12-14-2016

Exploring HIV Integrase 3'-processing Using Designed DNA Substrates and Structural Study of HIV DNA Hairpins

Qiushi Li
qli15@student.gsu.edu

EXPLORING HIV INTEGRASE 3'-PROCESSING USING DESIGNED DNA SUBSTRATES AND STRUCTURAL STUDY OF HIV DNA HAIRPINS

by

QIUSHI LI

Under the Direction of Markus W. Germann, PhD

ABSTRACT

In the HIV viral integration procedure, 3'-processing of the viral DNA by the integrase enzyme is an essential first step which is followed by the integration of viral DNA into the host genome. In 3'-processing, the integrase cleaves the backbone of the DNA substrate on the 3' end of a conserved CA dinucleotide motif and inserts a helix between the two DNA strands, forcing them apart (Hare, S., 2012). Our study confirms that the presence of a G-amino group is crucial for 3'-processing. Substituting inosine for G in the CA step removes this amino group and results in loss of enzyme activity. Further work showed that the presence of a terminal duplex segment is not required for 3'-processing. Additional substrate modifications are studied in order to evaluate the actual importance of the CA step.

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

EXPLORING HIV INTEGRASE 3'-PROCESSING USING DESIGNED DNA SUBSTRATES AND STRUCTURAL STUDY OF HIV DNA HAIRPINS

by

QIUSHI LI

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

2016

EXPLORING HIV INTEGRASE 3'-PROCESSING USING DESIGNED DNA SUBSTRATES AND STRUCTURAL STUDY OF HIV DNA HAIRPINS

by

QIUSHI LI

Committee Chair: Markus Germann

Committee: W. David Wilson

Gregory Poon

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2016

ACKNOWLEDGEMENTS

I'm very grateful that I could join in Dr. Germann's lab and spend three years for my graduate study. Those were really wonderful days. I would like to thank Dr. Markus Germann for letting me join this lab and enjoy my time here. Thank you for helping and teaching me in my work and being patient when I was faced with problems. Thank you for sharing your knowledge with us and I learned a lot for both study and life from my time in lab. I would like to thank Dr. Jin Zhang for helping me so much when I just came here and started at a new school in a new country. Thank you for teaching me a lot of skills for research work and also helping me to adjust to my new study in college. I would like to thank my lab mates for being such good friends. I would like thank Marina Evich for teaching me on NMR techniques and helping me in my project. Sarah, Bello, Chris, Beatrice, Chris, Zach and Jessica thank you for being great friends and I spent a lot of awesome time with you. I would also like to thank Georgia State University and the chemistry department for giving me an opportunity to study here. Last but not least, I would like to thank my family who love me and support me to achieve my goal all the time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 HIV Integrase

HIV-1 integrase is a multidomain enzyme with 288 amino acids. It is essential in the life cycle of the virus. This enzyme catalyzes the integration of blunt-ended linear viral cDNA into the host genome[(1)]. Retroviral INs belongs to a diverse family of polynucleotide transferases and consists of 3 functional domains: an amino-terminal domain, a catalytic core that coordinates two divalent metal ions, and a carboxy-terminal DNA-binding domain[(1)(4)]. The structure of a PFV-IN bound to a cognate recessed DNA substrate revealed the requirement of all 3 domains for IN activity. The HIV U3 and U5 terminus of the cDNA contain the recognition sites for the integrase[(2)]. IN recognizes the ends of its cognate long terminal repeat (LTR) sequences and cleaves 2-3 nucleotides from each 3' termini, in which the integrase cuts in the target DNA and ligates the recessed 3' ends of the viral DNA to the 5' ends of the target DNA at the site of cleavage, resulting in a gapped intermediate. Finally, the gap between the viral and the target DNA is repaired by the DNA polymerase and ligase[(1)(3)].

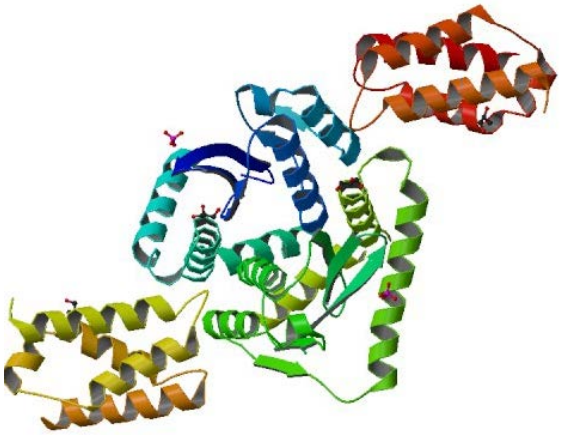


Figure 1.1 The structure of HIV integrase. (Cherepanov, P., 2005, Proc.Natl.Acad.Sci.USA 102: 17308-17313)

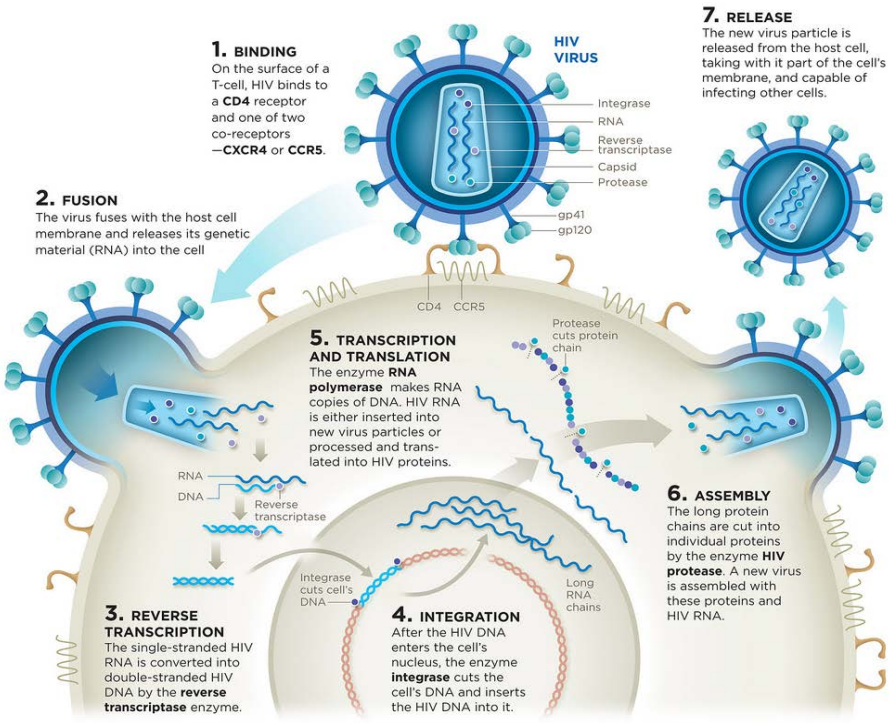


Figure 1.2 The procedure of integration of HIV integrase. (https://c2.staticflickr.com/4/3797/9037451756_ac0514c07b_b.jpg)

1.2 3'-processing

Much of the current knowledge of integrase 3'-processing is based on the crystal structures of full-length prototype foamy virus Integrase (PFV-IN). The enzyme inserts a helix between the two DNA strands that disrupts the 3' terminal DNA base pairs[(5)]. HIV integrase is not a sequence specific enzyme and common to all viral DNA 3' processing substrates is a conserved CA dinucleotide motif and corresponding GT pairing nucleotides[(9)]. Analysis of our DNA substrate structures reveals conserved but also different helical features of the CA motifs in these substrates. However, our study suggests that the local DNA structure at the CA motif is not the main determinant for retroviral recognition, and we demonstrate that by replacing the conserved guanosine (base partner of the conserved cytosine) with inosine which removes the exocyclic amino group abolishes integrase 3'-processing.

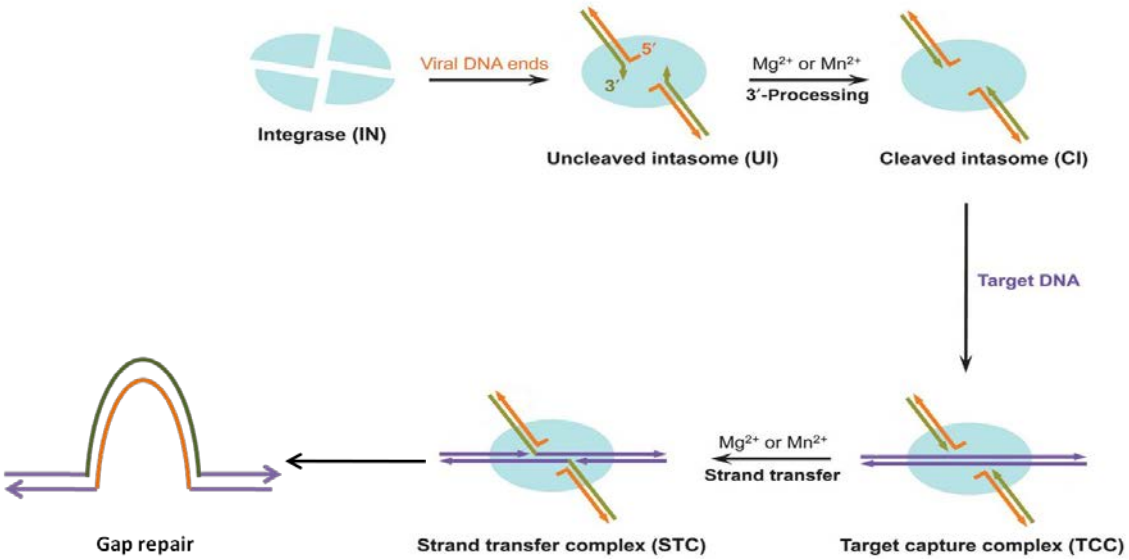


Figure 1.3 The 3'-processing and integration procedure of HIV integrase. (Hare, S., Maertens, G. N., 2012, E.M.B.O. J. 31, 3020-3028)

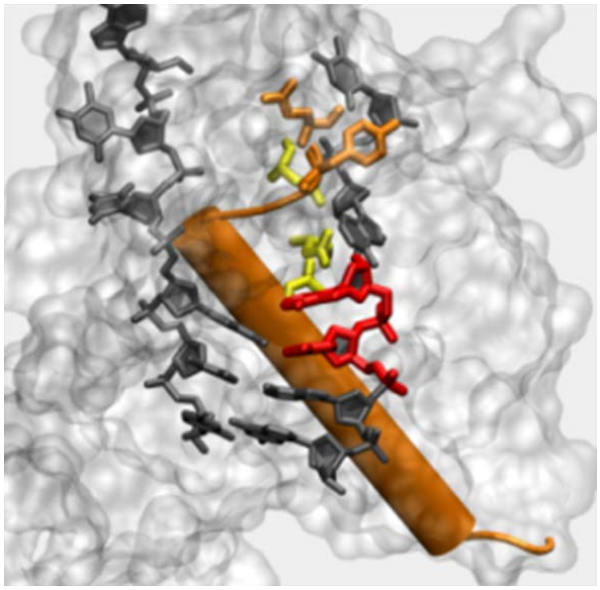


Figure 1.4 The crystal structure of prototype foamy virus Integrase (PFV-IN) bound to an uncleaved DNA substrate. *The PFV-IN enzyme inserts a helix (orange cylinder) in-between the two DNA strands (gray licorice). (Hare, S., Maertens, G. N., Cherepanov, P., 2012, E.M.B.O. J. 31, 3020-3028)*

1.3 DNA substrates for Integrase

Previous studies showed that the nucleotides close the conserved CA motif could influence the 3'-processing. In a wild type HIV DNA substrate sequence, if a nucleotide close to the CA motif was replaced by other nucleotides, the 3'-processing would be affected[(6)(7)]. The more close to CA motif the replaced nucleotide is, the more 3'-processing was affected[(7)]. From the structure of the PFV integrase bound to a DNA substrate (figure 1.4) [(5)], we know that the integrase disrupts the terminal base pairs of DNA substrates during the 3'-processing step. To provide more insight of 3'-processing of HIV integrase, various DNA substrates (table 2.2) were designed to study the quality of different substrates for HIV integrase 3'-processing and to find out the preference for 3'-processing. All the DNA substrates were reacted with purified HIV integrase under the exactly same condition, and the same reaction with wild-type sequence from a long terminal repeat of HIV DNA substrates was used as the control. The enzymatic study suggested that the amino group of the conserved guanosine is significant for 3'-processing, while a terminal duplex is not a requirement for 3'-processing reaction.

1.4 DNA G-hairpin and I-hairpin

The enzymatic study showed that the substitution of conserved G by I (inosine) rendered the DNA substrate not suitable for 3'-processing. We need to understand the structural consequences of the substitution of inosine to demonstrate that the abolishment of 3'-processing is not because of the inosine disturbed the structure of the substrate. An 18-mer hairpin with inosine (figure 1.5) was designed for the study since we already have extensive of structural data for a hairpin with normal guanosine (figure 1.5). A number of NMR experiments were applied and the results showed that the structure of I-hairpin is very similar to the G-hairpin. The

substitution of inosine abolished the 3'-processing without disturbing the whole structure of the substrate.

