

12-13-2018

# Understanding Integrase-DNA Interactions in Retroviruses Through 3'-processing

Zachary E. Ferris  
zferris2@student.gsu.edu

Markus W. Germann  
*Georgia State University*

Follow this and additional works at: [https://scholarworks.gsu.edu/chemistry\\_theses](https://scholarworks.gsu.edu/chemistry_theses)

---

## Recommended Citation

Ferris, Zachary E. and Germann, Markus W., "Understanding Integrase-DNA Interactions in Retroviruses Through 3'-processing." Thesis, Georgia State University, 2018.  
[https://scholarworks.gsu.edu/chemistry\\_theses/125](https://scholarworks.gsu.edu/chemistry_theses/125)

This Thesis is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

UNDERSTANDING INTEGRASE-DNA INTERACTIONS IN RETROVIRUSES  
THROUGH 3'-PROCESSING

by

ZACHARY E. FERRIS

Under the Direction of Markus W. Germann, PhD

ABSTRACT

Retroviral integrase is one of the key enzymes needed to integrate viral DNA into a host cell's genome for many retroviruses including HIV. Integrase's role is three-fold. It prepares the ends of the DNA so that they can successfully bind to the target genomic DNA via 3'-processing, it creates a complex with the viral DNA that is capable of transporting it into the nucleus, and it facilitates the insertion of the viral DNA into the host genome. The goal of this research is to help determine what sequence and structural characteristics of the viral DNA terminus are responsible for successful integrase binding and 3'-processing. Through the use of polyacrylamide gel electrophoresis (PAGE) and  $^{32}\text{P}$  end labeling, different substrates are introduced to integrase and the effectiveness of the enzyme in binding to the DNA and carrying out 3'-processing is observed. The importance of terminal structural characteristics as well as individual nucleotides are then determined through a combination of PAGE results, modeling, and NMR-based structural comparisons.

INDEX WORDS: Integrase, HIV, Retrovirus, 3'-processing, DNA Substrates, Inosine

UNDERSTANDING INTEGRASE-DNA INTERACTIONS IN RETROVIRUSES  
THROUGH 3'-PROCESSING

by

ZACHARY E. FERRIS

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2018

Copyright by  
Zachary Evanoff Ferris  
2018

UNDERSTANDING INTEGRASE-DNA INTERACTIONS IN RETROVIRUSES  
THROUGH 3'-PROCESSING

by

ZACHARY E. FERRIS

Committee Chair: Markus Germann

Committee: Hamed Laroui

W. David Wilson

Gregory Poon

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

July 2018

## TABLE OF CONTENTS

<b>LIST OF FIGURES .....</b>	<b>IX</b>
<b>1 INTRODUCTION.....</b>	<b>1</b>
<i>1.1 Retroviruses.....</i>	<i>1</i>
<i>1.2 Human Immuno Deficiency Virus.....</i>	<i>1</i>
<i>1.3 HIV Integrase and 3'-processing .....</i>	<i>2</i>
<i>1.4 Purpose of the Study .....</i>	<i>4</i>
<b>2 MATERIALS AND METHODS .....</b>	<b>6</b>
<b>2.1 Synthesis and Purification of DNA Substrates.....</b>	<b>6</b>
<b>2.2 DNA Substrate Labeling .....</b>	<b>7</b>
<b>2.3 Integrase Expression and Purification.....</b>	<b>7</b>
<b>2.4 3'-Processing.....</b>	<b>8</b>
<b>2.5 Modeling .....</b>	<b>8</b>
<b>3 RESULTS AND DISCUSSION .....</b>	<b>9</b>
<b>3.1 Homology Model.....</b>	<b>9</b>
<b>3.2 List of Substrates.....</b>	<b>12</b>
<b>3.3 Base References.....</b>	<b>14</b>
<b>3.4 Gel Results .....</b>	<b>15</b>
<b>3.5 Discussion.....</b>	<b>21</b>
<b>4 CONCLUSIONS .....</b>	<b>26</b>

**REFERENCES..... 29**

**APPENDIX..... 56**



**LIST OF TABLES**

Table 1.1 Buffers for the Purification of HIV Integrase.....	6
Table 3.1.1 Substrates with Terminal Structural Changes.....	12
Table 3.1.2 Substrates with Base Substitutions .....	13

## LIST OF FIGURES

<b>Figure 1: Wild Type Viral DNA Sequence</b> .....	3
<b>Figure 2: HIV/PFV Homology Model</b> .....	9
<b>Figure 3: PFV Integrase (Paired Down)</b> .....	10
<b>Figure 4: PFV Integrase (Full)</b> .....	11
<b>Figure 5: Altered Bases Used</b> .....	14
<b>Figure 6: Results Key</b> .....	15
<b>Figure 7: Terminal Structure Mutations - T-overhangs</b> .....	15
<b>Figure 8: Terminal Structure Mutations - T-tails</b> .....	16
<b>Figure 9: Terminal Structure Mutations - Closed Ends</b> .....	16
<b>Figure 10: Base Mutations and Substitutions - Side Specificity</b> .....	17
<b>Figure 11: Base Mutations and Substitutions - Importance of the G</b> .....	17
<b>Figure 12: Base Mutations and Substitutions - Importance of the A</b> .....	18
<b>Figure 13: Base Mutations and Substitutions - The Importance of N7 on the A</b> .....	18
<b>Figure 14: Non-complementary Bases</b> .....	19
<b>Figure 15: Terminal G Substrates</b> .....	19
<b>Figure 16: GA-5' and GT-5' Film Results</b> .....	20
<b>Figure 17: Summarized Gel Results</b> .....	21
<b>Figure 18: PFV with Pre-cleaved DNA</b> .....	22
<b>Figure 19: Non-complementary pairs</b> .....	25
<b>Figure 20: Hairpin Sequences</b> .....	32
<b>Figure 21: Solving the Structures</b> .....	34
<b>Figure 22: Final 10 Structures for the I-HP</b> .....	39

<b>Figure 23: Final 10 Structures for the G-HP</b> .....	40
<b>Figure 24: Trajectory RMSD Values</b> .....	41
<b>Figure 25: RMSD Values Shown Graphically</b> .....	42
<b>Figure 26: Backbone angle and sugar puckering definitions</b> .....	43
<b>Figure 27: Helical Analysis: Difference in Backbone Parameters</b> .....	44
<b>Figure 28: Helical Analysis: Difference in Intra-BP Parameters</b> .....	45
<b>Figure 29: Aligned Full Structure</b> .....	46
<b>Figure 30: Side by Side Comparison of I/G Local Effects</b> .....	47
<b>Figure 31: Aligned I/G Nucleotides</b> .....	48
<b>Figure 32: <math>T_m</math> Values and Enthalpy</b> .....	49
<b>Figure 33: Imino Proton Spectra</b> .....	50
<b>Figure 34: Phosphorous Spectra</b> .....	50
<b>Figure 35: Base Pair Lifetimes</b> .....	51

## 1 INTRODUCTION

### 1.1 Retroviruses

Retroviruses are single stranded positive sense RNA viruses (+ssRNA viruses). This means that their genetic information is stored in the form of RNA instead of DNA. Retroviruses insert their genetic information into a host cell's genome which then causes the cell to produce more of the virus. In the case of HIV, virions contact the outside of the cell membrane via CD4 receptors (typically T4 helper cells) and inject their genetic material into the cytoplasm. In the case of a retrovirus this will consist of two copies of single stranded RNA as well as a host of proteins including reverse transcriptase and integrase<sup>(1,13)</sup>. Reverse transcriptase uses host nucleotides to produce a single strand of viral cDNA from each RNA strand. This is the reverse order of a typical DNA virus thus the name retrovirus (retro is Latin for 'backwards'). The single stranded cDNA is then reverse transcribed into double stranded cDNA (dsDNA).

Once the dsDNA is created, it works as a substrate for the enzyme integrase. Integrase forms a pre-integration complex (PIC), which is capable of transporting the viral DNA through a nuclear pore where it makes contact with the host DNA<sup>(13)</sup>. The integrase then creates a nick in one strand of the host DNA that allows the viral DNA to insert itself. Host repair enzymes then fully integrate the viral DNA into the genome. This integration of the viral DNA into the genome is what allows the virus to potentially establish lifelong infection. To complete the lifecycle, RNA polymerase will create viral mRNA from the inserted sequence. These mRNA encode for various viral proteins and will result in the production and maturation of new virions.

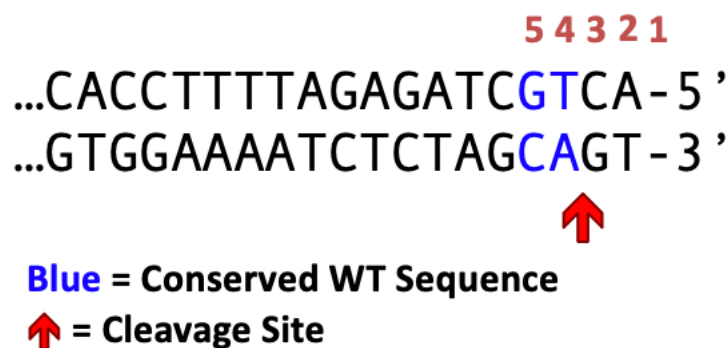
### 1.2 Human Immuno Deficiency Virus

The Human Immuno Deficiency Virus (HIV) is a retrovirus that infects human cells and leads to Acquired Immunodeficiency Syndrome (AIDS). AIDS is a condition caused by the

weakening and eventual failure of the immune system due to critically low levels of CD4+ T cells, allowing the proliferation of various infections and cancers. This drop in CD4+ T cells is caused by a combination of programmed cell death, direct viral death, and death by CD8+ cytotoxic lymphocytes, all of which are caused by the progressive spread of HIV<sup>(14)</sup>. HIV is classified as a lentivirus, which is a subgroup of retroviruses characterized by long incubation periods. HIV infects immune system cells such as helper T cells (CD4+ T cells), dendritic cells, and macrophages. There are two known species of HIV called HIV-1 and HIV-2. HIV-1 is the more prevalent form and has higher virulence and infectivity.

### **1.3 HIV Integrase and 3'-processing**

Integrase contains 288 amino acids and is divided into 3 domains: a Zn<sup>2+</sup>-stabilized three-helix bundled domain called the N-terminal, a central catalytic core domain that coordinates two divalent magnesium or manganese ions essential for DNA substrate processing, and the C-terminal domain which is responsible for both viral DNA binding and non-specific host DNA binding<sup>(1,2)</sup>. Integrase forms a tetramer (dimer-of-dimers), which binds viral DNA upon recognition of characteristic blunt-ended long terminal repeat (LTR) sequences<sup>(3)</sup>. These sequences are imperfect mirror images of each other (inverted repeats) and contain a highly conserved CA sequence preceding the terminal 2-3 nucleotides<sup>(4,5)</sup>. This is the beginning of an essential step in the HIV life cycle called 3'-processing.



**Figure 1: Wild Type Viral DNA Sequence**

*This figure shows the terminus of a representative wild type sequence. The CA and GT shown in blue are consistently maintained (conserved) between different variations of the viral DNA sequence. During 3'-processing the nucleotides after the conserved A are removed (red arrow).*

3'-processing is the process by which integrase prepares the viral DNA for integration into the host genome. It does this by cleaving 2-3 nucleotides off of each 3' end of the LTR. It cleaves by using metal ions to hydrolyze a phosphodiester bond. Although manganese has been shown to sufficiently manage this function, magnesium is generally considered to be the metal ion used *in vivo*. The enzyme cleaves between the highly conserved CA motif and the terminal nucleotides, which leaves a 5' overhang and an exposed hydroxyl group on each of the terminal 3' adenosines. This hydroxyl group is critical later on for a transesterification reaction with the host genome<sup>(3, 6, 7)</sup>. Integrase, with the help of other viral and host proteins, forms a pre-integration complex (PIC). The PIC carries out 3'-processing in the cytoplasm and then transports the viral DNA into the nucleus where it will start the integration process<sup>(6, 7)</sup>.

## 1.4 Purpose of the Study

3'-processing is a key step in the replication of HIV and other retroviruses. This makes it an excellent target for retroviral drugs. This is hindered by the fact that the specifics of 3'-processing are poorly understood. Previous studies on this subject have discovered important motifs within each domain that help the integrase perform its functions, however details concerning how the DNA-binding C-domain interact with the viral DNA are limited. One of the bigger discoveries in this area was the existence of a highly conserved CA sequence at the 3' ends of the viral DNA <sup>(1,8,9,10)</sup>. This is a major clue to how 3'-processing works, but the reasons why these nucleotides are important and exactly how the enzyme interacts with them remains elusive.

One of the better ways to get an actual look at how integrase is performing 3'-processing is by seeing the crystal structure of the enzyme. HIV integrase however, has proven very difficult to crystallize due mostly to low stability. Individual domains have been crystallized to a limited extent, but there is no successful crystallization of HIV integrase in complex with viral DNA before the enzyme has carried out 3'-processing (pre-cleaved DNA). The crystal structure of prototype foamy virus (PFV) integrase, however, was solved in 2010 and shows the uncut DNA in complex with the enzyme <sup>(5)</sup>. By looking at how the highly similar PFV integrase interacts with the DNA, insight into how HIV integrase performs 3'-processing can be gained. This is done through the creation of a homology model and through sequence alignment that allows for the comparison of matching residues at points of contact between the enzyme and the DNA.

Other features of HIV integrase-DNA interactions including the actual source of LTR recognition, the substrate's structural limitations, and the terminal and non-terminal binding characteristics of integrase are all still being debated. In order to shed some light on these traits,

this project aims to determine what basic structural substrate characteristics are required for efficient 3'-processing. To do this many different DNA substrates, each with a variation of some unique physical structure at the terminus (see results section), are allowed to react with integrase. By studying the extent to which the integrase is able to process the substrates relative to a standard, the preferences and limitations of integrase can be found. The combination of these two methods allows for a clarification of the effects of both general structural features and individual nucleotide preference in the terminal sequence of the viral substrate.

Previous studies on sequence specificity have shown that both single-point mutations and multi-point mutations have the greatest effect on 3'-processing rates when either the conserved C or A is affected <sup>(1,8,9,10)</sup>. It has also been shown that if the conserved C-G and A-T pairs are both flipped at the same time then 3'-processing is drastically reduced <sup>(8)</sup>. Due to the proven sensitivity of these 4 nucleotides and the lack of consistent sensitivity elsewhere near the terminus, many of the substrates tested will focus either specifically on these 4 nucleotides or on the area around them.



## 2 MATERIALS AND METHODS

*Table 1.1 Buffers for the Purification of HIV Integrase*

Buffers	I	II	III	IV	V	VI
NaCl (M)	1	1	1	1	0.5	1
HEPES (20mM, pH=7.5)	✓	✓	✓	✓	✓	✓
EDTA (mM)	0	0	0	0	2	1
Imidazole (mM)	20	60	60	1000	300	0
Glycerol (10%)			✓	✓	✓	✓
DTT (2mM)						✓
β-Mercaptoethanol (2mM)	✓	✓	✓	✓	✓	
PMSF (1mM)	✓					
ZnCl <sub>2</sub> (100μM)						✓
Total Volume (mL)	500	500	500	500	1000	1000

### 2.1 Synthesis and Purification of DNA Substrates

The U5 terminus of the HIV-1 genome was used as a template for the chemical structure during synthesis of the oligodeoxyribonucleotides. Standard synthesis protocols were used. Once deprotected, purification of the oligonucleotides was achieved by ion exchange using a Pharmacia 15Q PE column. They were then desalted using a GE HiTrap desalting column (gel filtration). Oligonucleotide concentrations were found using extinction coefficients calculated from the sum of its mononucleotides. The 5' ends of the oligonucleotides were labeled with T4 DNA kinase and  $\gamma$ -<sup>32</sup>P ATP (Dupont-NEN, Boston, MA).

## 2.2 DNA Substrate Labeling

<sup>32</sup>P ATP labeling of the 5' termini was achieved using T4 polynucleotide kinase. The reaction mixture contained 0.05 mg/mL BSA, 25 pmol hot/cold ATP mixture, 50 pmol DNA substrates, 1x T4 kinase reaction buffer, and 10 units of T4 kinase. The reaction mixture had a total volume of 20 µl. It was incubated at 37 °C for 1 hour followed by being held at 85 °C for 15 minutes (to stop the enzyme).

## 2.3 Integrase Expression and Purification

HIV-1 integrase (F185H/C280S) was expressed in the *Escherichia coli* BL-21 (DE3) cell with a hexahistidine-tag in the plasmid pET15. A 3 ml start culture was grown overnight before being transferred to a 300 ml LB medium. At a culture OD of 0.6-0.8 1 mM IPTG was added. It was then left for 3 hours. Next, the culture was spun at 6,000 rpm for 30 minutes and the pellet was washed with 10 ml of buffer I. Then the cell was sonicated for 5 seconds (pause), 10 seconds (pause), and finally for 3 minutes. It was then put in a centrifuge set at 13,000 rpm for 30 minutes. The supernatant was removed and put onto a 5 ml Ni-NTA affinity column, and washed with 50 ml of buffer II. The protein was then washed out with a gradient of 10 mM to 1 M imidazole. The resulting product was checked with a 12% SDS-PAGE before being dialyzed with 1 L of buffer V overnight. The 300 ml culture yielded roughly 3 mg of ~11.8 µM protein with an extinction coefficient of 50,670. Thrombin was added (10 units/mg protein) and then allowed to react overnight at 4 °C. This separated the his-tag and was check via 12% SDS-PAGE. The thrombin was then removed by passing the mixture through a 1 ml benzamidine sepharose column. The resulting protein was dialyzed with buffer VI and stored at -20 °C.

## 2.4 3'-Processing

Each 3'-processing reaction contained 2  $\mu$ M integrase, 10 mM  $\beta$ -mercaptoethanol, 10% glycerol, 25 mM MOPS (pH 7.2), 7.5 mM  $\text{MnCl}_2$ , 50 nM  $^{32}\text{P}$ -radiolabeled substrates (hairpin and duplex DNA), and 0.1 mg/ml BSA. The total volume for each reaction mixture was 20  $\mu$ l. The DNA was reacted for 3.5 hours at 37  $^{\circ}\text{C}$ . The reactions were stopped by placing them in a -20  $^{\circ}\text{C}$  freezer. The reaction products were separated using 15% polyacrylamide denaturing gels (8 M Urea). They were then detected using a phosphor storage screen and imaged using the GE Healthcare Typhoon 9400.

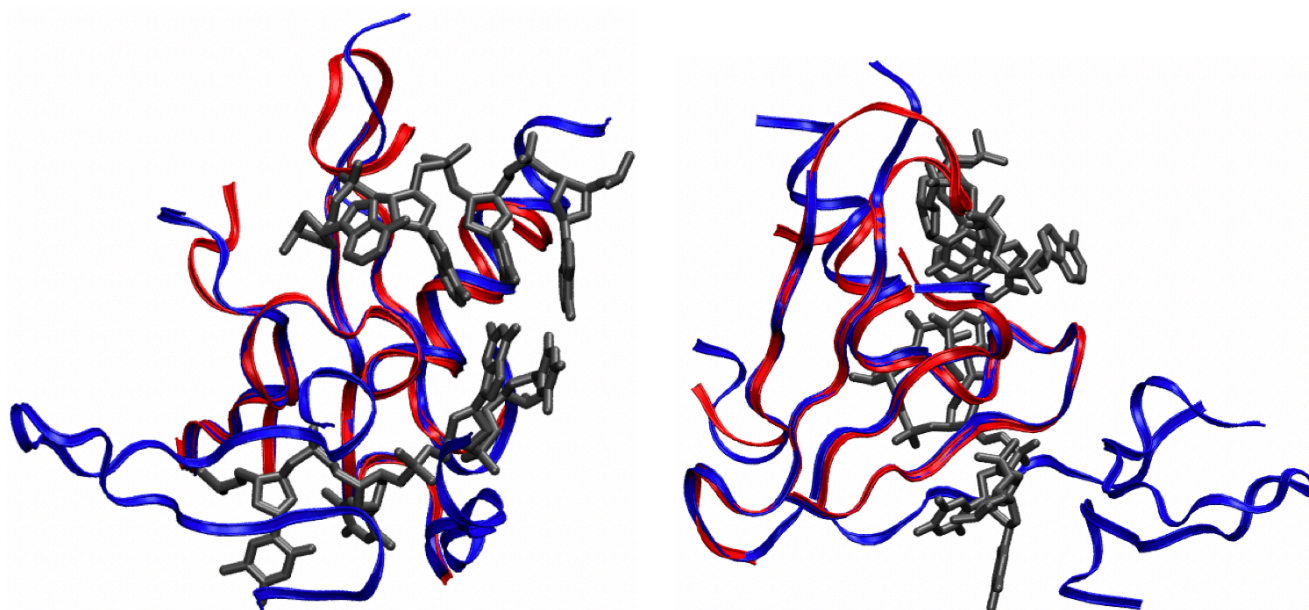
## 2.5 Modeling

The homology model between PFV and HIV integrase was created using the online SWISS-MODEL program from Biozentrum. The model was created based on target-template alignment using ProMod3. The template structure file for the PFV intasome was taken from [www.rcsb.org](http://www.rcsb.org) and is classified as 4E7H (DOI: 10.2210/pdb4E7H/pdb). The model was visualized using VMD software.

