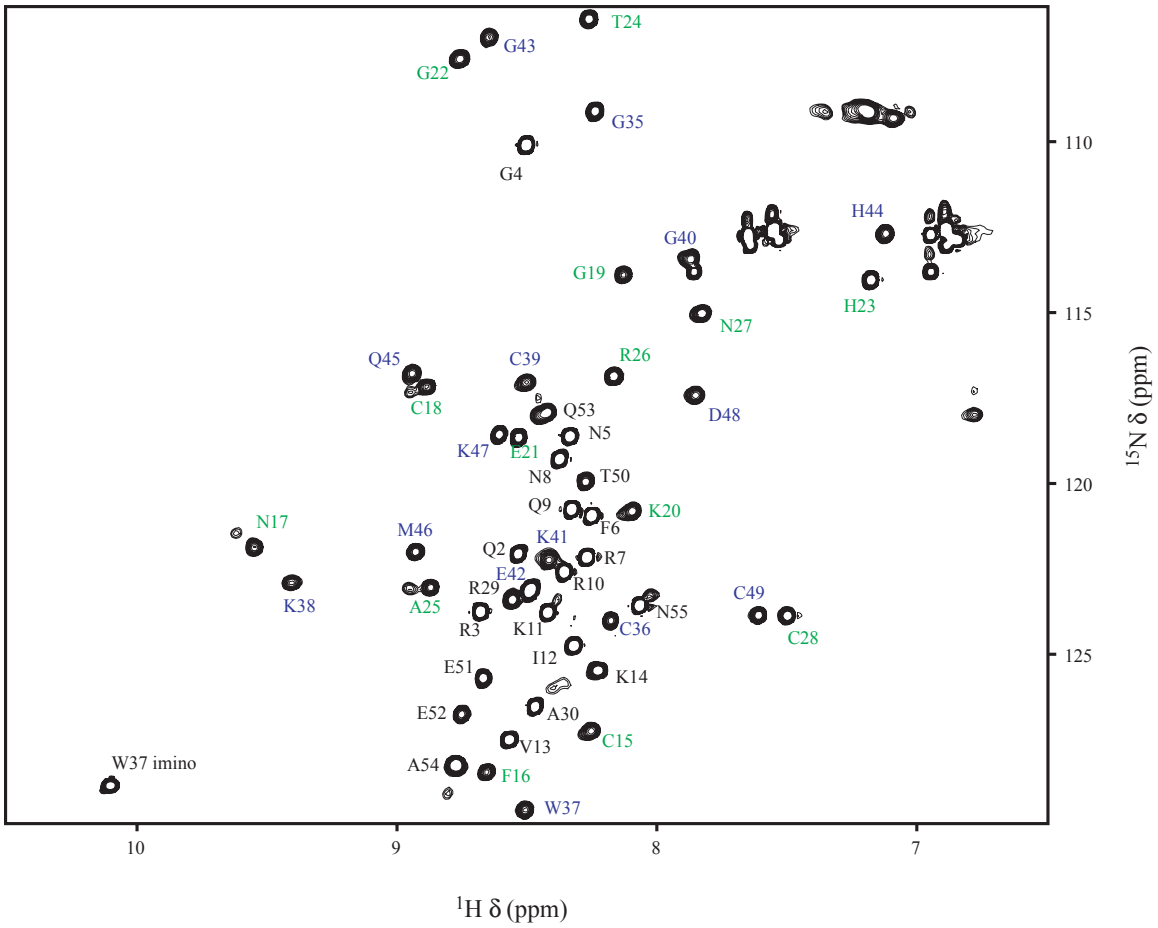


Figure 4.4: 2D ^{1}H-^{15}N HSQC of free ^{15}N-labeled NCp7. The amide signals are labeled with their assignments. Labels from residues in the amino-terminal zinc-binding domain are shown in green and those from the carboxyl-terminal zinc-binding domain are shown in blue. The spectrum was collected with 1 mM NCp7 at 25 ˚C.

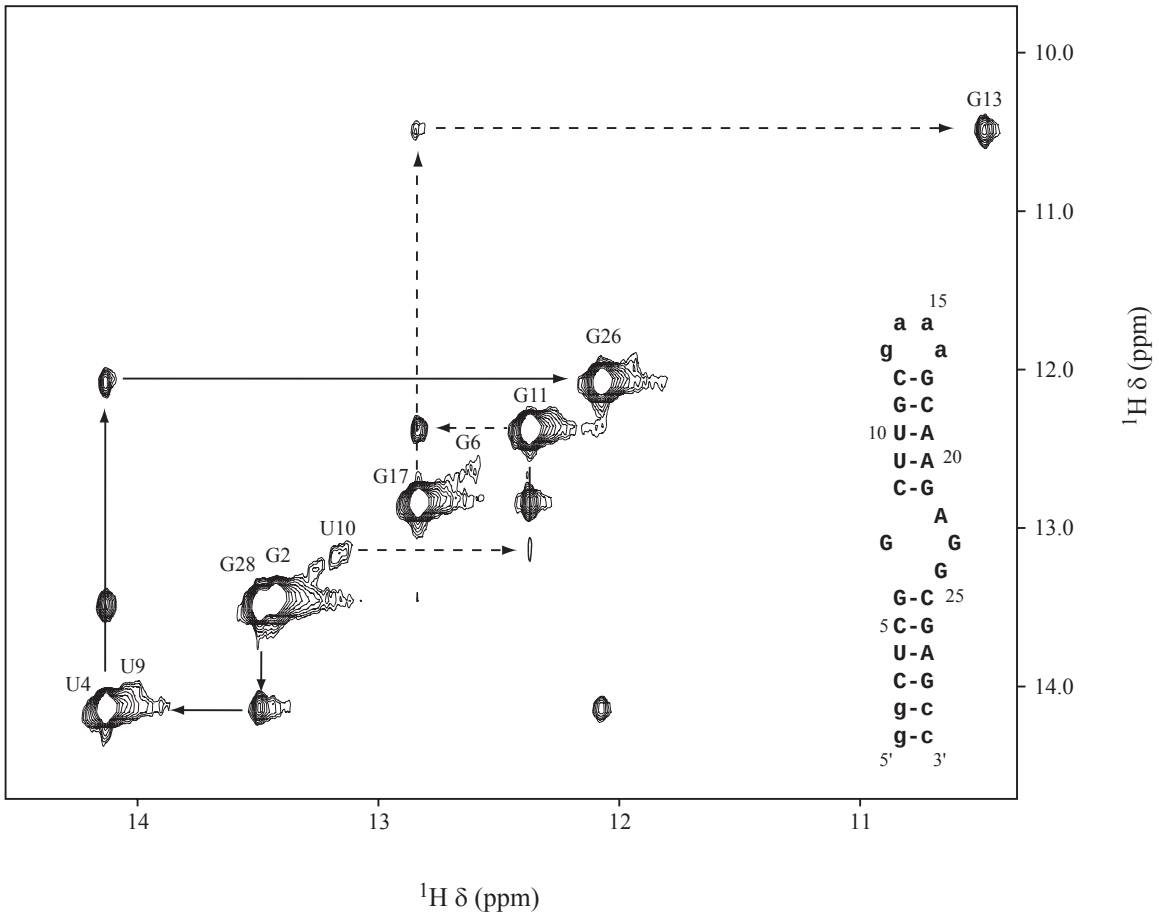


Figure 4.5 Imino-imino region of the 2D ^{1}H-^{1}H flipback watergate-NOESY of free SL1i. The assignment of the imino protons is shown, and lines connect each diagonal peak to the NOE crosspeak. Solid lines connect the signals from the lower stem and dashed lines connect signals from the upper stem. Also shown is the sequence and secondary structure of SL1i. The spectrum was collected at 5 ºC in 40 mM sodium phosphate pH 7.0 and 25 mM NaCl.

imino ^{1}H and adjacent amino ^{1}H signals. Imino signals were observed for 10 of the 12 predicted base pairs in the SL1i stem-loop. The imino signal of the terminal base pair is rarely observed in NMR spectra of RNA due to breathing at this position. The other base pair for which no imino proton is observed is the C_8-G_{21} adjacent to the internal loop. Characteristic NOEs were observed for the $G_{13}A_{14}A_{15}A_{16}$ tetraloop, indicating that the standard GNRA structure formed (data not shown).

To analyze the overall stability of the SL1i stem-loop structure and confirm the imino proton assignments, 1D ^{1}H spectra were collected for the SL1i RNA, both at low and high salt concentrations, at 5, 15, 25, 35, and 45 ˚C (Figure 4.6). At low salt concentrations, the first signal to disappear was the imino signal from the G_{13}-A_{16} base pair in the GAAA tetraloop (Figure 4.6A). At 25 ˚C, there was a significant loss in signal intensity for G_6, U_9, and U_{10}, and G_{13} at 35 ˚C, these signals had completely disappeared (Figure 4.6A).

The temperature profile of SL1i at high salt concentration is very similar to the profile at low salt concentration (Figure 4.6B). However G_2 and G_6 are relatively broader at high salt concentration than they were at low salt. This indicates that stem II is less stable than stem I, and this effect is more pronounced at high salt concentration.

High temperature, high RNA and protein concentrations, and high salt concentration yield optimal NMR spectra of the complex- Having established that NCp7 binds to SL1i at low protein and RNA concentrations by gel mobility shift assay, we wanted to determine if a complex could be formed at the high protein and RNA concentrations needed for NMR spectroscopy studies. ^{15}N-labeled NCp7 was titrated into SL1i while

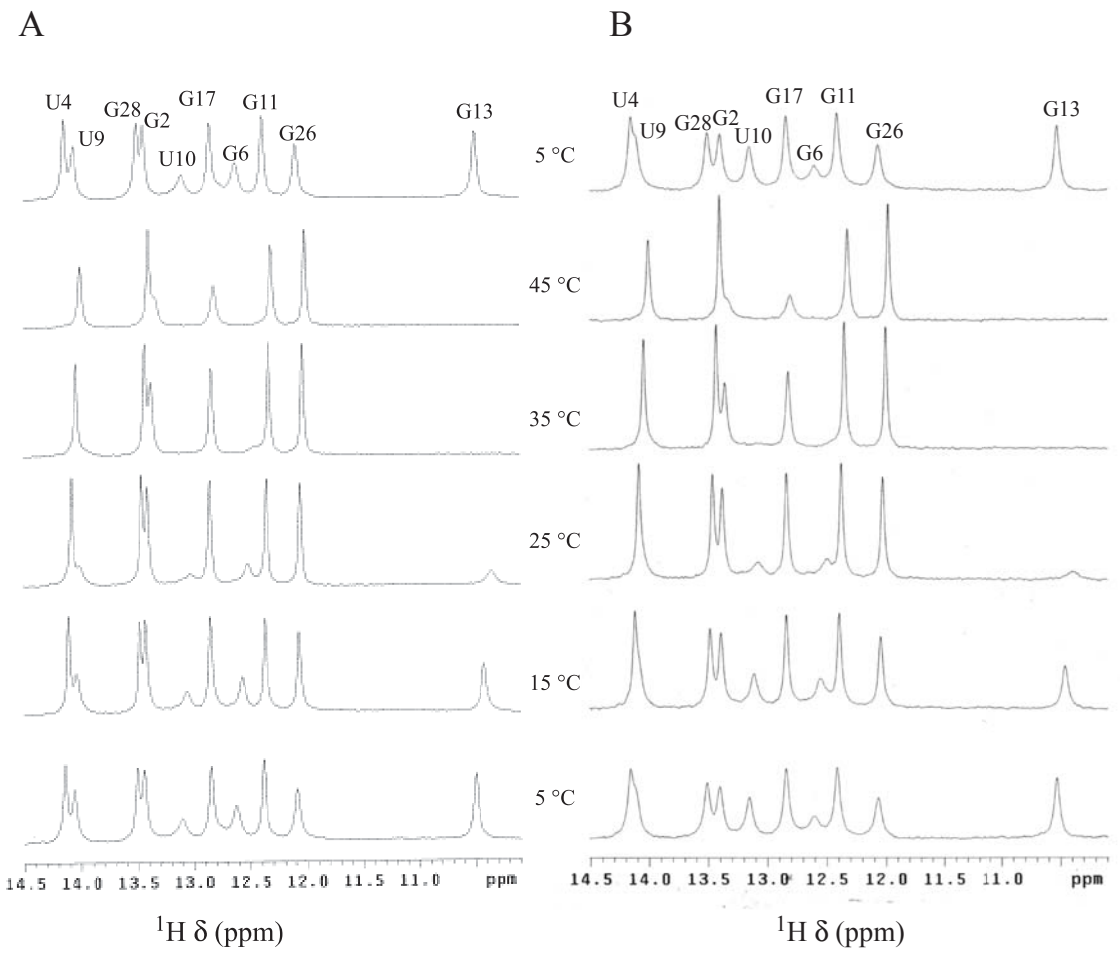


Figure 4.6: Temperature profile of SL1i at 25 mM NaCl (A) and 200 mM NaCl (B). In both A and B, a 1D ^{1}H spectrum was collected first at 5 ˚C, then at 25, 35, and 45 ˚C in 40 mM sodium phosphate pH 7.0. A final spectrum was collected at 5 ˚C to determine that the same structure was obtained after the temperature gradient.

monitoring complex formation using 1D ^{1}H flipback watergate and 2D ^{1}H-^{15}N HSQC spectra until a 1:1 complex was formed. The amide signals in the ^{1}H-^{15}N HSQC spectrum of the NCp7/SL1i complex were significantly broader than those of the free protein (Figure 4.7). A series of optimizations were next performed to find the conditions under which the linewidths of the protein amide signals and RNA imino signals were narrower.