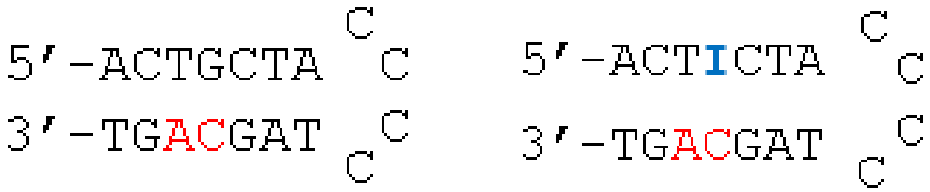


Figure 1.5 The DNA substrates for the structural study. *The 18-mer HIV-G-hairpin (left) and the HIV-I-hairpin (right) were used for NMR experiments.*

2 MATERIALS AND METHODS

2.1 Expression and purification of HIV Integrase

The HIV-1 integrase (F185H/C280S) is expressed in *Escherichia coli* BL 21 (DE3) cell with hexahistidine-tag in plasmid pET15 from Dr. Robert Craigie[(2)(8)]. A 5 mL start culture was grown overnight and transferred to 500 mL LB medium. When the OD of culture reaches 0.6-0.8, induce the expression with 1 mM IPTG for 3 hours. The expression was checked by 12% SDS-PAGE. Then the cell culture was spin for 30 minutes at 6000 rpm by Beckman J6-HC centrifuge with the JA-10 rotor, and the pellet was resuspended with 20 mL buffer I. The cell was opened by sonication (5 seconds burst, 10 seconds pause, sonicate for 5 minutes) and centrifuged at 13000 rpm for 30 minutes Beckman Avanti JXN-26 centrifuge with the JA-25.50 rotor. The supernatant was taken out and the pellet was resuspended the by 20 mL buffer I. Then the cell was opened by 20 minutes sonication (5 seconds burst, 10 seconds pause) and centrifuged at 13000 rpm for 30 minutes by Beckman Avanti JXN-26 centrifuge with the JA-25.50 rotor. The supernatant was loaded on a 5 mL Ni-NTA affinity column, and washed by 50 mL buffer II. The protein was eluted with a linear gradient from 60 mM to 1 M imidazole (buffer III and buffer IV). The collection from the column was checked by 12% SDS-PAGE. The purified protein was dialyzed against 1 L of buffer V overnight. About 2.2 mg protein was obtained from 500 mL culture and the concentration of the protein was ~24.8 μM with extinction coefficient number of 50670 $M^{-1}cm^{-1}$. The his-tag was cleaved by adding 10 units thrombin per mg of protein and reacted overnight at 4 °C. The cleavage was checked by 12% SDS-PAGE. Then the mixture was passed through a 1mL benzamidine sepharose column to remove the tag. After that, the protein was dialyzed against the final buffer VI and stored at -80 °C. The final concentration of the cleaved HIV integrase was ~23 μM.

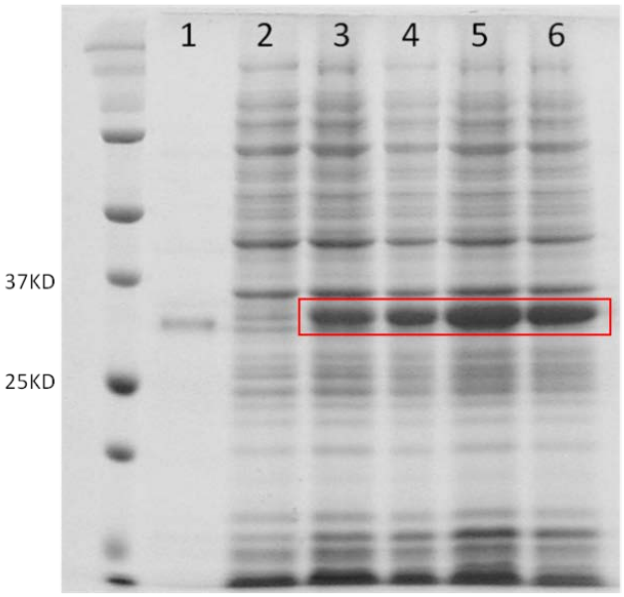


Figure 2.1 The cell expression of HIV integrase. *Lane 1 is the integrase purified by former lab member as a reference; lane 2 is uninduced cell and lanes 3-6 are cells induced by 1 mM IPTG for 1,2,3,4 hours. The gel shows that we obtained the optimal expression after inducing for 3 hours.*

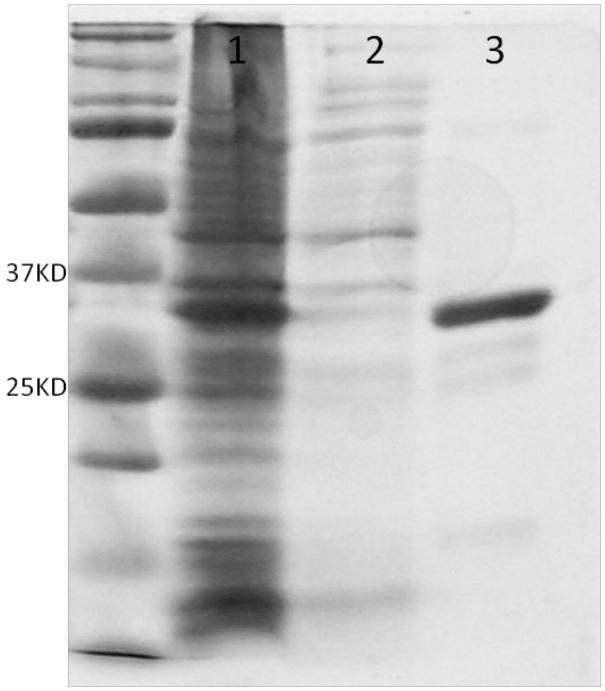


Figure 2.2 The purification of HIV integrase with Ni-NTA column. *Lane 1 is the cell lysate; lane 2 is the flow-through and lane 3 is the protein eluted by linear gradient of 60 mM to 1 M imidazole. After the purification, the HIV integrase was about 90% pure.*

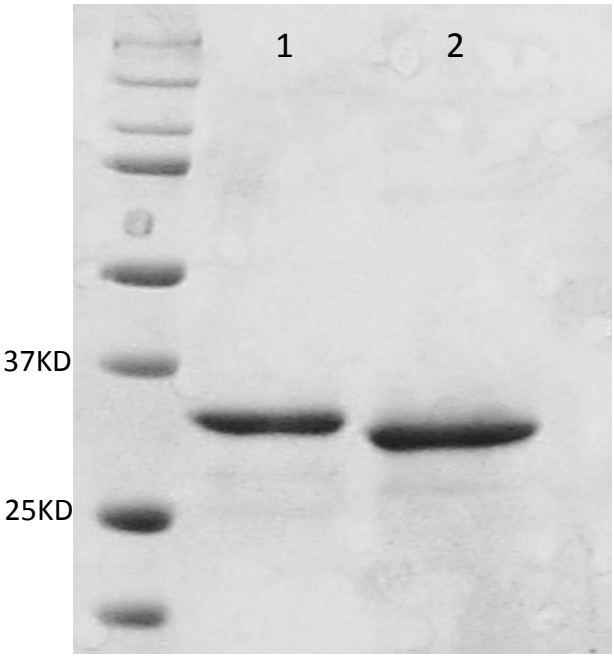


Figure 2.3 The cleavage of the his-tag of HIV integrase by thrombin. *Lane 1 is the integrase before cleavage and lane 2 is the integrase after the his-tag cleavage.*

Buffers for Chromatography

Table 2.1 The buffer for the purification of HIV integrase

Buffers	I	II	III	IV	V	VI
NaCl (M)	1	1	1	1	0.5	1
HEPES (20 mM, pH = 7.5)	√	√	√	√	√	√
EDTA (mM)	0	0	0	0	2	1
Imidazole (mM)	20	60	60	1000	300	0
Glycerol (10%)			√	√	√	√
DTT (2 mM)						√
β-Mercaptoethanol (2 mM)	√	√	√	√	√	
PMSF (1 mM)	√					
$ZnCl_2$ (100 μM)						√
Total volume (mL)	500	500	500	500	1000	1000

2.2 Reaction of integrase and substrates

The 5'-labeling reaction was performed by T4 polynucleotide kinase and γ–^{32}P-ATP in a final volume of 20 μL. The reaction contained 1x T4 kinase reaction buffer, 0.05 mg/mL BSA, 50 pmole DNA substrates, 25 pmol ^{32}P-ATP /cold ATP mixture and 10 units of T4 kinase. The reaction was incubated at 37 °C for 30 minutes and followed by heat inactivation of the enzyme at 85 °C for 15 minutes.

3'-processing reactions for HIV integrase substrates were made in a final volume of 20 μL. Reactions contained 25 mM MOPS pH 7.2, 7.5 mM $MnCl_2$, 0.1 mg/mL BSA, 10 mM β-mercaptoethanol, 10% glycerol, 50 nM ^{32}P-radiolabeled substrates and 2 μM integrase. The duplex DNA substrate was reacted at 37 °C for 3 hours. The 3'-processing products were separated by 15% polyacrylamide denaturing gels with 8 M urea and detected by being exposed to a phosphorescent screen. The screen was scanned by Typhoon 9400.

Table 2.2 The substrates of the 3'-processing reaction with HIV integrase

1(control)	3′-CACCTTTTAGAGATCGTCA-5′ 5'-GTGGAAAATCTCTAGCAGT-3'
2	3′-CACCTTTTAGAGATCITCA-5′ 5'-GTGGAAAATCTCTAGCAGT-3'
3	3′-CACCTTTTAGAGATCXTCA-5′ 5'-GTGGAAAATCTCTAGTAGT-3' X=2-amino purine
4	3′-CACCTTTTAGAGATCGTCATTTTT-5′ 5'-GTGGAAAATCTCTAGCAGTTTTTT-3′
5	3′-CACCTTTTAGAGATCGTTTTTT-5′ 5'-GTGGAAAATCTCTAGCATTTTT-3′

6	3′-CACCTTTTAGAGATCGTCA-5′ 5'-GTGGAAAATCTCTAGCAGTTTTTT-3′
7	3′-CACCTTTTAGAGATCGTCATTTTT-5′ 5'-GTGGAAAATCTCTAGCAGT-3'
8	3′-CACCTTTTAGAGATCCTCA-5′ 5'-GTGGAAAATCTCTAGGAGT-3'
9	3′-CACCTTTTAGAGATCGACA-5′ 5'-GTGGAAAATCTCTAGCTGT-3'
10	3′-CACCTTTTAGAGATCGTTTTTT-5′ 5'-GTGGAAAATCTCTAGCTTTTTT-3′
11	3′-CACCTTTTAGAGATCGTCA T 5'-GTGGAAAATCTCTAGCAGT T T
12	3′-CACCTTTTAGAGATCGTCA TTT 5'-GTGGAAAATCTCTAGCAGT TTT
13	T T CACTTAGGTCATTTTTCACC-5′ T GTGAATCCAGTTTTTTGTGG-3′ T

Table 2.3 The reaction condition of the 3'-processing

Reaction (20 µL)	
MOPS (pH 7.2)	25 mM
BSA	0.1 mg/mL
$MnCl_2$	7.5 mM
Glycerol	10%
β-Mercaptoethanol	10 mM
DNA substrates	50 nM
Integrase	2 µM

2.3 NMR experiments

2.3.1 Structure determination

NMR experiments were performed on a Bruker Avance 600 spectrometer, using a 5 mm IDTG triple resonance (Nalorac Corp) and QXI probe heads[10]. For experiments in D_2O is under the condition with 50 mM NaCl, 10 mM NaP, 0.1 mM EDTA, pH* 6.86 at 298 K: NOESY spectra were collected with mixing times of 75 ms, 150 ms, and 250 ms in 50 mM to get the distance restraint. The ^{31}P decoupled low flip angle COSY spectra was performed to get the $^{3}JH1'$-H2'1, $^{3}JH1'$-H2'2, $^{3}JH2'1$-H2'2, Σ H1', Σ H2'1 and Σ H2'2 couplings for the sugar puckering analysis[10]. Constant time NOESY (CT NOESY) experiments were collected using a 12 ms REBURP pulse to get the epsilon torsion angle[11]. ^{31}P spectra were performed in 300 K for phosphodiester backbone analysis. For water experiments (90% H_2O : 10% D_2O), imino proton spectra was performed in 50 mM NaCl, 10 mM NaP, 0.1 mM EDTA, pH 6.6 from 283 K to 308 K for base pairs. Assignment and integration of 2D spectra were done using SPARKY 3.33 U.C.S.F.20. ^{1}H was referenced to internal DSS.

2.3.2 Base pair opening

The base pair lifetime was measured for HIV-G hairpin and HIV-I hairpin. The base pair lifetime was determined by the relaxation time T_1 of the sample in ammonia catalyst with different concentration, which was measured by the intensity of the peaks of the DNA sample with different delay time and curve fit[13]. The base pair exchange rate τ_{ex} can be calculated by the equation[13]:

$$1/T_1 = 1/\tau_{ex} + 1/T_{1^\circ}$$

The T_1 experiment was performed at 278 K and 293 K. The DNA sample was prepared with the concentration of ~1 mM, in the buffer of 50 mM NaCl, 10 mM NaP, 0.1 mM EDTA, pH 7.8. The pH was measured each time after more catalyst was added. The NH_4OH was prepared 6 M and the concentration of ammonia catalyst was calculated by the equation[13]:

$$[NH_3] = [NH_4OH] \times 10^{(-pKa)} / (10^{(-pH)} + 10^{(-pKa)})$$

For the HIV-G-hairpin, we did the titration by the catalyst with a concentration of 2 mM, 2.5 mM, 3.5 mM, 5 mM, 10 mM, 20 mM, 40 mM and 100 mM. For the HIV-I hairpin, the titration was performed by the catalyst with a concentration of 5 mM, 10 mM, 15 mM, 30 mM and 60 mM.

After T_1 was measured for all concentrations, the base pair lifetime of infinite catalysts can be determined by the plot of 1/[cat] vs τ_{ex} with the equation[13]:

$$\tau_{ex} = \tau_{open} + 1/(K_d \alpha [B] k_i)$$

α is a correction factor for the accessibility difference of imino protons between mononucleotide and open state of an oligonucleotide and the value is about 1. k_i is the imino proton transfer rate with a value of ~ 2×10^8 s^{-1} M^{-1}[13]. K_d is the equilibrium constant and [B] is the concentration of

the catalyst. The base pair exchange rate is proportional to the 1/[cat], and the number of intercept with Y axis is the value of the base pair lifetime.