### 3 RESULTS AND DISCUSSION

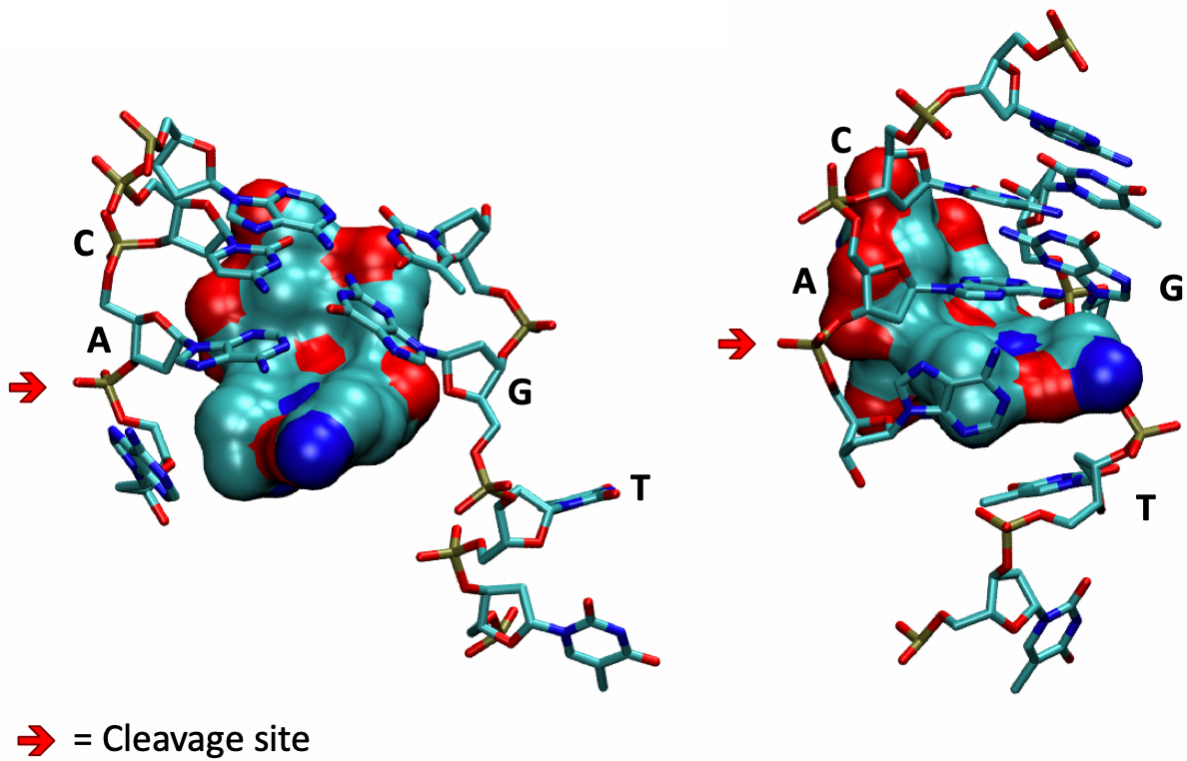
#### 3.1 Homology Model

HIV Integrase (**red**) and PFV Integrase (**blue**) bound to viral DNA (**grey**)

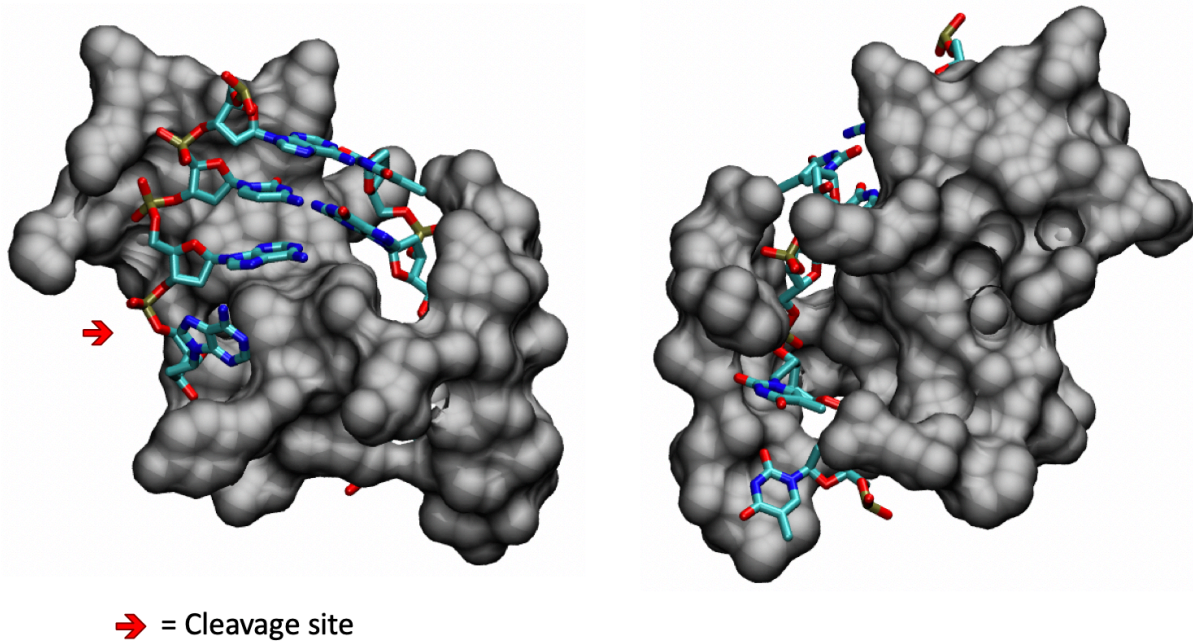


**Figure 2: HIV/PFV Homology Model**

*Here HIV integrase (red) is aligned with PFV integrase (blue) in complex with viral DNA (grey). The model on the right has been rotated about 90°. Both enzymes share the helix shown sitting between terminal sections of the two DNA strands. The homology model was created using SWISS-MODEL and visualized using VMD.*

**Figure 3: PFV Integrase (Paired Down)**

Here the PFV Integrase residues that push between the DNA strands are shown. The integrase forms a pocket for the conserved A, which is kept in close proximity to the conserved C and G. The conserved T is not found nearby.



**Figure 4: PFV Integrase (Full)**

All integrase residues near the DNA ends are shown here. The 3' strand containing the cleavage site (red arrow) is pushed to the side by a proline and a tyrosine (image on left). The 5' strand containing the conserved T is wrapped around the back (image on right) and stabilized separately.

### 3.2 List of Substrates

**Blue** = Conserved WT Sequence

**Red** = Mutation

**I** = Inosine

**X** = 2-Aminopurine

**D** = 7-Deaza-2'-Deoxyadenosine

*Table 2.1.1 Substrates with Terminal Structural Changes*

Substrate Name	Full Sequence
<b>Wild Type</b>	3' - CACCTTTTAGAGATCGTCA - 5' 5' - GTGGAAAATCTCTAGCAGT - 3'
<b>5'-Overhang</b>	3' - CACCTTTTAGAGATCGTCATTTTT - 5' 5' - GTGGAAAATCTCTAGCAGT - 3'
<b>3'-Overhang</b>	3' - CACCTTTTAGAGATCGTCA - 5' 5' - GTGGAAAATCTCTAGCAGTTTTT - 3'
<b>24-mer T-tails</b>	3' - CACCTTTTAGAGATCGTCATTTTT - 5' 5' - GTGGAAAATCTCTAGCAGTTTTT - 3'
<b>22-mer T-tails</b>	3' - CACCTTTTAGAGATCGTTTTT - 5' 5' - GTGGAAAATCTCTAGCATT - 3'
<b>3T-Hairpin</b>	3' - CACCTTTTAGAGATCGTCA <sup>T</sup> <sub>T</sub> 5' - GTGGAAAATCTCTAGCAGT <sup>T</sup>
<b>6T-Hairpin</b>	3' - CACCTTTTAGAGATCGTCA <sup>TTT</sup> <sub>T</sub> 5' - GTGGAAAATCTCTAGCAGT <sup>TTT</sup>
<b>Bulge</b>	3' - CACCTTTTAGAGATCGTCATTTTTCACC - 5' 5' - GTGGAAAATCTCTAGCAGTTTTTGTGG - 3'

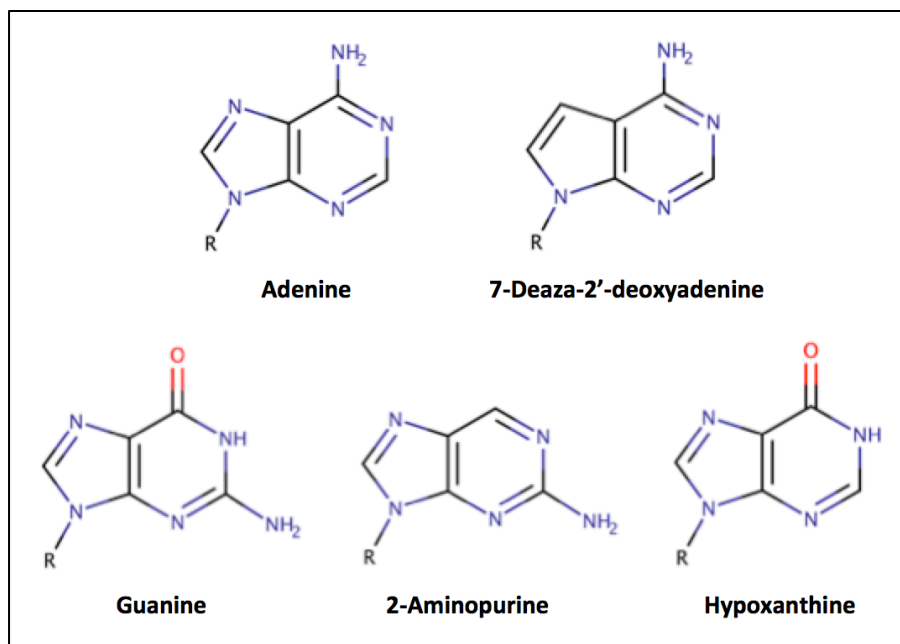
Table 3.1.2 Substrates with Base Substitutions

Substrate Name	Full Sequence
<b>Wild Type</b>	3' -CACCTTTTAGAGATCGTCA-5' 5' -GTGGAAAATCTCTAGCAGT-3'
<b>Inosine</b>	3' -CACCTTTTAGAGATCITCA-5' 5' -GTGGAAAATCTCTAGCAGT-3'
<b>2-Aminopurine</b>	3' -CACCTTTTAGAGATCXTC A-5' 5' -GTGGAAAATCTCTAGCAGT-3'
<b>G/C Switch</b>	3' -CACCTTTTAGAGATCCTCA-5' 5' -GTGGAAAATCTCTAGGAGT-3'
<b>A/T Switch</b>	3' -CACCTTTTAGAGATCGACA-5' 5' -GTGGAAAATCTCTAGCTGT-3'
<b>CT T-tails</b>	3' -CACCTTTTAGAGATCGTTTTT-5' 5' -GTGGAAAATCTCTAGCTTTTT-3'
<b>Deaza Bot w/ T</b>	3' -CACCTTTTAGAGATCGTCA-5' 5' -GTGGAAAATCTCTAGCDGT-3'
<b>Deaza Top w/ T</b>	3' -CACCTTTTAGAGATCGDCA-5' 5' -GTGGAAAATCTCTAGCTGT-3'
<b>Deaza Bot w/ C</b>	3' -CACCTTTTAGAGATCGCCA-5' 5' -GTGGAAAATCTCTAGCDGT-3'
<b>Deaza Top w/ C</b>	3' -CACCTTTTAGAGATCGDCA-5' 5' -GTGGAAAATCTCTAGCCGT-3'
<b>C Top w/ A</b>	3' -CACCTTTTAGAGATCGCCA-5' 5' -GTGGAAAATCTCTAGCAGT-3'
<b>G Bot w/ T</b>	3' -CACCTTTTAGAGATCGTCA-5' 5' -GTGGAAAATCTCTAGCGGT-3'
<b>Terminal G</b>	3' -CACCTTTTAGAGATCG-5'



	5' -GTGGAAAATCTCTAGCAGT- 3'
<b>A→C</b>	3' -CACCTTTTAGAGATCG- 5' 5' -GTGGAAAATCTCTAGCCGT- 3'
<b>A→G</b>	3' -CACCTTTTAGAGATCG- 5' 5' -GTGGAAAATCTCTAGCGGT- 3'
<b>A→T</b>	3' -CACCTTTTAGAGATCG- 5' 5' -GTGGAAAATCTCTAGCTGT- 3'
<b>GA-5'</b>	3' -CACCTTTTAGAGATCGA- 5' 5' -GTGGAAAATCTCTAGCTGT- 3'
<b>GT-5'</b>	3' -CACCTTTTAGAGATCGT- 5' 5' -GTGGAAAATCTCTAGCTGT- 3'

### 3.3 Base References

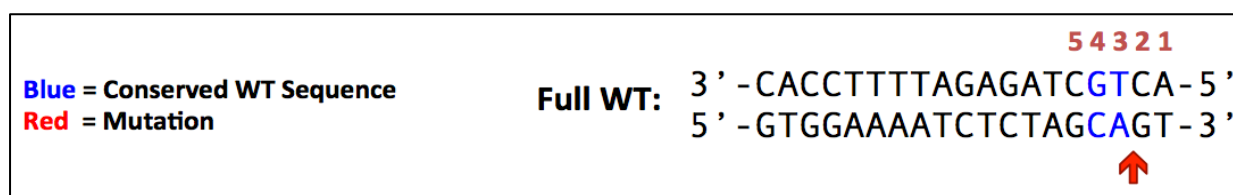


**Figure 5: Altered Bases Used**

*Altered bases used are shown here next to the adenine and guanine bases they mimic. Base references were made using MarvinSketch.*

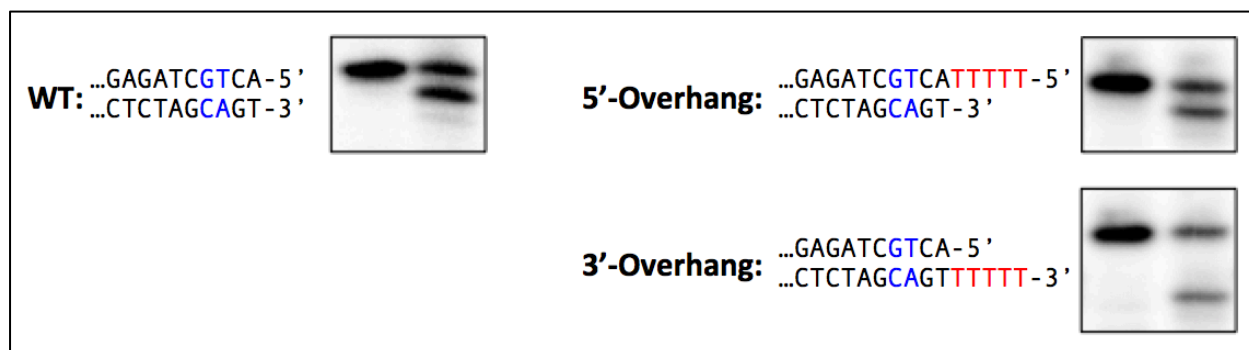
### 3.4 Gel Results

The left lane of each gel contains labeled substrate but no enzyme. The right lane is exactly the same but with added enzyme. All reactions were incubated for the same period of time, which yields about 50% cleavage on the WT as shown on the left in each figure. Each substrate was tested multiple times to ensure consistency. Note that there is a very light third band present below the cleaved product in the WT that may be visible in the mutations as well.



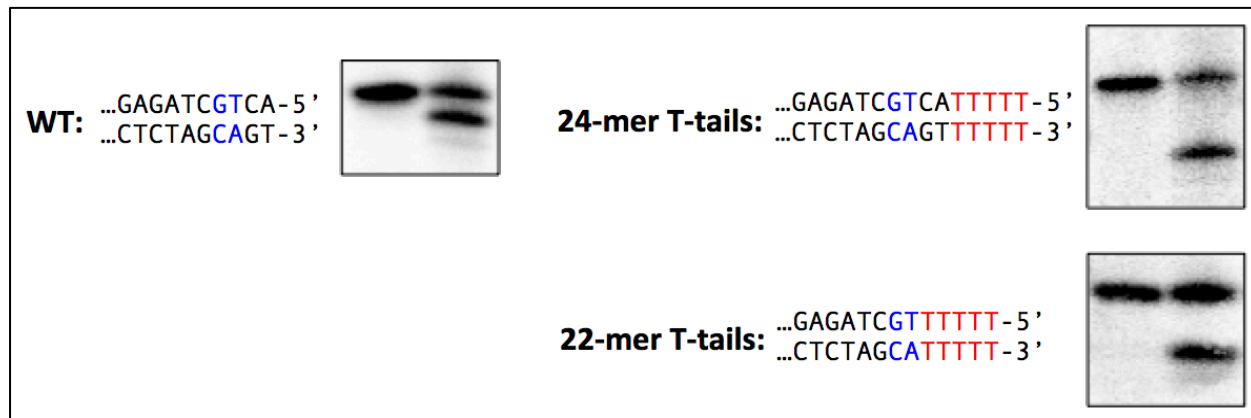
#### Figure 6: Results Key

The red arrow indicates the location of 3'-cleavage by integrase (just after the conserved adenosine). The red numbers on top of the WT sequence are position references.



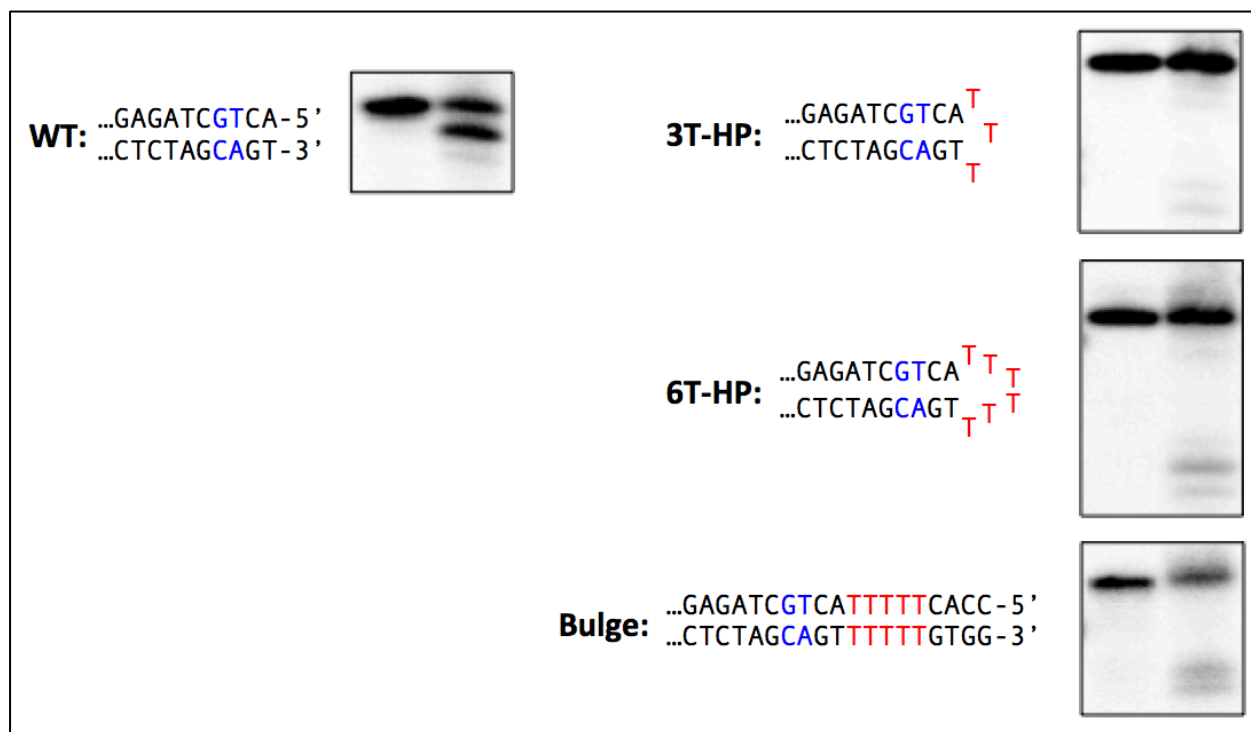
#### Figure 7: Terminal Structure Mutations - T-overhangs

T-overhangs located on either strand did not affect 3'-processing in any major way.



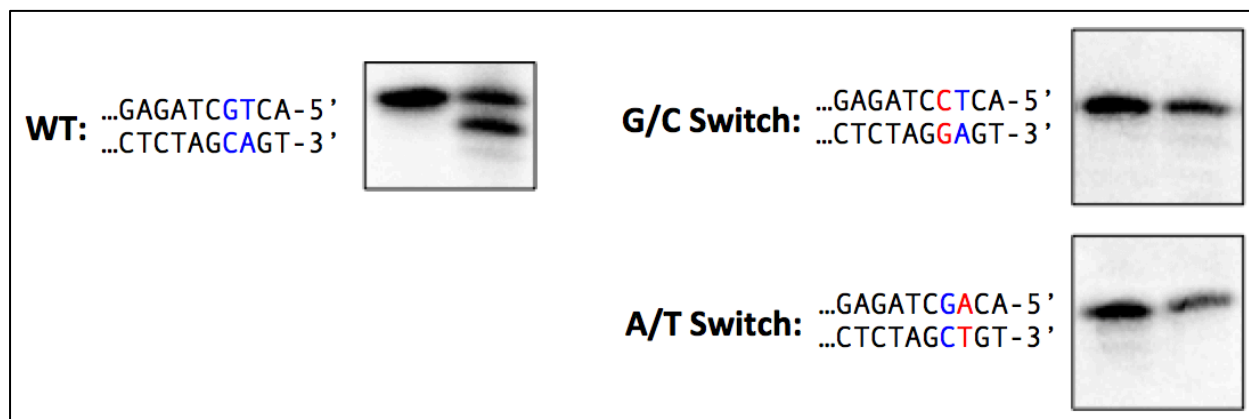
**Figure 8: Terminal Structure Mutations - T-tails**

*T-tails located either on the ends of the WT strands or directly after the conserved dinucleotide pairs do not have any major effect on 3'-processing.*



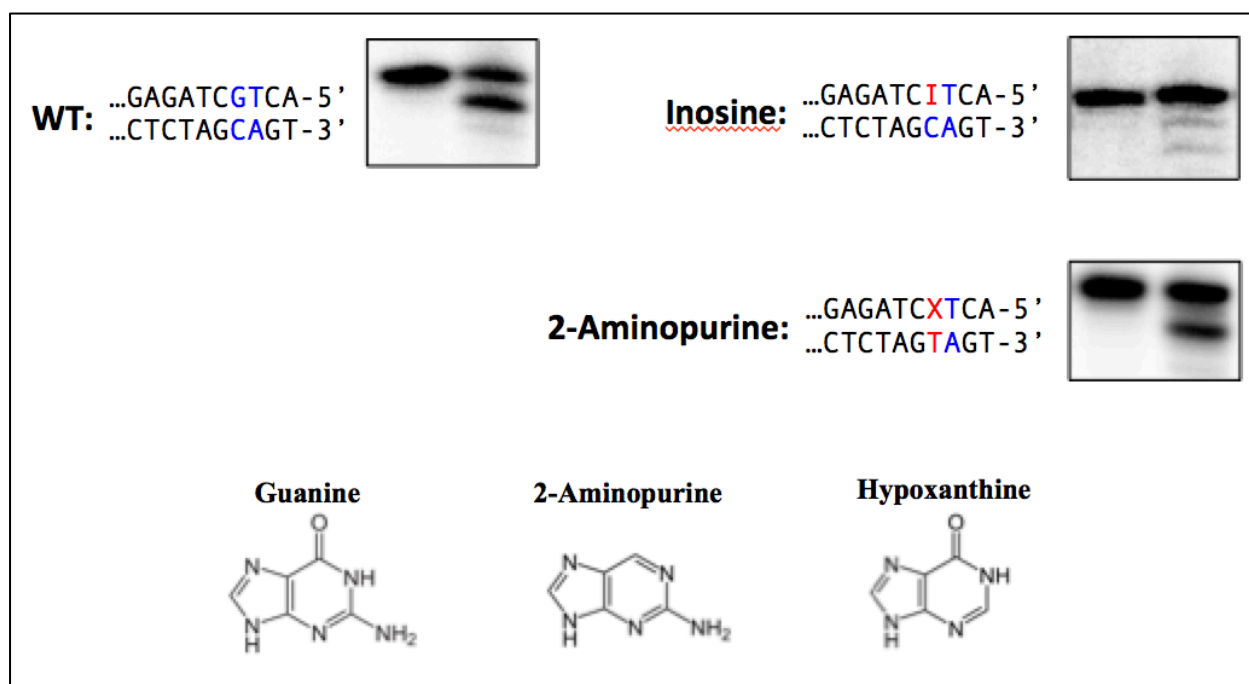
**Figure 9: Terminal Structure Mutations - Closed Ends**

*Unlike the T-tails and T-overhangs, the hairpin and bulge substrates substantially reduce the amount of cleaved product. This is likely due to the closed or blocked off nature of the strand ends in these substrates, which may hinder any attempt to force the strands apart.*