The first condition to be varied was the temperature: 1D ^{1}H spectra were recorded on a complex of ^{15}N-labeled NCp7 and unlabeled SL1i at 5 ˚C, 15 ˚C, and 25 ˚C (data not shown). At 5 ˚C, the exchange of the non-base paired imino signals is lowered, so signals were observed for most of the base pairs in SL1i (data not shown). Between 15 ˚C and 25 ˚C, the imino signals of SL1i remained the same, but the amide signals of NCp7 became narrower (data not shown). As the imino protons were not largely affected by the increase in temperature, the protein signals were more closely monitored. The temperature of the sample was then varied from 25 ˚C to 37.5 ˚C and 2D ^{1}H-^{15}N HSQC spectra were collected. Between 25 ˚C and 35 ˚C, an overall increase in the number and intensity of the amide signals of NCp7 was observed (Figure 4.8). This increase in signal intensity is possibly due to a decrease in the line width of these peaks. At 37.5 ˚C, however, there was no improvement in the quantity or intensity of the signals, and instead some signals became weaker or were lost (Figure 4.8). Thus, it was determined that 35 ˚C was the optimal temperature for the NMR spectroscopy studies of the NCp7/SL1i complex.

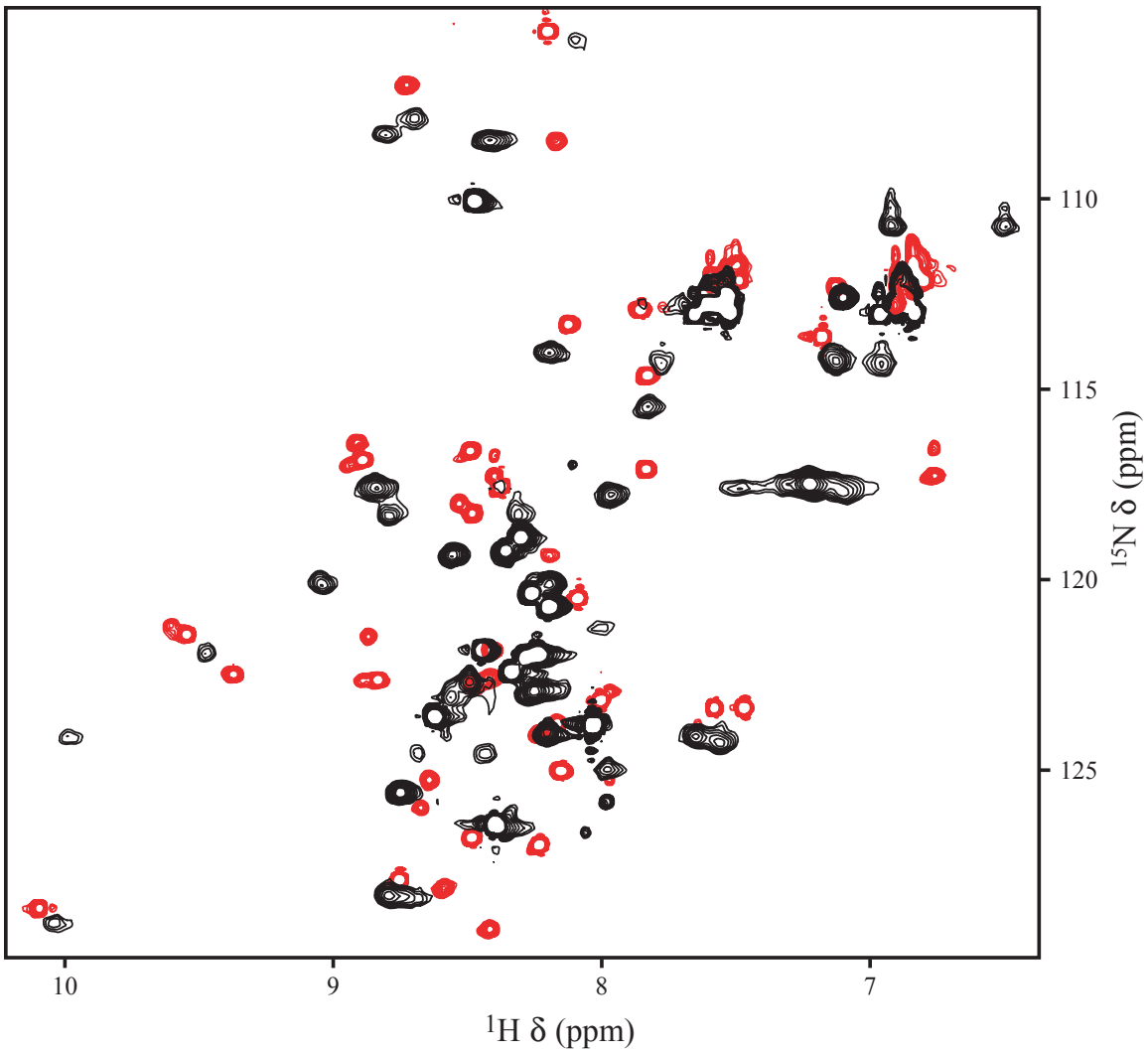
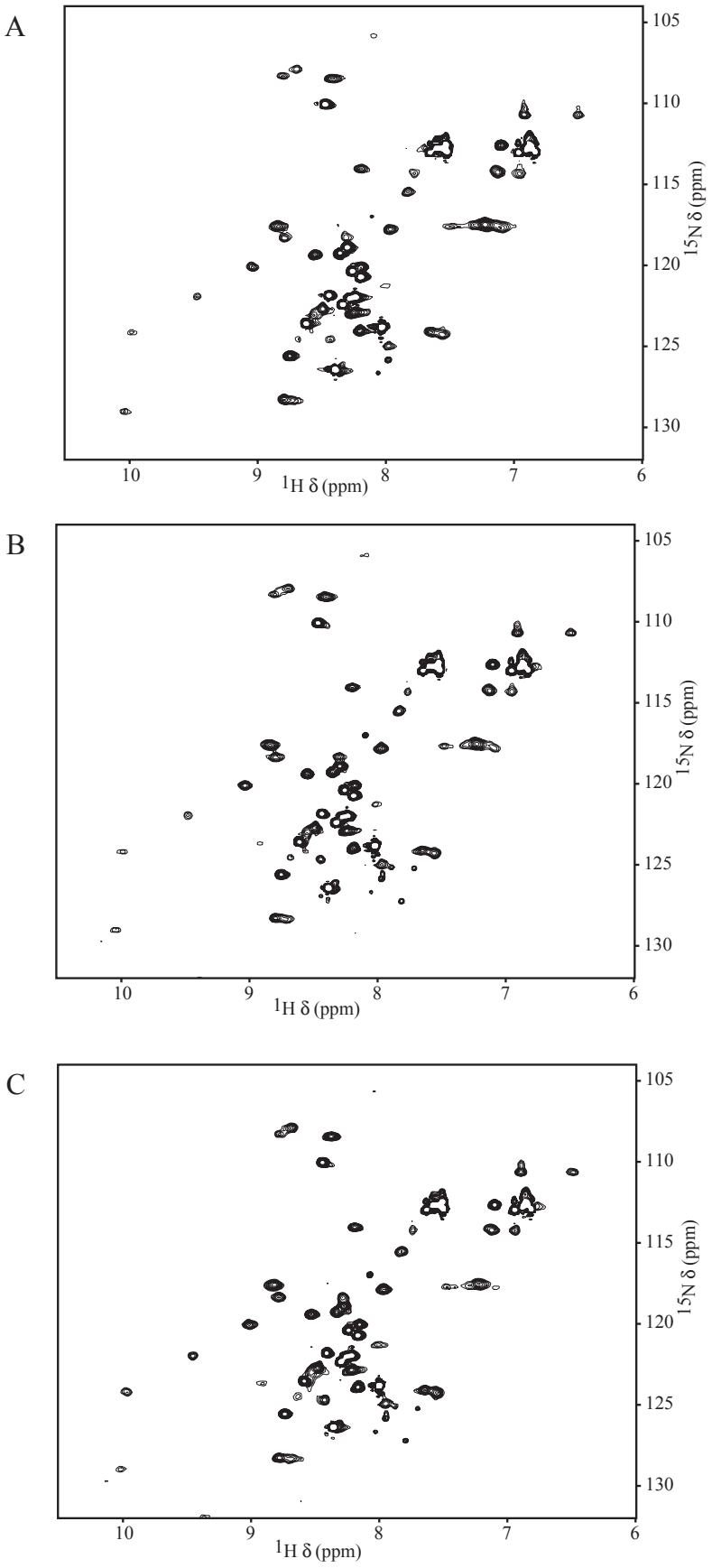
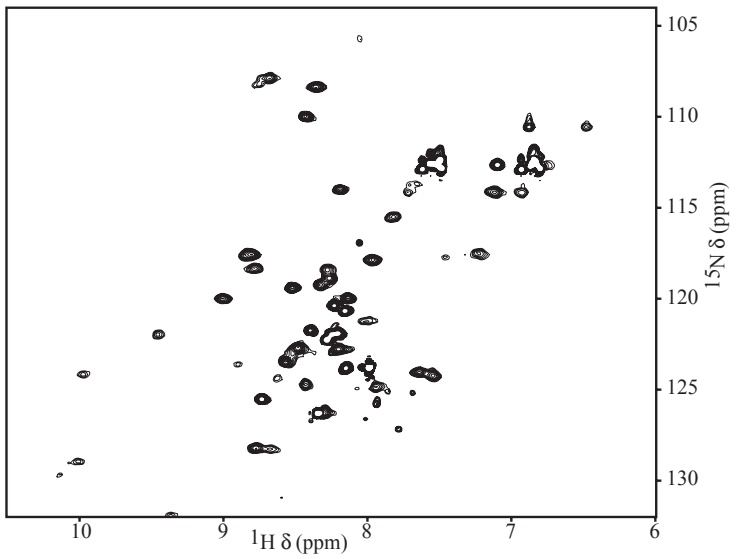


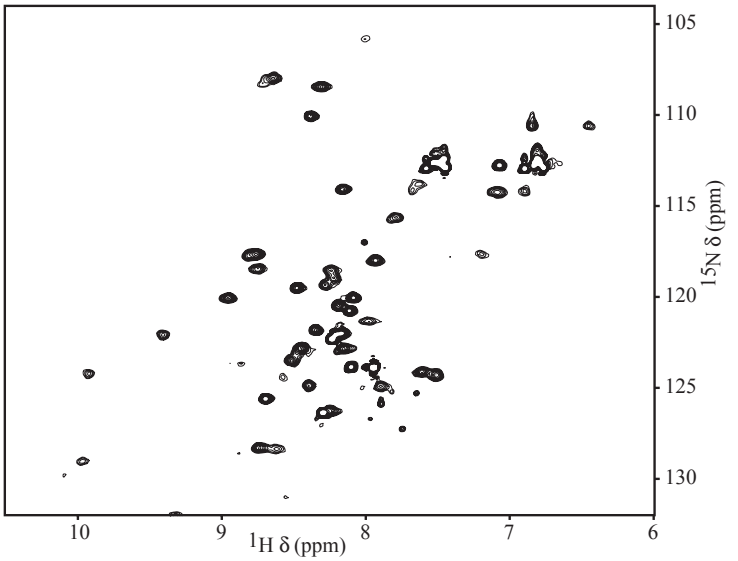
Figure 4.7: 2D ^{1}H-^{15}N HSQC spectrum of NCp7 free (red) superimposed on the 2D ^{1}H-^{15}N HSQC spectrum of the NCp7/SL1i complex (black). The spectra of both the free NCp7 and the NCp7/SL1i complex were collected at 25 ºC, 25 mM sodium phosphate pH 7.0.


A
B
C
105
110
115
120
125
130
^{15}N δ (ppm)
10
9
8
7
6
^{1}H δ (ppm)

D


105
110
115
120
125
130
15N δ (ppm)
10
9
8
7
6
1H δ (ppm)

E


105
110
115
120
125
130
15N δ (ppm)
10
9
8
7
6
1H δ (ppm)

F

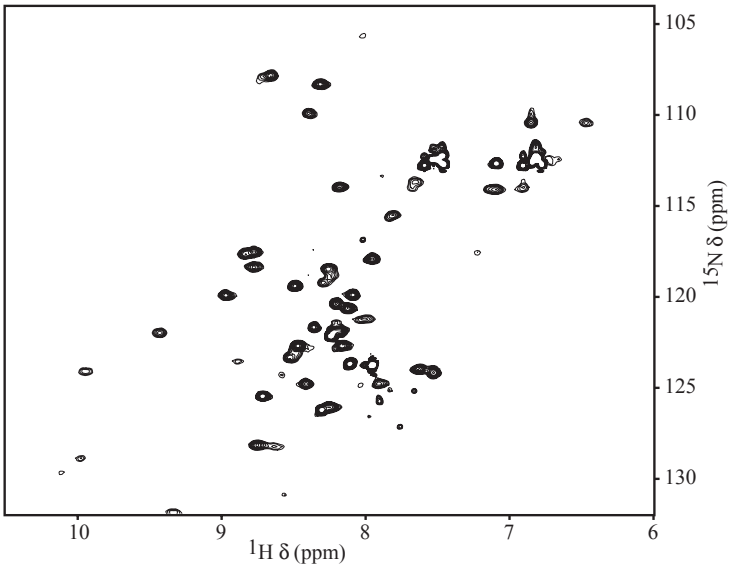

105
110
115
120
125
130
15N δ (ppm)
10
9
8
7
6
1H δ (ppm)

Figure 4.8: 2D ^{1}H-^{5}N HSQC spectra of ^{15}N-labeled NCp7 complexed with SL1i (1 mM complex) collected at various temperatures in 25 mM NaCl, 20 mM sodium phosphate buffer pH 7. The 2D ^{1}H-^{15}N HSQC spectra were collected at 25 (A), 27.5 (B), 30 (C), 32.5 (D), 35 (E), and 37.5 ˚C (F) and plotted at the same level.