3 RESULT AND DISCUSSION

3.1 3’-processing reaction

3.1.1 DNA duplex and hairpin with inosine

In the 3’-processing reaction of G-duplex, two nucleotides after the conserved CA motif was cleaved by the HIV integrase. In the 3’-processing reaction of I-duplex under the same condition, the cleavage was greatly reduced. The hairpin substrate is not very suitable for the 3’-processing reaction because of its short stem. However, we can still tell that the cleavage for the I-hairpin is much less that the G-hairpin. The result proved that the amino group of the guanosine base paired with conserved C is crucial for the 3’-processing reaction, even for the hairpin substrates.

Figure 3.1 HIV Integrase 3'-processing reaction of G-duplex and I-duplex. *Lane 1: 3' processing reaction of G-duplex with no integrase. Lane 2: the reaction of the G-duplex with integrase. Lane 3: the reaction of I-duplex with no integrase. Lane 4: the reaction of the I-duplex with integrase. Lane 5: the reaction of the mixture substrate of G-duplex and I-duplex with integrase.*

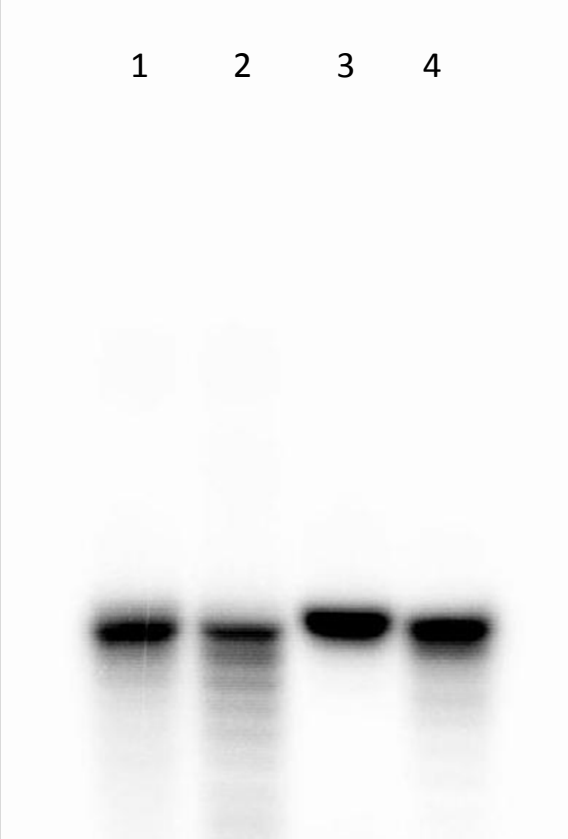


Figure 3.2 HIV Integrase 3'-processing reaction of G-hairpin and I-hairpin. *Lane 1: 3' processing reaction of G-hairpin with no integrase. Lane 2: the reaction of the G-hairpin with integrase. Lane 3: the reaction of I-hairpin with no integrase. Lane 4: the reaction of the I-hairpin with integrase.*

3.*1.2 Reaction with various DNA substrates*

The 3'-processing reaction study was performed with different DNA substrates (table 2.2). The results suggested that for the substrates of HIV integrase, the amino group of the G in the CA motif is crucial for the 3'-processing step. The substitution of G by I will abolish the cleavage, while the substrate with 2-amino purine is suitable for the enzyme. The 3'-processing step does not require substrates with a terminal duplex. The substrates with the frayed end or internal loops retain some enzyme activity. The results of the reaction with different DNA substrates were listed in table 3.1.

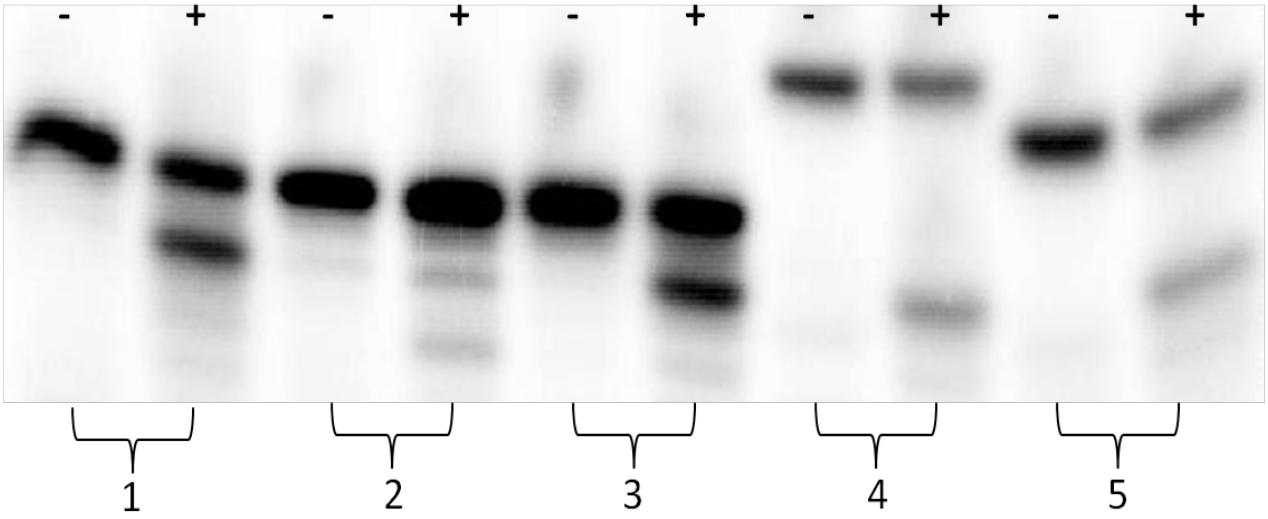


Figure 3.3 HIV Integrase 3'-processing reaction of DNA substrates 1-5 (table 2.2). *"-" means the reaction with no integrase. "+"means the reaction with integrase. There is almost no 3'-processing for inosine (lane 2); the cleavage of 2-amino purine (lane 3) and T-frayed end (lane 4 and 5) is similar to the control (lane 1).*

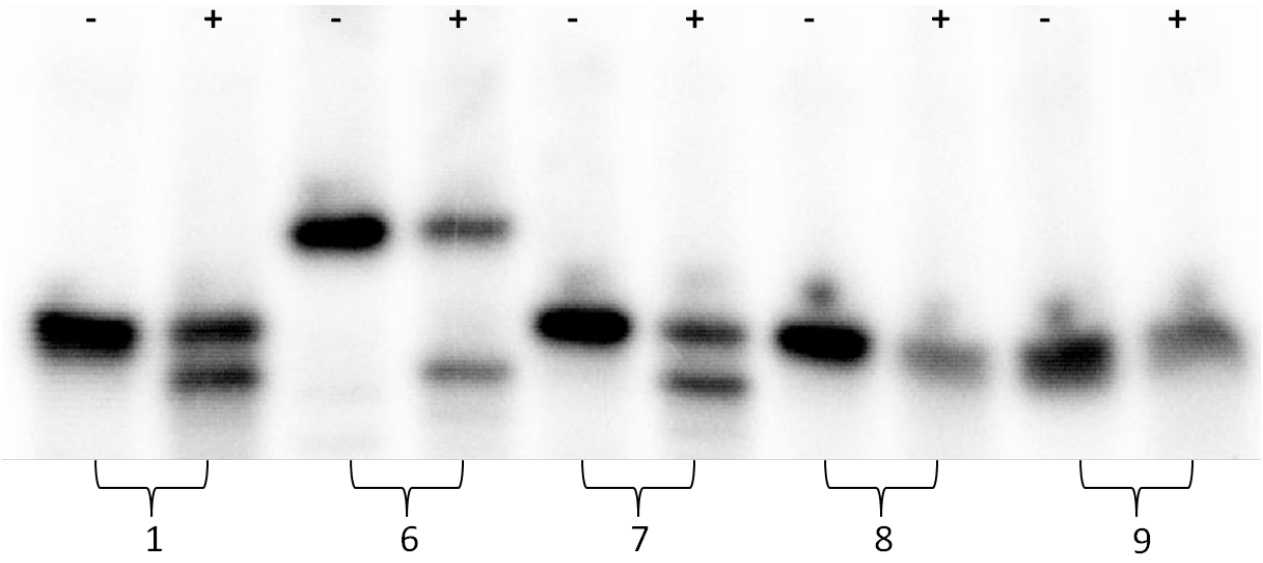


Figure 3.4 HIV Integrase 3'-processing reaction of DNA substrates 1and 6-9 (table 2.2). *"-" means the reaction with no integrase. "+"means the reaction with integrase. The cleavage of substrates with single T-tail on one side (lane 6 and 7) is similar to the control (lane 1). There is no cleavage for the substrates with a G4-C15 switch (lane 8) and with a T3-A16 switch (lane 9).*

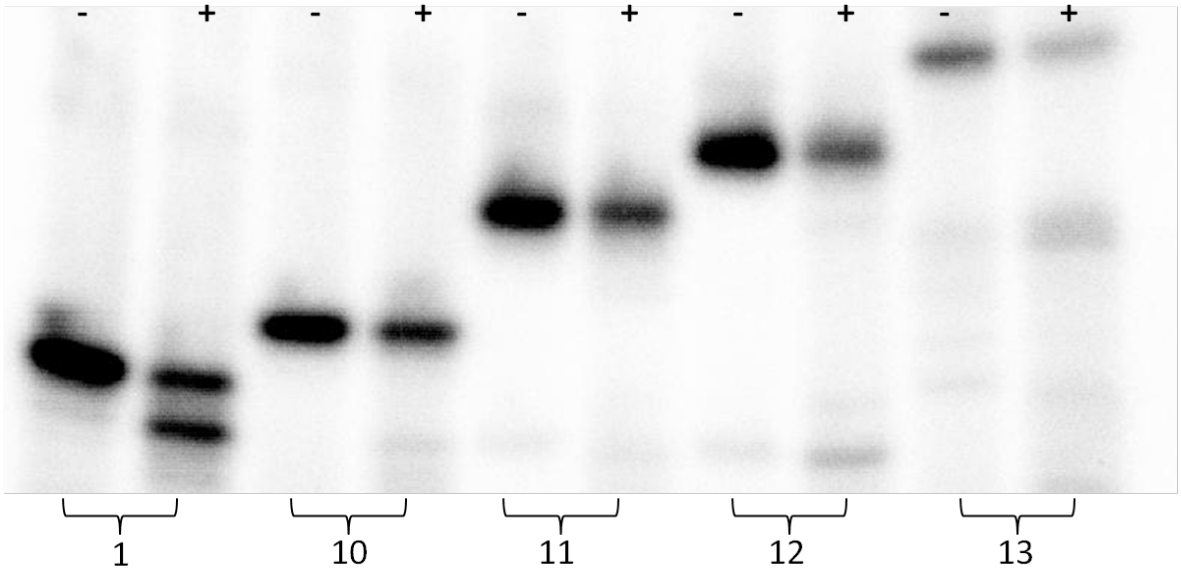


Figure 3.5 HIV Integrase 3'-processing reaction of DNA substrates 1and 10-13 (table 2.2). *"-" means the reaction with no integrase. "+"means the reaction with integrase. There is no cleavage of the substrates with T-tail directly following the conserved G-C base pair (lane 10). There is a little cleavage of the substrates with a terminal loop (lane 11 and 12). There is some cleavage of substrates with an internal loop (lane 13).*

Table 3.1 The 3'-processing result of different substrates

substrates	3'-processing
3′-CACCTTTTAGAGATCGTCA-5′ 5'-GTGGAAAATCTCTAGCAGT-3'	cleavage +++
3′-CACCTTTTAGAGATCITCA-5′ 5'-GTGGAAAATCTCTAGCAGT-3'	no cleavage -
3′-CACCTTTTAGAGATCGTCATTTTT-5′ 5'-GTGGAAAATCTCTAGCAGTTTTTT-3′ 3′-CACCTTTTAGAGATCGTTTTTT-5′ 5'-GTGGAAAATCTCTAGCATTTTT-3′	cleavage +++

3′-CACCTTTTAGAGATCGTCA-5′ 5'-GTGGAAAATCTCTAGCAGTTTTTT-3′ 3′-CACCTTTTAGAGATCGTCATTTTT-5′ 5'-GTGGAAAATCTCTAGCAGT-3'	cleavage +++
3′-CACCTTTTAGAGATCGTCA T T 5'-GTGGAAAATCTCTAGCAGT T 3′-CACCTTTTAGAGATCGTCA T T T 5'-GTGGAAAATCTCTAGCAGT T T T	a light cleavage +
3′-CACCTTTTAGAGATCCTCA-5′ 5'-GTGGAAAATCTCTAGGAGT-3' 3′-CACCTTTTAGAGATCGACA-5′ 5'-GTGGAAAATCTCTAGCTGT-3'	no cleavage -
3′-CACCTTTTAGAGATCGTTTTTT-5′ 5'-GTGGAAAATCTCTAGCTTTTTT-3′	no cleavage -
T T CACTTAGGTCATTTTTCACC-5′ T T GTGAATCCAGTTTTTTGTGG-3′ T	some cleavage ++
3′-CACCTTTTAGAGATCXTCA-5′ 5'-GTGGAAAATCTCTAGTAGT-3'	cleavage +++

3.2 NMR analysis of Inosine-hairpin

3.2.1 Base pairing and backbone structure

The NMR spectra of imino protons were performed at a temperature range from 283 K to 308 K, and the spectrum for HIV-I-hairpin is very similar with that of HIV-G-hairpin, except for

a chemical shift of the inosine at 4th position. The phosphorus spectra of HIV-I-hairpin and HIV-G-hairpin are very similar with each other as well.