**Figure 10: Base Mutations and Substitutions - Side Specificity**

*Flipping either pair of conserved nucleotides resulted in a dramatic loss of cleaved product.*



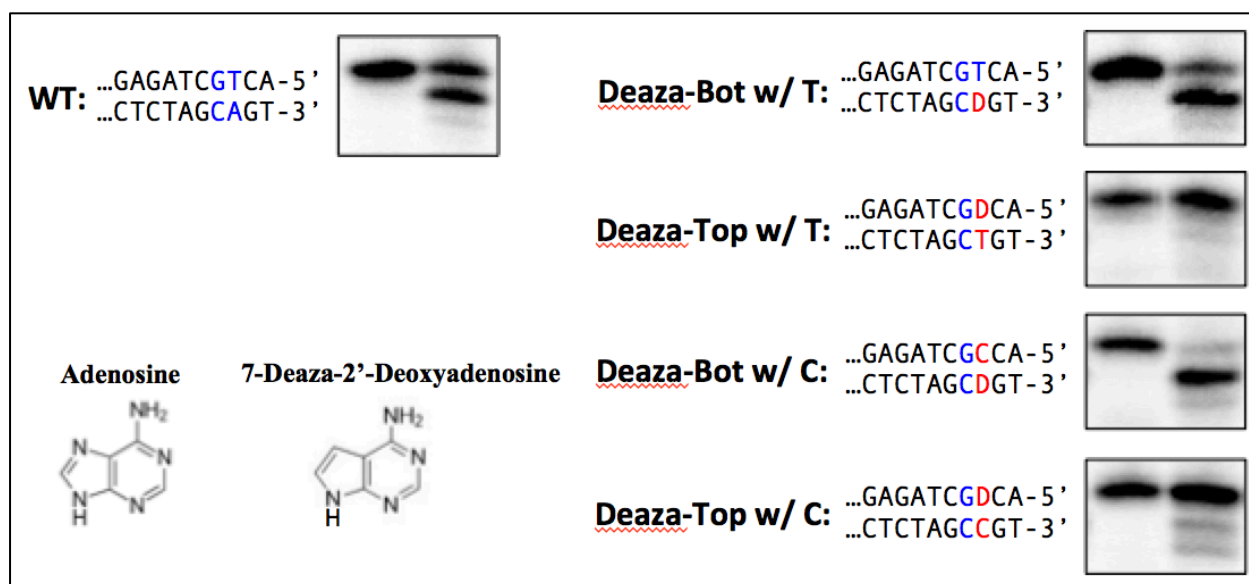
**Figure 11: Base Mutations and Substitutions - Importance of the G**

*In order to test the importance of the functional groups on the conserved G, the base was replaced with a 2-Aminopurine (lacking a carbonyl and imino proton) and a Hypoxanthine (lacking the amino group). As shown the amount of cleaved product was reduced only when the amino group was missing.*



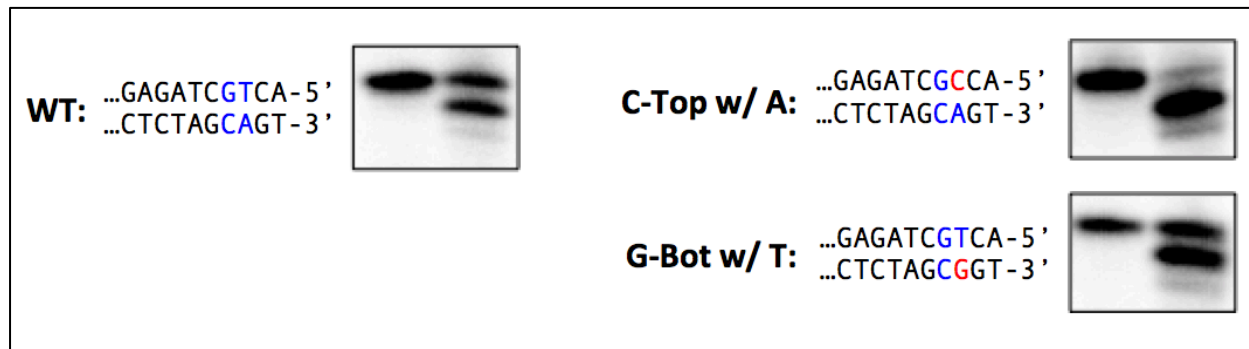
**Figure 12: Base Mutations and Substitutions - Importance of the A**

The CT T-tails substrate contains the same sequence as the 22-mer T-tails substrate except the conserved A is replaced by a T. Cleavage was drastically reduced when the A was not present.



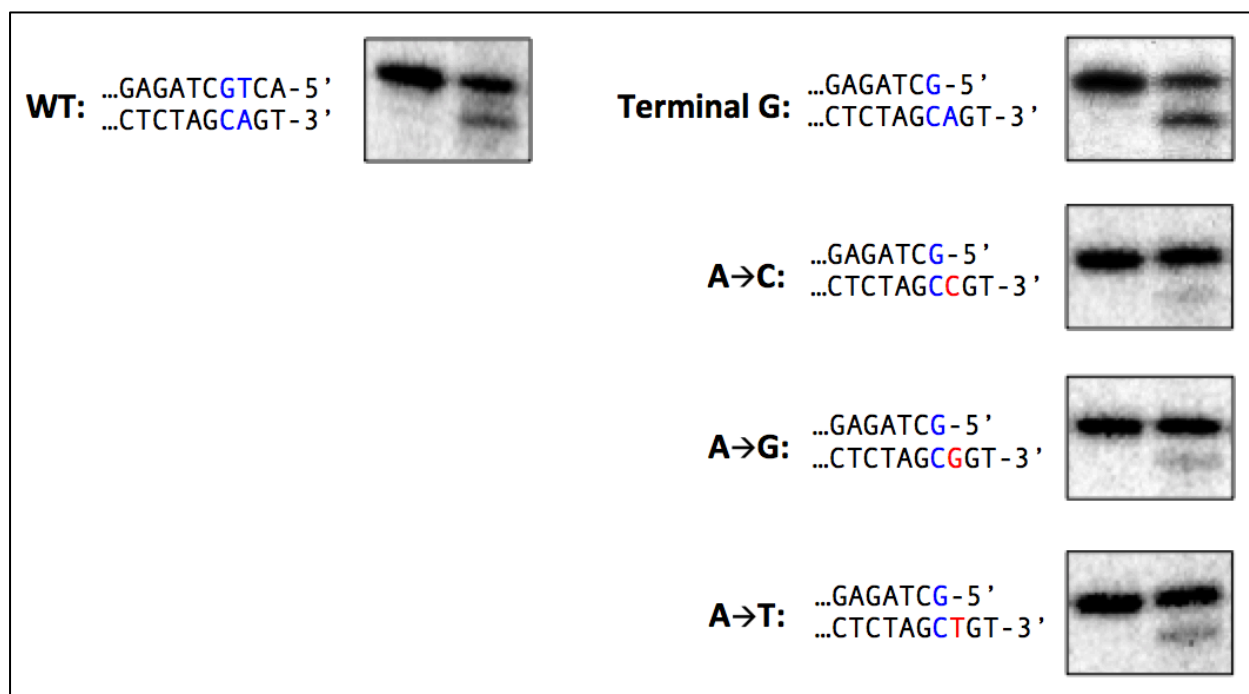
**Figure 13: Base Mutations and Substitutions - The Importance of N7 on the A**

The deaza compound was tested in both complementary and non-complementary scenarios in both the top and bottom strands as shown. The Deaza-Bot w/ T substrate is an exact copy of the WT sequence but without the N7 on the A. The success of this substrate suggests that the N7 is not required for 3'-processing.



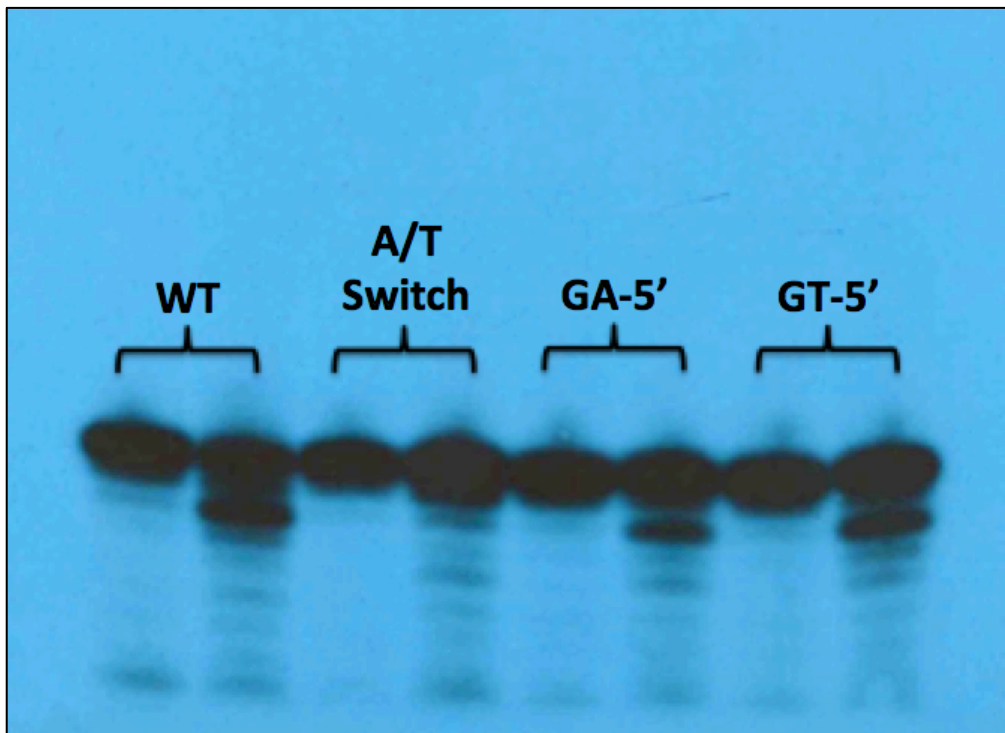
#### Figure 14: Non-complementary Bases

*These substrates are designed to test whether substituting the conserved A-T pair for a non-complementary pair has a detrimental effect on 3'-processing. The enzyme was able to cleave the substrate even better than the standard when the 3' base was an adenine and just as well as the standard when it was a guanine.*



#### Figure 15: Terminal G Substrates

*These substrates contain 5'-ends that stop after the conserved G. The success of the Terminal G substrate indicates that the conserved T is likely not very important. The other 3 substrates test the same scenario but with different bases in place of the A. Of the 3 substitutions, only the T is comparable to the WT.*



**Figure 16: GA-5' and GT-5' Film Results**

Gel lanes shown are in pairs with the negative (no enzyme) on the left and the positive (enzyme added) on the right. Substrates used for this gel from left to right are WT, A/T Switch, GA-5', and GT-5'. The results show a dramatic decrease in cleaved product for the A/T Switch compared to the GA-5' which is identical except for the lack of 2 terminal nucleotides on the 5'-side after the A. The GT-5' produced more cleaved product than either of the other 2 which is expected since it is a non-complementary pair. This gel has a blue tint because it was imaged on film instead of a GE Typhoon 9400.

Substrate	Result	Terminal Sequence	Gel	Substrate	Result	Terminal Sequence	Gel
Wild Type	✓	...GAGATCGTCA-5' ...CTCTAGCAGT-3'		Inosine	~	...GAGATCITCA-5' ...CTCTAGCAGT-3'	
5'-Overhang	✓	...GAGATCGTCATTTTT-5' ...CTCTAGCAGT-3'		2-Aminopurine	✓	...GAGATCXTCA-5' ...CTCTAGTAGT-3'	
3'-Overhang	✓	...GAGATCGTCA-5' ...CTCTAGCAGTTTTT-3'		G/C Switch	✗	...GAGATCCTCA-5' ...CTCTAGGAGT-3'	
22-mer T-Tails	✓	...GAGATCGTTTTT-5' ...CTCTAGCATT-3'		A/T Switch	✗	...GAGATCGACA-5' ...CTCTAGCTGT-3'	
24-mer T-Tails	✓	...GAGATCGTCATTTTT-5' ...CTCTAGCAGTTTTT-3'		Deaza Bot w/ T	✓	...GAGATCGTCA-5' ...CTCTAGCDGT-3'	
CT T-Tails	✗	...GAGATCGTTTTT-5' ...CTCTAGCTTTTT-3'		Deaza Top w/ T	✗	...GAGATCGDCA-5' ...CTCTAGCTGT-3'	
3T-HP	~	...GAGATCGTCA <sup>T</sup> ...CTCTAGCAGT <sup>T</sup>		Deaza Bot w/ C	✓	...GAGATCGCCA-5' ...CTCTAGCDGT-3'	
				Deaza Top w/ C	~	...GAGATCGDCA-5' ...CTCTAGCCGT-3'	
6T-HP	~	...GAGATCGTCA <sup>T T</sup> ...CTCTAGCAGT <sup>T T</sup>		C Top w/ A	✓	...GAGATCGCCA-5' ...CTCTAGCAGT-3'	
				G Bot w/ T	✓	...GAGATCGTCA-5' ...CTCTAGCGGT-3'	
Bulge	~	...GATCGTCA <sup>T T T T T</sup> CACC-5' ...CTAGCAG <sup>T T T T T</sup> GTGG-3'		Terminal G	✓	...GAGATCG-5' ...CTCTAGCAGT-3'	
				A→C	~	...GAGATCG-5' ...CTCTAGCCGT-3'	
				A→G	~	...GAGATCG-5' ...CTCTAGCGGT-3'	
				A→T	~	...GAGATCG-5' ...CTCTAGCTGT-3'	
				GA-5'	✓	...GAGATCGA-5' ...CTCTAGCTGT-3'	
				GT-5'	✓	...GAGATCGT-5' ...CTCTAGCTGT-3'	

**Figure 17: Summarized Gel Results**

*All substrate results are shown with corresponding sequences and gel results.*

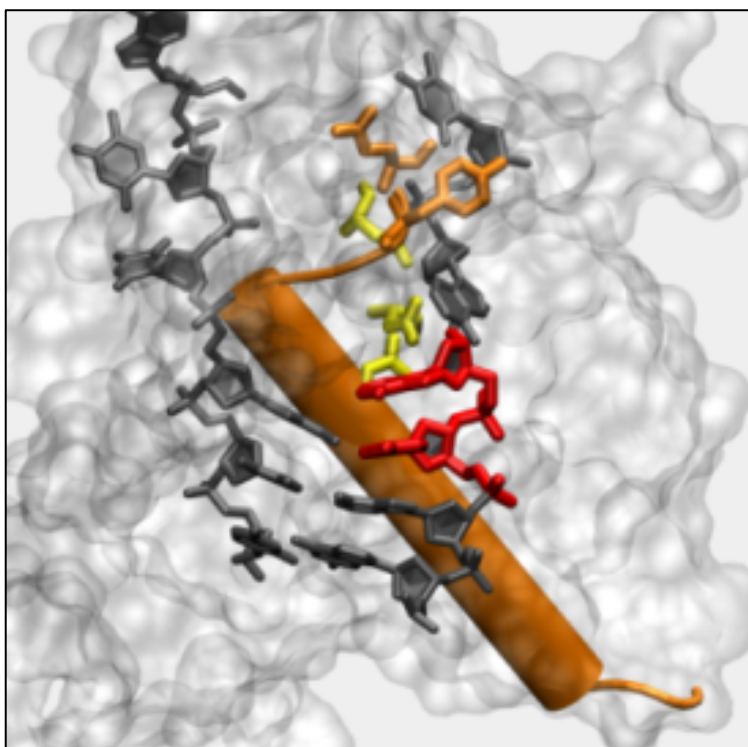
### 3.5 Discussion

The first set of substrates tested with integrase was designed to test the effect of having a bulky sequence on the end of the WT substrate. This was accomplished by adding T-sequence(s) onto the end of the substrate in a variety of ways. Figure 7 shows that substrates containing an individual T-sequence on either side does not majorly affect substrate cleavage. Figure 8 shows a similar result when T-sequences are present on both sides at once both after the typical terminal dinucleotides and when occurring directly after the conserved dinucleotides (shown in blue in each figure). These results would suggest that having a bulky object on the end of either strand or both simultaneously does not significantly affect DNA binding or 3'-processing. However, all of



these cases involve open-ended termini and do not necessarily block access to the conserved nucleotides any more than the WT.

The next set of substrates was designed with this in mind. They include one with a small 3-T hairpin, one with a larger 6-T hairpin, and one with a 9-mer sequence containing a 5-T non-complimentary bulge after the representative WT sequence. These sequences and their corresponding results can be seen in Figure 9. They show that with a closed end the amount of cleaved product is severely reduced. Even when comparing the 3-T and 6-T substrates, the 6-T, which has a larger more open loop has more cleaved product than the smaller 3-T loop. This would make sense if the enzyme needed to force the ends apart in order to cleave.



**Figure 18: PFV with Pre-cleaved DNA**

*PFV integrase is shown forcing the DNA ends apart using the helix shown in orange. The DNA is shown in dark grey with the conserved CA dinucleotide shown in red.*

PFV integrase, a very similar enzyme to HIV integrase, has actually been crystallized in complex with pre-cleaved viral DNA, and in the case of PFV the ends of the DNA are indeed forced apart by the insertion of a helix between the two strands by the integrase. An example of this is shown in Figure 15. In order to further investigate the extent of the similarities between PFV and HIV integrase, a homology model was made between the unbound HIV integrase and the PFV integrase in complex with the pre-cleaved DNA as shown in Figure 2.

The similarities between PFV and HIV integrase including the helix positioned between the two DNA strands support the idea that the closed-end substrates are reducing the 3'-processing efficiency by making it harder for the enzyme to separate the DNA ends. The difference in the results between the T-tail substrates and the hairpin substrates also support this theory.

Subsequent substrate sets focus more on the effects of individual base substitutions or mutations. Figure 10 shows the result of flipping the conserved GC or TA. In both cases 3'-processing is abolished. This confirms the importance of certain characteristics within these nucleotides and suggests possible side specificity. The next step then is to investigate why the individual bases are important. The substrates shown in Figure 11 aim to discover the importance of the G. The strategy here was to figure out if either side group on the G or the imino proton were essential at this position. This was tested by replacing the G with both 2-Aminopurine and Inosine. 2-Aminopurine contains the amino group but not the carbonyl or the imino proton while Inosine contains the carbonyl group and the imino proton but not the amino group. The results clearly show a drastic reduction in cleaved product if the amino group is missing and almost no change in cleaved product when the amino group is present but the carbonyl and imino group are not.

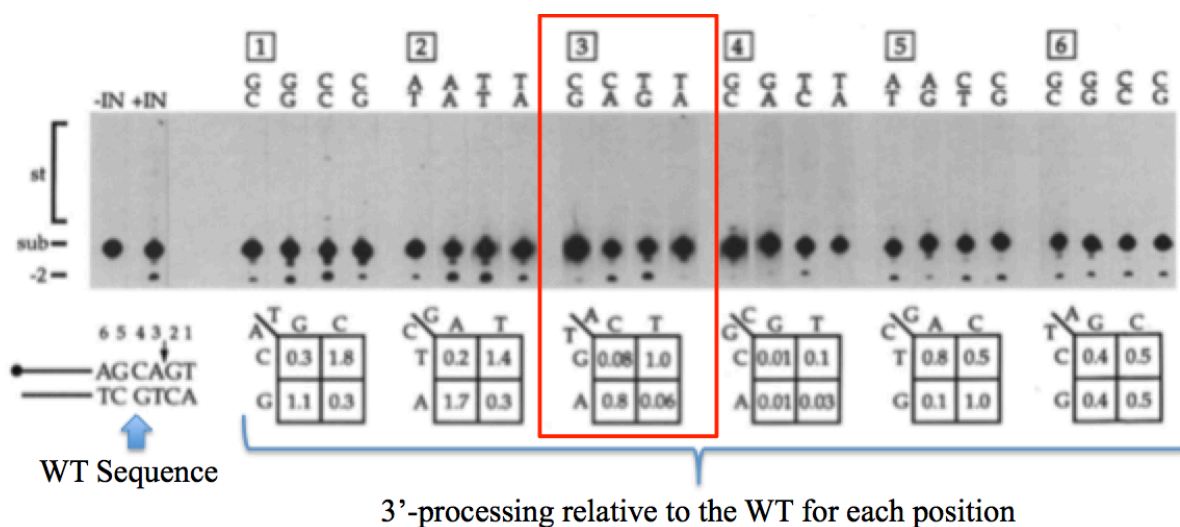
After the G, the conserved A was tested. This was done by comparing a substrate that is known to work with the same substrate containing a T in place of the conserved A. This is shown in Figure 12 and demonstrates that the A is important in some way. The reason for the importance of the A is a matter of some debate. One theory is that the N7 on the A is an important point of contact between the integrase and the DNA. A structure analysis and photocrosslinking study done by Dr. Robert Craigie suggests that Lys159 creates one of the final points of contact along the DNA strand and is partly responsible for stabilizing one of the two separated DNA ends <sup>(9)</sup>. In addition, it has been shown in other studies that the positively charged amino group on lysine can form a hydrogen bond with the N7 on adenine <sup>(11)</sup>. For this reason, a series of substrates were made containing 7-Deaza-2'-Deoxyadenine in place of the conserved A. This base is identical to adenine except it has a carbon in place of the N7.

The deaza compound was tested on both DNA strands in both complementary and non-complementary base pairs. The results, shown in Figure 13, reveal that the N7 is not crucial for binding and 3'-processing. When the deaza is on the 3'-side (the same position as the A in the WT) and across from a complementary base, there is cleaved product comparable to the WT. When those two bases are flipped, there is virtually no cleaved product, which is consistent with the A/T Switch substrate in Figure 10. Interestingly, when the bases are non-complementary the relative amount of cleaved product increased regardless of which side the deaza is on. This raises the question of how important the bases are at this position if they are non-complementary.

Work done by Dr. Patrick Brown shows some examples of non-complementary pairs being substituted (Figure 19) for the conserved pairs. At position 3 (the conserved A-T) the non-complementary pairs show successful cleavage even when the 3' residue is not an A. The complementary pairs shown are consistent with our data (little or no cleavage when there is no A

on the 3'-side). To further test non-complementary pairs more combinations were tried (Figure 14). In all cases non-complementary pairs at this position lead to successful cleavage. It is worth noting that among the combinations shown in Figures 14 and 19, the one with an A on the 3'-side still cleaves the best even when they are non-complementary. This ability to work with mismatched bases is not shared by position 4 (the conserved G-C) as shown in Figure 19.

The next batch of substrates was designed to test whether the base on the 3'-side at position 3 (the conserved A in the WT) needs anything across from it at all in order to work. Figure 15 shows that 3'-processing is still successful even when the 5'-strand ends after the conserved G. The amount of cleaved product is slightly reduced if the A is then replaced by a T and noticeably reduced if it is replaced by a G or C. However, in all cases it is able to cleave to some extent.



### Figure 19: Non-complementary pairs

Numbers along the top indicate the base pair position starting from the end as shown on the WT sequence in the lower left corner. The control is shown on the left. Successful 3'-processing is marked by the appearance of a second lower band. The red box indicates the position of interest that contains the conserved A. At this position both non-complementary pairs show successful cleavage while both of the complementary pairs that differ from the WT show a lack of cleaved product. This figure is adapted from Disruption of the terminal base pairs of retroviral DNA during integration<sup>(12)</sup>. Figure adapted from P. Brown<sup>(12)</sup>.

