

Syracuse University

SURFACE

Chemistry - Dissertations

College of Arts and Sciences

2011

Mutational Analysis of HIV-1 Nucleocapsid Protein and Methods Development for Aptamer Discovery

Wei Ouyang
Syracuse University

Follow this and additional works at: https://surface.syr.edu/che_etd

 Part of the [Chemistry Commons](#)

Recommended Citation

Ouyang, Wei, "Mutational Analysis of HIV-1 Nucleocapsid Protein and Methods Development for Aptamer Discovery" (2011). *Chemistry - Dissertations*. 178.

https://surface.syr.edu/che_etd/178

This Dissertation is brought to you for free and open access by the College of Arts and Sciences at SURFACE. It has been accepted for inclusion in Chemistry - Dissertations by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Abstract

This dissertation focuses on structural and functional studies related to the interaction between proteins and nucleic acids and their applications. The work reported here includes two projects. The first part of the thesis is on the mutational and combinatorial analysis of HIV-1 nucleocapsid protein (NCp7). Sixteen mutants of NCp7 were constructed and over-expressed in *E. coli*. The dissociation constants (K_d) between the mutant proteins and the major packaging signal in HIV-1, SL3 RNA, were measured using tryptophan fluorescence titration assay and/or isothermal titration calorimetry (ITC) assay. Most mutants show tolerance of the mutation but some mutants, including F16A, I24A and K14E-E21K, suffered moderate to significant loss of affinity for SL3 RNA, indicating the roles of the affected residues in the interaction between NCp7 and SL3. The second part focuses on the development of methods for aptamer discovery. Two prototype platforms were set up to imitate the two main-stream next-generation sequencing (NGS) technologies, bridge amplification based and emulsion PCR based NGS technologies. Proof-of-principle experiments were carried out and optimized to evaluate the practicality of these platforms in high-throughput and multiplexed aptamer discovery.

Mutational Analysis of HIV-1 Nucleocapsid Protein and Methods Development for Aptamer Discovery

By

Wei Ouyang

M.S. Nanjing University 2001

B.S. Nanjing University 1998

DISSERTATION

Submitted in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in Structural Biology, Biochemistry and Biophysics
in the Graduate School of Syracuse University

July 2011

UMI Number: 3461833

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent on the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3461833

Copyright 2011 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Copyright 2011 Wei Ouyang

All rights Reserved

Committee Approval Page (place holder – to be generated by GME Center and
forwarded to the Chair of the Oral Examination)

Table of Content

Abstract	i
Acknowledgement	ix
Part I. Mutational and Combinatorial Analysis of HIV-1 Nucleocapsid Protein (NCp7) ·	1
Chapter 1. Introduction	2
1.1. AIDS and HIV	2
1.2. HIV Replication cycle	3
1.3. Gag Polyprotein and NC (Nucleocapsid) Domain	7
1.4. SL3 Packaging Signal	13
1.5. Interactions within the NCp7-SL3 complex	15
1.6. Previous mutational studies	18
1.7. Specific Aims	19
Chapter 2. Construction, Expression, Purification and Characterization of Variants of NCp7 Protein	29
2.1. Introduction	29
2.1.1. Site-directed mutagenesis	29
2.1.2. Over-Expression and Purification of Recombinant NCp7	31
2.1.3. Protein characterization assays	32
2.2. Materials and Methods	34
2.3. Results	41
2.4. Discussion	43
Chapter 3. Quantitative Binding Analysis of NCp7 Variant Proteins with SL3 RNA ·	59
3.1. Introduction	59
3.1.1. Tryptophan fluorescence titration	59
3.1.2. Isothermal titration calorimetry	61
3.2. Materials and Methods	63
3.3. Results	65
3.4. Discussion	67
3.4.1. Single mutants	68
3.4.2. Salt-bridge switching double mutants	72
3.4.3. Comparison of Trp titration and ITC titration results	73
3.4.4. ITC data analysis	75
3.5. Conclusion	77
3.6. Suggestions for future work	79
Part II. Methods Development for Aptamer Discovery	92
Chapter 4. Introduction	93
4.1. Aptamers	93
4.2. SELEX	96
4.3. Acyclic identification of aptamers	99
4.4. Next-generation sequencing (NGS) technologies	101