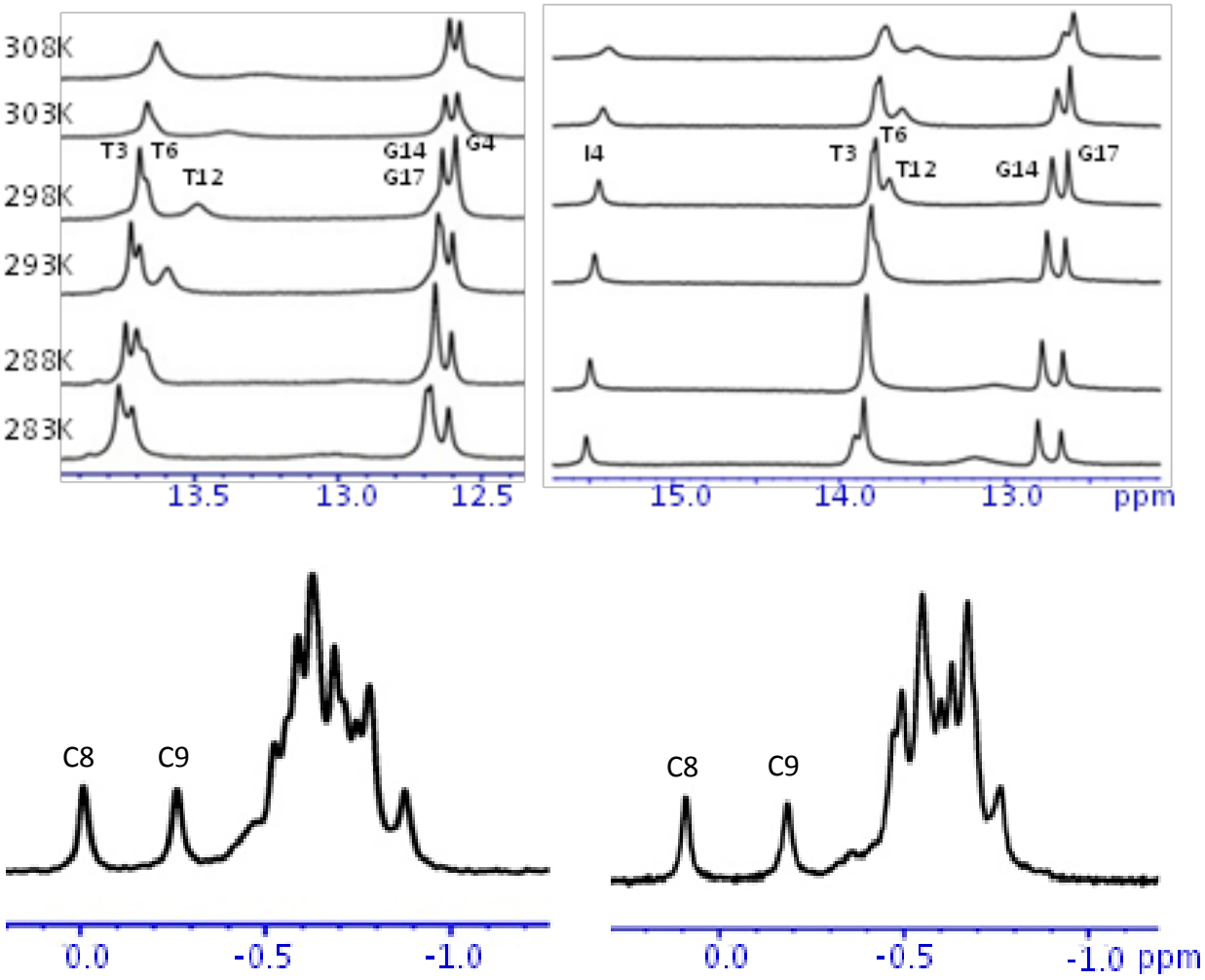


Figure 3.6 NMR spectra of imino protons (base pair) and phosphodiester backbone for HIV-G-hairpin and HIV-I hairpin. *The base pair and backbone of G-hairpin and I-hairpin are similar with each other.*

3.2.2 Sugar conformation

The sugar pucker of the HIV-I-hairpin was determined by the LF COSY NMR experiment[11]. Most of the nucleotides of both HIV-I-hairpin and HIV-G-hairpin are in a high south sugar conformation.

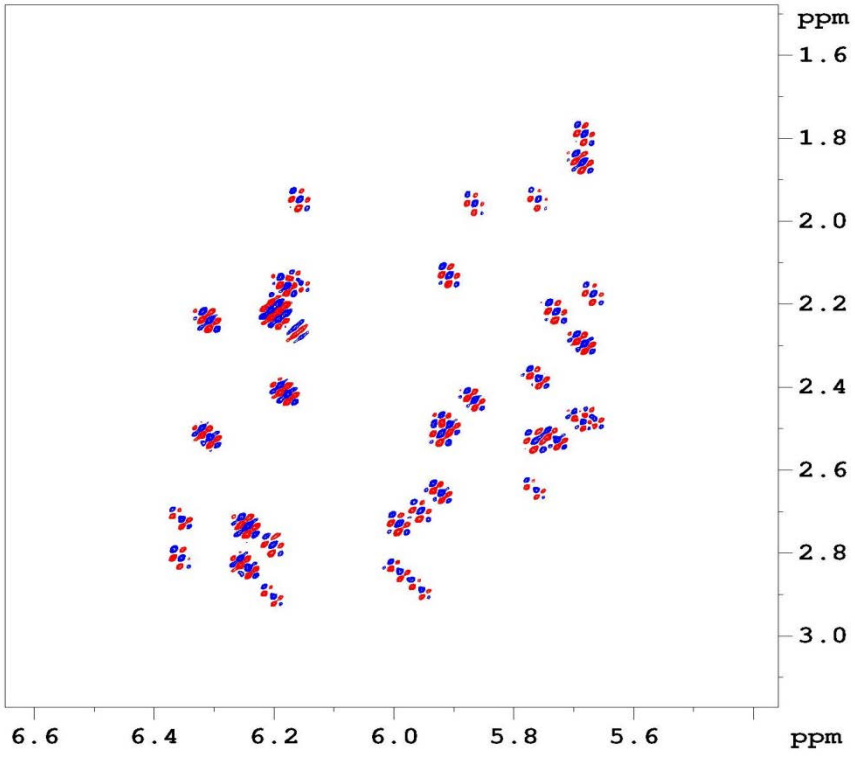


Figure 3.7 The spectrum of low flip angle COSY NMR experiment.

Table 3.2 The sugar conformation of I-hairpin

sugar form (0-100% South)			
HIV-I-hairpin		HIV-G-hairpin	
A1	60-100%	A1	78-100%
C2	65-87%	C2	78-100%
T3	80-100%	T3	89-98%
I4	65-100%	G4	90-98%
C5	65-100%	C5	53-73%
T6	70-100%	T6	75-92%
A7	50-100%	A7	92-100%
C8	65-90%	C8	66-83%
C9	70-100%	C9	90-100%
C10	65-90%	C10	87-96%
C11	65-90%	C11	78-98%
T12	70-80%	T12	75-99%
A13	80-100%	A13	96-100%
G14	95-100%	G14	95-100%
C15	65-90%	C15	82-98%
A16	75-90%	A16	93-100%
G17	80-90%	G17	81-100%

3.2.3 Epsilon torsion angle

The epsilon torsion angle is determined through the constant time NOESY NMR experiment, by measuring the ratio of the cross peak intensity of attenuated spectrum and reference spectrum using the equations[12]:

$$I_{attenuated}/I_{refernce} = \cos[\pi({}^{3}J_{H3'P} + D_{H3'P})T_{eff}]$$

$${}^{3}J_{(HCOP)} = 15.3\cos^{2}\Phi - 6.1\cos\Phi + 1.6$$

$$\varepsilon CCOP = \varepsilon HCOP - 120°$$

$J_{H3'P}$ is the J-coupling of H3' and ^{31}P. $D_{H3'P}$ is the dipolar coupling of H3' and ^{31}P with a value of 0 under our conditions. T_{eff} is a duration time which equals T + 2 x τ_{180} x 0.9. T is the constant time evolution period, τ_{180} is the duration of H3' REBURP pulse and 0.9 is a scale factor determined experimentally for ^{31}P coupling during the REBURP pulse [12]. The value of epsilon torsion angle of HIV-I-hairpin is very close to HIV-G-hairpin. The biggest difference is only ~10 degree, which happened to I4 and A16 which is in a position close to I4.

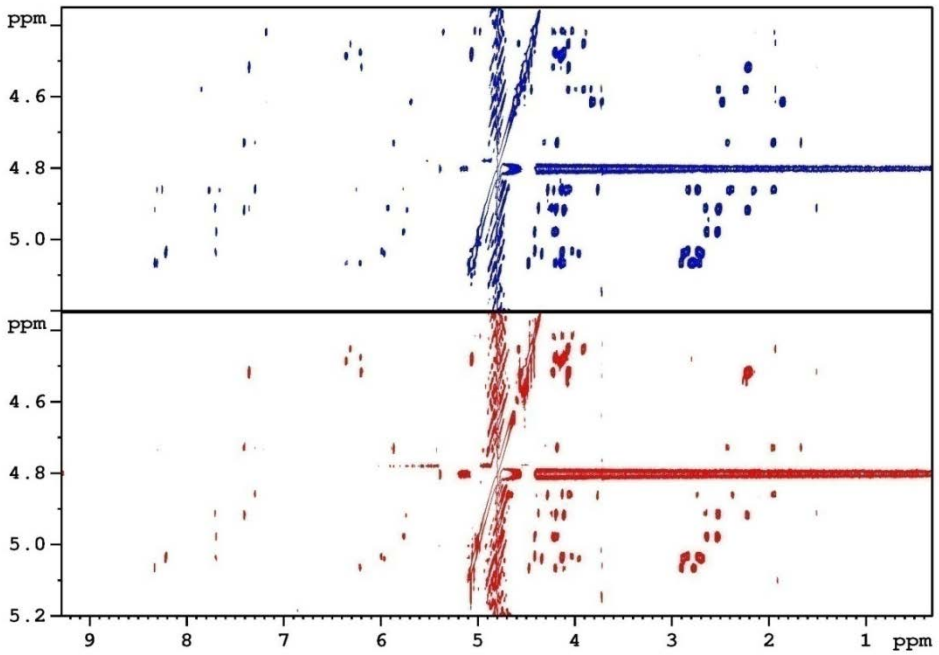


Figure 3.8 The spectrum of constant time NOESY NMR experiment. *The top spectrum is the reference spectrum and the bottom one is the attenuated spectrum. The experiment was run under 298 K with a mixing time of 300 ms.*

Table 3.3 The epsilon torsion angle of G- hairpin and I-hairpin

G-hp	εCCOP C	I-hp	εCCOP C	Δ
A1	-167	A1	-164.4	2.6
C2	-174	C2	-	-
T3	-174	T3	-169	5
G4	-184	I4	-173.4	10.6
C5	-172	C5	-168.9	3.1
T6	-175	T6	-167.6	7.4
A7	-157	A7	-	-
T12	-170	T12	-	-
A13		A13	-171.8	-
G14	-176	G14	-170.7	5.3
C15	-166	C15	-	-
A16	-181	A16	-170.9	10.1
G17	-169	G17	-166.2	2.8

3.2.4 NOESY analysis

Distance information of HIV hairpins was obtained from the cross peak volumes of NOESY experiment. The NOE volume showed in table 3.4 is the primary data from Sparky. At 250 ms mixing time spin diffusion affects the accuracy of distances calculated directly from cross peak volumes using a two spin approximation. Therefore, distances were computed from an iterative Mardigras procedure which takes account the spin diffusion for both HIV-I and HIV-G hairpins. The base (H8/H6) to H1’ sugar pathways for the HIV-I and HIV-G hairpins showed similar patterns, with the exception of the inosine base (I4) and base-paired residue C15, which was expected. A comparison of the cross peak volumes also shows high similarity. The error of NOE volumes are about 10% - 20% however, since the NOE volume is proportional to $1/r^6$, the

error of distance will be very small. Since the difference in the cross peak volumes are also very small between HIV-I and HIV-G hairpins it suggests that the 3D structures must also be highly similar.