#### 4 CONCLUSIONS

The main initiative in this project is to elucidate structure and sequence-related features on viral DNA substrate termini necessary for successful binding and 3'-processing by HIV integrase. The initial substrate sets were designed to test the ability of the enzyme to overcome large obstructive structural features on the terminus of the DNA. The idea was to see whether or not HIV integrase mimics the strand separation of PFV integrase and how resilient the process is to varying structural features on the end of the DNA. These bulky features on the end also reveal the effect of placing the conserved CA motif farther from the end. The results (Figures 7-9) show that moving the CA motif 5 base pairs farther from the end does not hurt the enzyme's ability to bind and cleave. In addition, having a bulky object on the end of the substrate did not reduce the amount of cleaved product in and of itself. Reduced cleavage was only seen when the ends of the DNA had a closed nature such as a hairpin ending (Figure 9).

The negative effect of having a closed end supports the idea that like PFV integrase, HIV integrase wants to separate the two DNA strands at the end before cleaving the 3'-side. Even among the hairpin substrates, the one with a larger (looser) end produced more cleaved product than the one with a smaller (tighter) loop. The homology model shown in Figure 2 also showed that both PFV and HIV integrase have a glycine residue in very close proximity to the conserved G's amino group. Given these similarities and the results from Figures 7-9, it is likely that HIV integrase uses a similar method to PFV integrase, which is to separate and stabilize the DNA ends individually before cleaving the 3'-side.

For substrates with sequence mutations the focus was on the importance of the conserved CA and corresponding GT since this has been shown to be the most important motif for successful 3'-processing. The G/C and A/T switch substrates (Figure 10) show a complete lack

of cleaved product. This could indicate a strong side preference for these bases. It also shows the importance of these 4 bases by demonstrating the process's sensitivity to their placement. So why are they so important? For the G, the amino group is a strong possibility as shown in Figure 11. Since both PFV integrase and HIV integrase contain a glycine in close proximity to this amino group and since the lack of the amino group leads to a lack of cleaved product, it is likely an important point of contact with the integrase.

When it comes to the A, the N7 is ruled out as an essential feature of the DNA (Figure 13), but there is something about the A that makes it key for 3'-processing as evidenced by the CT T-Tails substrate (Figure 12). It is possible that the N7 is still a typical point of contact with Lys159, but that the lack of an N7 does not prevent the lysine from interacting with the adenine. Other research has shown that lysine commonly interacts with adenine through both hydrogen bonds with the N7 and cation-pi interactions with the ring<sup>(12)</sup>. It is possible that these cation-pi interactions (possibly in conjunction with other interactions) are enough to maintain the proposed Lys159 contact and thereby facilitate 3'-processing even without the N7. It has also become clear that even though complementary bases at position 3 are required to be the conserved A-T in order to produce 3'-processing comparable to the WT, non-complementary bases will also work even when there is no A present (Figure 14 and 19). This would support the theory that the importance of the A has to do with the ability of the enzyme to separate the DNA strands, but if this were the only reason the A were important then you would expect to see roughly equal cleavage among non-complementary pairs. Looking at Figure 14, it is clear that even among non-complementary pairs there is a preference for an A on the 3'-side at this position. The deaza substrates in Figure 13 also show a preference for the deaza on the bottom when the bases are non-complementary. In addition, the Terminal G substrates shown in Figure 15 show that there is

a hierarchy of base preference even when there is nothing across from position of the conserved A and that the A is most preferred. So even if the A is used in separating the strands, there is something else causing the A to be desired for 3'-processing.

## REFERENCES

- (1) Esposito, D., and Craigie, R. "Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction". *The EMBO Journal*, 1998, 5832-5843.
- (2) Dolan, J., Chen, A., Weber, I., Harrison, R., and Leis, J. "Defining the DNA Substrate Binding Sites on HIV-1 Integrase". *J. of Mol. Biol.*, 2008, 568-579.
- (3) Li, M., and Craigie, R. "Processing of Viral DNA Ends Channels the HIV-1 Integration Reaction to Concerted Integration". *J. Biol. Chem.* 2005, 280:29334-29339
- (4) Vora, A. C., Chiu, R., McCord, M., Goodarzi, G., Stahl, S. J., Mueser, T.C., Hyde, C.C., and Grandgenett, D. P. "Avian retrovirus U3 and U5 DNA inverted repeats. Role of nonsymmetrical nucleotides in promoting full-site integration by purified virion and bacterial recombinant integrases". *J. of Biol. Chem.*, 1997, 272:23938-23945.
- (5) Hare, S., Maertens, G., & Cherepanov, P. "3'-Processing and strand transfer catalysed by retroviral integrase in crystallo". *The EMBO Journal*, 2012, 3020-3028.
- (6) Zennou V., Petit C., Guetard D., Nerhbass U., Montagnier L., Charneau P. "HIV-1 genome nuclear import is mediated by a central DNA flap". *Cell*. 2000, 101: 173-85.  
10.1016/S0092-8674 (00) 80828-4.
- (7) Aramini, J. M., Cleaver, S. H., Pon, R. T., Cunningham, R. P., and Germann, M. W. "Solution structure of a DNA duplex containing an alpha-anomeric adenosine: insights into substrate recognition by endonuclease". *J. Mol. Biol.*, 2004, 338, 77-91.
- (8) Hazuda, D.J., Wolfe, A.L., Hastings, J.C., Robbins, H.L., Graham, P.L., LaFemina, R.L., Emini, E.A. "Viral Long Terminal Repeat Substrate Binding Characteristics of the Human Immunodeficiency Virus Type 1 Integrase". *J. Biol. Chem.*, 11 Feb. 1994, 3999-4004



- (9) Jenkins, T. M., et al. “Critical Contacts between HIV-1 Integrase and Viral DNA Identified by Structure-Based Analysis and Photo-Crosslinking”. *The EMBO Journal*, EMBO Press, 15 Nov. 1997, [emboj.embopress.org/content/16/22/6849](http://emboj.embopress.org/content/16/22/6849).
- (10) Mazumder, A., and Pommier, Y. “Processing of Deoxyuridine Mismatches and Abasic Sites by Human Immunodeficiency Virus Type-1 Integrase”. *Nucleic Acids Research* 23.15 (1995): 2865–2871.
- (11) Mao, L., Wang, Y., Liu, Y., Hu, X. “Intermolecular Interaction Modes of Positively Charged Residues with Adenine in ATP-Binding Proteins”. *Journal of the American Chemical Society* 2003 125 (47), 14216-14217. DOI: 10.1021/ja036096p
- (12) Scottoline, B.P., Chow, S., Ellison, V., Brown, P.O. “Disruption of the Terminal Base Pairs of Retroviral DNA During Integration”. *Genes Dev.* 1 Feb. 1997; 11(3): 371-382
- (13) Sebastien, N., Saib, A. “Early steps of retrovirus replicative cycle”. *Retrovirology*. 14 May 2004; 1:9. Doi:10.1186/1742-4690-1-9
- (14) Centers for Disease Control and Prevention. HIV/AIDS. 23 July, 2018.  
<https://www.cdc.gov/hiv/basics/index.html>

# Substituting Inosine for Guanosine in DNA; Structural and Dynamic Consequences

## ABSTRACT

Inosine differs from guanosine only by the absence of the N2 amino group. In addition, both bases have similar electrostatic potentials. Therefore, substituting I for G has been used to probe various properties of nucleic acids and to facilitate the interpretation of binding studies. In particular, the absence of the N2 amino group permits the assessment of its importance in the binding of ligands to the minor groove of duplex DNA. It has been known for some time that an I-C base pair is of lower stability than a regular G-C base pair, which needs to be considered when making DNA constructs containing inosine. However, it is generally assumed that both base pairs are structurally highly similar. In order to test this assumption, we have determined the fine structure of two hairpin DNA substrates that differ only in the substitution of an I-C base pair for a G-C base pair. The structures have been solved using NOESY-NMR data in conjunction with molecular dynamics. The structural data will be compared and complemented with thermodynamic and dynamic information to get a more comprehensive appraisal of G-C vs. I-C base pair substitutions.

## 1 INTRODUCTION

Inosine is a nucleotide with a hypoxanthine base. It is chemically identical to guanosine with only one exception. It does not have an amino group attached to C2. This allows for the comparison of duplex DNA structures that differ only by the lack of an N2 group in the minor groove. By studying substrate binding behavior when the amino group is both present and not present, evidence for the importance of the amino group and for potential binding strategies via inosine substitutions can be found.

This does not however, take into account any structural or dynamic changes that might occur when an inosine is substituted for guanosine. So before this strategy can be used the broader effects of an inosine substitution must be studied. This project compares two 18-mer DNA hairpins that differ only in the substitution of an inosine for a guanosine. The structure for each hairpin was solved from NOESY spectra using Sparky, AMBER, MARDIGRAS, and CORMA software. Final structures were then analyzed visually using VMD. RMSD values were obtained through the VMD RMSD trajectory tool and helical analysis was performed using Curves+. Each structure also underwent 1  $\mu$ s of free MD followed by RMSD calculations on the resulting trajectories. In addition, the base pair lifetimes,  $T_m$  values, imino proton spectra, and phosphorous spectra were obtained for each structure. The visual and analytical structure comparisons were then paired with the dynamic and thermodynamic data in order to get a more complete picture of the effects of the G $\rightarrow$ I substitution.



**Figure 20: Hairpin Sequences**

*The complete hairpin sequences differ only by the I/G substitution at position 4 shown in red. The G and the I differ only by the N2 amino group present on the G.*

## 2 MATERIALS AND METHODS

### 2.1 – Using NMR to Obtain NOESY Distance Restraints

All NMR experiments were performed using a Bruker Avance 600 spectrometer with IDTG triple resonance and QXI probe heads. NMR samples were loaded in D<sub>2</sub>O with 10 mM NaP, 50 mM NaCl, 0.1 mM EDTA, and had a pH\* of 6.86. The temperature was set to 298K. Distance restraints were determined using NOESY data, which was collected using mixing times of 75 ms, 150 ms, and 250 ms with 50 mM samples. The NOESY data was then used to compile a list of cross peak volumes using Sparky software. This was done by first assigning the peaks manually and then integrating the peaks using Gaussian fit and sum over box integration methods. Distance restraints were then computed from the cross-peak volumes using MARDIGRAS software. NOESY experiments and cross peak assignments were performed by Qiushi Li.

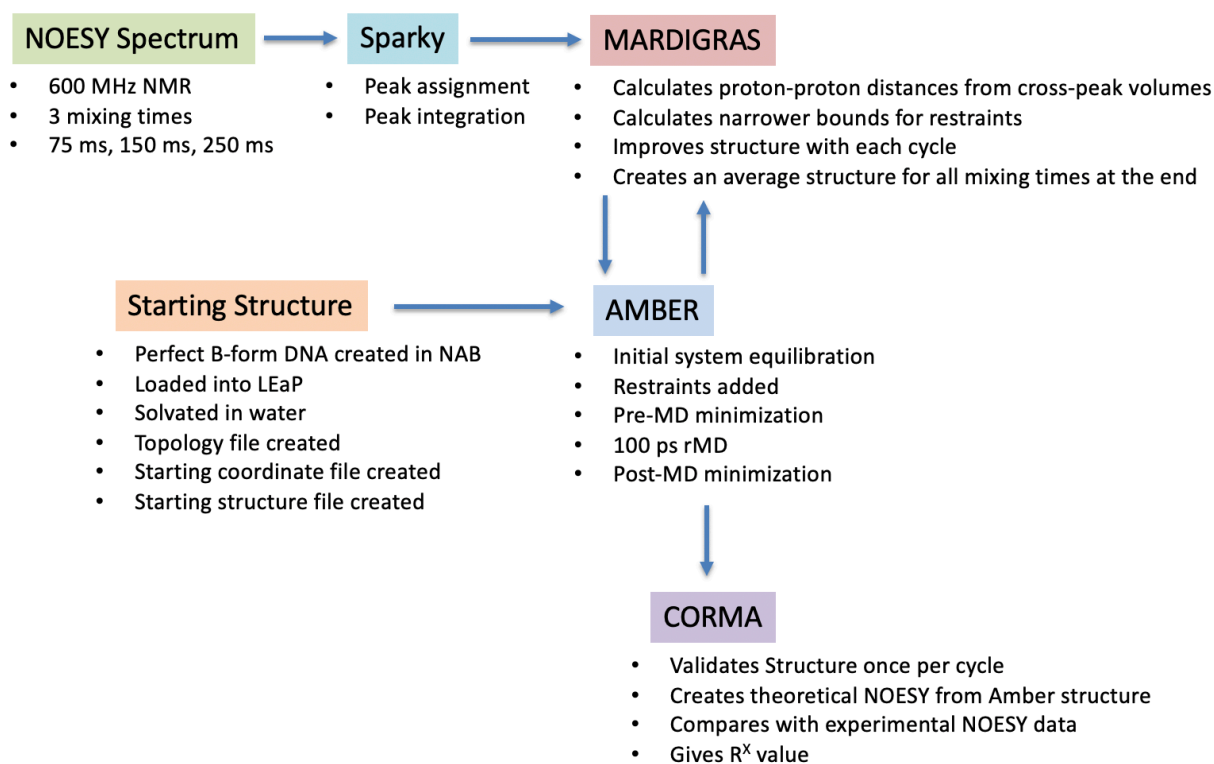
### 2.2 –Initial Structure and Equilibration

A perfect B form version of the hairpin structure was created using NAB in AMBER. This was used with LEaP in AMBER to create a topology file, a starting coordinate file, and a starting structure file. This was done using the PARMBSC0 force field and a TIP3P water model in a 10 Å truncated octahedral periodic box. A total of 3,452 solvent molecules were present in the box. The DNA was then held rigid while the solvent and counter-ions were minimized. The entire system including the DNA was then minimized.

## 2.3 – Implementation of NOESY Distance Restraints

Before the NOESY distance restraints were added, the system was first restrained with broad and qualitative restraints such as Watson-Crick base pairing restraints, broad backbone torsion angles, and qualitative restraints for the hairpin loop. Because these restraints are all fairly broad they were all implemented at the same time. All non-NOESY restraints used are shown in the appendix. NOESY distance restraints were added after this in small batches starting with the most intense and least ambiguous peaks.

## 2.4 – The AMBER/MARDIGRAS/CORMA Cycle



### Figure 21: Solving the Structures

*Here a diagram of the overall process used to solve each structure from NMR NOESY data is given.*

As a preface to this section a basic description of the function of each of these programs will be given. AMBER (Assisted Model Building with Energy Refinement) is a software suite that includes many tools used for running molecular dynamics on nucleic acid and protein systems. It includes the MD engine used here called Sander. MARDIGRAS (Matrix Analysis of Relaxation for Discerning the Geometry of an Aqueous Structure) calculates proton-proton distances and error-bounds from cross-peak intensities produced by a 2D NOE experiment. The distance bounds produced can then be implemented as distance restraints. MARDIGRAS also narrows existing restraints using an AMBER output structure file and an intensity file. CORMA (Complete Relaxation Matrix Analysis) is a program used to calculate a dipole-dipole relaxation matrix for a group of protons. It translates this into intensities expected of a 2D NOE experiment.

Initial intensities were converted into distance restraints using MARDIGRAS. For each batch of restraints, the system went through at least one cycle of AMBER, MARDIGRAS, and CORMA. In AMBER, the structure underwent an initial minimization, a 100 ps rMD step, and a final minimization. The resulting output file was converted to a pdb and used for both CORMA and MARDIGRAS. MARDIGRAS uses the structure file to produce narrower bounds for the restraints. This allows for the improvement of the structure after each cycle, and because it uses an iterative process called RANDMARDI it is able to account for spin diffusion that would normally affect longer mixing times (such as 250 ms here). The resulting output file from MARDIGRAS was then used as the initial structure for the next AMBER cycle. CORMA was used to validate the structure once per cycle. It creates a theoretical NOESY intensity file based on your AMBER structure and then compares it to your experimental data. Structures in this project were validated using CORMA-derived  $R^x$  values with a correlation time of 3.5 ns for

base and sugar protons. This correlation time was chosen because it typically produces the lowest  $R^x$  values (closest agreement with experimental data) <sup>(2)</sup>.

This cycle was repeated for data obtained using mixing times of 75 ms, 150 ms, and 250 ms. In each case the cycle was repeated until all restraints were incorporated. If individual restraints caused abnormally large distance or torsion penalties in AMBER or caused a spike in the  $R^x$  value then they were investigated in Sparky. If the peak in question was found to be abnormal in some way such as hiding another peak inside of it, having a particularly bizarre shape, or overlapping too heavily with another peak then it was either turned into a qualitative restraint or removed entirely depending on the severity of the abnormality.

## **2.5 – Obtaining a Final Structure**

After the structure for each mixing time was solved independently the restraints were combined and used to create an average set of restraints. This was done using MARDIGRAS and AMBER. MARDIGRAS can combine a series of bounds files (one from each mixing time) to create an average bounds file. This was then implemented in AMBER using the 250 ms (NOESY mixing time) structure file to create an average structure. The average structure then underwent additional AMBER/MARDIGRAS/CORMA cycles. The final cycle performed ran rMD for 10 ns in order to allow for any changes that might occur on a longer time scale. Using cpptraj in AMBER, 10 structures were taken from the final 100 ps (one structure every 10 ps) after the rMD. These structures were minimized with restraints and run through CORMA. A final structure was chosen from these based on their total AMBER distance penalties and CORMA  $R^x$  values.

## 2.6 – Helical Analysis

Helical analysis was performed on each final structure using Curves+. The output file will not include base opening parameters if the hairpins are run as a single strand of DNA. So, the hairpin loop was first removed from each structure file and the residue names were made contiguous so that Curves+ would read the nucleotides in the correct order. An example input file is shown in Figure A2 in the appendix.

## 2.7 – RMSD

Trajectory RMSD Calculations were done for each final trajectory. In each case the final trajectories (10 ns) were loaded into VMD. The RMSD trajectory tool was then used to perform RMSD calculations with an atom selection that specified residues 1-7 and 12-18 (the whole structure minus the hairpin loop). The noh selection modifier was used which excludes hydrogens from the atom selection. The calculation used every frame along the entire trajectory. This was repeated using only residues 3, 4, 5, 14, 15, and 16 and also one with only residues 4 and 15. Structures were aligned according to the atom selection in each case prior to RMSD calculations.

Comparative RMSD calculations were also done for the two final (static) structures. One calculation specified atoms in residues 3, 4, 5, 14, 15, and 16 while another specified only residues 4 and 15. Keep in mind that residue 4 is the I/G substitution and residue 15 is the cytosine base pairing with it. The numbering for the entire structure is shown in Figure 1. For each calculation the structures were first aligned according to the residues specified for the RMSD calculation. For comparative RMSD calculations the atoms must match exactly, so for



residue 4 the H2 on the inosine and the N2 and its hydrogens on the guanosine were purposefully excluded from the atom selection.

## 2.8 – Free MD

Each structure also underwent free MD. In each case the structure and solvent were equilibrated as described in section 2.2 followed by a 1  $\mu$ s fMD (no restraints). The resulting trajectories were loaded into VMD and trajectory RMSD calculations were performed as described in section 2.7.

## 2.9 – Base Pair Opening

The base pair lifetimes for the I and G hairpins were measured by monitoring the  $T_1$  relaxation times of samples containing an ammonium catalyst at different concentrations. The relaxation time was determined by measuring the intensity of the peaks with different delay times and curve fits. This was done at both 278 K and 293 K. Once the  $T_1$  was determined, the formula  $1/T_1 = 1/\tau_{\text{ex}} + 1/T_{1^\circ}$  was used to find the base pair exchange rate,  $\tau_{\text{ex}}$ .  $T_{1^\circ}$  represents  $T_1$  without any added catalyst.