4.4.1. Bridge amplification based NGS technology (Illumina)	102
4.4.2. Emulsion PCR based NGS technology (454/SOLiD/Polonator)	102
4.5. Potential of NGS in aptamer discovery	103
4.6. Specific aims	105
Chapter 5. Experimental Design and Preliminary Data	109
5.1. Experimental design	109
5.1.1. Bridge amplification	110
5.1.2. Emulsion PCR	111
5.2. Methods and Materials	112
5.2.1. Methods for bridge amplification based detection	112
5.2.2. Methods for emulsion PCR based detection	113
5.3. Preliminary Results	117
5.3.1. Bridge amplification	117
5.3.2. Emulsion PCR	119
5.4. Discussion	121
5.5. Conclusions and suggestions for future work	123
5.5.1. Bridge amplification	124
5.5.2. Emulsion PCR	124
Appendices	141
Appendix A. NCp7 Site-Mutagenesis Sequence information	142
A-1. Map of pET-3a plasmid	142
A-2. Full sequence of the pRD2 construct (from HIV-1 pNL4-3 strain)	143
Appendix B. Recipes and Protocols Used in NCp7 Mutagenesis	145
B-1. Recipes for medium, buffers and solutions	145
B-2. Protocol for plasmid preparation	153
B-3. Protocol for site-directed mutagenesis	155
B-4. Protocol for transformation into XL1-Blue competent cells	157
B-5. Protocol for transformation into BL21 competent cells	159
B-6. Protocol for making or renewing frozen stock of <i>E. coli</i> strains	161
B-7. Protocol for NCp7 protein over-expression	162
B-8. Protocol for cell lysis	165
B-9. Protocol for FPLC purification of NCp7	167
B-10. Protocols for cleaning and storing FPLC columns	169
B-11. Protocol for FPLC column packing	171
Appendix C. NC-SL3 Titration Data	173
C-1. Compilation of NCp7-SL3 Trp titration data	173
C-2. Individual NCp7-SL3 Trp Titration Curves	182
C-3. Individual NCp7-SL3 ITC Curves	224
Appendix D. Recipes and Protocols Used in the Development of Methods for Aptamer Discovery	267
D-1. Recipes for buffers and solutions	267
D-2. Protocol for bridge amplification and detection	271

D-3. Protocol for emulsion PCR and detection	274
D-4. Protocol for aminosilane treatment of glass slides	284
D-5. Protocol for labeling protein with Cy3 or Cy5	286
References	288
Biographical Data	302

List of Tables

Table 2-1. Site-directed mutagenesis primers for construction of NCp7 mutants.....	47
Table 2-2. Observed and calculated molecular weight of purified NCp7 proteins	48
Table 3-1. Trp titration results of NCp7 wild type and mutant proteins.....	81
Table 3-2. ITC titration results of NCp7 wild type and mutant proteins.....	82
Table 5-1. Template and primer sequences used in aptamer discovery method development	125

List of Figures

Figure 1-1. General features of HIV-1 replication cycle	22
Figure 1-2. Schematic representation of the HIV-1 genome and virion.....	23
Figure 1-3. The amino acid sequence of HIV-1 NCp7 protein.....	24
Figure 1-4. Variation in 1700 NCp7 sequences.....	25
Figure 1-5. Secondary RNA structure model of HIV-1 5' un-translated region (5'-UTR).....	26
Figure 1-6. NMR structure of NCp7-SL3 complex	27
Figure 1-7. Hydrophobic cleft on NCp7-SL3 complex	28
Figure 2-1. Schematic procedure of site-directed mutagenesis	49
Figure 2-2. HIV-1 NCp7 variants characterized in this study.	50
Figure 2-3. Sequence alignment of NCp7 wild type and mutants	51
Figure 2-4. Typical FPLC chromatogram of NCp7 purification	52
Figure 2-5. SDS-PAGE gel image of NCp7 protein sample (Tris-Tricine)	53
Figure 2-6. SDS-PAGE gel image of NCp7 protein sample (Invitrogen Precast Gel).....	54
Figure 2-7. SDS-PAGE gel image of NCp7 protein sample (Bio-Rad Precast Gel).....	55
Figure 2-8. SDS-PAGE gel image of NCp7 protein sample (Lonza Precast Gel)	56
Figure 2-9. Sample MALDI-TOF spectrum of NCp7 protein variants	57
Figure 2-10. Sedimentation velocity profile for NCp7 wild type.....	58
Figure 3-1. Schematic representation of isothermal titration calorimetry (ITC)	83
Figure 3-2. HPLC chromatogram of SL3 RNA.....	84
Figure 3-3. Tryptophan fluorescence titration curves of NCp7 wild type and mutant proteins with SL3 RNA.....	85
Figure 3-4. The effect of site-directed mutagenesis on the K_d values of NCp7 mutant proteins for SL3 RNA	86
Figure 3-5. Comparison of the Trp Titration and ITC Data	87
Figure 3-6. Representative ITC titration curves of NCp7 WT and mutant proteins.....	88