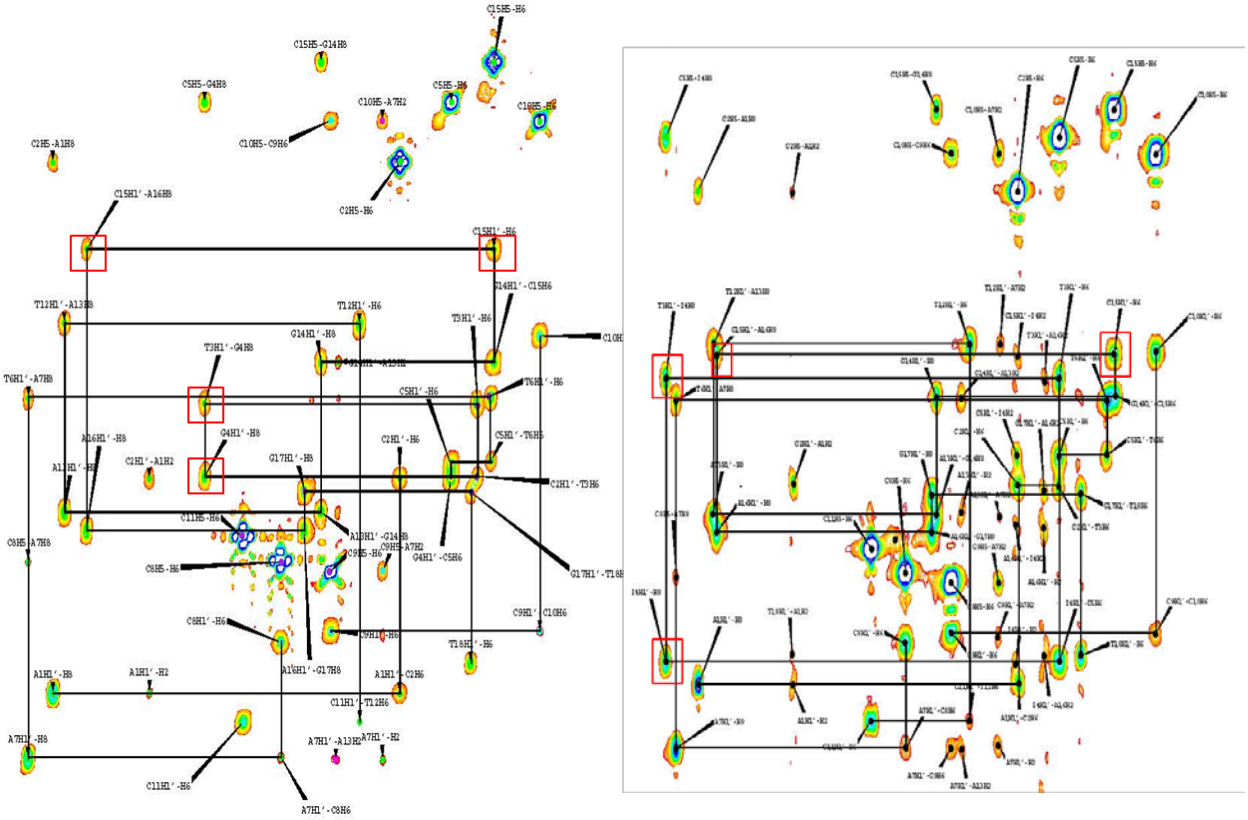


Figure 3.9 The H8/H6-H1' pathway of HIV-G-hairpin (left) and HIV-I-hairpin (right). *The two pathways are very similar, the cross peaks in red squares are position 4 and 15, which are shifted.*

Table 3.4 NOESY cross peak volumes of G and I-hairpins

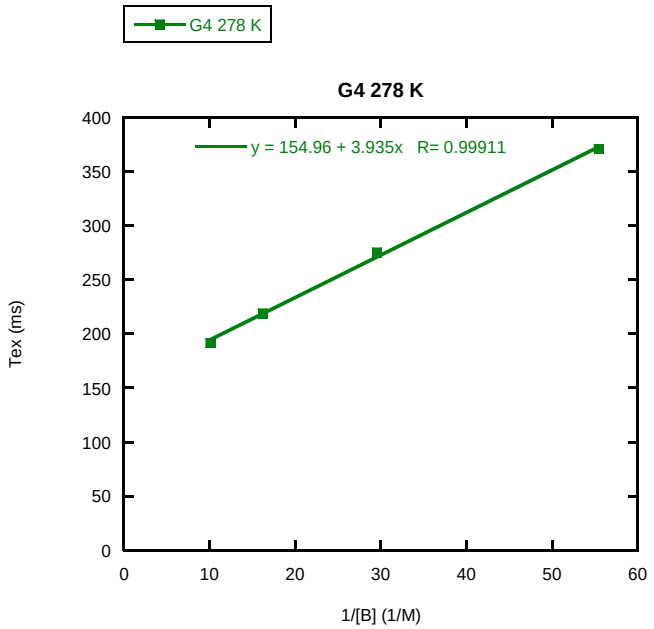
HIV-I			HIV-G			
Assignment	Volume	normalized to C9 H5-H6	Assignment	Volume	normalized to C9 H5-H6	|Δnormalized|
A1H1'-H2	2.21E+06	5.51E-02	A1H1'-H2	7.34E+06	5.69E-02	0.00
A1H1'-H8	9.61E+06	2.40E-01	A1H1'-H8	48200000	3.74E-01	0.13
A1H1'-C2H6	4.27E+06	1.06E-01	A1H1'-C2H6	15800000	1.22E-01	0.02
C2H1'-A1H2	3.03E+06	7.56E-02	C2H1'-A1H2	11000000	8.53E-02	0.01
C2H1'-H6	6.38E+06	1.59E-01	C2H1'-H6	21300000	1.65E-01	0.01
C2H1'-T3H6	3.83E+06	9.55E-02	C2H1'-T3H6	8.75E+06	6.78E-02	0.03
C2H5-A1H2	7.15E+05	1.78E-02				
C2H5-A1H8	2.08E+06	5.19E-02	C2H5-A1H8	9.83E+06	7.62E-02	0.02
C2H5-H6	4.43E+07	1.10E+00	C2H5-H6	2.3E+08	1.78E+00	0.68
T3H1'-H6	6.79E+06	1.69E-01	T3H1'-H6	20100000	1.56E-01	0.01
T3H1'-I4H8	5.18E+06	1.29E-01	T3H1'-G4H8	17100000	1.33E-01	0.00
T3H1'-A16H2	1.69E+06	4.21E-02				
I4H1'-H2	1.67E+06	4.16E-02				
I4H1'-H8	6.13E+06	1.53E-01	G4H1'-H8	24500000	1.90E-01	0.04
I4H1'-C5H6	9.92E+06	2.47E-01	G4H1'-C5H6		0.00E+00	
I4H1'-A16H2	6.22E+06	1.55E-01				
C5H1'-I4H2	3.02E+06	7.53E-02				
C5H1'-H6	7.25E+06	1.81E-01	C5H1'-H6		0.00E+00	
C5H1'-T6H6	3.65E+06	9.10E-02	C5H1'-T6H6	1.16E+07	8.99E-02	0.00
C5H5-I4H8	5.42E+06	1.35E-01	C5H5-G4H8	1.88E+07	1.46E-01	0.01
C5H5-H6	4.38E+07	1.09E+00	C5H5-H6	1.39E+08	1.08E+00	0.01
T6H1'-H6	8.86E+06	2.21E-01	T6H1'-H6	1.56E+07	1.21E-01	0.10
T6H1'-A7H8	3.27E+06	8.15E-02	T6H1'-A7H8	1.27E+07	9.84E-02	0.02
A7H1'-H2	1.50E+06	3.74E-02	A7H1'-H2	5.45E+06	4.22E-02	0.00
A7H1'-H8	9.50E+06	2.37E-01	A7H1'-H8	4.25E+07	3.29E-01	0.09
A7H1'-C8H6	2.91E+06	7.26E-02	A7H1'-C8H6	7.40E+06	5.74E-02	0.02
A7H1'-C9H6	2.32E+06	5.79E-02				
A7H1'-A13H2	1.43E+06	3.57E-02	A7H1'-A13H2	1.22E+07	9.46E-02	0.06
C8H1'-H6	8.98E+06	2.24E-01	C8H1'-H6	2.67E+07	2.07E-01	0.02
C8H5-A7H8	1.01E+06	2.52E-02	C8H5-A7H8	4.67E+06	3.62E-02	0.01
C8H5-H6	3.38E+07	8.43E-01	C8H5-H6	1.10E+08	8.53E-01	0.01
C9H1'-A7H2	1.69E+06	4.21E-02				

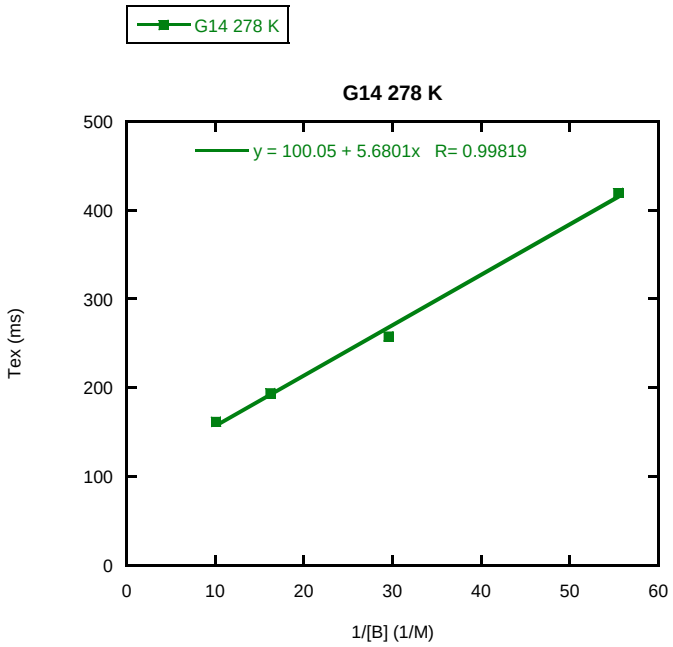
HIV-I			HIV-G			
Assignment	Volume	normalized to C9 H5-H6	Assignment	Volume	normalized to C9 H5-H6	|Δnormalized|
C9H1'-H6	8.98E+06	2.24E-01	C9H1'-H6	2.80E+07	2.17E-01	0.01
C9H1'-C10H6	2.50E+06	6.23E-02	C9H1'-C10H6	7.28E+06	5.64E-02	0.01
C9H5-A7H2	2.64E+06	6.58E-02	C9H5-A7H2	1.06E+07	8.22E-02	0.02
C9H5-H6	4.01E+07	1.00E+00	C9H5-H6	1.29E+08	1.00E+00	0.00
C10H1'-H6	7.88E+06	1.97E-01	C10H1'-H6	2.71E+07	2.10E-01	0.01
C10H5-A7H2	2.71E+06	6.76E-02	C10H5-A7H2	1.04E+07	8.06E-02	0.01
C10H5-C9H6	4.93E+06	1.23E-01	C10H5-C9H6	1.46E+07	1.13E-01	0.01
C10H5-H6	3.83E+07	9.55E-01	C10H5-H6	1.34E+08	1.04E+00	0.08
C11H1'-H6	8.96E+06	2.23E-01	C11H1'-H6	3.12E+07	2.42E-01	0.02
C11H1'-T12H6	1.10E+06	2.74E-02	C11H1'-T12H6	5.89E+06	4.57E-02	0.02
C11H5-H6	4.05E+07	1.01E+00	C11H5-H6	1.19E+08	9.22E-01	0.09
T12H1'-A7H2	1.33E+06	3.32E-02				
T12H1'-H6	6.63E+06	1.65E-01	T12H1'-H6	2.76E+07	2.14E-01	0.05
T12H1'-A13H8	4.01E+06	1.00E-01	T12H1'-A13H8	1.76E+07	1.36E-01	0.04
A13H1'-A7H2	8.50E+05	2.12E-02				
A13H1'-H2	2.22E+06	5.54E-02				
A13H1'-H8	5.05E+06	1.26E-01	A13H1'-H8	2.70E+07	2.09E-01	0.08
A13H1'-G14H8	5.03E+06	1.25E-01	A13H1'-G14H8	2.62E+07	2.03E-01	0.08
G14H1'-A13H2	2.22E+06	5.54E-02	G14H1'-A13H2	7.14E+06	5.53E-02	0.00
G14H1'-H8	5.09E+06	1.27E-01	G14H1'-H8	2.18E+07	1.69E-01	0.04
G14H1'-C15H6	7.31E+06	1.82E-01	G14H1'-C15H6	2.95E+07	2.29E-01	0.05
C15H1'-I4H2	1.76E+06	4.39E-02				
C15H1'-H6	6.69E+06	1.67E-01	C15H1'-H6	2.13E+07	1.65E-01	0.00
C15H1'-A16H8	2.14E+06	5.34E-02	C15H1'-A16H8	1.35E+07	1.05E-01	0.05
C15H5-G14H8	4.19E+06	1.04E-01	C15H5-G14H8	1.79E+07	1.39E-01	0.03
C15H5-H6	4.07E+07	1.01E+00	C15H5-H6	1.95E+08	1.51E+00	0.50
A16H1'-I4H2	2.56E+06	6.38E-02				
A16H1'-H2	3.15E+06	7.86E-02				
A16H1'-H8	6.78E+06	1.69E-01	A16H1'-H8	2.53E+07	1.96E-01	0.03
A16H1'-G17H8	8.18E+06	2.04E-01	A16H1'-G17H8	3.47E+07	2.69E-01	0.07
G17H1'-A16H2	2.12E+06	5.29E-02				
G17H1'-H8	6.55E+06	1.63E-01	G17H1'-H8	3.09E+07	2.40E-01	0.08
G17H1'-T18H6	4.81E+06	1.20E-01	G17H1'-T18H6	9.74E+06	7.55E-02	0.04
T18H1'-A1H2	7.42E+05	1.85E-02				
T18H1'-H6	5.98E+06	1.49E-01	T18H1'-H6	2.02E+07	1.57E-01	0.01

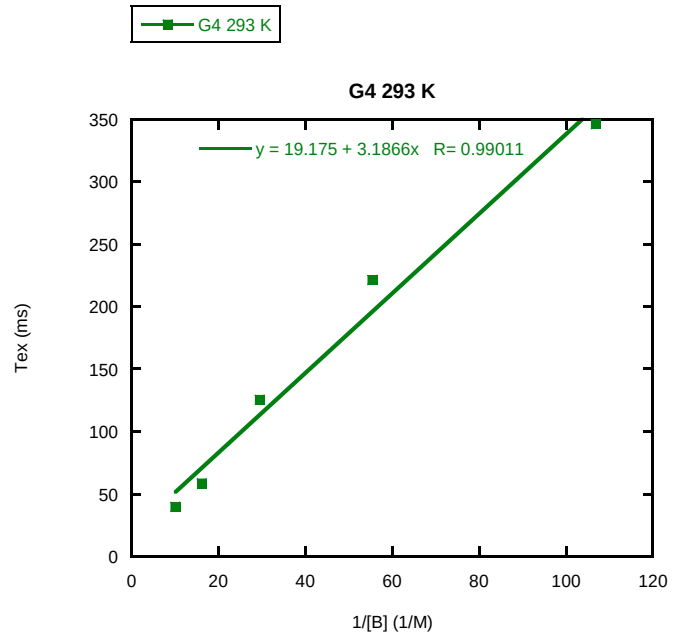
3.3 Base pair opening study

3.3.1 Base pair lifetime of G-hairpin

The base pair lifetime was measured for I4-C15 in HIV-I-hairpin, G4-C15 in HIV-G-hairpin and C5-G14 in both substrates at two temperatures 278 K and 293 K. The base pair lifetime of I4-C15 in I-hairpin is much smaller than that for G4-C15 in G-hairpin, while the base pair lifetimes of C5-G14 in both substrates are similar to each other.