### 3 RESULTS AND DISCUSSION

#### 3.1 – AMBER Distance Penalties and CORMA Scores for the Final 10 Structures

I-HP Structure #	AMBER	CORMA		
	Total Distance Penalty (kcal/mol)	Intra-residue R <sup>x</sup> Value (%)	Inter-residue R <sup>x</sup> Value (%)	Total R <sup>x</sup> Value (%)
1	77.662	3.85	6.02	4.64
2	79.710	3.85	6.27	4.73
3	73.549	4.16	6.48	5.01
4	74.778	3.90	6.02	4.67
5	77.988	4.02	6.18	4.81
6	74.461	3.83	5.97	4.61
7	72.854	4.05	6.52	4.95
8	77.245	3.87	6.28	4.75
9	75.035	3.52	5.82	4.36
10	74.239	3.75	6.08	4.60

**Figure 22: Final 10 Structures for the I-HP**

*Structure 9 was selected as the final structure due to having the lowest CORMA score and a normal AMBER distance penalty. Total AMBER torsion penalties were less than 2.1 kcal/mol and total angle penalties were less than 1.0 kcal/mol.*

<b>G-HP</b>	<b>AMBER</b>	<b>CORMA</b>		
<b>Structure #</b>	<b>Total Distance Penalty (kcal/mol)</b>	<b>Intra-residue R<sup>x</sup> Value (%)</b>	<b>Inter-residue R<sup>x</sup> Value (%)</b>	<b>Total R<sup>x</sup> Value (%)</b>
<b>1</b>	<b>64.108</b>	<b>4.00</b>	<b>6.04</b>	<b>4.73</b>
<b>2</b>	<b>66.790</b>	<b>4.05</b>	<b>6.23</b>	<b>4.82</b>
<b>3</b>	<b>64.665</b>	<b>4.17</b>	<b>5.87</b>	<b>4.78</b>
<b>4</b>	<b>66.525</b>	<b>4.07</b>	<b>6.00</b>	<b>4.75</b>
<b>5</b>	<b>64.926</b>	<b>4.05</b>	<b>6.29</b>	<b>4.85</b>
<b>6</b>	<b>64.326</b>	<b>4.10</b>	<b>6.06</b>	<b>4.80</b>
<b>7</b>	<b>64.558</b>	<b>4.10</b>	<b>6.06</b>	<b>4.80</b>
<b>8</b>	<b>66.296</b>	<b>4.10</b>	<b>6.06</b>	<b>4.80</b>
<b>9</b>	<b>65.115</b>	<b>4.10</b>	<b>6.06</b>	<b>4.80</b>
<b>10</b>	<b>65.030</b>	<b>3.97</b>	<b>5.84</b>	<b>4.63</b>

**Figure 23: Final 10 Structures for the G-HP**

*Structure 10 was selected as the final structure due to having the lowest CORMA score and a normal AMBER distance penalty. Total AMBER torsion penalties were less than 1.5 kcal/mol and total angle penalties were less than 1.0 kcal/mol.*

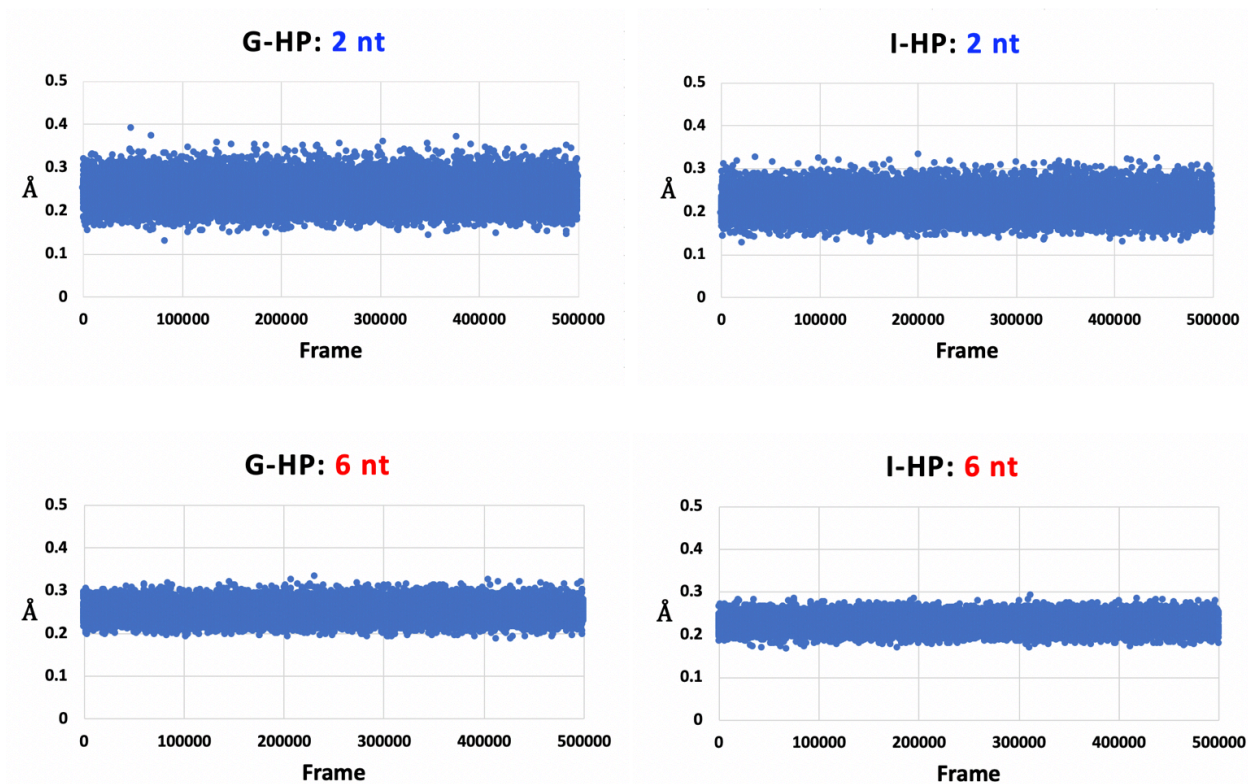
### 3.2 – RMSD Calculations

MD Type	Atom Selection	Structure	Average (Å)	S.D. (Å)	Min (Å)	Max (Å)
fMD (1 $\mu$ s)	2 nt	G-HP	0.24	0.03	0.13	0.40
		I-HP	0.22	0.03	0.12	0.38
	6 nt	G-HP	0.25	0.02	0.18	0.35
		I-HP	0.22	0.02	0.16	0.31
	Full (no loop)	G-HP	0.26	0.01	0.20	0.32
		I-HP	0.24	0.01	0.19	0.30



**Figure 24: Trajectory RMSD Values**

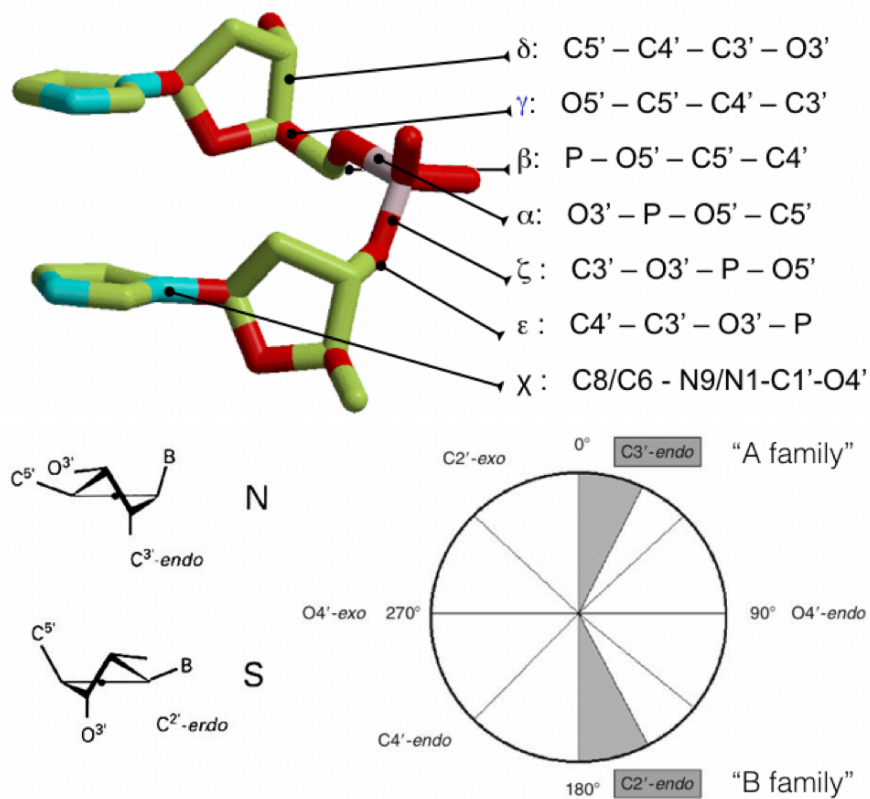
RMSD values were calculated for each structure along the entire free MD trajectory. Structures were aligned prior to RMSD calculations according to the atom selection shown. 2 nt RMSD values were calculated only using residues 4 and 15 for alignment and RMSD calculations. 6 nt values used residues 3, 4, 5, 14, 15, and 16. Results show little difference between the structures with similar averages, low standard deviations, and low maximum values.



**Figure 25: RMSD Values Shown Graphically**

Here the trajectory RMSD values are shown with respect to time, indicated by the frame number which progress chronologically. Each frame represents a snapshot of the structure during the trajectory. The band widths between the graphs (compare left to right) are very similar.

### 3.3 – Helical Analysis



**Figure 26: Backbone angle and sugar pucker definitions**

Picture from the Curves+ web server: <http://curvesplus.bsc.es/bbpar>

**$\Delta$ Value  
|G-I|**

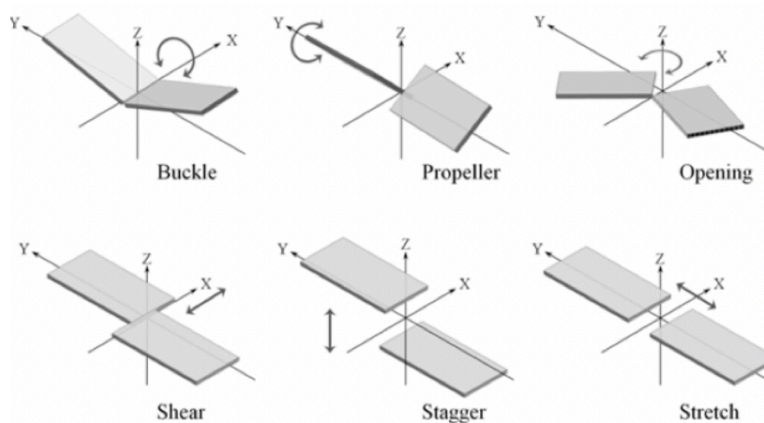
Residues	Alpha (°)	Beta (°)	Gamma (°)	Delta (°)	Epsilon (°)	Zeta (°)	Pucker (G/I)
A1	----	----	1.1	1.3	6.7	3.0	C2' en/C2' en
C2	12.2	13.6	12.9	5.3	3.4	2.9	C1' ex/C1' ex
T3	14.0	8.3	6.3	23.2	5.4	12.8	C2' en/C1' ex
<b>I/G4</b>	7.2	3.4	2.3	21.9	8.3	46.3	C2' en/C1' ex
C5	14.9	10.6	2.6	15.8	10.1	8.2	O1' en/C1' ex
T6	7.6	9.8	2.4	4.0	2.4	8.7	C1' ex/C1' ex
A7	12.8	1.0	9.5	15.4	----	----	C1' ex/C2' en
T12	----	----	12.1	14.8	5.6	22.3	C1' ex/C4' ex
A13	11.1	4.5	10.4	16.1	0.6	5.6	C2' en/C1' ex
G14	0.8	0.9	17.2	6.1	3.7	8.7	C1' ex/C1' ex
C15	3.4	5.1	0.6	19.1	1.5	2.6	C1' ex/C1' ex
A16	7.1	4.4	4.2	9.8	3.0	7.2	C1' ex/C1' ex
G17	0.3	2.0	6.6	6.7	10.5	4.9	C1' ex/C1' ex
T18	5.6	13.1	0.3	21.1	----	----	O1' en/C2' en

**Figure 27: Helical Analysis: Difference in Backbone Parameters**

*Values over 20° are highlighted in yellow. The I/G substitution is highlighted in red.*

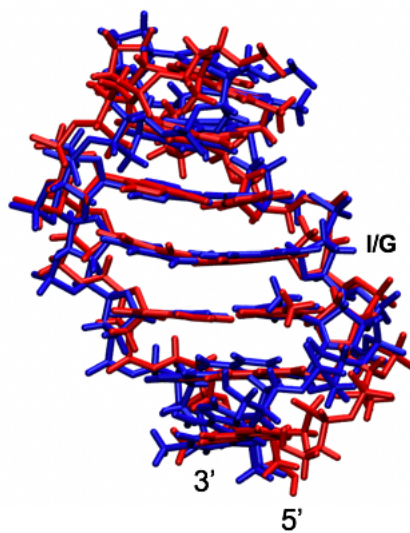
**$\Delta$ Value  
|G-I|**

Residues	Translational Parameters			Rotational Parameters		
	Shear (Å)	Stretch (Å)	Stagger (Å)	Buckle (°)	Propel (°)	Opening (°)
A1-T18	0.18	0.07	0.16	2.6	14.4	12.3
C2-G17	0.30	0.07	0.34	3.1	4.4	3.7
T3-A16	0.36	0.06	0.25	2.5	7.4	15.0
<b>I/G4-C15</b>	<b>1.12</b>	0.16	0.08	9.8	9.9	9.5
C5-G14	0.31	0.06	0.48	10.9	4.9	7.5
T6-A13	0.68	0.18	0.54	1.4	9.7	17.6
A7-T12	0.53	0.03	0.08	<b>18.1</b>	4.3	10.6
<b>Average:</b>	0.50	0.09	0.28	6.9	7.9	10.9



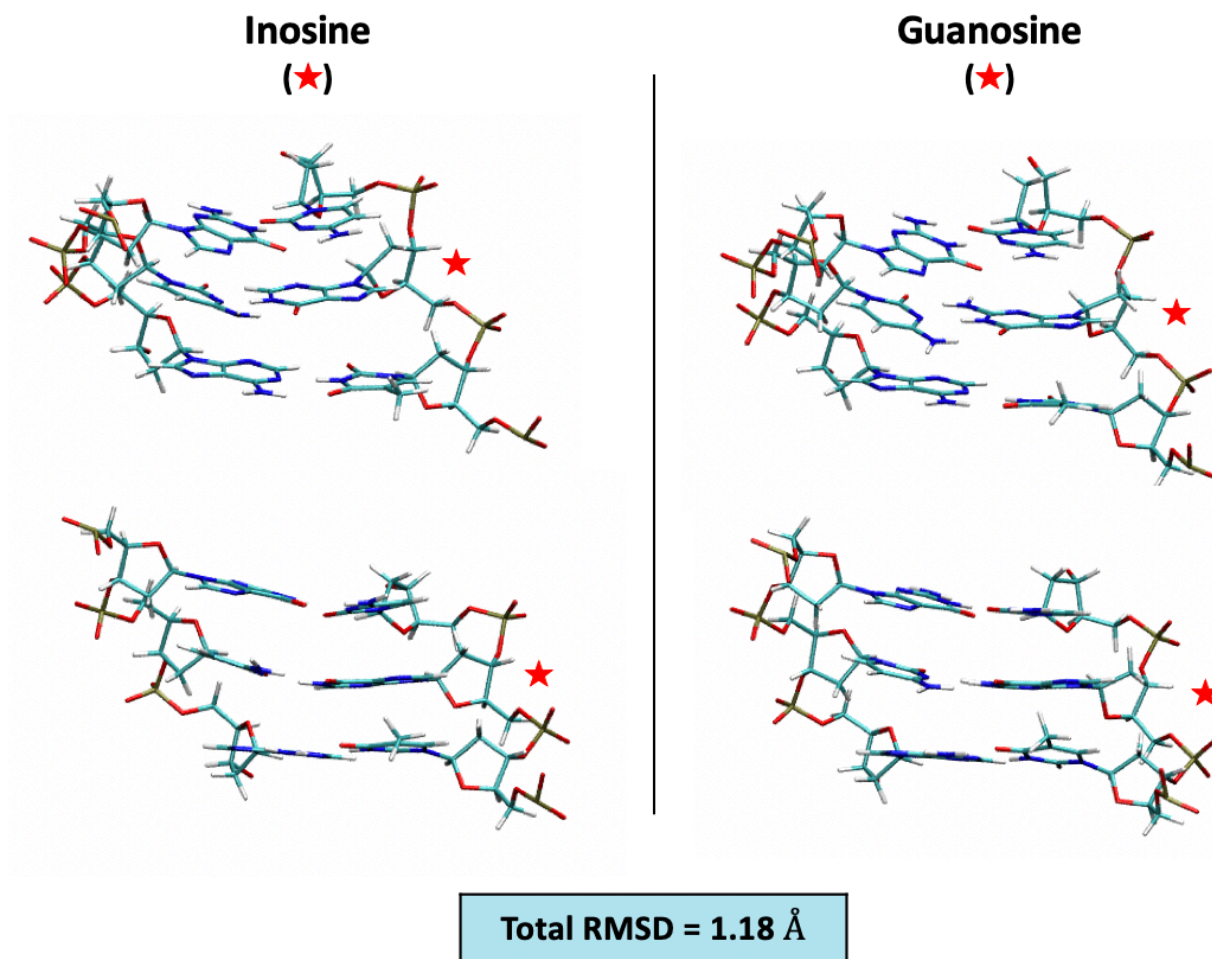
**Figure 28: Helical Analysis: Difference in Intra-BP Parameters**  
*Values more than twice the average value were highlighted in yellow. The I/G substitution is shown in red.*





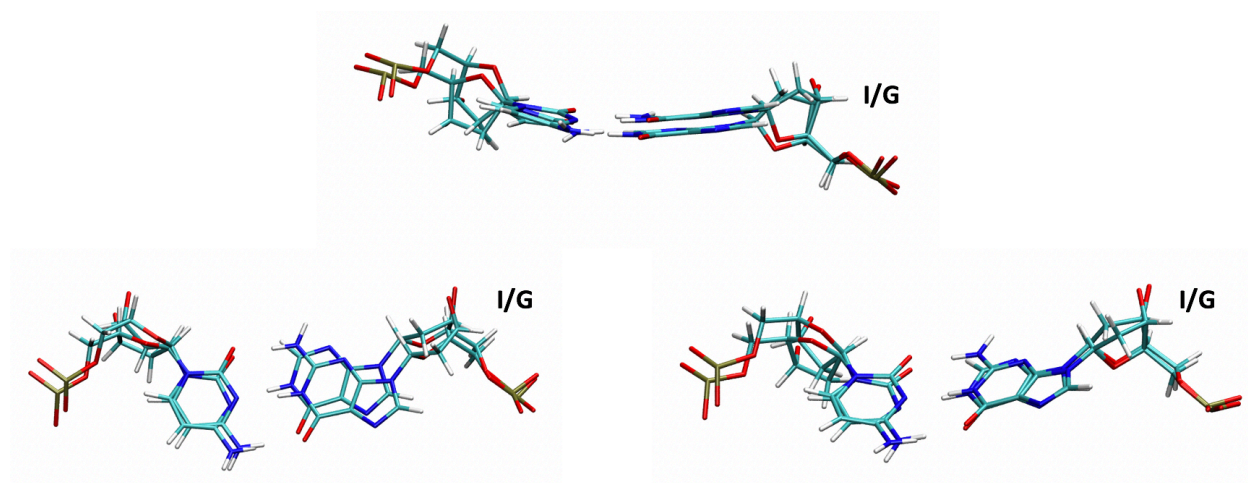
**Figure 29: Aligned Full Structure**

*Here the 2 final structures are shown minus the hairpin loop. The blue structure is the inosine hairpin and the red structure is the guanosine hairpin. The model has been rotated so the area around the substitution is most visible. The two structures were aligned using the RMSD calculator in VMD and were aligned by residues 3, 4, 5, 14, 15, and 16 (those in the area of the substitution).*



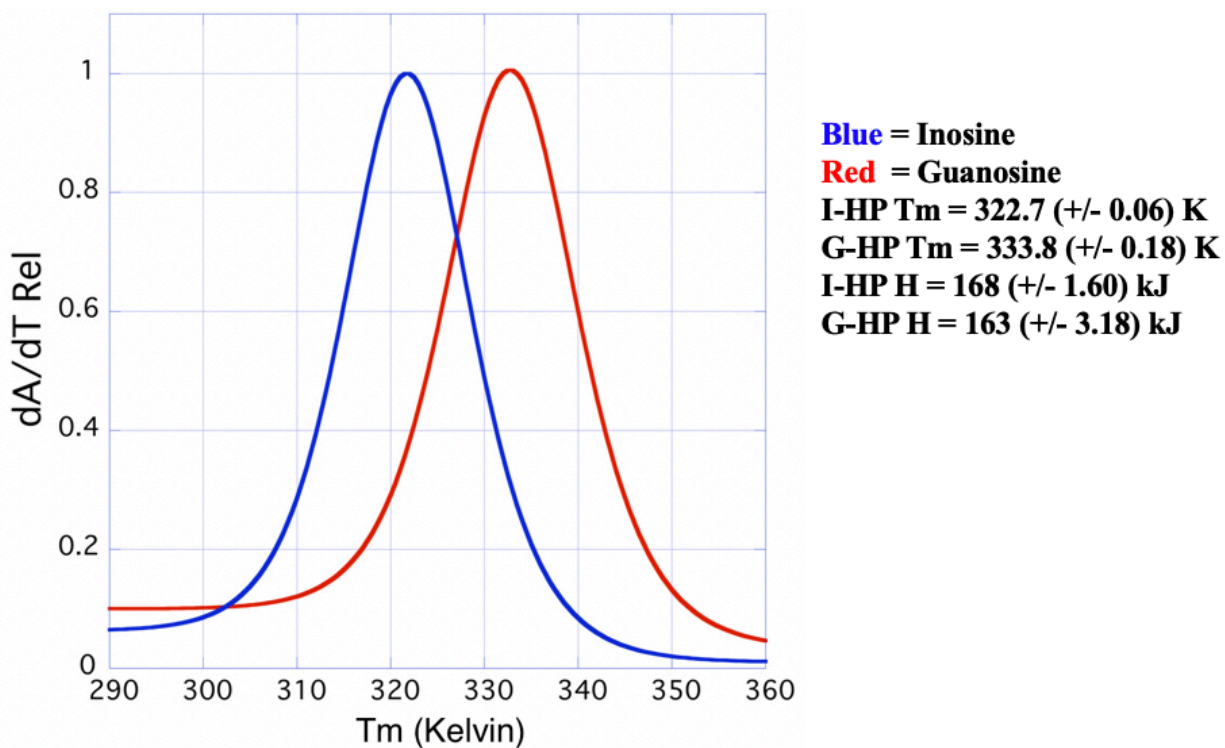
**Figure 30: Side by Side Comparison of I/G Local Effects**

*Here the IC and GC nucleotides are shown with their neighboring nucleotide pairs in the final structures. No major differences were observed in the position or orientation of the nucleotides. The comparative RMSD value was found using the two final structures. Residues 3, 4, 5, 14, 15, and 16 (shown) were used for alignment and RMSD calculations.*



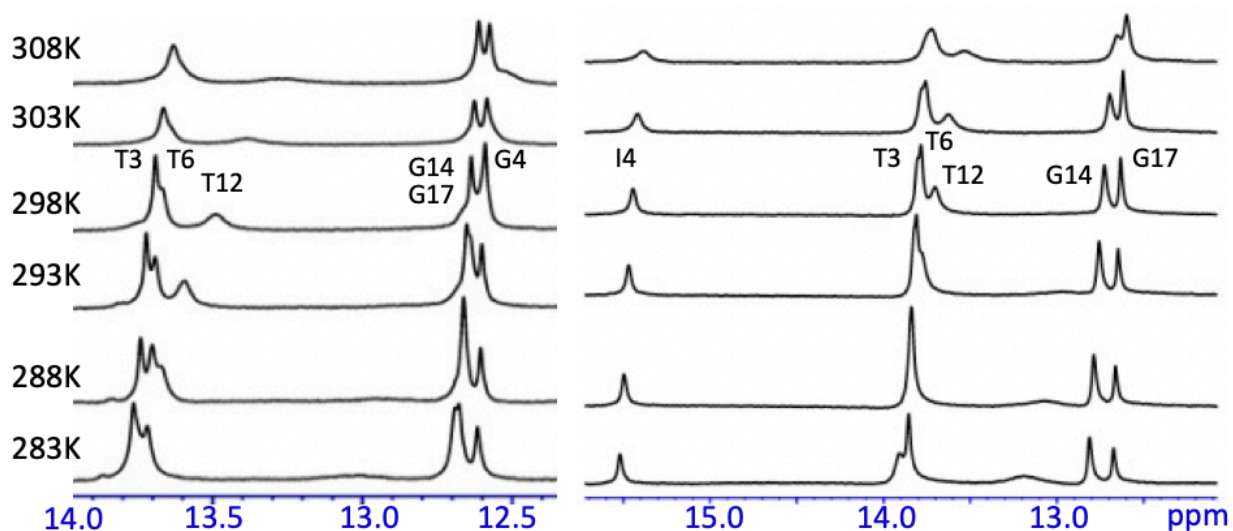
**Figure 31: Aligned I/G Nucleotides**

*The IC and GC base pairs were aligned using the RMSD calculator in VMD and were aligned by residues 3, 4, 5, 14, 15, and 16 (those in the area of the substitution).*



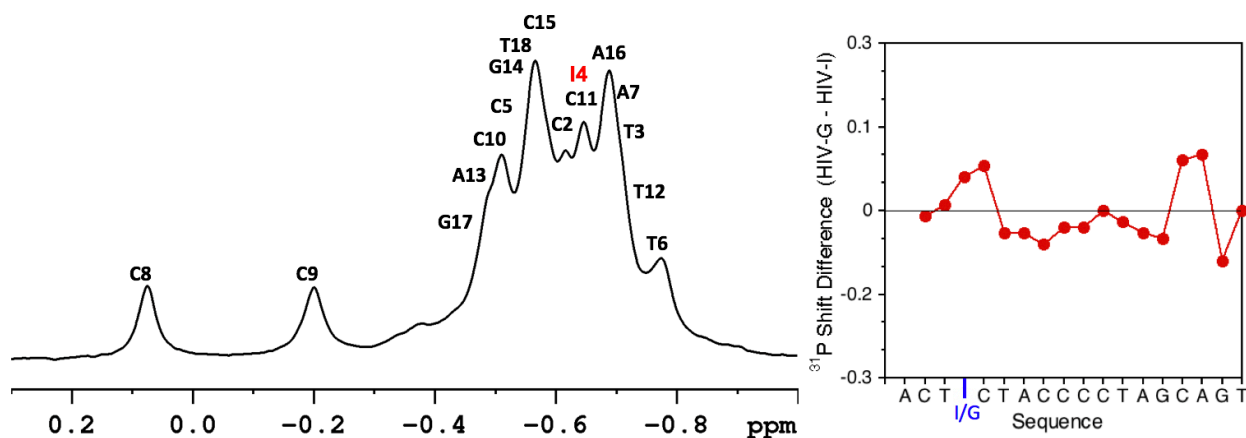
### Figure 32: $T_m$ Values and Enthalpy

The inosine structure is shown in blue and the guanosine structure is shown in red. This graph shows the derivative of the absorbance with respect to temperature. The apex of the peaks represents the  $T_m$  of each structure and the width at half height is proportional to the enthalpy.  $T_m$  data was derived from UV melting curves at different DNA concentrations and was calculated using a 6-parameter fit. Figure adapted from data collected by Qiushi Li<sup>(15)</sup>.