The salt concentration of the complex was varied from 25 mM NaCl to 200 mM NaCl, specifically 25 mM, 50 mM, 100 mM, and 200 mM NaCl were used. The imino signals of the RNA did not change much over this range, but the signals in the protein seemed to get significantly less broad and more intense at higher salt concentration (Figure 4.9). It is possible that the higher salt concentration minimized non-specific protein:RNA interactions and stabilized the specific NCp7 interaction with the internal loop of the SL1i RNA. We also observed that increasing the salt concentration in a complex between NCp7 and SL3 also decreased the linewidth of the signals (LMJ, JGO, PL unpublished data). It was decided that the optimal salt concentration was 200 mM NaCl in a buffer containing 20 mM sodium phosphate pH 7.0.

The pH of the complex was varied from pH 6.5 to pH 7.5. There were no significant changes in either the NCp7 amide signals or the SL1i imino proton signals over this range of pH (data not shown). It was not possible to go to a much lower pH, as NCp7 does not stably coordinate zinc below pH 6. As there was no optimal pH, pH 7 was chosen.

The two NCp7/SL1i concentrations studied were 1 and 2 mM. The higher complex concentration was clearly the best, yielding better signal:noise (data not shown). This result was not surprising, as higher concentration samples can give better spectra. Therefore, the final sample conditions chosen were: 2 mM NCp7/SL1i complex, 40 mM sodium phosphate pH 7.0, 200 mM NaCl, and 35 ˚C.

Although these sample conditions provided optimal signal intensity, we observed that NCp7 was degraded over time. Indeed, after only one week at 35 ˚C, we observed some loss of signal intensity and the appearance of new signals indicative of sample

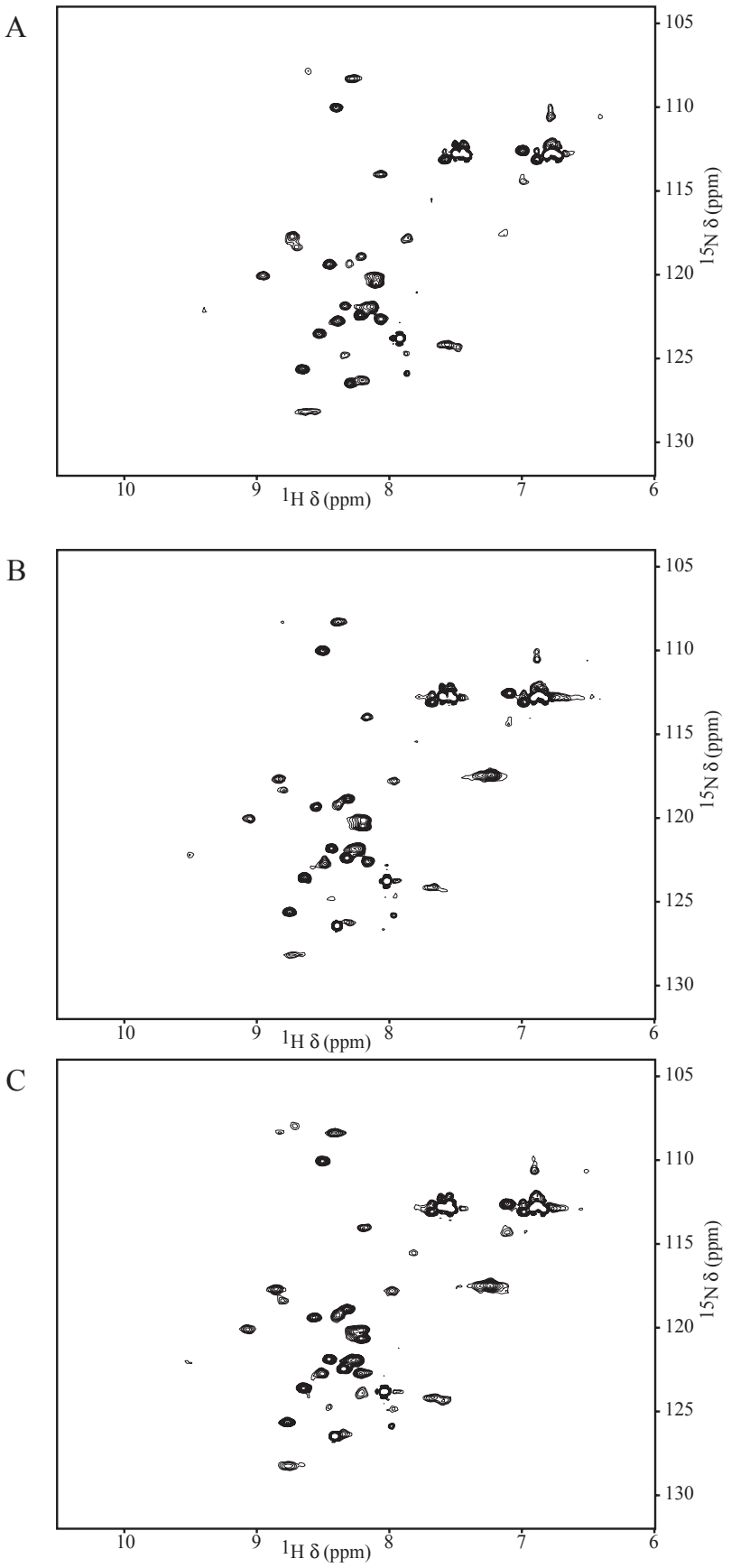

A
B
C
105
110
115
120
125
130
15N δ (ppm)
10
9
8
7
6
1H δ (ppm)

Figure 4.9: 2D ^{1}H-^{5}N HSQC spectra of ^{15}N-labeled NCp7 complexed with SL1i (1 mM complex) collected at various salt concentrations at 25 ˚C. The 2D ^{1}H-^{15}N HSQC spectra were collected with 25 mM NaCl and 20 mM sodium phosphate buffer pH 7.0 (A), 50 mM NaCl and 20 mM sodium phosphate buffer pH 7.0 (B), and 100 mM NaCl and 20 mM sodium phosphate buffer pH 7.0 (C). The spectrum collected with 200 mM NaCl and 40 mM sodium phosphate buffer pH 7.0 is shown in Figure 4.8A. All spectra were plotted at the same level.

degradation (data not shown). We observed by denaturing gel that the SL1i in a sample that was two months old showed a large amount of degradation (data not shown). The SL1i degradation could have been caused by free zinc in solution from the degraded NCp7 (data not shown). From our experiments, it seemed that each complex sample could not be used more than one month (data not shown), or the degradation became too great. The NOESY experiments were more sensitive to the sample degradation than the through-bond experiments. The sample degradation caused the most problem with the filtered NOESY used to connect the protein to the RNA which was collected on a sample that had been used for two weeks. No intermolecular NOEs were observed in this experiment, resulting in the inability to link NCp7 and SL1i in space. The lack of signals was likely due to degradation of the sample, leading to loss of signal intensity in the crucial signals. In addition, we observed small inconsistencies between different complex samples, such as the relative intensities of the imino proton signals of the SL1i. The cause of these inconsistencies is unknown.

Resonance assignments of the NCp7/SL1i complex- Assignments of the exchangeable ^{1}H signals and their attached ^{15}N signals of SL1i in the NCp7/SLi complex were made as described in Materials and Methods (Supplementary Table 4.3). The 2D ^{1}H-^{15}N HSQC spectrum of the ^{15}N-labeled NCp7/SL1i complex was characterized by good chemical shift dispersion, though the signals were significantly broadened compared to the free protein (Figure 4.7, black peaks). The first two amino-terminal residues were not observed in the 2D ^{1}H-^{15}N HSQC spectrum, likely due to fast amide proton exchange and/or signal broadening. For the remaining residues (3-55), nearly complete assignment

was achieved for the backbone: 100% of the $^{1}H^{N}$, 98% of the ^{15}N, 98% of the $^{13}C^{\alpha}$, 98% of the $^{1}H^{\alpha}$, and 96% of the $^{13}C'$ (Supplementary Table 4.4). For residues 2-55 in the NCp7/SL1i complex, the extent of assignment for the side chains was 65%: 65% of the side chains were completely assigned and 35% were partially assigned (Supplementary Table 4.4).

Binding of SL1i induces a conformational change in NCp7- The 2D ^{1}H-^{15}N HSQC of the NCp7/SL1i complex revealed many chemical shift changes in the amide signals of the protein (Figure 4.7). The chemical shift dispersion indicates that NCp7 retains a folded conformation upon binding to SL1i. The changes in amide chemical shift suggest that the structure of complexed NCp7 is significantly different from the free protein.

A comparison of the chemical shifts of NCp7 in the free and bound states revealed a number of large changes in the amide resonances of NCp7 (Figure 4.10A). The largest chemical shift changes (0.75-1.0 ppm) were observed for the amides of Val_{13}, Lys_{14}, Cys_{15}, Phe_{16}, Asn_{17}, Gly_{19}, Ala_{25}, Arg_{26}, and Cys_{28} in the amino-terminal ZBD (ZD1) and Trp_{37} and Met_{46} in the carboxyl-terminal ZBD (ZD2) (Figure 4.10). The amide signals of Cys_{18} and Thr_{24} in ZD1, Arg_{29} in the linker, and Glu_{42} in ZBD2 underwent moderate changes in chemical shift (0.5-0.74 ppm) (Figure 4.10). Small, but significant (0.25-0.49 ppm), amide chemical shift changes were observed for Gly_{20} and Glu_{21} in ZD1, Cys_{36}, Lys_{38}, Cys_{39}, Gly_{40}, Gly_{43}, and Lys_{47} in ZD2, and Gln_{53} and Ala_{54} in the carboxyl terminus (Figure 4.10). Overall, 25 of the 45 amide signals (56%) for which

assignments were available in both the free and bound states showed a significant change (> 0.25 ppm) in chemical shift.

The residues that showed significant amide chemical shift changes are primarily located in the two ZBDs of NCp7 (Figure 4.10B). For example, the residues between the first two cysteines of each domain (Phe_{16}, Asn_{17}, Trp_{37}, and Lys_{38}) show large amide chemical shift changes. Also, the amide signals for the glutamate residues between the second cysteine and the histidine in each domain (Glu_{21} and Glu_{42}) move considerably (Figure 4.10B). Most of the amide chemical shift changes that were not localized to the ZBDs were found immediately before and after ZD1: Val_{13}, Lys_{14}, Arg_{29}, and Ala_{30} (Figure 4.10B). Interestingly, the equivalent residues before and after ZD2 (Gly_{35}, Thr_{51}, and Glu_{51}) did not undergo significant chemical shift changes (Figure 4.10B). The backbone signal for Lys_{34} was not observed in the free NCp7, so no chemical shift comparison could be made.