G14 278 K
G14 278 K
y = 100.05 + 5.6801x R= 0.99819
Tex (ms)
1/[B] (1/M)
0
100
200
300
400
500
0
10
20
30
40
50
60


G4 293 K
G4 293 K
y = 19.175 + 3.1866x R= 0.99011
Tex (ms)
1/[B] (1/M)
0
50
100
150
200
250
300
350
0
20
40
60
80
100
120

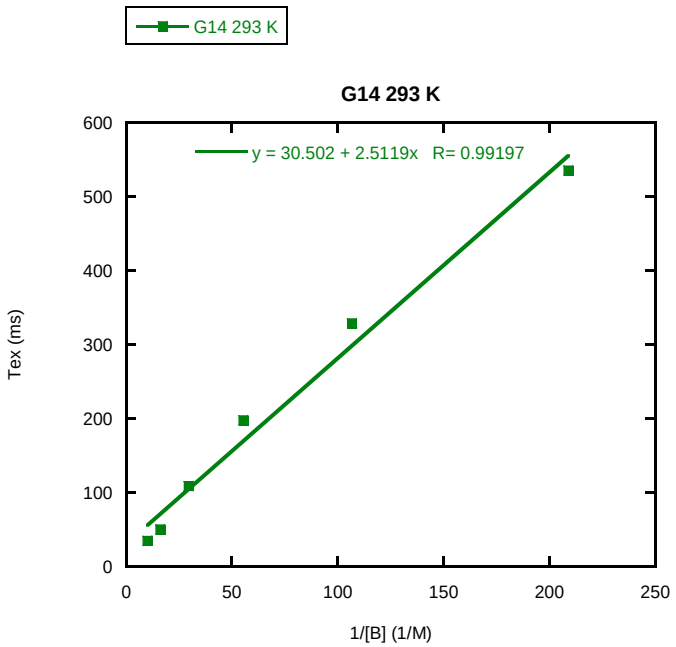
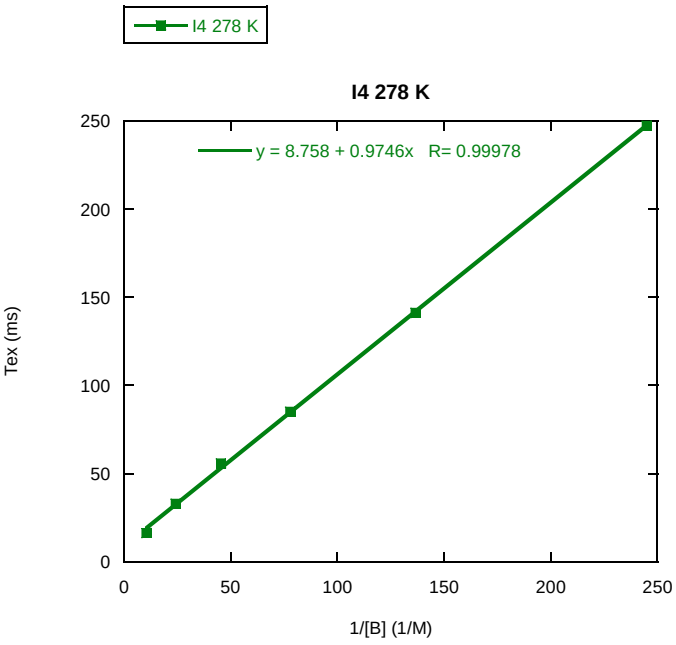
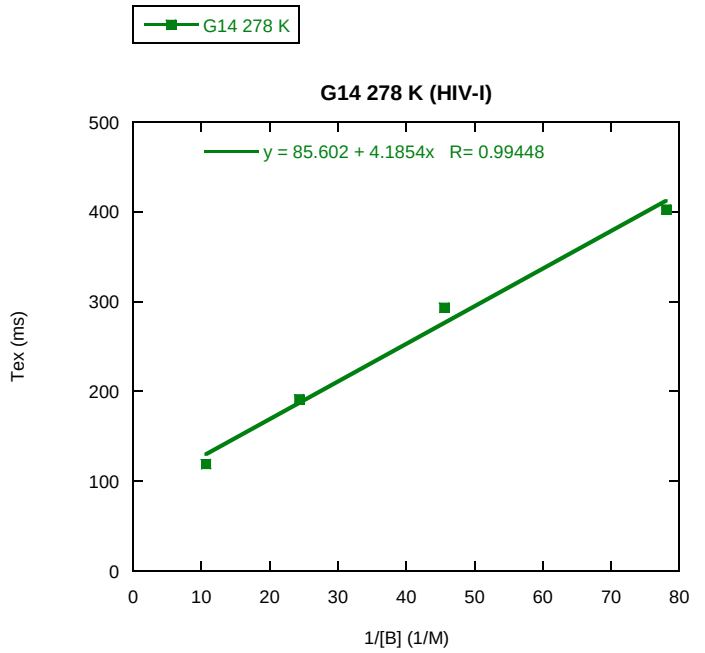
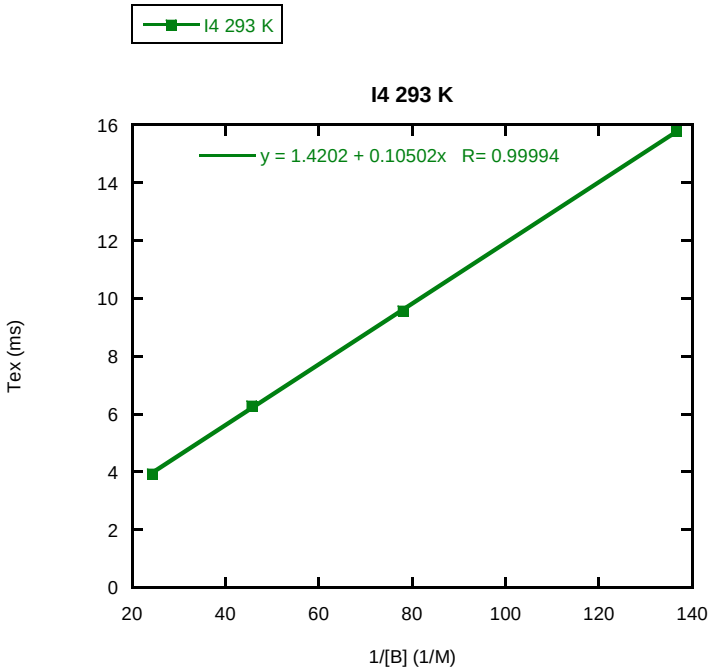


Figure 3.10 The base pair lifetime for the G-hairpin*. The base pair lifetime of G4-C15 is 155 ms in 278 K and 19 ms in 293 K. The base pair lifetime of C5-G14 is 100 ms in 278 K and 31 ms in 293 K.*

3.3.2 Base pair lifetime of I-hairpin




G14 278 K
G14 278 K (HIV-I)
y = 85.602 + 4.1854x R= 0.99448
Tex (ms)
1/[B] (1/M)
0
100
200
300
400
500
0
10
20
30
40
50
60
70
80


I4 293 K
I4 293 K
y = 1.4202 + 0.10502x R= 0.99994
Tex (ms)
1/[B] (1/M)
0
2
4
6
8
10
12
14
16
20
40
60
80
100
120
140

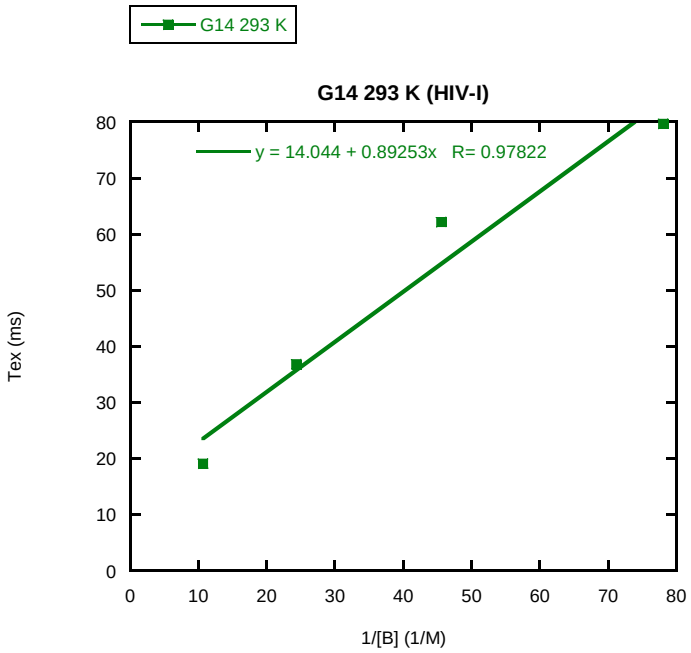


Figure 3.11 The base pair lifetime for the I-hairpin*. The base pair lifetime of I4-C15 is 8.8 ms in 278 K and 2.6 ms in 293 K. The base pair lifetime of C5-G14 is 86 ms in 278 K and 14 ms in 293 K.*

Table 3.5 Base pair lifetime (ms) of G-hairpin and I-hairpin (green for control and red for inosine)

	278 K	293 K
G4	155	19
I4	8.8	2.6
G14	100	31
G14	86	14

3.4 Thermal stability of HIV hairpin substrates

The Tm experiment was performed for HIV-I-hairpin and HIV-G-hairpin. The DNA samples were prepared in ~1 OD and the temperature was from 17 °C to 85 °C, with an increase of 0.5 °C per minute. The Tm of HIV-I-hairpin is about 10 °C lower that the HIV-G-hairpin.