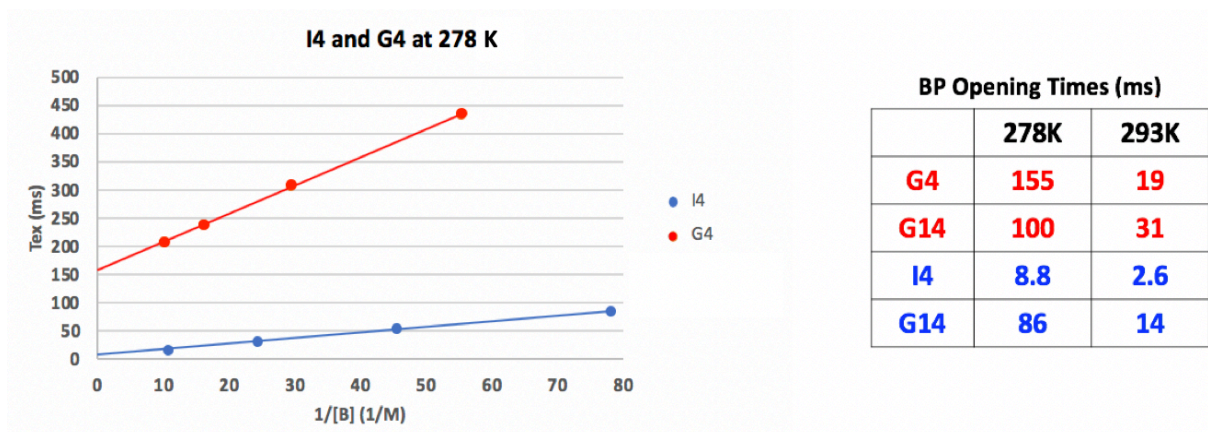
**Figure 33: Imino Proton Spectra**

The imino proton spectrum for the guanosine HP is shown on the left and the inosine HP is shown on the right. The peaks shown represent imino proton peaks which typically result from the formation of a stable base pair formation suggesting that the nucleotides labeled above are in a stable base pair. Figure by Qiushi Li and Marina Evich <sup>(15)</sup>.



**Figure 34: Phosphorous Spectra**

The phosphorous spectrum of the inosine hairpin is shown on the left and the differences in chemical shifts between the two hairpins is shown on the right. None of the chemical shift differences were particularly large. The greatest differences were in the immediate vicinity of the I4/G4 substitution site. This would suggest that there is no major change in



the backbone of the structure caused by the substitution and that the torsions in the backbone are relatively normal. Figure by C. Johnson<sup>(14)</sup>.

### Figure 35: Base Pair Lifetimes

Base pair lifetimes for the I4/G4 and the G14 of each structure. “B” on the x-axis represents the ammonia catalyst and  $\tau_{ex}$  represents the base pair lifetime on the y-axis. The base pair lifetimes were slower for the G than the I. This makes sense since the I only has 2 hydrogen bonds connecting it to the C and the G has 3 hydrogen bonds. The bp opening rate of G14 is faster when an I is present suggesting a minor destabilization caused by the presence of the I. Figure by Qiushi Li and Marina Evich<sup>(15)</sup>.

### 3.4 - Discussion:

The final structures each achieved a total CORMA  $R^x$  value of less than 5% indicating a strong agreement with the NMR data. AMBER penalties were negligible for torsions and angles. The distance penalty for the I-HP totaled 75.0 kcal/mol with no individual penalty greater than 3.5 kcal/mol. The G-HP had a total distance penalty 65.0 kcal/mol with no individual penalty greater than 6.3 kcal/mol. Based on this data it is then reasonable to assume that the final 2 structures are good representations of the actual structures.

The trajectory-based RMSD values shown in Figure 24 and 25 show that the structures are very consistent throughout the trajectory with average values below 0.3 Å and standard deviations below 0.05 Å. The maximum values reached for each structure are fairly close to the average suggesting there is no point during the trajectory that the structure changes drastically from the final structure. The results of the free MD show that those two structures are nearly identical.

Helical analysis was done using Curves+ software on the final hairpin structures. The results are shown in Figure 27 and 28. The highlighted values for A7 and T12 are likely due to their proximity to the hairpin loop. It is interesting however that the sheer distance and two of the backbone angles (zeta and delta) are noticeably different for residue 4, the I/G substitution. This could be indicative of slightly different dynamics or stability.

The visual structural comparisons between the two hairpins in Figure 29, 30, and 31 show very little difference in position or orientation of the bases, sugars, or backbone components. Any broad structural changes would really stand out here so despite the lack of the N2 amino group on the inosine, and therefore the third hydrogen bond, it is likely that structures of the two

hairpins do not differ in any fundamental way. This missing hydrogen bond could also explain the lower  $T_m$ . With fewer bonds between the I and the C, it would make sense then that the overall stability of the base pair would be lower. This was further investigated by looking at the base pair lifetimes (Figure 35).

Base pair lifetimes allow us to study the dynamics of individual base pairs on a millisecond timescale. By introducing a catalyst for the exchange of the imino proton with water, a reduction in signal intensity can be seen for base pairs that are farther apart. This is because base pairs achieve an open state before exchanging protons with water. The catalyst is necessary because the rate limiting step of this mechanism is normally the proton exchange itself. The catalyst introduces a proton acceptor that accelerates the proton exchange to conditions that are at or close to opening-limited conditions. The results shown in Figure 35 show a great difference between the rate at which the I4 and G4 are opening. The I4 is able to open and close much faster than the G4 at both temperatures. This agrees with the idea that the stability is lower for the IC than the GC. The base pair lifetimes are also shown for the G14 in each hairpin to give an idea of how the substitution affects the adjacent nucleotides. The G14 at 278/298 K has a base pair lifetime of 86/14 ms in the I-HP and 100/31 ms in the G-HP. This difference suggests that the inosine substitution does have a destabilizing effect on the surrounding base pairs although it is significantly muted compared to the IC vs GC (14% decrease for G14 vs 94% decrease for I/G4 at 278 K).



#### 4 CONCLUSIONS

From the overlays and side by side comparisons of the final I and G hairpin structures it is clear that there is no major difference in nucleotide position or orientation of I4 and G4 and that the overall structure of the hairpins are nearly identical. In addition, the position and orientation of nucleotides across from and on either side of the I/G nucleotide seem unperturbed by the substitution despite the slight drop in base pair lifetime.

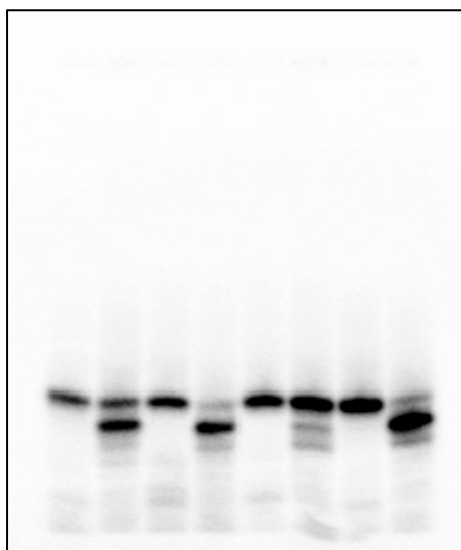
Although they are structurally similar, Figure 32 ( $T_m$ ) and 35 (bp lifetime) show a sizeable difference in the stability and dynamics of the I/G4. This is further supported by the difference in shear and backbone parameters for residue 4. This should be considered when using inosine as a substitute for guanosine. If these factors don't play a major role in a given experiment then all other indications given here would suggest that it is acceptable to make this substitution.

## REFERENCES

- (15) Pratip K. Bhattacharya, Julie Cha, Jacqueline K. Barton; “<sup>1</sup>H NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches”, *Nucleic Acids Research*, Volume 30, Issue 21, 1 November 2002, Pages 4740–4750, <https://doi.org/10.1093/nar/gkf601>
- (16) Christopher N. Johnson; “Nucleic Acid Substrates: Investigation of Structural and Dynamic Features that Influence Enzyme Activity”. Dissertation, Georgia State University, 2011. [http://scholarworks.gsu.edu/chemistry\\_diss/59](http://scholarworks.gsu.edu/chemistry_diss/59)
- (17) Qiushi Li; “Exploring HIV Integrase 3’-Processing Using Designated DNA Substrates and Structural Study of HIV DNA Hairpins”. Thesis, Georgia State University, 2016. [http://scholarworks.gsu.edu/chemistry\\_theses/96](http://scholarworks.gsu.edu/chemistry_theses/96)
- (18) Alexander M. Spring-Connell, Marina G. Evich, Markus W. Germann; “NMR Structure Determination for Oligonucleotides”. *Current Protocols in Nucleic Acid Chemistry*, 72, 7.28.1-7.28.39. 2018. <http://doi.org/10.1002/cpnc.48>
- (19) M. Gueron and J. Leroy; “Studies of base pair kinetics by NMR measurement of proton exchange”. *Methods Enzymology* 261, 383-413. 1995. [https://doi.org/10.1016/S0076-6879\(95\)61018-9](https://doi.org/10.1016/S0076-6879(95)61018-9)
- (20) M. Krepl, M. Otyepka, P. Banas, J. Sponer; “Effect of Guanine to Inosine Substitution on Stability of Canonical DNA and RNA Duplexes: Molecular Dynamics Thermodynamics Integration Study”. *The Journal of Physical Chemistry B*, 117(6), pp 1872-1879. 2013. <http://doi.org/10.1021/jp311180u>

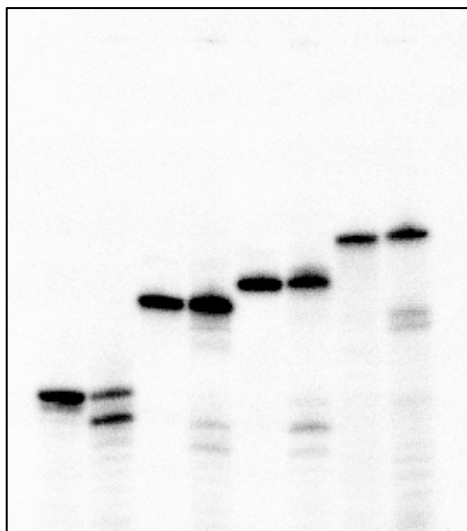
## APPENDIX

### Full Gel Examples

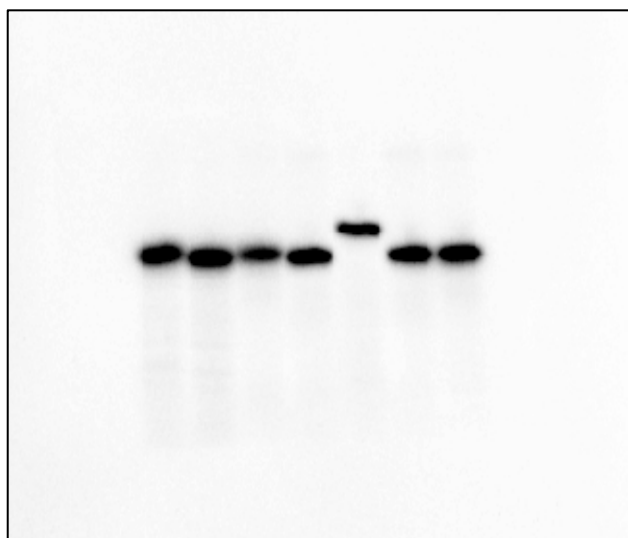


**Figure A1: Sample Gel 1**

*This is an 8M urea 15% PAGE gel. Each substrate contains a lane with no enzyme (left) and one with enzyme (right). From left to right this gel contains the WT, Deaza Bot w/ C, Deaza Top w/ C, and C Top w/ A substrates.*

**Figure A2: Sample Gel 2**

*This is an 8M urea 15% PAGE gel. Each substrate contains a lane with no enzyme (left) and one with enzyme (right). From left to right this gel contains the WT, 3T HP, 6T HP, and the Bulge substrates.*

**Figure A3: Sample Gel 3**

*This is an 8M urea 15% PAGE gel showing an example of a substrate check. These samples were run after the labeling process but before 3'-processing. It allows you to check the success of the labeling process, shows if anything besides the substrate is getting labeled, and lets you know if a particular substrate has poor labeling efficiency.*

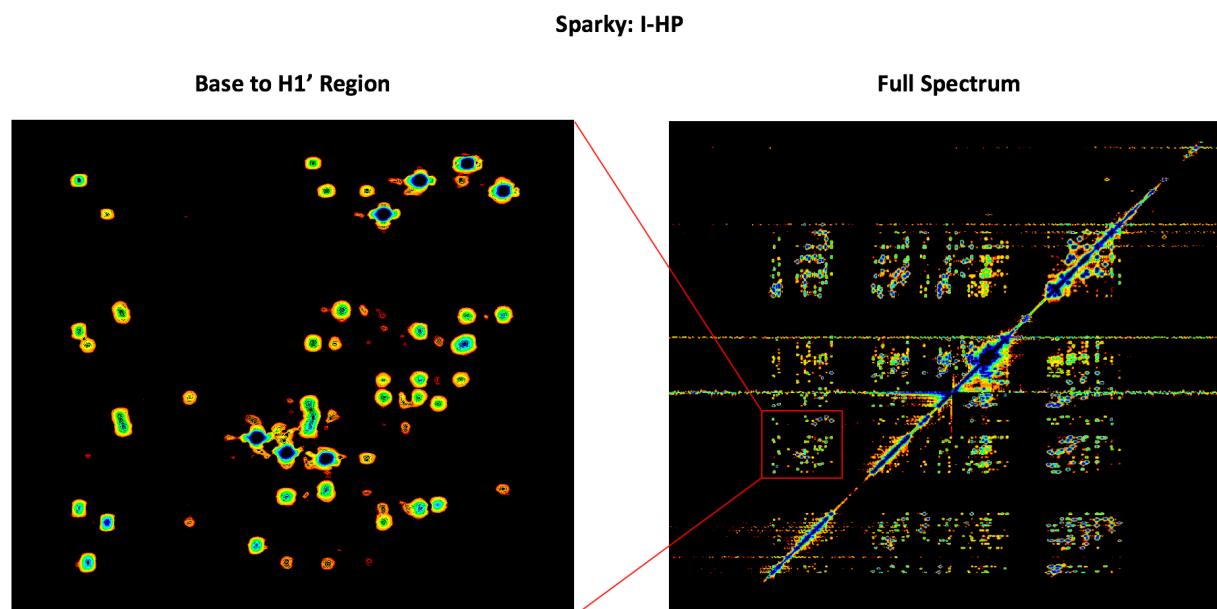


Figure A4: Sample NOESY Spectrum

```
/usr/local/curvesplus/Cur+ <<!  
&inp file=10nsI_ed_9.pdb, lis=ihp2, fit=.t.,  
lib=/usr/local/curvesplus/standard, &end  
2 1 -1 0 0  
1 2 3 4 5 6 7  
14 13 12 11 10 9 8  
0. 0. 0. 0.  
!  
█  
~  
~  
~
```

Figure A5: Sample Curves+ Input File

```
source "/usr/local/amber14/dat/leap/cmd/leaprc.ff14SB"
addAtomTypes {
  { "C1" "C" "sp2" }
  { "C2" "C" "sp2" }
  { "CI" "C" "sp3" }
  { "CE" "C" "sp3" }
}

#Load bsc1 params
loadoff parmBSC1.lib
loadamberparams parmBSC1.frcmod

#load inosine library, atom definitions and charges
loadamberprep ./Params-DNA-Inosine/di.in
in = loadamberparams ./Params-DNA-Inosine/frcmod-inosine

loadoff /usr/local/amber14/dat/leap/lib/atomic_ions.lib
mods2 = loadamberparams "/usr/local/amber14/dat/leap/parm/frcmod.ionsjc_tip3p"

mol = loadpdb HIV_i-hp.pdb
check mol
alignAxes mol

solvateoct mol TIP3PBOX 10.0
charge mol
addions mol Na+ 0
charge mol
saveamberparm mol HIVI_bsc1.prmtop HIVI_bsc1.crd
savepdb mol HIVI_bsc1.pdb
quit
█
~
~
~
```

**Figure A6: Sample LEaP Input File**

```
#Qualitative restraints for loop B-> H1' B->H2'1 B->H2'2
# distance constraint number 1
&rst iat = 8, 8, iresid = 1, atnam(1)='H1"',atnam(2)='H6',
    r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,
    rk2 =30.000, rk3 =30.000, &end
# distance constraint number 2
&rst iat = 9, 9, iresid = 1, atnam(1)='H1"',atnam(2)='H6',
    r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,
    rk2 =30.000, rk3 =30.000, &end
# distance constraint number 3
&rst iat = 10, 10, iresid = 1, atnam(1)='H1"',atnam(2)='H6',
    r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,
    rk2 =30.000, rk3 =30.000, &end
# distance constraint number 4
&rst iat = 11, 11, iresid = 1, atnam(1)='H1"',atnam(2)='H6',
    r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,
    rk2 =30.000, rk3 =30.000, &end
# distance constraint number 5
&rst iat = 8, 8, iresid = 1, atnam(1)='H2"1',atnam(2)='H6',
    r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,
    rk2 =30.000, rk3 =30.000, &end
# distance constraint number 5
&rst iat = 8, 8, iresid = 1, atnam(1)='H2"2',atnam(2)='H6',
    r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,
    rk2 =30.000, rk3 =30.000, &end
# distance constraint number 6
&rst iat = 9, 9, iresid = 1, atnam(1)='H2"1',atnam(2)='H6',
    r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,
```

```
rk2 =30.000, rk3 =30.000, &end

# distance constraint number 6

&rst iat = 9, 9, iresid = 1, atnam(1)='H2"2',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,

rk2 =30.000, rk3 =30.000, &end

# distance constraint number 7

&rst iat = 10, 10, iresid = 1, atnam(1)='H2"1',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,

rk2 =30.000, rk3 =30.000, &end

# distance constraint number 8

&rst iat = 10, 10, iresid = 1, atnam(1)='H2"2',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,

rk2 =30.000, rk3 =30.000, &end

# distance constraint number 9

&rst iat = 11, 11, iresid = 1, atnam(1)='H2"1',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,

rk2 =30.000, rk3 =30.000, &end

# distance constraint number 10

&rst iat = 11, 11, iresid = 1, atnam(1)='H2"2',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,

rk2 =30.000, rk3 =30.000, &end

# distance constraint number 11

&rst iat = 8, 9, iresid = 1, atnam(1)='H2"1',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,

rk2 =30.000, rk3 =30.000, &end

# distance constraint number 12

&rst iat = 8, 9, iresid = 1, atnam(1)='H2"2',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,
```



```
rk2 =30.000, rk3 =30.000, &end

# distance constraint number 13

&rst iat = 9, 10, iresid = 1, atnam(1)='H2"1',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,

rk2 =30.000, rk3 =30.000, &end

# distance constraint number 14

&rst iat = 9, 10, iresid = 1, atnam(1)='H2"2',atnam(2)='H6',

r1 = 1.5, r2 = 2.0, r3 = 5.0, r4 = 5.5,

rk2 =30.000, rk3 =30.000, &end

#=====

# 10935 atoms read from pdb file ../HIV-I-EZmin2.pdb. (Sugar Restraints)

# 1 ADE NU0: (1 DA5 C4')-(1 DA5 O4')-(1 DA5 C1')-(1 DA5 C2') -40.2 -10.2

&rst iat = 6, 8, 9, 27,

r1 = -41.2, r2 = -40.2, r3 = -10.2, r4 = -9.2,

rk2 = 50.0, rk3 = 50.0, &end

# 1 ADE NU1: (1 DA5 O4')-(1 DA5 C1')-(1 DA5 C2')-(1 DA5 C3') 18.1 48.1

&rst iat = 8, 9, 27, 25,

r1 = 17.1, r2 = 18.1, r3 = 48.1, r4 = 49.1,

rk2 = 50.0, rk3 = 50.0, &end

# 1 ADE NU2: (1 DA5 C1')-(1 DA5 C2')-(1 DA5 C3')-(1 DA5 C4') -38.5 -6.7

&rst iat = 9, 27, 25, 6,

r1 = -39.5, r2 = -38.5, r3 = -6.7, r4 = -5.7,

rk2 = 50.0, rk3 = 50.0, &end
```

# 1 ADE NU3: (1 DA5 C2')-(1 DA5 C3')-(1 DA5 C4')-(1 DA5 O4') -16.9 31.1

```
&rst iat = 27, 25, 6, 8,
      r1 = -17.9, r2 = -16.9, r3 = 31.1, r4 = 32.1,
      rk2 = 50.0, rk3 = 50.0,          &end
```

# 1 ADE NU4: (1 DA5 C3')-(1 DA5 C4')-(1 DA5 O4')-(1 DA5 C1') -11.8 34.0

```
&rst iat = 25, 6, 8, 9,
      r1 = -12.8, r2 = -11.8, r3 = 34.0, r4 = 35.0,
      rk2 = 50.0, rk3 = 50.0,          &end
```

# 2 CYT NU0: (2 DC C4')-(2 DC O4')-(2 DC C1')-(2 DC C2') -39.8 -9.8

```
&rst iat = 38, 40, 41, 57,
      r1 = -40.8, r2 = -39.8, r3 = -9.8, r4 = -8.8,
      rk2 = 50.0, rk3 = 50.0,          &end
```

# 2 CYT NU1: (2 DC O4')-(2 DC C1')-(2 DC C2')-(2 DC C3') 19.8 49.8

```
&rst iat = 40, 41, 57, 55,
      r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,
      rk2 = 50.0, rk3 = 50.0,          &end
```

# 2 CYT NU2: (2 DC C1')-(2 DC C2')-(2 DC C3')-(2 DC C4') -43.9 -13.9

```
&rst iat = 41, 57, 55, 38,
      r1 = -44.9, r2 = -43.9, r3 = -13.9, r4 = -12.9,
      rk2 = 50.0, rk3 = 50.0,          &end
```