Comparison of NCp7 conformation in the NCp7/SL1i complex with that in the NCp7/SL3 complex- A comparison of the change in chemical shift between free NCp7 and SL3-bound NCp7 reveals similar changes to the SL1i-bound protein (4.11). In both complexes, the amide signals of the following residues show significant changes (> 0.25 ppm) in amide chemical shift: Val_{13} in the amino terminus; Cys_{15}, Glu_{21}, and Ala_{25} in ZBD1; Arg_{29} and Ala_{30} in the linker; Cys_{36}, Trp_{37}, Lys_{38}, Gly_{43}, Met_{46}, and Lys_{47} in

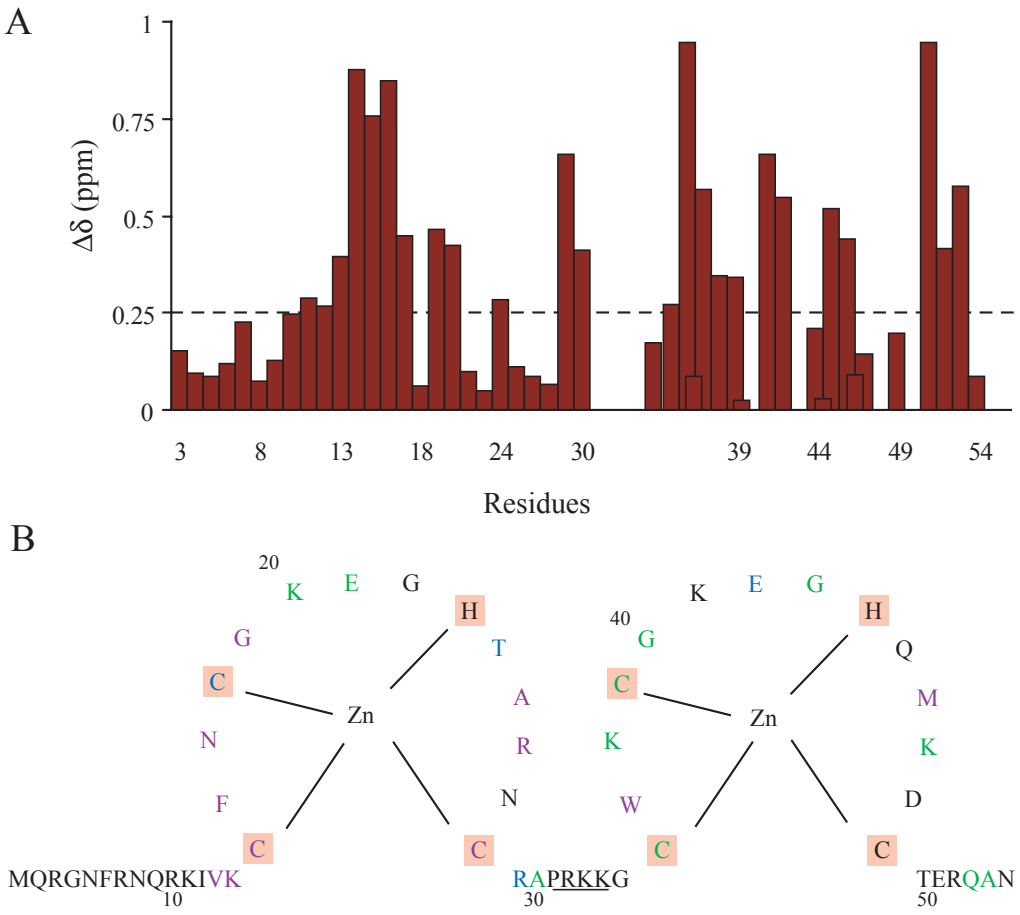


Figure 4.10: Observed changes in amide chemical shift of NCp7 upon complex formation with SL1i. A) Histogram displaying the amide chemical shift differences between the free and SL1i-bound NCp7. The change in amide chemical shift was calculated as: $\Delta\delta=[(0.17\Delta N_H)^2+(\Delta H_N)^2]^{1/2}$ [40]. B) Schematic diagram of residues in the sequence of NCp7 that show a significant change in amide chemical shift in the complex with SL1i. The residues with chemical shift changes between 0.75-1.0 ppm are shown in purple, 0.5-0.74 ppm are shown in blue, 0.25-0.49 ppm are shown in green. Residues that are not assigned in either the free or bound states are underlined (no Δδ was calculated).

ZBD2; and Gln_{53} and Ala_{54} at the carboxyl terminus (Figure 4.11). This result shows that there is a significant change in the amide chemical shift of 6 of the 14 residues (43%) of ZD2 in both complexes. A similar comparison in ZD1 is not possible due to lack of amide chemical shift assignment for several residues in ZD1 in the NCp7/SL3 complex.

The primary structural feature of NCp7 in complex with SL3 is the 3_{10} helix formed by the amino-terminal residues 3-10 that binds in the RNA minor groove[6]. Helical structure has been shown to alter the chemical shift of the C^{α}, H^{α}, C^{β}, and C′ in a predictable way compared with random coil values[41]. Negative differences between the random coil ^{1}H values and the sample chemical shifts are indicative of formation of a helix[41]. Chemical shift index analysis did not suggest that this 3_{10} helix formed in the NCp7/SL1i complex (data not shown). However, comparison of the H^{α} chemical shift of NCp7 with random coil values shows that, although no obvious helix is forming ($\Delta\delta$=-1), residues 3-9 have helical tendencies (Figure 4.12). In a 3_{10} helix, NOEs are expected between the alpha proton of residue *i* and the amide proton of residue *i+2*[42]. This NOE pattern is not observed for α-helices and helps to distinguish the two types of helices. A few such NOEs were observed, for example between Arg_{3HA} and Asn_{5HN} (data not shown), however there are not very many of these NOEs. Thus, it is possible that NCp7 is not stably bound to SL1i.

In the NCp7/SL3 complex, the ZBDs of NCp7 interact strongly with one another, forming a more compact structure[6]. NOEs were observed between residues from each of the domains that were not observed in the free protein[6]. Very few long-range NOEs were observed between the two ZBDs in the NCp7/SL1i structure. Therefore, there is little

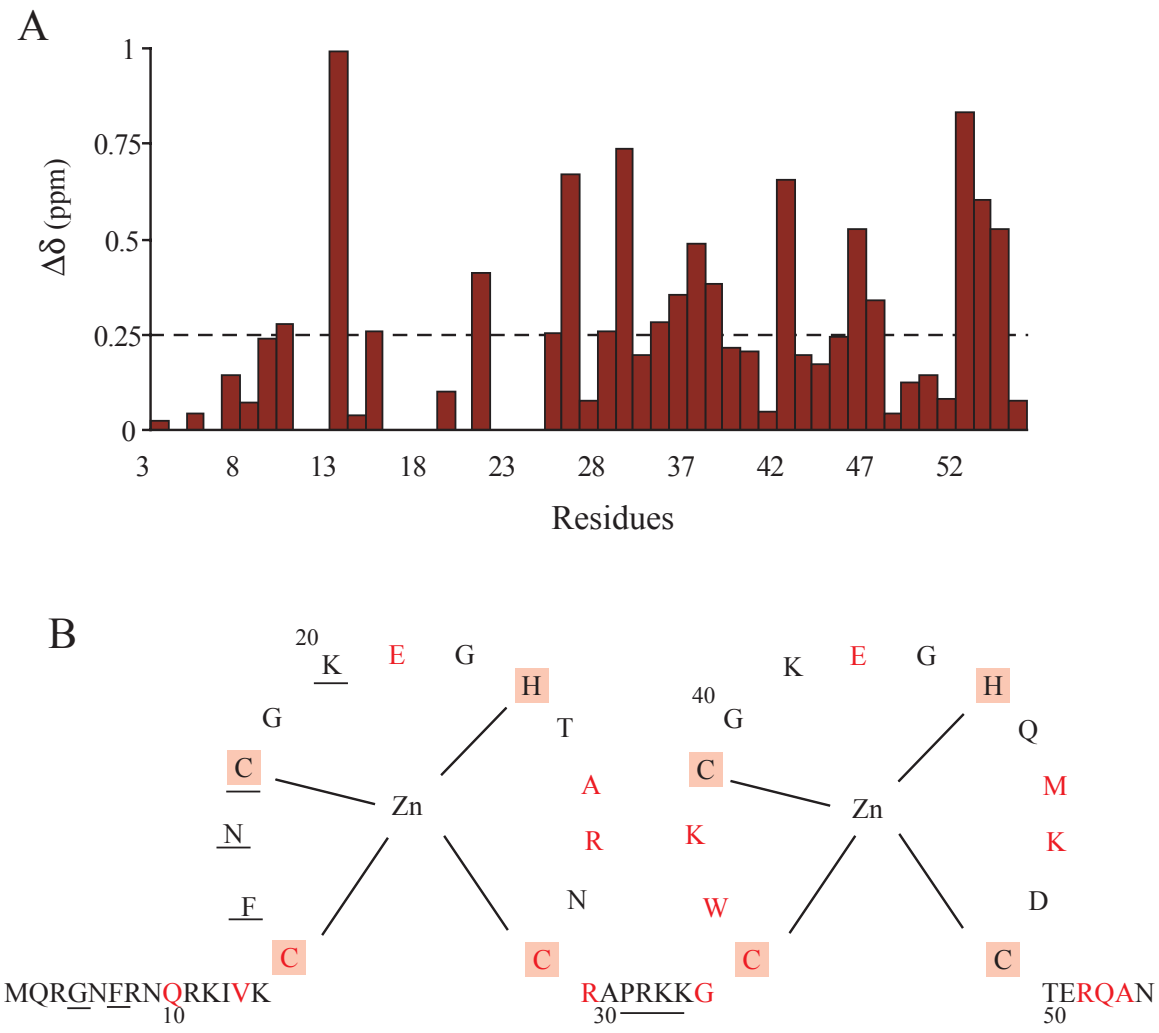


Figure 4.11: Observed changes in amide chemical shift in NCp7 upon complex formation with SL3. A) Histogram displaying the amide chemical shift differences between the free and SL3-bound NCp7. The change in amide chemical shift was calculated as: $\Delta\delta = [(0.17\Delta N_H)^2+(\Delta H_N)^2]^{1/2}$ [40]. B) Schematic diagram of residues in the sequence of NCp7 that show a significant change in amide chemical shift in the complex with SL3. The residues with chemical shift changes between 0.25-1 ppm are shown in red. Unassigned residues in either free of SL3-bound NCp7 are underlined.

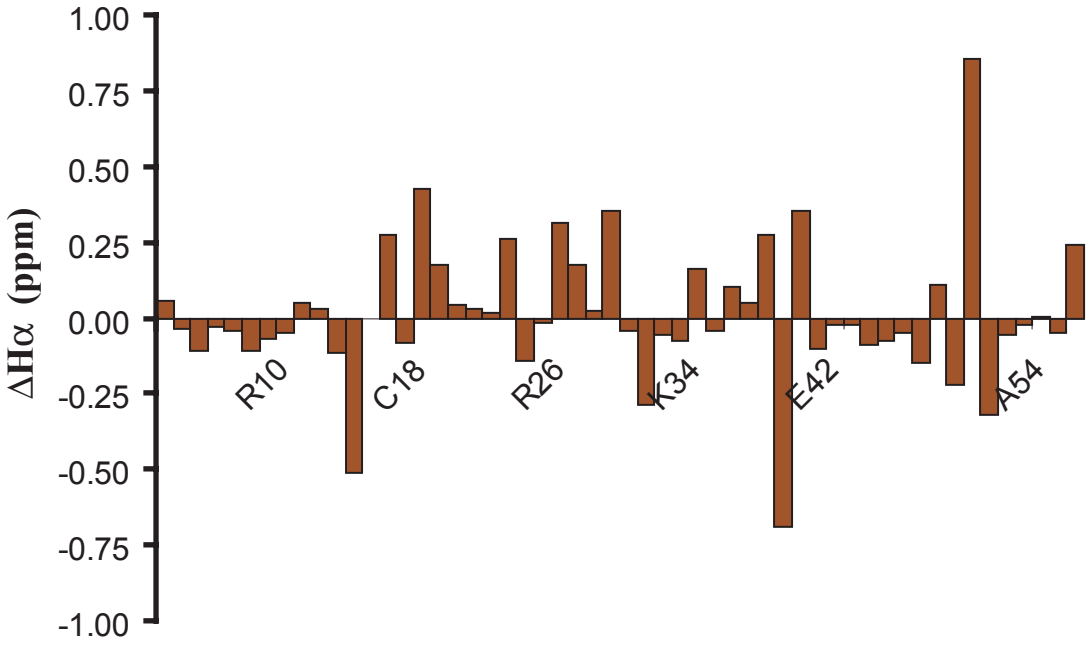


Figure 4.12: Comparison of the H^{α} chemical shifts in the NCp7/SL1i complex with random coil values. The difference in chemical shift was calculated as $\Delta H^{\alpha}=H^{\alpha}_{random}-H^{\alpha}_{NCp7}$. Random coil values have been previously reported[41].

structural evidence that indicates that the two ZBDs interact with one another in the NCp7/SL1i complex. Thus, NCp7 may form a somewhat different, less compact, structure when binding to SL1i than when binding to SL3 or SL2 RNAs.

In the solution structures of NCp7 complexed with SL3 and with SL2, the guanine bases bind in hydrophobic clefts formed by the side chains of Val_{13}, Phe_{16}, Ile_{24}, and Ala_{25} or Trp_{37}, Gln_{45}, and Met_{46}[6,43]. The amides of all of these residues underwent significant chemical shift changes in the NCp7/SL1i complex. This indicates that these residues are in a significantly different conformation in the complex than in the free protein. It is possible, therefore, that these hydrophobic clefts are also being formed in the NCp7/SL1i complex and that NCp7 recognizes the guanines in the internal loop through a similar mechanism.

Effect of NCp7 binding on SL1i- When NCp7 was titrated into SL1i, there were several changes in the imino region of the spectrum (Figure 4.13). The signals for G_{13}, G_{26}, and U_4 decreased in intensity (Figure 4.13). The imino protons for G_6 and U_9 were not observed in the bound state (Figure 4.13). In addition, the observed imino signals were significantly broadened compared to the free imino signals. The 2D imino-optimized ^{1}H-^{15}N HSQC demonstrated that only two A-U base pairs and five G-C base pairs were stable in the NCp7/SL1i complex (Figure 4.14). Also, signals from NCp7 were observed at 10.0 ppm that were not observed in the free protein (Figure 4.13).