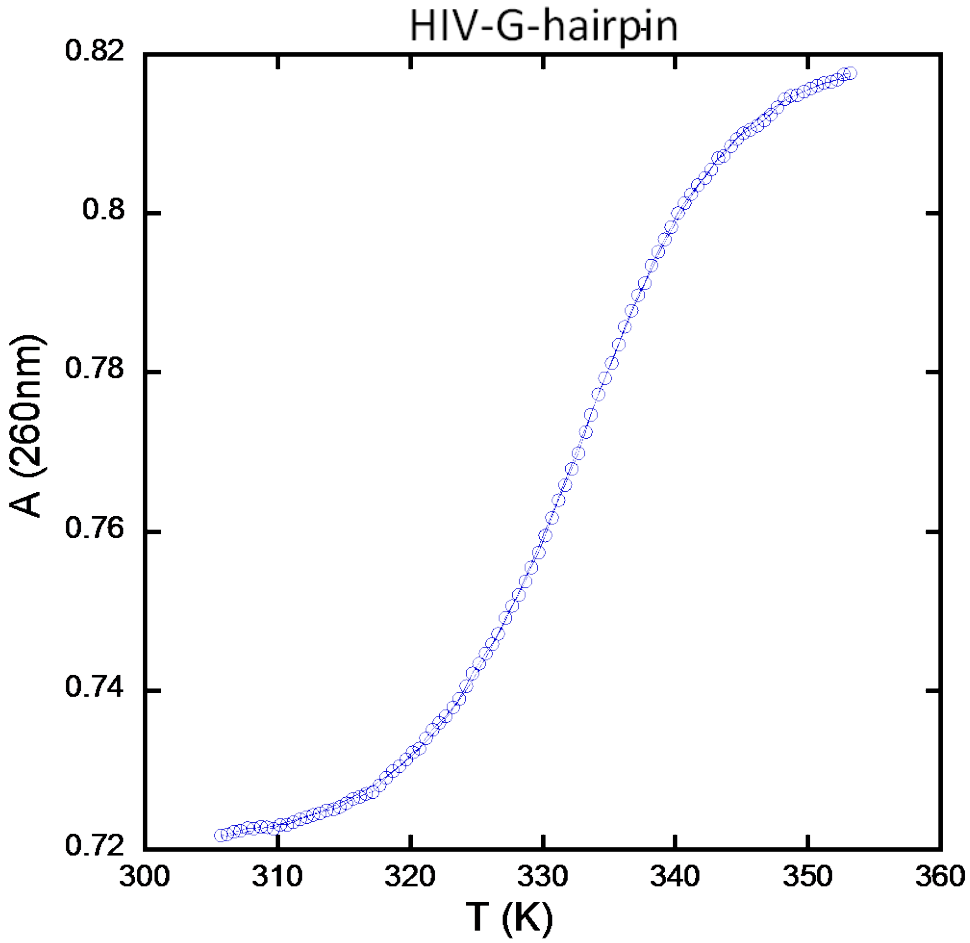


Figure 3.12 Tm curve for HIV-G-hairpin*. Tm is 333.8 K.*

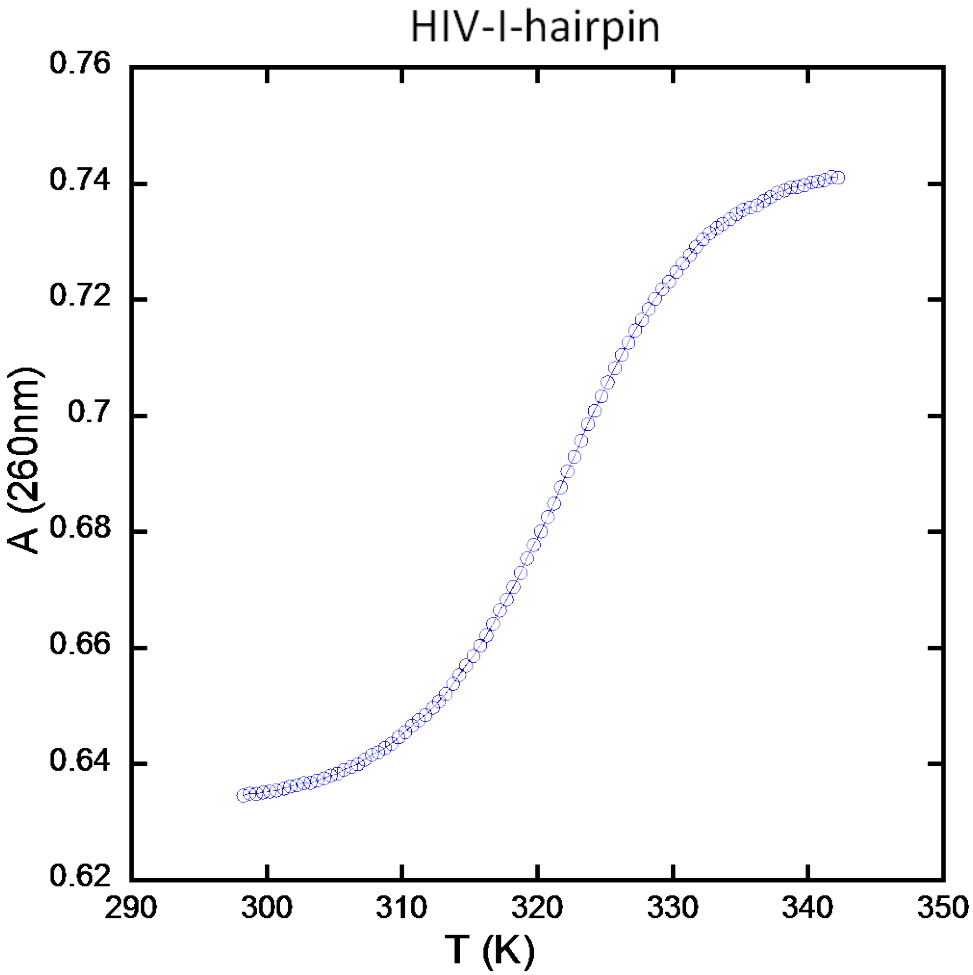


Figure 3.13 Tm curve for HIV-I-hairpin*. Tm is 322.7 K.*

4 CONCLUSIONS

The NMR structures of the HIV hairpins reveal several conserved features of the CA motif, however, there is no clear structural element that provokes recognition by the integrase. The HIV-G and HIV-I hairpin are structurally highly similar, including the CA motif. A major difference between the hairpin substrates is the base pair lifetimes which revealed that the HIV-I hairpin is more dynamic at the I4-C15 base pair location. This would be expected to aid in enzyme recognition. However, the substitution of G with I removes the exocyclic amino group and abolishes integrase 3'- processing. This highlights the importance of the amino group. In addition, the substrate with 2-amino purine is suitable for the enzyme. The absence of this amino group prohibits a sequence specific interaction of the DNA substrate with a glycine carbonyl group which is crucial for enzyme activity. The substrates with T-frayed end and with T-tail on single side are as suitable for the 3'-processing of HIV integrase as the control duplex; the substrates with G4-C15 switch or T3-A16 switch have no cleavage for the 3'-processing; the substrate with T-tail directly following the conserved G-C base pair is not suitable for the 3'-processing; there is a little cleavage of the substrates with terminal loop and the substrate with internal loop has some cleavage. The enzymatic study of HIV integrase suggested that the 3'-processing step does not require substrates with a terminal duplex.

REFERENCES

(1) Robert Craigie, HIV Integrase, a Brief Overview from Chemistry to Therapeutics, The Journal of Biological Chemistry, 2001, Vol. 276, No. 26, 23213–23216

(2) Min Li and Robert Craigie, Processing of Viral DNA Ends Channels the HIV-1 Integration Reaction to Concerted Integration, J. Biol. Chem. 2005, 280: 29334-29339

(3) Sylvia Lee-Huang, Paul L. Huang, Philip L. Huang, Inhibition of the integrase of human immunodeficiency virus (HIV) type 1 by anti-HIV plant proteins MAP30 and GAP31, Proc. Natl. Acad. Sci. USA September 1995, Vol. 92, 8818-8822

(4) Yehuda Goldgur, Fred Dyda, Robert Craigie, and David R. Davies, Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium, Proc. Natl. Acad. Sci. USA, 1998, Vol. 95, 9150–9154

(5) Hare, S., Maertens, G. N., Cherepanov, P., 3' Processing and strand transfer catalysed by retroviral integrase in crystallo, The E.M.B.O. J., 2012, 31, 3020-3028

(6) Brian P. Scottoline, Samson Chow, Viola Ellison and Patrick, O. Brown, Disruption of the terminal base pairs of retroviral DNA during integration, Genes & Development, 1997, 11: 371-382

(7) Dominic Esposito and Robert Craigie, Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein–DNA interaction, The EMBO Journal, 1998, Vol.17 No.19, 5832–5843

(8) Colleen B. Jonsson, George A. Donzella, and Monica J. Roth, Characterization of the Forward and Reverse Integration Reactions of the Moloney Murine Leukemia Virus Integrase Protein Purified from *Escherichia coli,* The Journal of Biological Chemistry, 1993, Vol. 268, No. 2, 1462-1469

(9) Tan Wang, Andrew J. Piefer, and Colleen B. Jonsson, Interactions of the Human T-cell Leukemia Virus Type-II Integrase with the Conserved CA in the Retroviral Long Terminal Repeat End, The Journal of Biological Chemistry, 2001, Vol. 276, No.18, 14710–14717

(10) Aramini, J. M., Cleaver, S. H., Pon, R. T., Cunningham, R. P., and Germann, M. W. (2004) Solution structure of a DNA duplex containing an alpha-anomeric adenosine: insights into substrate recognition by endonuclease IV, J. Mol. Biol. 338, 77–91

(11) Germann MW, Kalisch BW, Lundberg P, Perturbation of DNA hairpins containing the EcoRI recognition site by hairpin loops of varying size and composition: physical (NMR and UV) and enzymatic (EcoRI) studies, Nucleic Acids Res, 1990 Mar 25;18(6):1489-98

(12) Zhengrong Wu, Nico Tjandra, and Ad Bax, Measurement of ^{1}H3'-^{31}P dipolar couplings in a DNA oligonucleotide by constant-time NOESY difference spectroscopy, Journal of Biomolecular NMR, 2001, 19: 367–370

(13) James G. Moe and Irina M. RUSS, Kinetics and Energetics of Base-Pair Opening in 5'-d (CGCGAATTCGCG)-3' and a Substituted Dodecamer Containing GaT Mismatches, Biochemistry, 1992, 31, 8421-8428

APPENDIX

The purification of T7 polymerase

The purification of T7 polymerase is another minor project that I worked on, which independent from the research that discussed before in this thesis.

Introduction

T7 RNA Polymerase (T7 RNAP) is an extremely promoter-specific RNA polymerase from the T7 bacteriophage with a molecular weight of 99 kDa, which can catalyze the formation of RNA in the 5'→ 3' direction and transcribes only DNA downstream of a T7 promoter. T7 RNAP can be used to synthesize pure, single-strand RNA molecules of various length and sequence for *in vitro* transcription reaction. Large-scale RNAs generated by T7 RNAP has been applied widely to the investigation of RNA serving as a biologically active molecule. T7 RNAP is commercially available, however, the concentration is lower than required and the cost for milligram levels of RNA synthesize is prohibitively high. Therefore, the T7 RNAP is generally produced from cells containing a plasmid with the T7 RNA polymerase gene in many laboratories. The E. coli strain, BL21 is used as the host cell for the expression of T7 RNAP. The plasmid pAR1219 is a pBR322-based vector that expresses T7 RNA Polymerase under control of the IPTG-inducible lac UV5 promoter. After the transformation, the enzyme can be induced by isopropyl-β-thiogalactoside (IPTG). To obtain the homogenous and RNase-free T7 RNAP, the purification procedure is necessary to remove the proteins expressed along with the RNAP.

Several procedures have been applied for purifying the T7 RNAP by conventional column chromatography. Here we describe a simple and powerful method to purify the T7 RNAP fast using two columns. The first column is an ion-exchange column, which is used to separate the T7 RNAP from most of other proteins. The second column, an affinity

chromatography, was used for elimination of the RNase contamination and some other proteins that remain. After this procedure, the T7 RNAP can be pure enough to be used to synthesize RNA in large quantities.

Materials and Methods

pAR1219

Plasmid pAR1219 is a pBR322-based vector that expresses T7 RNA Polymerase under control of the IPTG-inducible lac UV5 promoter. Ampicillin or Carbenicillin is used to select for pAR1219.

Strain

Remove the BL21 (DE3) competent cells from -70 °C and thaw cells on the ice for 5 to 10 minutes. Then, take 100 μL of the competent cells into 14 mL BD Falcon polypropylene round bottom tubes which were placed on ice. Mix the cells gently by tapping tube. Add 1 μL of the pAR1219 plasmid into the cells tube. Swirl the pipet tip through the cells while dispensing the plasmid. Mix the cells and plasmid gently. Place the tube on ice and incubate the reaction for 30 minutes. Following that, put cells in a 42 °C water bath for a 45 seconds heat-shock. Add 450 μL preheated SOC medium (prepare immediately before use) to each transformation reaction. Then, incubate the reaction at 37 °C for one hour with shaking at the speed of 225 to 250 rpm. Spread the incubated mixture on LB agar plates containing 10mg/mL Carbenicillin. Incubate the plates overnight at 37 °C. 8 colonies are picked and added respectively into a 5 mL medium for cell growth. The cells are grown overnight at 37 °C with shaking at the speed of 200 rpm. 500 μL cell growth medium and 500 μL of autoclaved 80% glycerol were added in each reaction, which can be stored at -80 °C as T7 RNA polymerase strain. SOC Medium (per 100 mL): 2 mL of 20% (w/v) glucose. Add SOB medium to a final volume of 100 mL, filter sterilized SOB

Medium (per Liter): 20.0 g of tryptone; 5.0 g of yeast extract; 0.5 g of NaCl. Add deionized H_2O to a final volume of 1 L. After being autoclaved, add 10 mL 1 M $MgCl_2$ and 10 mL 1 M $MgSO_4$ prior to use.

Cell growth

A loop of T7 RNA polymerase strain was added into the 25 mL growth medium, which contains 50 μg/mL Carbenicillin. This start culture was shaken at 37 °C at 250 rpm overnight before added into 900 mL of growth medium. The growth medium was also shaken at 37 °C at 250 rpm, and the cell growth was monitored by checking the absorbance at 600 nm. When the absorbance reached 0.7, IPTG was added to a final concentration of 1 mM to induce the expression of T7 RNA polymerase. After adding the IPTG, the cells were grown for 4 hours and pelleted by centrifuging at 4 °C in the speed of 5000 rpm for 45 minutes. These pellets can be resuspended with buffer L and then frozen in the refrigerator at -80 °C for long-term storage.

Cell Lysis

The cell was resuspended with buffer L (see table 1) and lysed by cell disruptor, which breaks the cell and releases the nucleic acids and proteins. 0.2 g Spermidine HCl was added and incubated for 10 minutes to precipitate DNA. The mixture was centrifuged at 4 °C and the supernatant was collected. The supernatant was cleared further by ultracentrifugation at 9000 rpm for 1 hour.

Chromatography: DEAE-Sepharose fast flow

The pre-swollen suspension was mixed with buffer I (see table 1) to form a moderately thick slurry, which consists of 75% settled gel and 25% liquid. Then, the column was mounted vertically on a suitable stand and a small amount of buffer I was poured into the empty column and allow it to flow through spaces to eliminate air pockets. The prepared ion-exchange medium

was poured into the column and let it flow gently down. Buffer I should be added when the gel flows down to avoid dry. After the gel was settled (2.5 x 12 cm), the column was connected to a pump. The column was equilibrated with 300 mL buffer I at a flow rate of 1.0 mL/min, then the sample was loaded onto the column at flow rate of 1.0 mL/min. The column was then washed with 150 mL buffer I to remove the unbounded proteins at a flow rate of 1.5 mL/min. Collect the flow-through. Then the bound proteins were eluted by the total volume of 400 mL with a linear gradient of NaCl from 100 mM (buffer II, see table 1) to 500 mM (buffer III, see table 1), at a flow rate of 1.0 mL/min. The fractions were detected by the 10% polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS–PAGE). The sample and buffer should be on the ice to keep the low temperature.

Chromatography: Cibachron Blue 3GA (Type 3000-CL)

The matrices in suspension were poured into a column which has been washed with buffer IV (see table 1). The matrices then were washed with 3 to 5 column volumes of buffer IV to be settled (1 x 5 cm). After equilibrating with 60 mL buffer IV at a flow rate of 1.0 mL/min, the fractions that concentrated of T7 RNAP, as detected by the SDS-PAGE were applied into the column. The column was washed first with 100 mL of buffer IV and then with 50 mL of buffer IV containing 1 mM ATP and GTP at a flow rate of 1.0 mL/min. The T7 RNAP was then eluted with 100 mL buffer V (see table 1) at a flow rate of 1.0 mL/min. Flow-through and elutes were analyzed by the 10% SDS-PAGE. The sample and buffer should be on the ice to keep the low temperature.

Buffer exchange and RNA transcription

The fraction that concentrated of T7 RNAP was loaded into a 50 K Amicon Ultra centrifugal Millipore filter tube, which can eliminate molecules smaller than 50 kDa. The tube

should be rinsed with buffer VI (see table 1) prior to use. Buffer VI was added to the tube and centrifuged at 3000 rpm by Beckman Allegra 25R Centrifuge with TS-5.1-500 rotor for 15 minutes. The process was repeated three times to change the buffer V to buffer VI. 4.5 mL of T7 RNAP was obtained after buffer exchange. The concentration of T7 RNAP was determined by UV spectrophotometry. The extinction coefficient of T7 RNAP is 140760 $M^{-1}cm^{-1}$. The OD was 1.9 at the wavelength of 280 nm, and the concentration was 14 μM. The yield of the protein was 6.93 mg/L culture. Then, the purified enzyme was tested by RNA transcription. After 10 times dilution, 0.7 μL of the enzyme (~0.84 pmol) was added to 20 μL reaction and incubated at 37 °C for 4 hours, then checked by 12% denaturing polyacrylamide gel electrophoresis in the presence of urea. 0.3 μM DNA template was used.