# 2 CYT NU3: (2 DC C2')-(2 DC C3')-(2 DC C4')-(2 DC O4') -4.0 31.1

```
&rst iat = 57, 55, 38, 40,
      r1 = -5.0, r2 = -4.0, r3 = 31.1, r4 = 32.1,
```

rk2 = 50.0, rk3 = 50.0,                   &end

# 2 CYT NU4: (2 DC C3')-(2 DC C4')-(2 DC O4')-(2 DC C1') -11.8 25.8

&rst iat = 55, 38, 40, 41,

r1 = -12.8, r2 = -11.8, r3 = 25.8, r4 = 26.8,

rk2 = 50.0, rk3 = 50.0,                   &end

# 3 THY NU0: (3 DT C4')-(3 DT O4')-(3 DT C1')-(3 DT C2') -38.1 -8.1

&rst iat = 68, 70, 71, 89,

r1 = -39.1, r2 = -38.1, r3 = -8.1, r4 = -7.1,

rk2 = 50.0, rk3 = 50.0,                   &end

# 3 THY NU1: (3 DT O4')-(3 DT C1')-(3 DT C2')-(3 DT C3') 19.8 49.8

&rst iat = 70, 71, 89, 87,

r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,

rk2 = 50.0, rk3 = 50.0,                   &end

# 3 THY NU2: (3 DT C1')-(3 DT C2')-(3 DT C3')-(3 DT C4') -47.9 -17.9

&rst iat = 71, 89, 87, 68,

r1 = -48.9, r2 = -47.9, r3 = -17.9, r4 = -16.9,

rk2 = 50.0, rk3 = 50.0,                   &end

# 3 THY NU3: (3 DT C2')-(3 DT C3')-(3 DT C4')-(3 DT O4') 3.6 33.6

&rst iat = 89, 87, 68, 70,

r1 = 2.6, r2 = 3.6, r3 = 33.6, r4 = 34.6,

rk2 = 50.0, rk3 = 50.0,                   &end

# 3 THY NU4: (3 DT C3')-(3 DT C4')-(3 DT O4')-(3 DT C1') -12.2 17.8

```

&rst iat = 87, 68, 70, 71,
      r1 = -13.2, r2 = -12.2, r3 = 17.8, r4 = 18.8,
      rk2 = 50.0, rk3 = 50.0,          &end

# 4 DI NU0: (4 DI C4')-(4 DI O4')-(4 DI C1')-(4 DI C2') -39.8 -9.8
&rst iat = 100, 102, 103, 120,
      r1 = -40.8, r2 = -39.8, r3 = -9.8, r4 = -8.8,
      rk2 = 50.0, rk3 = 50.0,          &end

# 4 DI NU1: (4 DI O4')-(4 DI C1')-(4 DI C2')-(4 DI C3') 19.8 49.8
&rst iat = 102, 103, 120, 118,
      r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,
      rk2 = 50.0, rk3 = 50.0,          &end

# 4 DI NU2: (4 DI C1')-(4 DI C2')-(4 DI C3')-(4 DI C4') -43.9 -13.9
&rst iat = 103, 120, 118, 100,
      r1 = -44.9, r2 = -43.9, r3 = -13.9, r4 = -12.9,
      rk2 = 50.0, rk3 = 50.0,          &end

# 4 DI NU3: (4 DI C2')-(4 DI C3')-(4 DI C4')-(4 DI O4') -4.0 31.1
&rst iat = 120, 118, 100, 102,
      r1 = -5.0, r2 = -4.0, r3 = 31.1, r4 = 32.1,
      rk2 = 50.0, rk3 = 50.0,          &end

# 4 DI NU4: (4 DI C3')-(4 DI C4')-(4 DI O4')-(4 DI C1') -11.8 25.8
&rst iat = 118, 100, 102, 103,
      r1 = -12.8, r2 = -11.8, r3 = 25.8, r4 = 26.8,
      rk2 = 50.0, rk3 = 50.0,          &end

```

# 5 CYT NU0: (5 DC C4')-(5 DC O4')-(5 DC C1')-(5 DC C2') -39.8 -9.8

&rst iat = 131, 133, 134, 150,  
 r1 = -40.8, r2 = -39.8, r3 = -9.8, r4 = -8.8,  
 rk2 = 50.0, rk3 = 50.0, &end

# 5 CYT NU1: (5 DC O4')-(5 DC C1')-(5 DC C2')-(5 DC C3') 19.8 49.8

&rst iat = 133, 134, 150, 148,  
 r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,  
 rk2 = 50.0, rk3 = 50.0, &end

# 5 CYT NU2: (5 DC C1')-(5 DC C2')-(5 DC C3')-(5 DC C4') -43.9 -13.9

&rst iat = 134, 150, 148, 131,  
 r1 = -44.9, r2 = -43.9, r3 = -13.9, r4 = -12.9,  
 rk2 = 50.0, rk3 = 50.0, &end

# 5 CYT NU3: (5 DC C2')-(5 DC C3')-(5 DC C4')-(5 DC O4') -4.0 31.1

&rst iat = 150, 148, 131, 133,  
 r1 = -5.0, r2 = -4.0, r3 = 31.1, r4 = 32.1,  
 rk2 = 50.0, rk3 = 50.0, &end

# 5 CYT NU4: (5 DC C3')-(5 DC C4')-(5 DC O4')-(5 DC C1') -11.8 25.8

&rst iat = 148, 131, 133, 134,  
 r1 = -12.8, r2 = -11.8, r3 = 25.8, r4 = 26.8,  
 rk2 = 50.0, rk3 = 50.0, &end

# 6 THY NU0: (6 DT C4')-(6 DT O4')-(6 DT C1')-(6 DT C2') -38.1 -8.1

&rst iat = 161, 163, 164, 182,

r1 = -39.1, r2 = -38.1, r3 = -8.1, r4 = -7.1,  
rk2 = 50.0, rk3 = 50.0, &end

# 6 THY NU1: (6 DT O4')-(6 DT C1')-(6 DT C2')-(6 DT C3') 19.8 49.8

&rst iat = 163, 164, 182, 180,  
r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,  
rk2 = 50.0, rk3 = 50.0, &end

# 6 THY NU2: (6 DT C1')-(6 DT C2')-(6 DT C3')-(6 DT C4') -47.9 -17.9

&rst iat = 164, 182, 180, 161,  
r1 = -48.9, r2 = -47.9, r3 = -17.9, r4 = -16.9,  
rk2 = 50.0, rk3 = 50.0, &end

# 6 THY NU3: (6 DT C2')-(6 DT C3')-(6 DT C4')-(6 DT O4') 3.6 33.6

&rst iat = 182, 180, 161, 163,  
r1 = 2.6, r2 = 3.6, r3 = 33.6, r4 = 34.6,  
rk2 = 50.0, rk3 = 50.0, &end

# 6 THY NU4: (6 DT C3')-(6 DT C4')-(6 DT O4')-(6 DT C1') -12.2 17.8

&rst iat = 180, 161, 163, 164,  
r1 = -13.2, r2 = -12.2, r3 = 17.8, r4 = 18.8,  
rk2 = 50.0, rk3 = 50.0, &end

# 7 ADE NU0: (7 DA C4')-(7 DA O4')-(7 DA C1')-(7 DA C2') -40.2 -10.2

&rst iat = 193, 195, 196, 214,  
r1 = -41.2, r2 = -40.2, r3 = -10.2, r4 = -9.2,  
rk2 = 50.0, rk3 = 50.0, &end

# 7 ADE NU1: (7 DA O4')-(7 DA C1')-(7 DA C2')-(7 DA C3') 11.1 41.1

```
&rst iat = 195, 196, 214, 212,
      r1 = 10.1, r2 = 11.1, r3 = 41.1, r4 = 42.1,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 7 ADE NU2: (7 DA C1')-(7 DA C2')-(7 DA C3')-(7 DA C4') -38.5 10.0

```
&rst iat = 196, 214, 212, 193,
      r1 = -39.5, r2 = -38.5, r3 = 10.0, r4 = 11.0,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 7 ADE NU3: (7 DA C2')-(7 DA C3')-(7 DA C4')-(7 DA O4') -29.9 31.1

```
&rst iat = 214, 212, 193, 195,
      r1 = -30.9, r2 = -29.9, r3 = 31.1, r4 = 32.1,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 7 ADE NU4: (7 DA C3')-(7 DA C4')-(7 DA O4')-(7 DA C1') -11.8 38.4

```
&rst iat = 212, 193, 195, 196,
      r1 = -12.8, r2 = -11.8, r3 = 38.4, r4 = 39.4,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 12 THY NU0: (12 DT C4')-(12 DT O4')-(12 DT C1')-(12 DT C2') -38.1 -8.1

```
&rst iat = 345, 347, 348, 366,
      r1 = -39.1, r2 = -38.1, r3 = -8.1, r4 = -7.1,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 12 THY NU1: (12 DT O4')-(12 DT C1')-(12 DT C2')-(12 DT C3') 19.8 49.8

```
&rst iat = 347, 348, 366, 364,
      r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,
```

rk2 = 50.0, rk3 = 50.0, &end

# 12 THY NU2: (12 DT C1')-(12 DT C2')-(12 DT C3')-(12 DT C4') -47.9 -17.9

&rst iat = 348, 366, 364, 345,

r1 = -48.9, r2 = -47.9, r3 = -17.9, r4 = -16.9,

rk2 = 50.0, rk3 = 50.0, &end

# 12 THY NU3: (12 DT C2')-(12 DT C3')-(12 DT C4')-(12 DT O4') 3.6 33.6

&rst iat = 366, 364, 345, 347,

r1 = 2.6, r2 = 3.6, r3 = 33.6, r4 = 34.6,

rk2 = 50.0, rk3 = 50.0, &end

# 12 THY NU4: (12 DT C3')-(12 DT C4')-(12 DT O4')-(12 DT C1') -12.2 17.8

&rst iat = 364, 345, 347, 348,

r1 = -13.2, r2 = -12.2, r3 = 17.8, r4 = 18.8,

rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE NU0: (13 DA C4')-(13 DA O4')-(13 DA C1')-(13 DA C2') -38.1 -8.1

&rst iat = 377, 379, 380, 398,

r1 = -39.1, r2 = -38.1, r3 = -8.1, r4 = -7.1,

rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE NU1: (13 DA O4')-(13 DA C1')-(13 DA C2')-(13 DA C3') 19.8 49.8

&rst iat = 379, 380, 398, 396,

r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,

rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE NU2: (13 DA C1')-(13 DA C2')-(13 DA C3')-(13 DA C4') -47.9 -17.9



```

&rst iat = 380, 398, 396, 377,
      r1 = -48.9, r2 = -47.9, r3 = -17.9, r4 = -16.9,
      rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE NU3: (13 DA C2')-(13 DA C3')-(13 DA C4')-(13 DA O4') 3.6 33.6
&rst iat = 398, 396, 377, 379,
      r1 = 2.6, r2 = 3.6, r3 = 33.6, r4 = 34.6,
      rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE NU4: (13 DA C3')-(13 DA C4')-(13 DA O4')-(13 DA C1') -12.2 17.8
&rst iat = 396, 377, 379, 380,
      r1 = -13.2, r2 = -12.2, r3 = 17.8, r4 = 18.8,
      rk2 = 50.0, rk3 = 50.0, &end

# 14 GUA NU0: (14 DG C4')-(14 DG O4')-(14 DG C1')-(14 DG C2') -38.1 -8.1
&rst iat = 409, 411, 412, 431,
      r1 = -39.1, r2 = -38.1, r3 = -8.1, r4 = -7.1,
      rk2 = 50.0, rk3 = 50.0, &end

# 14 GUA NU1: (14 DG O4')-(14 DG C1')-(14 DG C2')-(14 DG C3') 19.8 49.8
&rst iat = 411, 412, 431, 429,
      r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,
      rk2 = 50.0, rk3 = 50.0, &end

# 14 GUA NU2: (14 DG C1')-(14 DG C2')-(14 DG C3')-(14 DG C4') -47.9 -17.9
&rst iat = 412, 431, 429, 409,
      r1 = -48.9, r2 = -47.9, r3 = -17.9, r4 = -16.9,
      rk2 = 50.0, rk3 = 50.0, &end

```

# 14 GUA NU3: (14 DG C2')-(14 DG C3')-(14 DG C4')-(14 DG O4') 3.6 33.6

```
&rst iat = 431, 429, 409, 411,
      r1 = 2.6, r2 = 3.6, r3 = 33.6, r4 = 34.6,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 14 GUA NU4: (14 DG C3')-(14 DG C4')-(14 DG O4')-(14 DG C1') -12.2 17.8

```
&rst iat = 429, 409, 411, 412,
      r1 = -13.2, r2 = -12.2, r3 = 17.8, r4 = 18.8,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 15 CYT NU0: (15 DC C4')-(15 DC O4')-(15 DC C1')-(15 DC C2') -39.8 -9.8

```
&rst iat = 442, 444, 445, 461,
      r1 = -40.8, r2 = -39.8, r3 = -9.8, r4 = -8.8,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 15 CYT NU1: (15 DC O4')-(15 DC C1')-(15 DC C2')-(15 DC C3') 19.8 49.8

```
&rst iat = 444, 445, 461, 459,
      r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 15 CYT NU2: (15 DC C1')-(15 DC C2')-(15 DC C3')-(15 DC C4') -43.9 -13.9

```
&rst iat = 445, 461, 459, 442,
      r1 = -44.9, r2 = -43.9, r3 = -13.9, r4 = -12.9,
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 15 CYT NU3: (15 DC C2')-(15 DC C3')-(15 DC C4')-(15 DC O4') -4.0 31.1

```
&rst iat = 461, 459, 442, 444,
```

r1 = -5.0, r2 = -4.0, r3 = 31.1, r4 = 32.1,  
rk2 = 50.0, rk3 = 50.0, &end

# 15 CYT NU4: (15 DC C3')-(15 DC C4')-(15 DC O4')-(15 DC C1') -11.8 25.8

&rst iat = 459, 442, 444, 445,  
r1 = -12.8, r2 = -11.8, r3 = 25.8, r4 = 26.8,  
rk2 = 50.0, rk3 = 50.0, &end

# 16 ADE NU0: (16 DA C4')-(16 DA O4')-(16 DA C1')-(16 DA C2') -38.1 -8.1

&rst iat = 472, 474, 475, 493,  
r1 = -39.1, r2 = -38.1, r3 = -8.1, r4 = -7.1,  
rk2 = 50.0, rk3 = 50.0, &end

# 16 ADE NU1: (16 DA O4')-(16 DA C1')-(16 DA C2')-(16 DA C3') 19.8 49.8

&rst iat = 474, 475, 493, 491,  
r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,  
rk2 = 50.0, rk3 = 50.0, &end

# 16 ADE NU2: (16 DA C1')-(16 DA C2')-(16 DA C3')-(16 DA C4') -47.9 -17.9

&rst iat = 475, 493, 491, 472,  
r1 = -48.9, r2 = -47.9, r3 = -17.9, r4 = -16.9,  
rk2 = 50.0, rk3 = 50.0, &end

# 16 ADE NU3: (16 DA C2')-(16 DA C3')-(16 DA C4')-(16 DA O4') 3.6 33.6

&rst iat = 493, 491, 472, 474,  
r1 = 2.6, r2 = 3.6, r3 = 33.6, r4 = 34.6,  
rk2 = 50.0, rk3 = 50.0, &end

# 16 ADE NU4: (16 DA C3')-(16 DA C4')-(16 DA O4')-(16 DA C1') -12.2 17.8

&rst iat = 491, 472, 474, 475,

r1 = -13.2, r2 = -12.2, r3 = 17.8, r4 = 18.8,

rk2 = 50.0, rk3 = 50.0, &end

# 17 GUA NU0: (17 DG C4')-(17 DG O4')-(17 DG C1')-(17 DG C2') -38.1 -8.1

&rst iat = 504, 506, 507, 526,

r1 = -39.1, r2 = -38.1, r3 = -8.1, r4 = -7.1,

rk2 = 50.0, rk3 = 50.0, &end

# 17 GUA NU1: (17 DG O4')-(17 DG C1')-(17 DG C2')-(17 DG C3') 19.8 49.8

&rst iat = 506, 507, 526, 524,

r1 = 18.8, r2 = 19.8, r3 = 49.8, r4 = 50.8,

rk2 = 50.0, rk3 = 50.0, &end

# 17 GUA NU2: (17 DG C1')-(17 DG C2')-(17 DG C3')-(17 DG C4') -47.9 -17.9

&rst iat = 507, 526, 524, 504,

r1 = -48.9, r2 = -47.9, r3 = -17.9, r4 = -16.9,

rk2 = 50.0, rk3 = 50.0, &end

# 17 GUA NU3: (17 DG C2')-(17 DG C3')-(17 DG C4')-(17 DG O4') 3.6 33.6

&rst iat = 526, 524, 504, 506,

r1 = 2.6, r2 = 3.6, r3 = 33.6, r4 = 34.6,

rk2 = 50.0, rk3 = 50.0, &end

# 17 GUA NU4: (17 DG C3')-(17 DG C4')-(17 DG O4')-(17 DG C1') -12.2 17.8

&rst iat = 524, 504, 506, 507,

r1 = -13.2, r2 = -12.2, r3 = 17.8, r4 = 18.8,

rk2 = 50.0, rk3 = 50.0, &end

#=====

# 10935 atoms read from pdb file ../HIV-I-EZmin2.pdb. (Backbone Restraints)

# 1 ADE EPSILN: (1 DA5 C4')-(1 DA5 C3')-(1 DA5 O3')-(2 DC P) 175.6 215.6

&rst iat = 6, 25, 30, 31,

r1 = 174.6, r2 = 175.6, r3 = 215.6, r4 = 216.6,

rk2 = 50.0, rk3 = 50.0, &end

# 3 THY EPSILN: (3 DT C4')-(3 DT C3')-(3 DT O3')-(4 DI P) 181.0 201.0

&rst iat = 68, 87, 92, 93,

r1 = 180.0, r2 = 181.0, r3 = 201.0, r4 = 202.0,

rk2 = 50.0, rk3 = 50.0, &end

# 4 DI EPSILN: (4 DI C4')-(4 DI C3')-(4 DI O3')-(5 DC P) 156.6 216.6

&rst iat = 100, 118, 123, 124,

r1 = 155.6, r2 = 156.6, r3 = 216.6, r4 = 217.6,

rk2 = 50.0, rk3 = 50.0, &end

# 5 CYT EPSILN: (5 DC C4')-(5 DC C3')-(5 DC O3')-(6 DT P) 181.1 201.1

&rst iat = 131, 148, 153, 154,

r1 = 180.1, r2 = 181.1, r3 = 201.1, r4 = 202.1,

rk2 = 50.0, rk3 = 50.0, &end

# 6 THY EPSILN: (6 DT C4')-(6 DT C3')-(6 DT O3')-(7 DA P) 172.4 212.4

&rst iat = 161, 180, 185, 186,

r1 = 171.4, r2 = 172.4, r3 = 212.4, r4 = 213.4,

rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE EPSILN: (13 DA C4')-(13 DA C3')-(13 DA O3')-(14 DG P) 168.2 208.2

&rst iat = 377, 396, 401, 402,

r1 = 167.2, r2 = 168.2, r3 = 208.2, r4 = 209.2,

rk2 = 50.0, rk3 = 50.0, &end

# 14 GUA EPSILN: (14 DG C4')-(14 DG C3')-(14 DG O3')-(15 DC P) 179.3 199.3

&rst iat = 409, 429, 434, 435,

r1 = 178.3, r2 = 179.3, r3 = 199.3, r4 = 200.3,

rk2 = 50.0, rk3 = 50.0, &end

# 16 ADE EPSILN: (16 DA C4')-(16 DA C3')-(16 DA O3')-(17 DG P) 169.1 209.1

&rst iat = 472, 491, 496, 497,

r1 = 168.1, r2 = 169.1, r3 = 209.1, r4 = 210.1,

rk2 = 50.0, rk3 = 50.0, &end

# 17 GUA EPSILN: (17 DG C4')-(17 DG C3')-(17 DG O3')-(18 DT3 P) 183.8 203.8

&rst iat = 504, 524, 529, 530,

r1 = 182.8, r2 = 183.8, r3 = 203.8, r4 = 204.8,

rk2 = 50.0, rk3 = 50.0, &end

# 10935 atoms read from pdb file ../HIV-I-EZmin2.pdb.

# 1 ADE ZETA: (1 DA5 C3')-(1 DA5 O3')-(2 DC P)-(2 DC O5') 150.0 315.0

&rst iat = 25, 30, 31, 34,

r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,

rk2 = 50.0, rk3 = 50.0, &end

# 2 CYT ZETA: (2 DC C3')-(2 DC O3')-(3 DT P)-(3 DT O5') 150.0 315.0

```
&rst iat = 55, 60, 61, 64,
      r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,
      rk2 = 50.0, rk3 = 50.0,                                &end
```

# 3 THY ZETA: (3 DT C3')-(3 DT O3')-(4 DI P)-(4 DI O5') 150.0 315.0

```
&rst iat = 87, 92, 93, 96,
      r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,
      rk2 = 50.0, rk3 = 50.0,                                &end
```

# 4 DI ZETA: (4 DI C3')-(4 DI O3')-(5 DC P)-(5 DC O5') 150.0 315.0

```
&rst iat = 118, 123, 124, 127,
      r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,
      rk2 = 50.0, rk3 = 50.0,                                &end
```

# 5 CYT ZETA: (5 DC C3')-(5 DC O3')-(6 DT P)-(6 DT O5') 150.0 315.0

```
&rst iat = 148, 153, 154, 157,
      r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,
      rk2 = 50.0, rk3 = 50.0,                                &end
```

# 6 THY ZETA: (6 DT C3')-(6 DT O3')-(7 DA P)-(7 DA O5') 150.0 315.0

```
&rst iat = 180, 185, 186, 189,
      r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,
      rk2 = 50.0, rk3 = 50.0,                                &end
```

# 7 ADE ZETA: (7 DA C3')-(7 DA O3')-(8 DC P)-(8 DC O5') 150.0 315.0

```
&rst iat = 212, 217, 218, 221,
      r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,
```

rk2 = 50.0, rk3 = 50.0, &end

# 12 THY ZETA: (12 DT C3')-(12 DT O3')-(13 DA P)-(13 DA O5') 150.0 315.0

&rst iat = 364, 369, 370, 373,

r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,

rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE ZETA: (13 DA C3')-(13 DA O3')-(14 DG P)-(14 DG O5') 150.0 315.0