In the NCp7/SL1i complex, there were seven imino ^{1}H signals, whereas in the free SL1i there were nine. This loss of imino ^{1}H signals indicates that addition of NCp7 to SL1i destabilizes base pairs around the internal loop (U_9-A_{20} and G_6-C_{25}), such that the

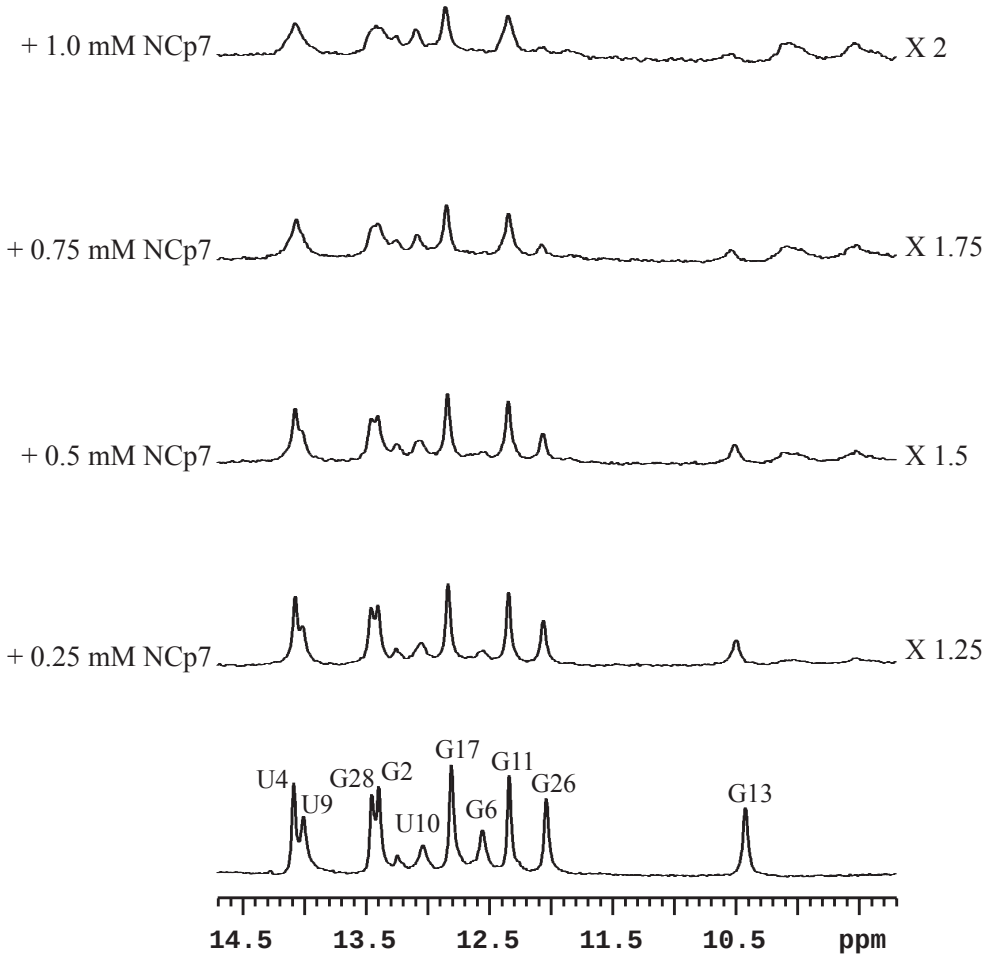


Figure 4.13: Imino proton region of the 1D ^{15}N-decoupled ^{1}H spectra of SL1i to follow the titration of NCp7. A 1D ^{1}H spectrum was collected after addition of ^{15}N-labeled NCp7 to unlabeled SL1i (1 mM) in 25 mM NaCl, and 20 mM sodium phosphate buffer pH 7.0, 90% H_2O/10% D_2O at 25 ˚C. The vertical scale of the spectra was adjusted by the factor indicated to correct for sample dilution.

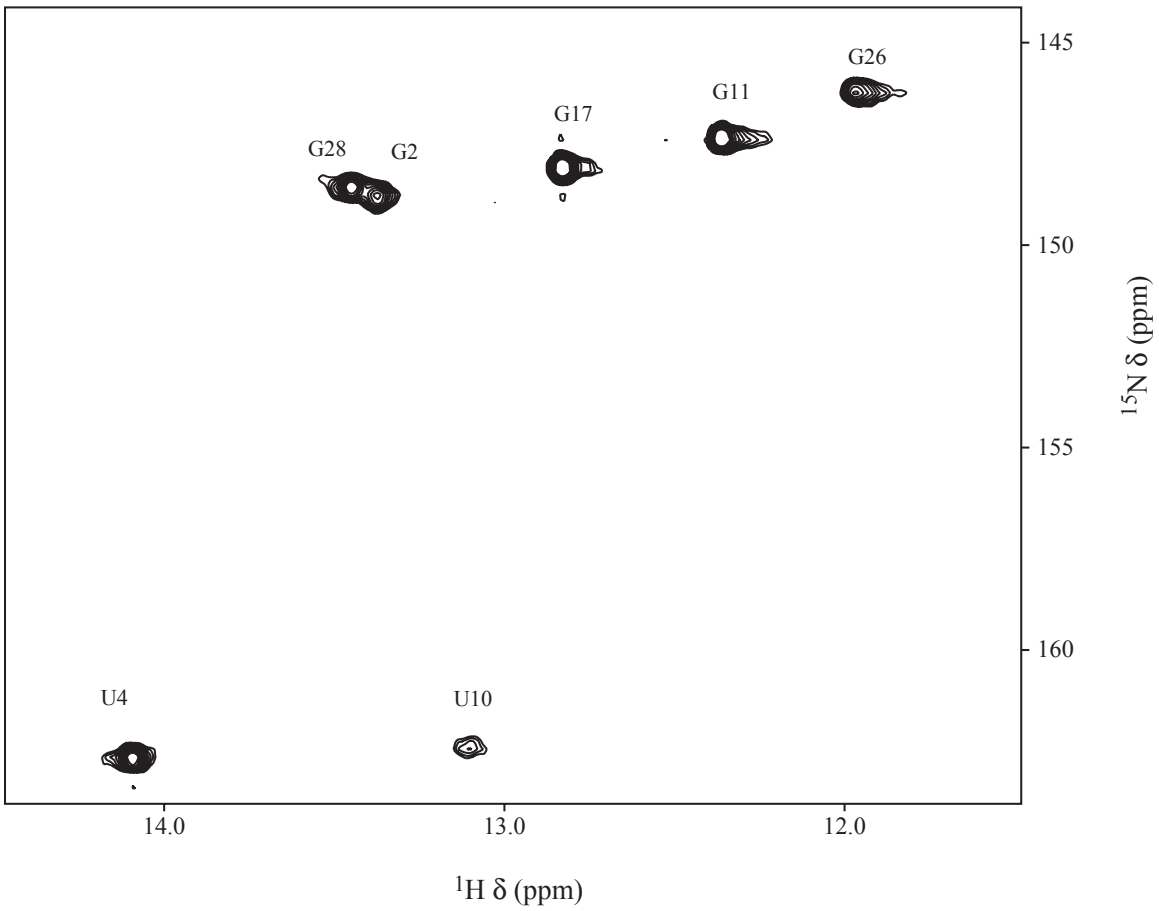


Figure 4.14: Portion of the imino-optimized 2D ^{1}H-^{15}N HSQC of the NCp7-bound SL1i. The spectrum was collected at 35 ºC in 40 mM sodium phosphate buffer pH 7.0 and 200 mM NaCl.

imino protons are able to exchange with water. Comparison of the imino proton chemical shifts of the SL1i stem in the free and bound states revealed few differences in chemical shift (Supplementary Tables 4.2 and 4.3). The chemical shift of the imino proton of G_{28}, however, was somewhat altered compared with the free RNA (12.05 ppm in the free form and 11.90 in the bound form) (Figure 4.13). This indicates that NCp7 does not induce significant conformational changes in the SL1i RNA. Therefore, the primary effect of NCp7 binding on SL1i is the destabilization of base pairs in stems I and II around the internal loop.

Surprisingly, the G_{13} imino proton signal from the G_{13}-A_{16} base pair in the GAAA tetraloop also lost intensity during the titration with NCp7. This loss in signal intensity suggests that NCp7 could interact with the GAAA tetraloop, even though gel mobility shift analysis demonstrated that it did not bind $SL1i_m$ (Figure 4.2). When salt was added to the sample, however, the intensity of this peak increased (data not shown). The effect of salt concentration suggests that NCp7 was able to non-specifically interact with G_{13} in the GAAA tetraloop, and the high salt concentration reduced this interaction. Indeed, it has been observed in fluorescence binding studies with NCp7 that 200 mM NaCl is required to minimize non-specific binding[19].

Discussion

In this report, the interaction of HIV-1 NCp7 with the internal loop of the first stem-loop in the Ψ site, SL1i, was analyzed using gel mobility shift assays and NMR spectroscopy. This study was severely limited by the stability of the sample that led to our inability to solve the solution structure of this complex. We observed degradation of

both the NCp7 and the SL1i after one and two months of experiments, respectively. In addition, we observed evidence of some non-specific interaction between NCp7 and the GAAA tetraloop. Interaction with this loop could have lead to line broadening, as there was a mixed population of NCp7 bound to the terminal GAAA loop and the internal loop. There was not enough NOE data to facilitate structure determination, both between the NCp7 and SL1i and within NCp7 itself. This issue may be remedied using an alternate tetraloop sequence, such as a UNCG tetraloop that is also stable and would facilitate formation of the hairpin conformation.

We observed that binding of SL1i induced some conformational changes in NCp7. There was a significant shift in the backbone of the protein, especially among residues in the two ZBDs. The ZBDs have been shown in other studies to specifically bind to RNA[6,43]. Phe_{16} and Trp_{37} are among the residues that show significant amide chemical shift change in the NCp7/SL1i complex. These residues made important contacts with the RNA in other NCp7/RNA structures[6,43], and it is possible that their large amide chemical shift change indicates a change in conformation in the NCp7/SL1i complex. It was observed that incubation of NCp7 with the 2-mercaptobenzamide thioester compound **1** was able to block the binding of NCp7 to SL1i. When the protein and RNA were preincubated, however, NCp7 did not react with compound **1** and remained bound to SL1i. The same result was observed for NCp7 binding to SL3; interaction of compound **1** with NCp7 can block binding to SL3, but the complex was not affected by the thioester compound[39]. The structure of NCp7 in the complex with SL1i does not seem to be the same as in the complex with SL3, though there may be some similar binding characteristics. Alternatively, the NCp7/SL1i complex is more dynamic

than the SL3 complex. NCp7 may not be completely bound and undergoes fast exchange in the NMR timescale.

In the NCp7/SL1i complex, we observed that two stem base pairs were destabilized compared to the free RNA, one in stem I and one in stem II. When NCp7 was titrated into SL1i, the signal for the imino protons of G_6 and U_9 of these base pairs disappeared as more NCp7 was added. Most of the other base pairs remained unchanged, though there was a change in the chemical shift of the imino proton of G_{26}. The loss of the imino signals suggests that the base pairs were significantly less stable when NCp7 was bound to the RNA, though they may still form. There was little change in chemical shift for the imino and amino protons of the other bases in the stem, indicating that NCp7 did not induce any conformational change in SL1i. The protons in the bases of the loop nucleotides were not assigned in the complex, so no assessment could be made about changes in the NCp7/SL1i complex.

In our study, complex formation was monitored using the imino region of the 1D ^{1}H spectrum. Once the complex was formed, we observed that binding of NCp7 resulted n the destabilization of two base pairs. Thus, the imino signals may not have been the best gauge of complex formation. Titration of more NCp7, beyond equimolar concentration could have revealed important changes in the complex. It is possible that higher concentrations of NCp7 would have led to further loss of imino signals and, thus, stem-destabilization. In addition, the complex could have been subjected to analytical ultracentrifugation sedimentation experiments to determine if a 1:1 complex was truly formed.

The temperature profile of free SL1i revealed that base pairs in stems I and II around the internal loop (G_6-C_{25} and U_9-A_{20}) were among the first to melt as the temperature increased (Figure 4.11). Interestingly, these same base pairs were destabilized upon NCp7 binding (4.12). Thus, high temperature appears to have the same effect on SL1i as binding of NCp7. Both destabilize base pairs around the internal loop. Interestingly, this similarity in the effect of temperature and NCp7-binding has also been observed for the dimerization reaction. Either NCp7 or high temperatures are required for efficient kissing loop to extended duplex conversion[14]. Thus, it is possible that binding of NCp7 to the internal loop of SL1i has a role in the dimerization reaction.

Based upon our NMR spectroscopy results, we propose that the internal loop functions as a nucleation point for the destabilization and melting of ΨSL1. This destabilization would be necessary for the conversion of the kissing loop to the extended duplex. As described above, following formation of the kissing loop, the ΨSL1 stem of each molecule of genomic RNA begins melt, allowing full base-pairing between the two strands of RNA, and finally forming the extended duplex. The internal loop is a disruption in the stem, which makes ΨSL1 less stable than if the stem were longer and uninterrupted. Indeed, we observed that the internal loop of ΨSL1 showed some instability even in the free form. For example, the C_8-G_{21} base pair that closed stem I was not observed in the free state (Figure 4.5). Thus, this loop could be a good starting point for stem destabilization.

Binding of NCp7 to the internal loop could greatly facilitate the melting of the stem, enhancing the formation of the extended duplex. It has previously been shown that NCp7 specifically enhances the conversion from the kissing loop to the extended duplex

conformation[4,5]. Thus, our proposal that destabilization of the stem begins at the internal loop is consistent with the observations of previous studies. Therefore, NCp7 could bind to the internal loop of SL1 to destabilize the stem for optimal extended duplex formation.