Reaction (20 μL):

10 x buffer	2 μL
rNTP mixture	4 μL
300 mM $MgCl_2$	1.6 μL
RNase secure	0.8 μL
1 M DTT	0.2 μL
3 μM DNA template	2 μL
DI H_2O	7.4 μL
0.1% Triton X-100	2 μL

Dialysis and storage

The T7 RNAP after the buffer exchange was dialyzed against 450 mL buffer VI diluted with an equal volume of glycerol. The purified T7 RNAP was stored at -20 °C in PH 8.0, 5 mM Potassium phosphate, 50 mM NaCl, 0.05 mM EDTA, 0.5 mM DTT, and 50% glycerol. After the

dialysis, the volume was 1 mL and concentration was 56 μM. After 10 times dilution, the final volume was 10 mL and concentration was 5.6 μM.

TABLE 1

Buffers for Chromatography

Buffers	L	I	II	III	IV	V	VI
NaCl (mM)	50	50	100	500	250	1500	100
EDTA (mM)	10	1	1	1	1	1	0.1
Potassium phosphate (10 mM, pH = 8.0)	√	√	√	√	√	√	√
Glycerol (5%)	√	√	√	√	√	√	√
DTT (1 mM)	√				√	√	√
β-Mercaptoethanol (10 mM)	√	√	√	√			
PMSF (20 mg/ mL)	√						
Total volume (mL)	500	1000	500	500	500	500	1000

The pH of the solution of EDTA and potassium phosphate and the final pH of the buffers should be adjusted to 8.0.

√ means that the component should be added, while blank means no.

The β-mercaptoethanol, DTT should be added after the buffers being autoclaved.

The PMSF should always be added fresh.

Results and Discussion

The purity of the T7 RNAP was detected by the SDS-PAGE. After the two columns, the enzyme was much purified. The final concentration of protein was 5.6 μM. The yield of the protein was about 6.93 mg/L.

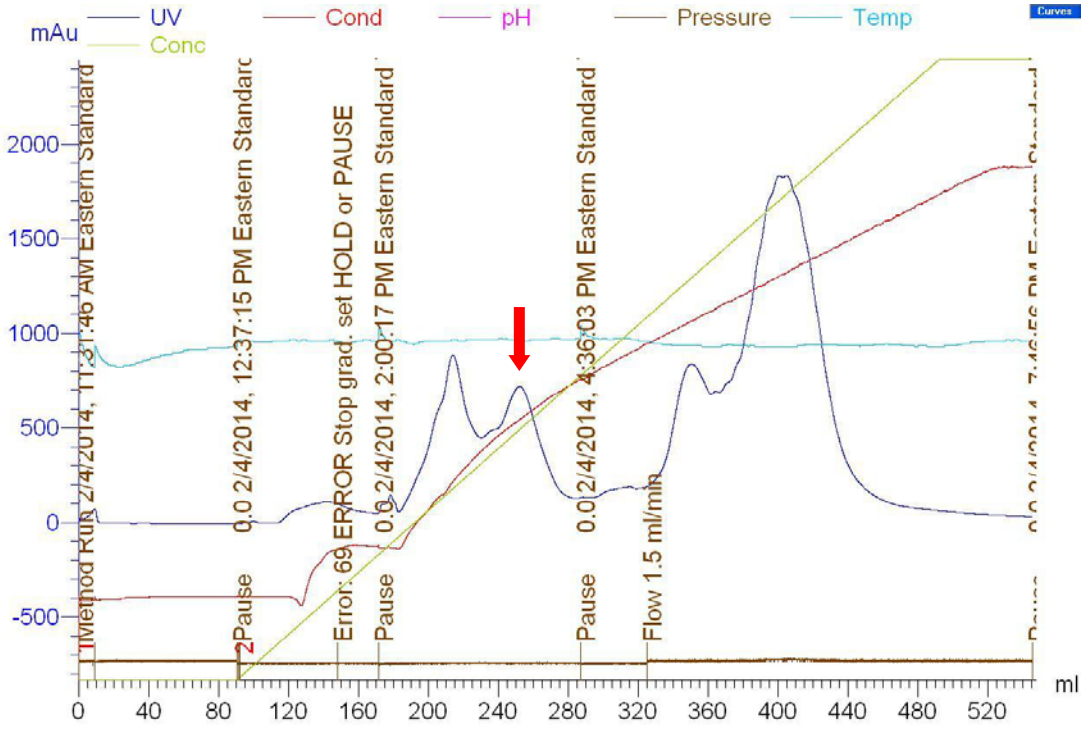


Figure 1 Elution of DEAE-Sepharose column was performed with a linear salt gradient from 100 mM NaCl to 500 mM NaCl at a flow rate of 1.0 mL/min. All peaks were collected and analyzed by 10% SDS-PAGE. The peak marked with the arrow from 210-228 mL contained T7 RNAP.

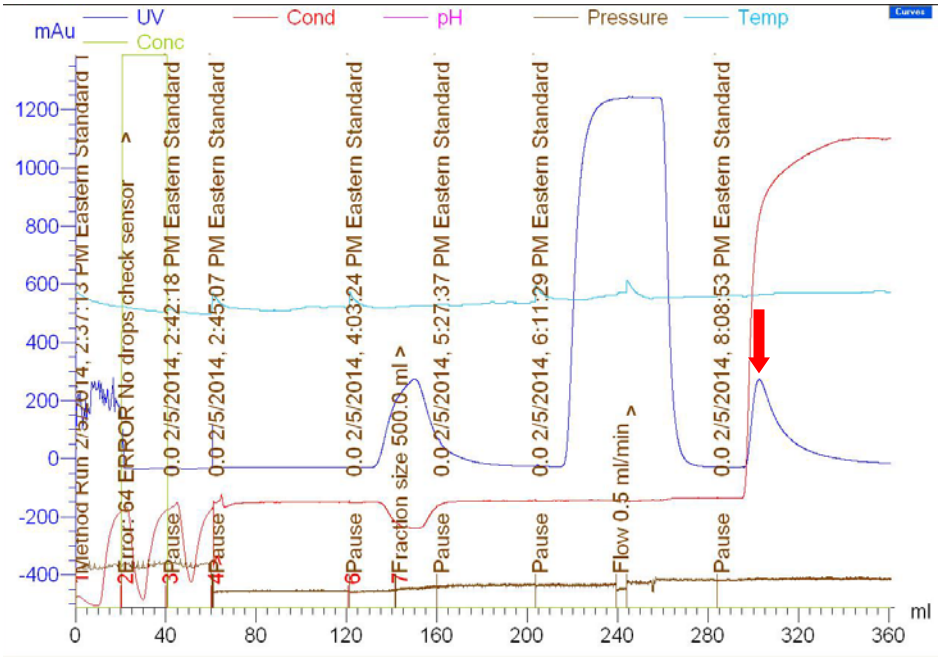


Figure 2 Elution of Cibachron Blue 3GA column was performed with 1.5 M NaCl at a flow rate of 1.0 mL/min. All the fractions of elutes with a peak was collected and detected by 10% SDS-PAGE. The peak marked with the arrow was T7 RNAP.

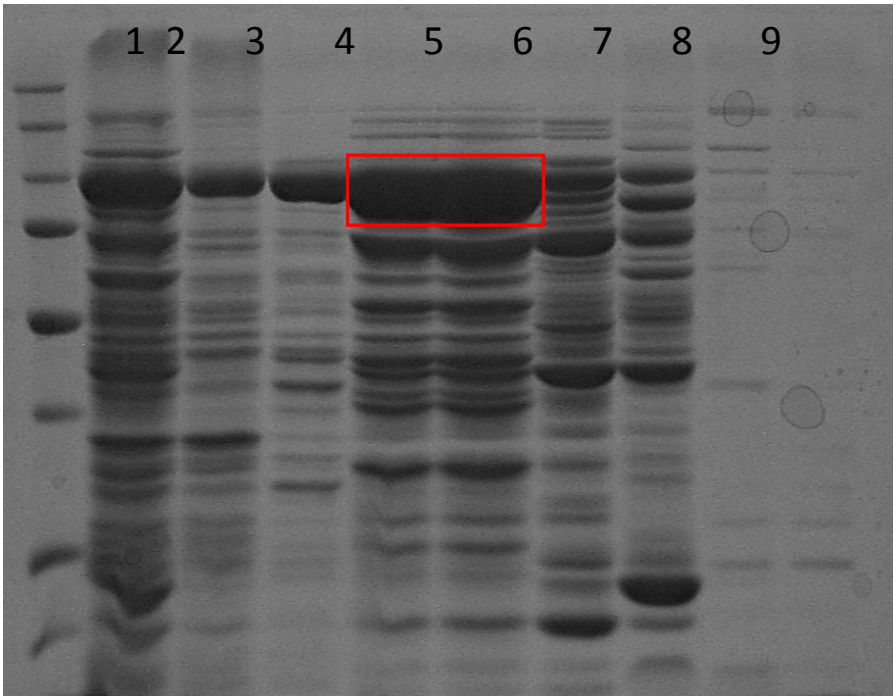


Figure 3 The collections of DEAE-Sepharose column was checked with 10% SDS-PAGE. Lane 1 is the cell lysate; lane 2 is the flow-through; lanes 3-9 are different collections of the elution. The T7 RNA polymerase (99 kDa), which is marked with the red rectangle accumulated in lane 4 and 5 and more pure compared to the crude cell lysate. However, the presence of other proteins indicated that further purification is required.

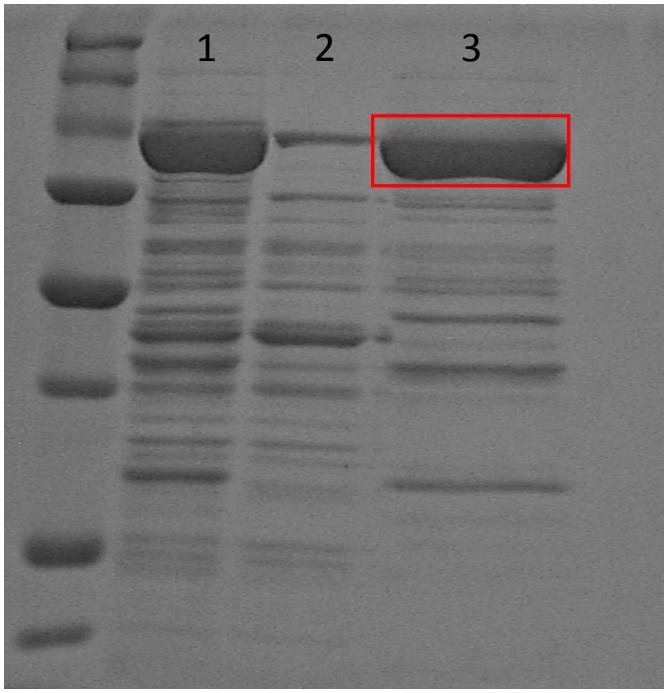


Figure 4 The collections of Cibachron Blue 3GA column was checked with 10% SDS-PAGE. Lane 1 is the elute that concentrated in T7 RNAP from DEAE column; lane 2 is the flow-through; lanes 3 is the elute. The T7 RNA polymerase (99 kDa), which is marked with the red rectangle accumulated in lane 3 and more purified.

To check the activity of the purified RNAP, the RNA transcription reaction was analyzed by 12% polyacrylamide gel electrophoresis in the presence of urea. The RNA synthesized named RREIIB-Tr with a sequence of 5'-GGUCUGGGCGCAGCGCAAGCUGACGGUACAGGCC-3'. The purified T7 could synthesize the RREIIB-Tr-RNA and it has a good efficiency. In 20 μL reaction, RNA with total OD of 3.7 was made.

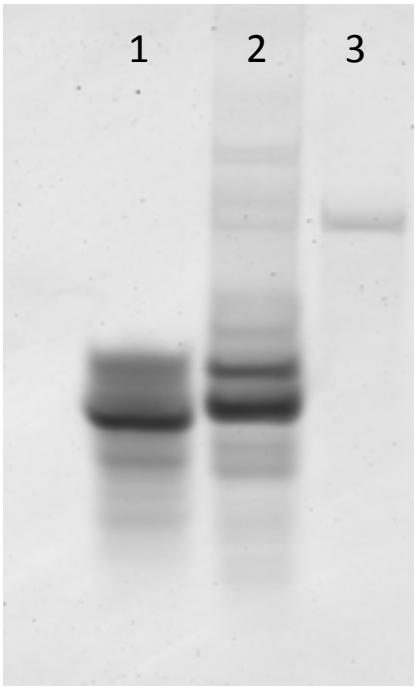


Figure 5 The RNA transcription reaction was checked by a 12% urea gel. Lane 1 is the RNA control; lane 2 is the reaction and lane 3 is the DNA template.

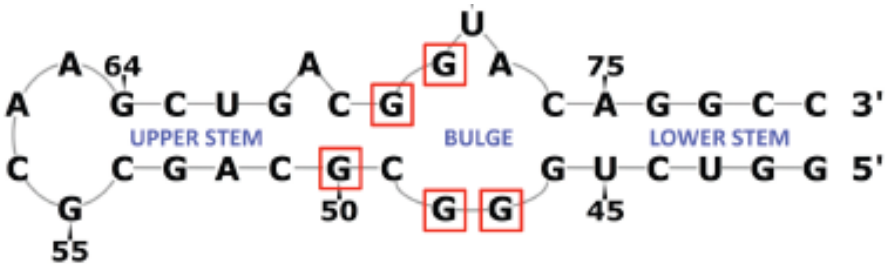


Figure 6 The RREIIB-Tr-RNA is a truncated version (Tr) of the IIB loop of rev responsive element (RRE) in HIV introns. It has 34 nucleotides.

Figure 7 5'-3' RREIIB-Tr-DNA template is used in the RNA transcription reaction to synthesize RREIIB-Tr-RNA by purified T7 RNA polymerase.