&rst iat = 396, 401, 402, 405,

r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,

rk2 = 50.0, rk3 = 50.0, &end

# 14 GUA ZETA: (14 DG C3')-(14 DG O3')-(15 DC P)-(15 DC O5') 150.0 315.0

&rst iat = 429, 434, 435, 438,

r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,

rk2 = 50.0, rk3 = 50.0, &end

# 15 CYT ZETA: (15 DC C3')-(15 DC O3')-(16 DA P)-(16 DA O5') 150.0 315.0

&rst iat = 459, 464, 465, 468,

r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,

rk2 = 50.0, rk3 = 50.0, &end

# 16 ADE ZETA: (16 DA C3')-(16 DA O3')-(17 DG P)-(17 DG O5') 150.0 315.0

&rst iat = 491, 496, 497, 500,

r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,

rk2 = 50.0, rk3 = 50.0, &end

# 17 GUA ZETA: (17 DG C3')-(17 DG O3')-(18 DT3 P)-(18 DT3 O5') 150.0 315.0



```
&rst iat = 524, 529, 530, 533,  
      r1 = 140.1, r2 = 150.0, r3 = 315.0, r4 = 325.1,  
      rk2 = 50.0, rk3 = 50.0,                               &end  
  
# 10935 atoms read from pdb file ../HIV-I-EZmin2.pdb.  
# 2 CYT ALPHA: (1 DA5 O3')-(2 DC P)-(2 DC O5')-(2 DC C5') 270.0 330.0  
&rst iat = 30, 31, 34, 35,  
      r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,  
      rk2 = 50.0, rk3 = 50.0,                               &end  
  
# 3 THY ALPHA: (2 DC O3')-(3 DT P)-(3 DT O5')-(3 DT C5') 270.0 330.0  
&rst iat = 60, 61, 64, 65,  
      r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,  
      rk2 = 50.0, rk3 = 50.0,                               &end  
  
# 4 DI ALPHA: (3 DT O3')-(4 DI P)-(4 DI O5')-(4 DI C5') 270.0 330.0  
&rst iat = 92, 93, 96, 97,  
      r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,  
      rk2 = 50.0, rk3 = 50.0,                               &end  
  
# 5 CYT ALPHA: (4 DI O3')-(5 DC P)-(5 DC O5')-(5 DC C5') 270.0 330.0  
&rst iat = 123, 124, 127, 128,  
      r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,  
      rk2 = 50.0, rk3 = 50.0,                               &end  
  
# 6 THY ALPHA: (5 DC O3')-(6 DT P)-(6 DT O5')-(6 DT C5') 270.0 330.0  
&rst iat = 153, 154, 157, 158,  
      r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,
```

rk2 = 50.0, rk3 = 50.0, &end

# 7 ADE ALPHA: (6 DT O3')-(7 DA P)-(7 DA O5')-(7 DA C5') 270.0 330.0

&rst iat = 185, 186, 189, 190,

r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,

rk2 = 50.0, rk3 = 50.0, &end

# 12 THY ALPHA: (11 DC O3')-(12 DT P)-(12 DT O5')-(12 DT C5') 270.0 330.0

&rst iat = 337, 338, 341, 342,

r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,

rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE ALPHA: (12 DT O3')-(13 DA P)-(13 DA O5')-(13 DA C5') 270.0 330.0

&rst iat = 369, 370, 373, 374,

r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,

rk2 = 50.0, rk3 = 50.0, &end

# 14 GUA ALPHA: (13 DA O3')-(14 DG P)-(14 DG O5')-(14 DG C5') 270.0 330.0

&rst iat = 401, 402, 405, 406,

r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,

rk2 = 50.0, rk3 = 50.0, &end

# 15 CYT ALPHA: (14 DG O3')-(15 DC P)-(15 DC O5')-(15 DC C5') 270.0 330.0

&rst iat = 434, 435, 438, 439,

r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,

rk2 = 50.0, rk3 = 50.0, &end

# 16 ADE ALPHA: (15 DC O3')-(16 DA P)-(16 DA O5')-(16 DA C5') 270.0 330.0

```
&rst iat = 464, 465, 468, 469,
      r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,
      rk2 = 50.0, rk3 = 50.0, &end

# 17 GUA ALPHA: (16 DA O3')-(17 DG P)-(17 DG O5')-(17 DG C5') 270.0 330.0
&rst iat = 496, 497, 500, 501,
      r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,
      rk2 = 50.0, rk3 = 50.0, &end

# 18 THY ALPHA: (17 DG O3')-(18 DT3 P)-(18 DT3 O5')-(18 DT3 C5') 270.0 330.0
&rst iat = 529, 530, 533, 534,
      r1 = 260.2, r2 = 270.0, r3 = 330.0, r4 = 340.2,
      rk2 = 50.0, rk3 = 50.0, &end

#=====
# 10935 atoms read from pdb file ../HIV-I-EZmin2.pdb. (Backbone Restraints)
# 1 ADE BETA: (2 DC P)-(1 DA5 O5')-(1 DA5 C5')-(1 DA5 C4') 135.0 215.0
# &rst iat = 31, 2, 3, 6,
#      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,
#      rk2 = 50.0, rk3 = 50.0, &end

# 2 CYT BETA: (2 DC P)-(2 DC O5')-(2 DC C5')-(2 DC C4') 135.0 215.0
&rst iat = 31, 34, 35, 38,
      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,
      rk2 = 50.0, rk3 = 50.0, &end

# 3 THY BETA: (3 DT P)-(3 DT O5')-(3 DT C5')-(3 DT C4') 135.0 215.0
&rst iat = 61, 64, 65, 68,
      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,
```

rk2 = 50.0, rk3 = 50.0, &end

# 4 DI BETA: (4 DI P)-(4 DI O5')-(4 DI C5')-(4 DI C4') 135.0 215.0

&rst iat = 93, 96, 97, 100,

r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,

rk2 = 50.0, rk3 = 50.0, &end

# 5 CYT BETA: (5 DC P)-(5 DC O5')-(5 DC C5')-(5 DC C4') 135.0 215.0

&rst iat = 124, 127, 128, 131,

r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,

rk2 = 50.0, rk3 = 50.0, &end

# 6 THY BETA: (6 DT P)-(6 DT O5')-(6 DT C5')-(6 DT C4') 135.0 215.0

&rst iat = 154, 157, 158, 161,

r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,

rk2 = 50.0, rk3 = 50.0, &end

# 7 ADE BETA: (7 DA P)-(7 DA O5')-(7 DA C5')-(7 DA C4') 135.0 215.0

&rst iat = 186, 189, 190, 193,

r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,

rk2 = 50.0, rk3 = 50.0, &end

# 12 THY BETA: (12 DT P)-(12 DT O5')-(12 DT C5')-(12 DT C4') 135.0 215.0

&rst iat = 338, 341, 342, 345,

r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,

rk2 = 50.0, rk3 = 50.0, &end

# 13 ADE BETA: (13 DA P)-(13 DA O5')-(13 DA C5')-(13 DA C4') 135.0 215.0

```
&rst iat = 370, 373, 374, 377,  
      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,  
      rk2 = 50.0, rk3 = 50.0,                                &end  
  
# 14 GUA BETA: (14 DG P)-(14 DG O5')-(14 DG C5')-(14 DG C4') 135.0 215.0  
&rst iat = 402, 405, 406, 409,  
      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,  
      rk2 = 50.0, rk3 = 50.0,                                &end  
  
# 15 CYT BETA: (15 DC P)-(15 DC O5')-(15 DC C5')-(15 DC C4') 135.0 215.0  
&rst iat = 435, 438, 439, 442,  
      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,  
      rk2 = 50.0, rk3 = 50.0,                                &end  
  
# 16 ADE BETA: (16 DA P)-(16 DA O5')-(16 DA C5')-(16 DA C4') 135.0 215.0  
&rst iat = 465, 468, 469, 472,  
      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,  
      rk2 = 50.0, rk3 = 50.0,                                &end  
  
# 17 GUA BETA: (17 DG P)-(17 DG O5')-(17 DG C5')-(17 DG C4') 135.0 215.0  
&rst iat = 497, 500, 501, 504,  
      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,  
      rk2 = 50.0, rk3 = 50.0,                                &end  
  
# 18 THY BETA: (18 DT3 P)-(18 DT3 O5')-(18 DT3 C5')-(18 DT3 C4') 135.0 215.0  
&rst iat = 530, 533, 534, 537,  
      r1 = 125.3, r2 = 135.0, r3 = 215.0, r4 = 225.3,  
      rk2 = 50.0, rk3 = 50.0,                                &end
```

# 10935 atoms read from pdb file ../HIV-I-EZmin2.pdb.

# 1 ADE GAMMA: (1 DA5 O5')-(1 DA5 C5')-(1 DA5 C4')-(1 DA5 C3') 30.0 90.0

&rst iat = 2, 3, 6, 25,

r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,

rk2 = 50.0, rk3 = 50.0, &end

# 2 CYT GAMMA: (2 DC O5')-(2 DC C5')-(2 DC C4')-(2 DC C3') 30.0 90.0

&rst iat = 34, 35, 38, 55,

r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,

rk2 = 50.0, rk3 = 50.0, &end

# 3 THY GAMMA: (3 DT O5')-(3 DT C5')-(3 DT C4')-(3 DT C3') 30.0 90.0

&rst iat = 64, 65, 68, 87,

r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,

rk2 = 50.0, rk3 = 50.0, &end

# 4 DI GAMMA: (4 DI O5')-(4 DI C5')-(4 DI C4')-(4 DI C3') 30.0 90.0

&rst iat = 96, 97, 100, 118,

r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,

rk2 = 50.0, rk3 = 50.0, &end

# 5 CYT GAMMA: (5 DC O5')-(5 DC C5')-(5 DC C4')-(5 DC C3') 30.0 90.0

&rst iat = 127, 128, 131, 148,

r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,

rk2 = 50.0, rk3 = 50.0, &end

# 6 THY GAMMA: (6 DT O5')-(6 DT C5')-(6 DT C4')-(6 DT C3') 30.0 90.0

```
&rst iat = 157, 158, 161, 180,  
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,  
      rk2 = 50.0, rk3 = 50.0,                &end  
  
# 7 ADE GAMMA: (7 DA O5')-(7 DA C5')-(7 DA C4')-(7 DA C3') 30.0 90.0  
&rst iat = 189, 190, 193, 212,  
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,  
      rk2 = 50.0, rk3 = 50.0,                &end  
  
# 12 THY GAMMA: (12 DT O5')-(12 DT C5')-(12 DT C4')-(12 DT C3') 30.0 90.0  
&rst iat = 341, 342, 345, 364,  
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,  
      rk2 = 50.0, rk3 = 50.0,                &end  
  
# 13 ADE GAMMA: (13 DA O5')-(13 DA C5')-(13 DA C4')-(13 DA C3') 30.0 90.0  
&rst iat = 373, 374, 377, 396,  
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,  
      rk2 = 50.0, rk3 = 50.0,                &end  
  
# 14 GUA GAMMA: (14 DG O5')-(14 DG C5')-(14 DG C4')-(14 DG C3') 30.0 90.0  
&rst iat = 405, 406, 409, 429,  
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,  
      rk2 = 50.0, rk3 = 50.0,                &end  
  
# 15 CYT GAMMA: (15 DC O5')-(15 DC C5')-(15 DC C4')-(15 DC C3') 30.0 90.0  
&rst iat = 438, 439, 442, 459,  
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,  
      rk2 = 50.0, rk3 = 50.0,                &end
```

# 16 ADE GAMMA: (16 DA O5')-(16 DA C5')-(16 DA C4')-(16 DA C3') 30.0 90.0

```
&rst iat = 468, 469, 472, 491,
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,
      rk2 = 50.0, rk3 = 50.0, &end
```

# 17 GUA GAMMA: (17 DG O5')-(17 DG C5')-(17 DG C4')-(17 DG C3') 30.0 90.0

```
&rst iat = 500, 501, 504, 524,
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,
      rk2 = 50.0, rk3 = 50.0, &end
```

# 18 THY GAMMA: (18 DT3 O5')-(18 DT3 C5')-(18 DT3 C4')-(18 DT3 C3') 30.0 90.0

```
&rst iat = 533, 534, 537, 556,
      r1 = 20.4, r2 = 30.0, r3 = 90.0, r4 = 100.4,
      rk2 = 50.0, rk3 = 50.0, &end
```

#=====

# WATSON CRICK DISTANCES FOR WC (AT) BONDS BETWEEN 1 18

```
&rst iat = 1, 18, iresid = 1, atnam(1)='N1',atnam(2)='N3',
      r1 = 0, r2 = 2.72, r3 = 2.92, r4 = 3.42,
      rk2 =25.00, rk3 =25.00, &end
```

```
&rst iat = 1, 18, iresid = 1, atnam(1)='N6',atnam(2)='O4',
      r1 = 0, r2 = 2.85, r3 = 3.05, r4 = 3.55,
      rk2 =25.00, rk3 =25.00, &end
```

# Watson Crick angles for WC (AT) bonds between 1 18

```
&rst iat = 1, 18, 18, iresid = 1, atnam(1)='N1',atnam(2)='H3',
      atnam(3)='N3', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
```



```
rk2 =10.00, rk3 =10.00, &end

&rst iat = 1, 1, 18, iresid = 1, atnam(1)='N6',atnam(2)='H61',
atnam(3)='O4', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
rk2 =10.00, rk3 =10.00, &end

# Watson Crick distances for WC (CG) bonds between 2 17

&rst iat = 17, 2, iresid = 1, atnam(1)='O6',atnam(2)='N4',
r1 = 0, r2 = 2.81, r3 = 3.01, r4 = 3.51,
rk2 =25.00, rk3 =25.00, &end

&rst iat = 17, 2, iresid = 1, atnam(1)='N1',atnam(2)='N3',
r1 = 0, r2 = 2.85, r3 = 3.05, r4 = 3.55,
rk2 =25.00, rk3 =25.00, &end

&rst iat = 17, 2, iresid = 1, atnam(1)='N2',atnam(2)='O2',
r1 = 0, r2 = 2.76, r3 = 2.96, r4 = 3.46,
rk2 =25.00, rk3 =25.00, &end

# Watson Crick angles for WC (CG) bonds between 2 17

&rst iat = 17, 2, 2, iresid = 1, atnam(1)='O6',atnam(2)='H41',
atnam(3)='N4', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
rk2 =10.00, rk3 =10.00, &end

&rst iat = 17, 17, 2, iresid = 1, atnam(1)='N2',atnam(2)='H21',
atnam(3)='O2', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
rk2 =10.00, rk3 =10.00, &end

&rst iat = 17, 17, 2, iresid = 1, atnam(1)='N1',atnam(2)='H1',
atnam(3)='N3', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
rk2 =10.00, rk3 =10.00, &end

# Watson Crick distances for WC (TA) bonds between 3 16
```

```
&rst iat = 16, 3, iresid = 1, atnam(1)='N1',atnam(2)='N3',
```

```
  r1 = 0, r2 = 2.72, r3 = 2.92, r4 = 3.42,
```

```
  rk2 =25.00, rk3 =25.00, &end
```

```
&rst iat = 16, 3, iresid = 1, atnam(1)='N6',atnam(2)='O4',
```

```
  r1 = 0, r2 = 2.85, r3 = 3.05, r4 = 3.55,
```

```
  rk2 =25.00, rk3 =25.00, &end
```

```
# Watson Crick angles for WC (TA) bonds between 3 16
```

```
&rst iat = 16, 3, 3, iresid = 1, atnam(1)='N1',atnam(2)='H3',
```

```
  atnam(3)='N3', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
```

```
  rk2 =10.00, rk3 =10.00, &end
```

```
&rst iat = 16, 16, 3, iresid = 1, atnam(1)='N6',atnam(2)='H61',
```

```
  atnam(3)='O4', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
```

```
  rk2 =10.00, rk3 =10.00, &end
```

```
# Watson Crick distances for WC (IC) bonds between 4 15
```

```
&rst iat = 4, 15, iresid = 1, atnam(1)='O6',atnam(2)='N4',
```

```
  r1 = 0, r2 = 2.81, r3 = 3.01, r4 = 3.51,
```

```
  rk2 =25.00, rk3 =25.00, &end
```

```
&rst iat = 4, 15, iresid = 1, atnam(1)='N1',atnam(2)='N3',
```

```
  r1 = 0, r2 = 2.85, r3 = 3.05, r4 = 3.55,
```

```
  rk2 =25.00, rk3 =25.00, &end
```

```
# &rst iat = 4, 15, iresid = 1, atnam(1)='N2',atnam(2)='O2',
```

```
  r1 = 0, r2 = 2.76, r3 = 2.96, r4 = 3.46,
```

```
  rk2 =25.00, rk3 =25.00, &end
```

```
# Watson Crick angles for WC (IC) bonds between 4 15
```

```
&rst iat = 4, 15, 15, iresid = 1, atnam(1)='O6',atnam(2)='H41',
```

```

    atnam(3)='N4', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
    rk2 =10.00, rk3 =10.00, &end

# &rst iat = 4, 4, 15, iresid = 1, atnam(1)='N2',atnam(2)='H21',
    atnam(3)='O2', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
    rk2 =10.00, rk3 =10.00, &end

&rst iat = 4, 4, 15, iresid = 1, atnam(1)='N1',atnam(2)='H1',
    atnam(3)='N3', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
    rk2 =10.00, rk3 =10.00, &end

# Watson Crick distances for WC (CG) bonds between 5 14
&rst iat = 14, 5, iresid = 1, atnam(1)='O6',atnam(2)='N4',
    r1 = 0, r2 = 2.81, r3 = 3.01, r4 = 3.51,
    rk2 =25.00, rk3 =25.00, &end

&rst iat = 14, 5, iresid = 1, atnam(1)='N1',atnam(2)='N3',
    r1 = 0, r2 = 2.85, r3 = 3.05, r4 = 3.55,
    rk2 =25.00, rk3 =25.00, &end

&rst iat = 14, 5, iresid = 1, atnam(1)='N2',atnam(2)='O2',
    r1 = 0, r2 = 2.76, r3 = 2.96, r4 = 3.46,
    rk2 =25.00, rk3 =25.00, &end

# Watson Crick angles for WC (CG) bonds between 5 14
&rst iat = 14, 5, 5, iresid = 1, atnam(1)='O6',atnam(2)='H41',
    atnam(3)='N4', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
    rk2 =10.00, rk3 =10.00, &end

&rst iat = 14, 14, 5, iresid = 1, atnam(1)='N2',atnam(2)='H21',
    atnam(3)='O2', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
    rk2 =10.00, rk3 =10.00, &end

&rst iat = 14, 14, 5, iresid = 1, atnam(1)='N1',atnam(2)='H1',

```

```

atnam(3)='N3', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
rk2 =10.00, rk3 =10.00, &end

# Watson Crick distances for WC (TA) bonds between 6 13
&rst iat = 13, 6, iresid = 1, atnam(1)='N1',atnam(2)='N3',
r1 = 0, r2 = 2.72, r3 = 2.92, r4 = 3.42,
rk2 =25.00, rk3 =25.00, &end
&rst iat = 13, 6, iresid = 1, atnam(1)='N6',atnam(2)='O4',
r1 = 0, r2 = 2.85, r3 = 3.05, r4 = 3.55,
rk2 =25.00, rk3 =25.00, &end

# Watson Crick angles for WC (TA) bonds between 6 13
&rst iat = 13, 6, 6, iresid = 1, atnam(1)='N1',atnam(2)='H3',
atnam(3)='N3', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
rk2 =10.00, rk3 =10.00, &end
&rst iat = 13, 13, 6, iresid = 1, atnam(1)='N6',atnam(2)='H61',
atnam(3)='O4', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
rk2 =10.00, rk3 =10.00, &end

# Watson Crick distances for WC (AT) bonds between 7 12
&rst iat = 7, 12, iresid = 1, atnam(1)='N1',atnam(2)='N3',
r1 = 0, r2 = 2.72, r3 = 2.92, r4 = 3.42,
rk2 =25.00, rk3 =25.00, &end
&rst iat = 7, 12, iresid = 1, atnam(1)='N6',atnam(2)='O4',
r1 = 0, r2 = 2.85, r3 = 3.05, r4 = 3.55,
rk2 =25.00, rk3 =25.00, &end

# Watson Crick angles for WC (AT) bonds between 7 12

```

```
&rst iat = 7, 12, 12, iresid = 1, atnam(1)='N1',atnam(2)='H3',
```

```
  atnam(3)='N3', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
```

```
  rk2 =10.00, rk3 =10.00, &end
```

```
&rst iat = 7, 7, 12, iresid = 1, atnam(1)='N6',atnam(2)='H61',
```

```
  atnam(3)='O4', r1 = 150, r2 = 170, r3 = 190, r4 = 210,
```

```
  rk2 =10.00, rk3 =10.00, &end
```

```
#=====
```

```
# IMINO SEQUENTIAL
```

```
# distance constraint number 1
```

```
&rst iat = 4, 3, iresid = 1, atnam(1)='H1',atnam(2)='H3',
```

```
  r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,
```

```
  rk2 =30.000, rk3 =30.000, /
```

```
# distance constraint number 2
```

```
&rst iat = 14, 6, iresid = 1, atnam(1)='H1',atnam(2)='H3',
```

```
  r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,
```

```
  rk2 =30.000, rk3 =30.000, /
```

```
# distance constraint number 3
```

```
&rst iat = 14, 4, iresid = 1, atnam(1)='H1',atnam(2)='H1',
```

```
  r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,
```

```
  rk2 =30.000, rk3 =30.000, /
```

```
# distance constraint number 4
```

```
&rst iat = 3, 17, iresid = 1, atnam(1)='H3',atnam(2)='H1',
```

```
  r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,
```

```
  rk2 =30.000, rk3 =30.000, /
```

```
# distance constraint number 5
```

```
&rst iat = 13, 6, iresid = 1, atnam(1)='H2',atnam(2)='H3',
```

```
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 6  
&rst iat = 13, 14, iresid = 1, atnam(1)='H2',atnam(2)='H1',  
r1 = 1.50, r2 = 2.00, r3 = 5.50, r4 = 6.50,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 7  
&rst iat = 16, 3, iresid = 1, atnam(1)='H2',atnam(2)='H3',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 8  
&rst iat = 16, 4, iresid = 1, atnam(1)='H2',atnam(2)='H1',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 9  
&rst iat = 2, 17, iresid = 1, atnam(1)='H42',atnam(2)='H1',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 10  
&rst iat = 2, 17, iresid = 1, atnam(1)='H41',atnam(2)='H1',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 11  
&rst iat = 5, 14, iresid = 1, atnam(1)='H42',atnam(2)='H1',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 12  
&rst iat = 5, 14, iresid = 1, atnam(1)='H41',atnam(2)='H1',
```

```
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 13  
&rst iat = 15, 4, iresid = 1, atnam(1)='H42',atnam(2)='H1',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 14  
&rst iat = 15, 4, iresid = 1, atnam(1)='H41',atnam(2)='H1',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 15  
&rst iat = 16, 3, iresid = 1, atnam(1)='H62',atnam(2)='H3',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 16  
&rst iat = 16, 3, iresid = 1, atnam(1)='H61',atnam(2)='H3',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 17  
&rst iat = 2, 2, iresid = 1, atnam(1)='H5',atnam(2)='H42',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 18  
&rst iat = 2, 2, iresid = 1, atnam(1)='H5',atnam(2)='H41',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 19  
&rst iat = 5, 5, iresid = 1, atnam(1)='H5',atnam(2)='H42',
```

```
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 20  
&rst iat = 5, 5, iresid = 1, atnam(1)='H5',atnam(2)='H41',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 21  
&rst iat = 15, 15, iresid = 1, atnam(1)='H5',atnam(2)='H42',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /  
# distance constraint number 22  
&rst iat = 5, 5, iresid = 1, atnam(1)='H5',atnam(2)='H41',  
r1 = 1.50, r2 = 2.00, r3 = 5.00, r4 = 6.00,  
rk2 =30.000, rk3 =30.000, /
```

**Figure A7: Non-NMR Restraints**

*Restraints added prior to the NMR restraints. These include WC base pair restraints, sugar restraints, qualitative restraints for the HP loop, backbone restraints, and imino sequential restraints.*