NCp7 has been proposed to act as a chaperone protein by destabilizing stable stem-loop structures at other stages of the HIV-1 lifecycle as well[44-52]. For example, NCp7 facilitates binding of the $tRNA_3^{Lys}$ primer to the primer binding site (PBS) in part by destabilizing the stable secondary structure of the tRNA and the (PBS)[44,45,53]. NCp7 is able to unwind the acceptor stem of the tRNA in the presence of an RNA template[53]. In addition, it destabilizes a four base pair helix in the PBS and facilitates formation of a new duplex between the tRNA and the PBS[45]. NCp7 is also thought to block self-priming reactions and reduce pausing during reverse transcription by destabilizing the stable secondary structures in the genomic RNA, including the DNA-complement to the HIV-1 transactivation response element (cTAR) stem-loop structure[46-50]. Fluorescence data suggests that NCp7 opens base pairs in cTAR beginning at the least stable parts of the stem[47]. This result is similar to what we observe with SL1i, where the destabilization occurs around the internal loop. Clearly, this function of NCp7 is crucial throughout the HIV-1 lifecycle.

Much more research is needed to fully understand the destabilization role of NCp7. Future experiments will more closely examine the destabilizing activity of NCp7. For example, UV melting studies will be performed with various RNAs, including SL1i and the internal loop of the TAR RNA, to measure the effect of NCp7 binding on their melting. Similar studies have already been conducted with duplex DNA molecules[51]. RNA molecules, however, can adopt a greater variety of secondary structures that might

impact the destabilization function. The stem and loop sequences of these RNAs will be altered to determine if there is a preference for NCp7-activated destabilization. As the dimerization reaction has been shown to be influenced by magnesium, its effect on the binding and destabilizing role of NCp7 will also be investigated.

We have shown that NCp7 is able to bind to the internal loop of ΨSL1 with a relatively strong affinity ($K_D \approx 100$ nM). This interaction has properties similar to that of other NCp7/RNA complexes, such as reaction with a 2-mercaptobenzamide thioester compound. To the best of our knowledge, this report is the first attempt to a structural study of the interaction of NCp7 with ΨSL1. Binding of NCp7 induces the destabilization of the stem of ΨSL1, as evidenced by the loss of base pairs around the internal loop. The destabilization is crucial for the conversion from kissing loop to extended duplex in the dimerization process[52]. Therefore, the internal loop of ΨSL1 likely functions as a nucleation site for the destabilization of the internal loop by NCp7 for optimal genomic RNA dimerization.

References

(1) Clever, J.; Sassetti, C.; Parslow, T. G. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J. Virol.* **1995***, 69*, 2101-2109.

(2) Amarasinghe, G. K.; Zhou, J.; Miskimon, M.; Chancellor, K. J.; McDonald, J. A. et al. Stem-loop SL4 of the HIV-1 Ψ RNA packaging signal exhibits weak affinity for the nucleocapsid protein. Structural studies and implications for genome recognition. *J. Mol. Biol.* **2001***, 314*, 961-970.

(3) Shubsda, M. F.; Paoletti, A. C.; Hudson, B. S.; Borer, P. N. Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry* **2002***, 41*, 5276-5282.

(4) Feng, Y. X.; Copeland, T. D.; Henderson, L. E.; Gorelick, R. J.; Bosche, W. J. et al. HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA *in vitro*. *Proc. Natl. Acad. Sci. USA* **1996***, 93*, 7577-7581.

(5) Muriaux, D.; de Rocquigny, H.; Roques, B. P.; Paoletti, J. NCp7 activates HIV-1 Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J. Biol. Chem.* **1996***, 271*, 33686-33692.

(6) De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N. et al. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **1998***, 279*, 384-388.

(7) Skripkin, E., Paillart, J.C., Marquet, R., Ehresmann, B., Ehresmann,C. Identification of the primary site of the HIV-1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. USA* **1994***, 91*, 4945-4949.

(8) Paillart, J. C.; Skripkin, E.; Ehresmann, B.; Ehresmann, C.; Marquet, R. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. USA* **1996***, 93*, 5572-5577.

(9) Laughrea, M.; Jette, L. 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*, 1589-1598.

(10) Paillart, J. C.; Westhof, E.; Ehresmann, C.; Ehresmann, B.; Marquet, R. Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J. Mol. Biol.* **1997***, 270*, 36-49.

(11) Laughrea, M.; Jette, L.; Mak, J.; Kleiman, L.; Liang, C. 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*, 3397-3406.

(12) Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., Dumas, P Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.* **2001***, 8*, 1064-1068.

(13) Mihailescu, M. R.; Marino, J. P. A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex. *Proc. Natl. Acad. Sci. USA* **2004***, 101*, 1189-1194.

(14) Takahashi, K. I.; Baba, S.; Chattopadhyay, P.; Koyanagi, Y.; Yamamoto, N. et al. Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA* **2000***, 6*, 96-102.

(15) Greatorex, J.; Gallego, J.; Varani, G.; Lever, A. Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal. *J. Mol. Biol.* **2002***, 322*, 543-557.

(16) Laughrea, M.; Shen, N.; Jette, L.; Wainberg, M. A. Variant effects of non-native kissing-loop hairpin palindromes on HIV replication and HIV RNA dimerization: role of stem-loop B in HIV replication and HIV RNA dimerization. *Biochemistry* **1999***, 38*, 226-234.

(17) Shen, N.; Jette, L.; Liang, C.; Wainberg, M. A.; Laughrea, M. Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging. *J. Virol.* **2000***, 74*, 5729-5735.

(18) Lawrence, D. C.; Stover, C. C.; Noznitsky, J.; Wu, Z. R.; Summers, M. F. Structure of the intact stem and bulge of HIV-1 Ψ-RNA stem-loop SL1. *J. Mol. Biol.* **2003***, 326*, 529-542.

(19) Yuan, Y. Q.; Kerwood, D. J.; Paoletti, A. C.; Shubsda, M. F.; Borer, P. N. Stem of SL1 RNA in HIV-1: structure and nucleocapsid protein binding for a 1 x 3 internal loop. *Biochemistry* **2003***, 42*, 5259-5269.

(20) Rist, M. J.; Marino, J. P. Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* **2002***, 41*, 14762-14770.

(21) Davanloo, P.; Rosenberg, A. H.; Dunn, J. J.; Studier, F. W. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **1984***, 81*, 2035-2039.

(22) Nikonowicz, E. P.; Sirr, A.; Legault, P.; Jucker, F. M.; Baer, L. M. et al. Preparation of 13C and 15N labelled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Res* **1992***, 20*, 4507-4513.

(23) Piotto, M.; Saudek, V.; Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **1992***, 2*, 661-665.

(24) Kay, L. E.; Keifer, P.; Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **1992***, 114*, 10663.

(25) Grzesiek, S.; Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **1992***, 99*, 201.

(26) Grzesiek, S.; Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **1992***, 114*, 6291.

(27) Wittekind, M.; Mueller, L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. Magn. Reson. Ser. B* **1993***, 101*, 201-205.

(28) Muhandiram, D. R.; Kay, L. E. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Ser. B* **1994***, 103*, 203-216.

(29) Bax, A.; Clore, G. M.; Driscoll, P. C.; Gronenborn, A. M.; Ikura, M. et al. Practical aspects of proton carbon carbon proton 3-dimensional correlation sepctroscopy of ^{13}C-labeled proteins. *J Magn. Reson.* **1990***, 87*, 620-627.

(30) Kay, L. E.; Ikura, M.; Bax, A. Proton-proton correlation via carbon-carbon couplings: a three dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with carbon-13. *J. Am. Chem. Soc.* **1990***, 112*, 888-889.

(31) Ikura, M.; Kay, L. E.; Bax, A. Improved three-dimensional ^{1}H-^{13}C-^{1}H correlation spectroscopy of a ^{13}C-labeled protein using constant-time evolution. *J. Biomol. NMR* **1991**, *1*, 299-304.

(32) Bax, A.; Clore, G. M.; Gronenborn, A. M. ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new 3-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.* **1990**, *88*, 425-431.

(33) Marion, D.; Kay, L. E.; Sparks, S. W.; Torchia, D. A.; Bax, A. 3-dimensional heteronuclear NMR of ^{15}N-labeled proteins. *J. Am. Chem. Soc.* **1989**, *111*, 1515-1517.

(34) Zuiderweg, E. R. P.; McIntosh, C. P.; Dalhquist, F. W.; Fesik, S. W. 3-dimensional ^{13}C-resolved protein NOE spectroscopy of uniformly ^{13}C-labeled proteins for the NMR assignment and structure determination of larger molecules. *J. Magn. Reson.* **1990**, *86*, 210-216.

(35) Lippens, G.; Dhalluin, C.; Wieruszeski, J. M. Use of a water flip-back pulse in the homonuclear NOESY experiment. *J. Biomol. NMR* **1995**, *5*, 327-331.

(36) Pascal, S. M.; Muhandiram, D. R.; Yamazaki, T.; Formankay, J. D.; Kay, L. E. Simultaneous acquisition of ^{15}N-edited and ^{13}C-edited NOE spectra of proteins dissolved in H_2O. *J. Magn. Reson. B* **1994**, *103*, 197-201.

(37) Dingley, A. J.; Grzesiek, S. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *J. Am. Chem. Soc.* **1998**, *120*, 8293-8297.

(38) Mueller, L.; Legault, P.; Pardi, A. Improved RNA structure determination by detection of NOE contacts to exchange-broadened amino groups. *J. Am. Chem. Soc.* **1995***, 117*, 11043-11048.

(39) Miller Jenkins, L. M.; Byrd, J. C.; Hara, T.; Srivastava, P.; Mazur, S. et al. Studies on the inactivation of the HIV-1 nucleocapsid protein NCp7 with 2-mercaptobenzamide thioesters. *J Med Chem* **2005***, 48*, 2487-2458.

(40) Farmer, B. T.; Constantine, K. L.; Goldfarb, V.; Friedrichs, M. S.; Wittekind, M. et al. Localizing the NADP+ binding site on the MurB enzyme by NMR. *Nat. Struct. Biol.* **1996***, 3*, 995-997.

(41) Wishart, D. S.; Sykes, B. D.; Richards, F. M. The chemical shift index: a fast and simple method of the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992***, 31*, 1647-1651.

(42) Wuthrich, K. *NMR of proteins and nucleic acids*; J. Wiley & Sons: New York, 1986.

(43) Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R. et al. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* **2000***, 301*, 491-511.

(44) Hargittai, M. R.; Mangla, A. T.; Gorelick, R. J.; Musier-Forsyth, K. HIV-1 nucleocapsid protein zinc finger structures induce $tRNA_{Lys,3}$ structural changes but are not critical for primer/template annealing. *J. Mol. Biol.* **2001***, 312*, 985-997.

(45) Hargittai, M. R.; Gorelick, R. J.; Rouzina, I.; Musier-Forsyth, K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.* **2004***, 337*, 951-968.

(46) Rein, A.; Henderson, L. E.; Levin, J. G. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci* **1998***, 23*, 297-301.

(47) Bernacchi, S.; Stoylov, S.; Piemont, E.; Ficheux, D.; Roques, B. P. et al. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.* **2002***, 317*, 385-399.

(48) Azoulay, J.; Clamme, J. P.; Darlix, J. L.; Roques, B. P.; Mely, Y. Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J. Mol. Biol.* **2003***, 326*, 691-700.

(49) Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J. P.; Ficheux, D. et al. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **2003***, 328*, 95-108.

(50) Beltz, H.; Piemont, E.; Schaub, E.; Ficheux, D.; Roques, B. et al. Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **2004***, 338*, 711-723.

(51) Urbaneja, M. A.; Wu, M.; Casas-Finet, J. R.; Karpel, R. L. HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing

properties, structural binding specificity, and annealing activity. *J. Mol. Biol.* **2002***, 318*, 749-764.

(52) Darlix, J. L.; Lapadat-Tapolsky, M.; de Rocquigny, H.; Roques, B. P. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol* **1995***, 254*, 523-537.

(53) Chan, B.; Weidemaier, K.; Yip, W. T.; Barbara, P. F.; Musier-Forsyth, K. Intra-tRNA distance measurements for nucleocapsid protein dependent tRNA unwinding during priming of HIV reverse transcription. *Proc. Natl. Acad. Sci. USA* **1999***, 96*, 459-464.

Supplementary Tables:

Supplementary Table 1. Chemical shift assignments (ppm) for free NCp7[a].

AA	HN	N	CA	CB	C'
M1	N/A	N/A	55.9	30.7	176.1
Q2	8.53	122.07	55.7	29.6	175.4
R3	8.68	123.76	56.2	31.0	176.6
G4	8.50	110.10	45.1	N/A	173.4
N5	8.33	118.62	52.9	38.7	175.0
F6	8.25	120.95	57.8	39.3	175.8
R7	8.27	122.16	56.3	30.6	175.8
N8	8.37	119.28	53.2	38.6	175.0
Q9	8.33	120.76	55.8	29.4	175.7
R10	8.36	122.60	56.0	30.7	175.8
K11	8.42	123.78	55.9	32.8	176.0
I12	8.32	124.75	60.5	38.7	175.8
V13	8.57	127.50	61.7	32.7	174.1
K14	8.23	125.48	54.8	35.4	174.5
C15	8.25	127.25	59.2	30.0	178.4
F16	8.65	128.45	59.0	38.7	175.7
N17	9.55	121.87	56.8	40.1	175.3
C18	8.89	117.18	58.5	32.6	176.5
G19	8.13	113.89	46.2	N/A	173.6
K20	8.10	120.81	56.6	33.6	176.3
E21	8.53	118.66	55.8	31.0	178.5
G22	8.76	107.58	45.6	N/A	173.6
H23	7.18	114.05	55.1	30.1	171.1
T24	8.26	106.42	59.2	71.7	176.3
A25	8.87	123.04	55.7	18.6	180.0
R26	8.17	116.86	58.3	29.8	176.6
N27	7.83	115.02	52.0	39.8	173.5
C28	7.50	123.87	62.7	31.8	176.9
R29	8.56	123.40	55.6	29.5	175.3
A30	8.47	126.54	52.3	19.5	N/A
P31	N/A	N/A	N/A	N/A	N/A
R32	N/A	N/A	N/A	N/A	N/A
K33	N/A	N/A	N/A	N/A	N/A
K34	N/A	N/A	56.5	33.3	176.4
G35	8.24	109.11	44.2	N/A	172.6

AA	HN	N	CA	CB	C'
C36	8.18	124.01	59.8	30.4	178.8
W37	8.51	129.55	58.4	29.6	176.3
K38	9.41	122.91	58.9	33.8	177.1
C39	8.50	117.04	58.5	32.3	176.3
G40	7.87	113.45	46.2	N/A	173.9
K41	8.41	122.24	56.5	33.1	176.3
E42	8.49	123.13	56.6	30.7	178.0
G43	8.65	106.94	45.7	N/A	173.5
H44	7.12	112.69	55.3	30.4	172.2
Q45	8.94	116.78	53.9	31.4	177.7
M46	8.93	122.01	60.1	32.6	178.4
K47	8.61	118.57	58.5	31.3	176.0
D48	7.85	117.42	42.5	53.5	174.9
C49	7.61	123.86	62.6	30.8	177.2
T50	8.27	119.95	61.5	69.3	174.9
E51	8.67	125.70	56.8	29.9	N/A
R52	N/A	N/A	55.9	33.8	175.5
Q53	8.42	117.93	55.3	30.4	176.6
A54	8.77	128.25	52.3	19.4	176.4
N55	8.07	123.58	54.6	40.2	N/A

[a]: The NCp7 assignments were made at 25 ºC at pH 6.0. The error on the proton chemical shifts is ± 0.02 ppm, the carbon chemical shifts is ± 0.2 ppm, and the nitrogen chemical shift is ± 0.06 ppm.

Supplementary Table 2. Chemical shift assignments (ppm) for NCp7 complexed with SL1i[a].

AA	HN	N	CA	HA	CB	HB	CG	HG	CD/ND	HD	CE/ NE	HE	C'
M1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	17.2	2.14	N/A
Q2	N/A	N/A	57.0	4.32	29.9	2.33	36.1	2.53	N/A	N/A	N/A	N/A	127.7
R3	8.25	122.13	55.2	4.42	31.3	1.90	27.4	1.66	43.6	3.24	N/A	N/A	176.8
G4	8.42	109.90	45.5	4.08, 3.96	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	173.8
N5	8.25	118.78	53.2	4.72	39.0	2.80, 2.71	N/A	N/A	N/A	N/A	N/A	N/A	175.2
F6	8.15	120.61	58.0	4.62	39.6	3.18, 3.05	N/A	N/A	N/A	7.24	N/A	7.25	176.0
R7	8.08	121.18	56.6	4.27	30.9	1.84, 1.76	27.3	1.57	43.5	3.20	N/A	N/A	176.2
N8	8.30	119.13	53.6	4.68	38.9	2.99, 2.80	177.2	N/A	112.2	7.61, 6.92	N/A	N/A	175.3
Q9	8.22	120.34	56.0	4.42	29.8	2.14, 2.05	34.0	2.43	180.6	N/A	111.9	7.55, 6.85	175.6
R10	8.56	123.42	56.3	4.33	31.0	1.92, 1.82	27.3	1.63	43.6	3.20	N/A	N/A	176.1
K11	8.20	122.70	56.3	4.33	33.3	1.72	25.0	1.41	29.3	1.82	43.3	3.01	176.1
I12	8.13	123.63	60.6	4.35	39.0	1.81	29.7	1.85, 1.18	12.9	0.83	17.7	0.83	175.8
V13	8.28	125.91	61.6	4.69	33.0	1.71	24.7	1.44, 1.68	N/A	N/A	N/A	N/A	178.5
K14	8.99	119.90	58.7	4.36	31.7	2.35, 1.82	28.0	1.79	N/A	N/A	N/A	N/A	177.1
C15	8.42	122.90	56.8	4.42	31.0	3.35, 3.01	N/A	N/A	N/A	N/A	N/A	N/A	176.3
F16	7.81	117.28	59.1	4.74	39.7	3.08	N/A	N/A	N/A	7.36	N/A	7.36	175.8
N17	8.55	119.54	N/A	4.32	37.8	3.27	N/A	N/A	N/A	N/A	N/A	N/A	175.6
C18	8.86	117.50	58.8	4.51	32.4	2.63, 2.58	N/A	N/A	N/A	N/A	N/A	N/A	178.7
G19	8.69	107.82	45.6	3.93	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	175.8
K20	8.29	118.34	56.6	4.33	32.6	1.81	24.9	1.46	29.3	1.63	42.3	3.03	177.0
E21	8.47	121.13	56.9	4.27	29.7	2.15, 2.05	36.9	2.37	N/A	N/A	N/A	N/A	178.2

AA	HN	N	CA	HA	CB	HB	CG	HG	CD/ND	HD	CE/ NE	HE	C'
G22	8.73	108.15	45.9	4.45, 3.71	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	173.7
H23	7.12	114.02	55.3	4.77	30.2	3.28, 3.22	N/A	N/A	N/A	13.72, 7.02	N/A	7.00	171.0
T24	8.03	105.45	62.6	4.36	70.2	4.27	21.9	1.23	N/A	N/A	N/A	N/A	174.6
A25	N/A	N/A	50.1	4.04	20.7	0.83, 0.56	N/A	N/A	N/A	N/A	N/A	N/A	174.2
R26	8.06	124.76	58.8	4.20	30.0	N/A	30.2	1.72	42.6	3.22	N/A	3.52	176.8
N27	7.84	115.52	52.4	4.73	40.1	2.79, 2.70	177.2	N/A	112.2	6.87	N/A	N/A	173.8
C28	7.54	124.51	62.5	4.34	31.1	3.04	N/A	N/A	N/A	N/A	N/A	N/A	174.0
R29	7.93	124.71	55.6	4.42	30.5	1.93	N/A	1.83, 1.70	43.3	3.13	N/A	3.42	176.1
A30	8.78	128.11	51.2	4.64	18.3	1.44	N/A	N/A	N/A	N/A	N/A	N/A	N/A
P31	N/A	N/A	62.5	4.50	30.9	2.43	N/A	2.17, 2.06	50.3	3.82, 3.59	N/A	N/A	176.8
R32	8.65	128.29	55.9	4.46	30.7	2.43, 2.18	N/A	2.06	N/A	3.27	N/A	N/A	175.6
K33	N/A	N/A	55.3	4.20	33.1	2.24	31.0	N/A	N/A	N/A	N/A	2.65	172.4
K34	8.82	124.02	56.7	4.40	33.4	1.94, 1.85	25.3	1.74	29.4	1.52	42.1	3.04	177.7
G35	8.34	108.29	43.9	3.87, 3.68	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	173.1
C36	8.43	124.63	48.2	4.64	30.6	3.32, 2.94	N/A	N/A	N/A	N/A	N/A	N/A	180.1
W37	9.37	131.83	58.3	4.43	29.3	1.45	N/A	N/A	N/A	7.29	128.9	10.05, 7.29	176.5
K38	9.94	123.98	59.1	5.05	34.3	3.31, 2.62	N/A	N/A	29.8	1.41	42.6	3.10	177.3
C39	8.78	118.25	59.2	4.33	33.0	2.49	N/A	N/A	N/A	N/A	N/A	N/A	177.0
G40	8.20	113.93	46.4	4.27, 4.04	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	174.4
K41	8.48	122.59	57.2	4.38	33.8	1.92, 1.82	N/A	1.48	N/A	1.76	42.7	3.11	176.7
E42	8.50	119.25	56.9	4.31	30.0	2.16, 2.00	36.0	2.34	N/A	N/A	N/A	N/A	177.1
G43	8.30	109.43	45.9	4.42, 3.70	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	173.8
H44	7.10	112.59	55.8	4.71	30.5	3.25, 3.20	N/A	N/A	N/A	7.08	N/A	7.08	172.2

AA	HN	N	CA	HA	CB	HB	CG	HG	CD/ND	HD	CE/ NE	HE	C'
Q45	8.77	117.44	54.3	4.42	31.6	1.45	N/A	N/A	N/A	N/A	N/A	N/A	177.4
M46	9.45	121.84	56.4	4.67	32.2	1.95	N/A	2.33	N/A	N/A	17.9	2.01	176.4
K47	8.99	119.90	58.8	4.25	31.6	1.90	N/A	1.34	N/A	1.76	N/A	3.08	176.3
D48	7.98	117.84	53.8	4.98	42.9	3.01, 2.57	N/A	N/A	N/A	N/A	N/A	N/A	175.2
C49	7.64	123.89	63.0	3.84	30.9	3.31, 3.02	N/A	N/A	N/A	N/A	N/A	N/A	177.4
T50	8.08	119.76	61.9	4.67	69.2	4.44	21.6	1.19	N/A	N/A	N/A	N/A	174.4
E51	8.28	125.32	56.2	4.35	29.7	2.16, 2.02	34.1	2.39	N/A	N/A	N/A	N/A	175.6
R52	8.24	121.89	56.2	4.40	31.2	1.89, 1.80	27.3	1.50, 1.17	43.5	3.24	N/A	N/A	176.2
Q53	8.39	120.36	56.0	4.36	29.7	2.16, 2.04	34.0	2.77, 2.41	178.3	N/A	112.2	7.48, 6.84	175.5
A54	8.32	126.12	52.6	4.40	19.8	1.44	N/A	N/A	N/A	N/A	N/A	N/A	176.6
N55	7.98	123.59	54.9	4.51	40.7	2.82, 2.73	178.7	N/A	112.7	7.48, 6.80	N/A	N/A	N/A

[a]: The chemical shift assignments were made at 35 ºC in 200 mM NaCl, 40 mM sodium phosphate buffer pH 7.0. The error on the proton chemical shifts is ± 0.02 ppm, the carbon chemical shifts is ± 0.2 ppm, and the nitrogen chemical shift is ± 0.06 ppm.

Supplementary Table 4.3. Chemical shifts assignments (ppm) for free SL1i.[a]

	NH	NH2	H8/H6	H1'	H5/H2
G1/C30	N/A	N/A	N/A	N/A	N/A
G2	13.39	N/A	N/A	N/A	N/A
C3	N/A	8.65, 6.91	N/A	5.55	5.19
U4	14.15	N/A	N/A	5.53?	N/A
C5	N/A	8.29, 6.86	N/A	5.92	N/A
G6	12.59	N/A	N/A	5.95	N/A
G7	N/A	N/A	N/A	N/A	N/A
C8	N/A	N/A	N/A	N/A	N/A
U9	14.07	N/A	N/A	N/A	N/A
U10	13.16	N/A	N/A	5.40	N/A
G11	12.40	N/A	N/A	5.85	N/A
C12	N/A	8.16, 6.73	N/A	5.85	5.12
G13	10.52	N/A	N/A	5.72	N/A
A14	N/A	N/A	8.36	N/A	N/A
A15	N/A	N/A	N/A	N/A	N/A
A16	N/A	N/A	8.16	N/A	N/A
G17	12.83	N/A	N/A	3.93	N/A
C18	N/A	8.45, 6.73	N/A	5.46	5.13
A19	N/A	N/A	N/A	5.48	6.68
A20	N/A	N/A	N/A	5.76	7.31
G21	N/A	N/A	N/A	5.61	N/A
A22	N/A	N/A	N/A	N/A	N/A
G23	N/A	N/A	N/A	N/A	N/A
G24	N/A	N/A	N/A	N/A	N/A
C25	N/A	8.28, 6.99	N/A	N/A	N/A
G26	12.05	N/A	N/A	N/A	N/A
A27	N/A	N/A	7.90	5.67	7.40
G28	13.50	N/A	N/A	5.62	N/A
C29	N/A	8.52, 6.91	N/A	5.53?	5.19

[a]: The assignements were made at 5 ºC in 200 mM NaCl, 40 mM sodium phosphate buffer pH 7.0. The error on the proton chemical shifts is ± 0.02 ppm.

Supplementary Table 4.4. Chemical shift assignments (ppm) for SL1i complexed with NCp7[a].

	NH	N1/N3	N4/N6	NH2	H5/H2
G2	13.39	148.72	N/A	N/A	N/A
C3	N/A	N/A	99.07	8.61, 6.82	N/A
U4	14.15	167.73	N/A	N/A	N/A
C5	N/A	N/A	97.74	8.28, 6.90	N/A
G6	N/A	N/A	N/A	N/A	N/A
G7	N/A	N/A	N/A	N/A	N/A
C8	N/A	N/A	N/A	N/A	N/A
U9	N/A	N/A	N/A	N/A	N/A
U10	13.11	162.42	N/A	N/A	N/A
G11	12.40	147.51	N/A	N/A	N/A
C12	N/A	N/A	99.09	8.13, 6.57	N/A
G13	10.52	N/A	N/A	N/A	N/A
A14	N/A	N/A	N/A	N/A	N/A
A15	N/A	N/A	N/A	N/A	N/A
A16	N/A	N/A	N/A	N/A	N/A
G17	12.83	148.13	N/A	N/A	N/A
C18	N/A	N/A	97.77	8.40, 6.67	N/A
A19	N/A	N/A	N/A	N/A	N/A
A20	N/A	N/A	N/A	N/A	N/A
G21	N/A	N/A	N/A	N/A	N/A
A22	N/A	N/A	N/A	N/A	N/A
G23	N/A	N/A	N/A	N/A	N/A
G24	N/A	N/A	N/A	N/A	N/A
C25	N/A	N/A	N/A	N/A	N/A
G26	11.95	146.48	N/A	N/A	N/A
A27	N/A	N/A	79.08	N/A	7.44
G28	13.50	N/A	N/A	N/A	N/A
C29	N/A	N/A	98.92	8.51, 6.73	N/A

[a]: The assignments were made at 35 ºC in 200 mM NaCl, 40 mM sodium phosphate buffer pH 7.0. The error on the proton chemical shifts is ± 0.02 ppm and that of the nitrogen chemical shifts is ± 0.05 ppm.

CHAPTER 5

CONCLUSIONS

This study has examined the interactions of the HIV-1 nucleocapsid protein (NCp7) with 2-mercaptobenzamide thioesters and with HIV-1 ΨSL1. NCp7 is a nucleic acid binding protein with crucial roles throughout the viral lifecycle. The most important function of NCp7 is to facilitate the dimerization and packaging of the viral genomic RNA into a new virion. The dimerization reaction requires that NCp7 binds ΨSL1 to enhance the kissing loop to extended duplex conversion. The many essential roles of NCp7 have made it a new target for antiviral therapy. The first developed antiviral compounds targeted to NCp7 ejected the coordinated zinc of NCp7. The mechanism of new such compounds, such as the 2-mercaptobenzamide thioesters, had not been thoroughly examined. Nor had the specificity of these compounds been tested.

Mechanism of Action and Specificity of 2-Mercaptobenzamide Thioester Compounds

The first aim of this research was to determine the mechanism of action of the 2-mercaptobenzamide thioester compounds with NCp7. It was found that these compounds are able to react with NCp7 specifically in the carboxyl-terminal domain (ZD2), ejecting the coordinated zinc. Cys_{39} was found to be the first site of reaction with these compounds, though the other zinc-coordinating cysteines in ZD2 are later sites of modification. As the interaction with the thioester compounds resulted in the loss of zinc coordination, NCp7 was no longer able to bind target RNA sequences after interaction with the compounds. Interestingly, however, the

thioester compounds were unable to interact with NCp7 when it was already bound to RNA. The RNA was thus able to protect the NCp7 from the thioester compounds.

Following reaction with NCp7, a free thiol group is released from the 2-mercaptobenzamide thioesters. It is possible that *in vivo* this group could undergo a subsequent reaction with acyl coenzyme A or glutathione. Either of these reactions would result in formation of a new thioester compound. Indeed, an acylated-thiol compound was shown to eject cobalt from NCp7 at a rate similar to that of the 2-mercaptobenzamide thioester compounds. It is also possible that reaction with acyl coenzyme A or glutathione could be part of the metabolism of the 2-mercaptobenzamide thioester compounds. Similar reactions have been proposed in the metabolism of non-steroidal anti-inflammatory drugs, such as ibuprofen[1]. Further experiments performed with a nicotinamide glutathione thioester showed that this compound was also able to eject cobalt from NCp7 (LMMJ, PL, JGO unpublished data). Thus, it is possible that reaction with acyl coenzyme A or glutathione, in the process or metabolism or post-reaction, could lead to new, reactive thioester compounds.

The second aim of this research was to examine the specificity of the 2-mercaptobenzamide thioester compounds by studying their effect on other structural zinc-binding proteins. The interaction of the thioester compounds was analyzed with six different proteins that represented five types of zinc-binding motifs. It was found that proteins containing the "classical" zinc finger motif (Cys_2His_2) were resistant to reaction with the thioester compounds. The interleaved zinc-binding domains ($Cys_2HisCys$ and $HisCys_3$) tested also did not interact with the thioester compounds. The Cys_4 zinc-binding protein (GATA-1) was very reactive with the thioester compounds; however this protein was unreactive when it was bound to its target DNA sequence. A protein with a zinc-coordinating motif and fold similar to NCp7

($Cys_2HisCys$, MMTV NCp10) showed similar reactivity towards the thioester compounds as did NCp7 – only one of the two zinc-binding domains (ZBDs) was reactive whereas the other remained relatively unreactive. A second $Cys_2HisCys$ protein with a very different structure than NCp7 (FOG-1) was also able to interact with the thioester compounds. Thus, the $Cys_2HisCys$ proteins with very different sequences and structures were all able to interact with the 2-mercaptobenzamide thioester compounds.

The inability of the 2-mercaptobenzamide thioester compounds to interact with the amino-terminal zinc-binding domain (ZD1) of NCp7 suggests that these compounds do show some specificity. There is no difference in the fold or zinc-coordinating motif of the two ZBDs, so the interaction with the thioester compounds must be related to the sequence of the ZBDs themselves. The sequence of NCp7 between the second cysteine and the histidine is identical in both ZBDs, suggesting that the determinant residues are found between the first and second cysteines (residues 16-17 and 37-38 in ZD1 and ZD2, respectively) and/or between the histidine and the final cysteine (residues 24-27 and 45-48 in ZD1 and ZD2, respectively). This is further highlighted by the MMTV NCp10 ZBDs in which only ZD1 interacted with the thioester compounds. It is possible that differences at positions around the zinc-coordinating residues alter the reactivity of the cysteine residues such that they are more likely to interact with the thioester compounds. It is also possible that the residues around the zinc affect the strength of zinc coordination such that zinc is chelated more tightly in ZD1 than in ZD2. The latter explanation is more likely, as both Cys_4 ZBDs in the protein we studied were equally reactive with the thioester compounds, despite having different sequences. These results support theoretical studies showing that the sequences around ZBDs affect the charge and reactivity of the domains[2-4].

The high degree of reactivity observed with the cellular zinc-binding proteins *in vitro* suggests that the 2-mercaptobenzamide thioester compounds would not be successful as an antiviral therapy due to problems with toxicity. NCp7 is highly up-regulated in HIV-1 infected cells, and there are ~1500 copies of NCp7 in each virion. It is possible that in many cells, the amount of NCp7 in the cytosol would make it a preferred target of the thioester compounds. Even though protection of NCp7 was observed when bound to RNA, with the large amount of protein translated, is very likely that some NCp7 would not be bound to RNA at any given time. Also, the degree of protection afforded by the RNA may vary with the specific RNA sequence to which NCp7 binds. However, GATA-1, the Cys_4 ZBD we studied, is highly expressed in red blood cells and could be affected by the thioester compounds. Though some GATA-1 would be protected when bound to DNA, the sheer quantity of GATA-1 in the red blood cells makes it likely that some protein would not be bound to DNA. Thus, it is very likely that the thioester compounds would have toxicity problems if used to treat HIV-1.

It is possible that the specificity of the thioester compounds could be improved by linking the 2-mercaptobenzamide thiol to a compound that was able to specifically interact with NCp7. Preliminary studies have been performed to identify compounds that specifically bind NCp7 by mimicking RNA[5]. The primary research that should be done next with the 2-mercaptobenzamide thioesters is to improve their specificity for NCp7. If this could be done, it is possible that the resultant thioester compound could be used to treat HIV-1. In addition, the efficacy of the thioester compounds could be improved by using them in conjunction with protease inhibitors. Studies have shown that the full gag polyprotein does not bind RNA as well as the cleaved NCp7[6]. The NCp7 domain within the gag polyprotein might be less protected and, therein, more reactive towards the 2-mercaptobenzamide thioesters.

Interaction of NCp7 with ΨSL1

The third aim of this research was to investigate the interaction between NCp7 and the internal loop of ΨSL1. Gel mobility shift assays demonstrated that NCp7 binds specifically to the internal loop. This interaction was blocked by incubation of NCp7 with a 2-mercaptobenzamide thioester compound, as observed for the NCp7/ΨSL3 complex. However, binding of the internal loop protected NCp7 from interaction with this compound. NMR structural studies showed that NCp7 underwent a significant conformational change upon binding the internal loop, possibly forming a structure similar to that found in the NCp7/ΨSL3 complex. In the complex, NCp7 destabilized two of the base pairs in the stem, one on each side of the internal loop. The base pairs in the remainder of the stem did not appear to undergo any conformational change.

The NMR spectroscopy results have suggested a function for the internal loop of ΨSL1. The internal loop could act as the nucleation point for the melting of the stem in the conversion of the kissing loop to the extended duplex. Binding of NCp7 to this loop could facilitate the destabilization and melting of the stem. NCp7 has previously been observed to enhance the formation of the extended duplex[7,8]. Binding of the internal loop could be an important part of the conversion and a function of NCp7 that has, thus far, been unreported.

NCp7 has been suggested to destabilize other stable RNA structures. It is thought that during reverse transcription, NCp7 facilitates the binding of the tRNA primer to the primer binding site by destabilizing the double-stranded regions in both the primer binding site and the tRNA[9-11]. Also, it has been suggested that NCp7 blocks self-priming reactions by destabilizing the TAR stem-loop RNA[12-16]. Thus, destabilization of double-stranded structures could be a general function of NCp7 in the HIV-1 lifecycle, part of its the nucleic-acid binding roles.

Internal loops could function as an important site of initiation for this destabilization, such as observed for ΨSL1. Indeed, the TAR RNA has an internal loop to which NCp7 could bind. The ability of NCp7 to bind internal loops has not previously been studied, and much more research is required to fully understand this binding mechanism.

References

(1) Li, C. Z.; Olurinde, M. O.; Hodges, L. M.; Grillo, M. P.; Benet, L. Z. Covalent binding of 2-phenylpropionyl-S-acyl-CoA thioester to tissue proteins in vitro. *Drug Metab. Dispos.* **2003***, 31*, 727-730.

(2) Konrat, R.; Weiskirchen, R.; Bister, K.; Kräutler, B. Bispheric coordinative structuring in a zinc finger protein: NMR analysis of a point mutant of the carboxy-terminal LIM domain of quail cysteine- and glycine-rich protein CRP2. *J. Am. Chem. Soc.* **1998***, 120*, 7127-7130.

(3) Maynard, A. T., Covell, D.G. Reactivity of zinc finger cores: analysis of protein packing and electrostatic screening. *J. Am. Chem. Soc.* **2001***, 123*, 1047-1058.

(4) Dudev, T.; Lin, Y.-L.; Dudev, M.; Lim, C. First-second shell interactions in metal binding sites in proteins: a PDB survey and DFT/CDM calculations. *J. Am. Chem. Soc.* **2003***, 125*, 3168-3180.

(5) Stephen, A. G.; Rein, A.; Fisher, R. J.; Shoemaker, R. H. The nucleocapsid protein as a target for novel anti-HIV drugs. *Curr. Drug Disc.* **2003**, 33-36.

(6) Berkowitz, R. D.; Luban, J.; Goff, S. P. Specific binding of human immunodeficiency virus type 1 *gag* polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J. Virol.* **1993***, 67*, 7190-7200.

(7) Feng, Y. X.; Copeland, T. D.; Henderson, L. E.; Gorelick, R. J.; Bosche, W. J. et al. HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA *in vitro*. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7577-7581.

(8) Muriaux, D.; de Rocquigny, H.; Roques, B. P.; Paoletti, J. NCp7 activates HIV-1 Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J. Biol. Chem.* **1996**, *271*, 33686-33692.

(9) Chan, B.; Weidemaier, K.; Yip, W. T.; Barbara, P. F.; Musier-Forsyth, K. Intra-tRNA distance measurements for nucleocapsid protein dependent tRNA unwinding during priming of HIV reverse transcription. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 459-464.

(10) Hargittai, M. R.; Mangla, A. T.; Gorelick, R. J.; Musier-Forsyth, K. HIV-1 nucleocapsid protein zinc finger structures induce $tRNA_{Lys,3}$ structural changes but are not critical for primer/template annealing. *J. Mol. Biol.* **2001**, *312*, 985-997.

(11) Hargittai, M. R.; Gorelick, R. J.; Rouzina, I.; Musier-Forsyth, K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.* **2004**, *337*, 951-968.

(12) Rein, A.; Henderson, L. E.; Levin, J. G. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci* **1998**, *23*, 297-301.

(13) Bernacchi, S.; Stoylov, S.; Piemont, E.; Ficheux, D.; Roques, B. P. et al. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.* **2002**, *317*, 385-399.

(14) Azoulay, J.; Clamme, J. P.; Darlix, J. L.; Roques, B. P.; Mely, Y. Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J. Mol. Biol.* **2003***, 326*, 691-700.

(15) Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J. P.; Ficheux, D. et al. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **2003***, 328*, 95-108.

(16) Beltz, H.; Piemont, E.; Schaub, E.; Ficheux, D.; Roques, B. et al. Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **2004***, 338*, 711-723.