Figure 3-7. Comparison of the changes in NCp7 protein's affinity for SL3 RNA by alanine mutations on different residues.....	89
Figure 3-8. Tryptophan fluorescence titration curve of K14E-E21K for SL3 RNA	90
Figure 3-9. Positions of the three salt bridges in NC-SL3 complex	91
Figure 4-1. Schematic SELEX (Systematic Evolution of Ligands by Exponential Enrichment) procedure.....	106
Figure 4-2. Comparison of SELEX and AIA.....	107
Figure 4-3. Template preparation strategies in NGS	108
Figure 5-1. Formation of DNA clusters by bridge amplification	126
Figure 5-2. TBA 55mer.....	127
Figure 5-3. TBA scrambled 55mer	128
Figure 5-4. Detection of DNA clusters after bridge amplification	129
Figure 5-5. Schematic setup of emulsion PCR	130
Figure 5-6. TBA clusters formed by bridge amplification	131
Figure 5-7. Open loop template design.....	132
Figure 5-8. Detection of emulsion PCR beads by TBA probe (6-FAM).....	133
Figure 5-9. Detection of emulsion PCR beads by Thrombin (Cy3)	134
Figure 5-10. Detection of emulsion PCR beads by Thrombin (Cy5)	135
Figure 5-11. Detection of control beads by Thrombin (Cy5)	136
Figure 5-12. Detection of emulsion PCR beads by Thrombin (Cy3)	137
Figure 5-13. New emulsion PCR template design.....	138
Figure 5-14. Detection of emulsion PCR beads by TBA probe (6-FAM).....	139
Figure 5-15. Detection of emulsion PCR beads by Thrombin (Cy5)	140

Acknowledgement

I would like to thank my advisor, professors, colleagues, friends and family for their guidance, support and encouragement during my graduate studies at Syracuse University.

My utmost gratitude goes to my advisor, Prof. Philip N. Borer, for his insightful guidance and inspirations, for his selfless support and help and for his great patience and optimism. The work reported in this dissertation would not have been possible without his constant support and motivation. What I learned in the Borer lab will benefit me for life.

I would like to thank Prof. Michael S. Cosgrove, Prof. Melissa E. Pepling, Prof. Robert P. Doyle, Prof. James Hougland and Prof. Stephan Wilkens, for taking time to serve in my oral exam committee. Special thanks also go to Prof. Michael S. Cosgrove, Prof. Stephan Wilkens, Prof. Bruce Hudson, Prof. Ramesh Raina, Prof. Roy D. Welch, Prof. Mark S. Braiman, Prof. H. Ernest Hemphill, Prof. John Belote and Prof. Tom Duncan for lending their equipments and advices I needed for completing this work.

I'm also very grateful for the chance to have known and worked with the members of the Borer lab. They are the most hard-working and fun-loving people I've ever worked with. Especially I want to thank Stephen Okaine for his great dedication and enormous

help at the very beginning of this project. I'd also like to acknowledge Dr. Mark P. McPike, Dr. Christopher DeCiantis, Dr. Deborah J. Kerwood, Dr. Michael Shubsda, Dr. Lingchun Yang, Dr. Shreyas S. Athavale, Dr. Lei Chen, Dr. Gillian V. Kupakuwana, Dr. Michelle F. Homsher, Dr. Damian Allis, Dr. Troy M. Lam, James Crill II, Collin Fischer, Raghuvaran M. Iyer, Caitlin M. Jenne and Hoi Cheung. They made my life in graduate school joyous and fruitful. I'll always cherish the happy moments I shared with them in the Borer lab.

I would also like to extend my gratitude to all my friends for their support and help in my life and career, to the friendly and helpful staff members in the Department of Chemistry, the Department of Biology, the Program of Structural Biology, Biochemistry and Biophysics (SB3) and the Center for Advanced Systems and Engineering (CASE) at Syracuse University, for their acceptance and assistance in the past years. I am grateful to Syracuse University for providing me the opportunity and resources to complete my graduate study.

Last but not least, I want to thank my family, for their unconditional love and care, for always having faith in me and always being there for me. Especially, I want to thank my beloved wife, Jinyuan Yan, for her love, support and sacrifice, for being the better half of me and for all the happiness and sadness we shared and will be sharing for many years to come.

**Part I. Mutational and Combinatorial
Analysis of HIV-1 Nucleocapsid Protein
(NCp7)**

Chapter 1. Introduction

This dissertation focuses on structural and functional studies related to the interaction between proteins and nucleic acids and their applications. Work on two projects is reported in the following chapters in two parts. The first part of the thesis is on the mutational and combinatorial analysis of HIV-1 nucleocapsid protein, also known as NCp7. It covers the topics of construction of HIV-1 NCp7 mutants via site-directed mutagenesis, characterization of the mutant and wild type proteins, affinity assays for NCp7 variants with HIV-1 SL3 RNA, including tryptophan fluorescence titration and isothermal titration calorimetry. The following is a general introduction to these topics.

1.1. AIDS and HIV

Acquired Immunodeficiency Syndrome (AIDS) is a disease that affects human immune system. It progressively breaks down the patient's immune system and makes the victims more vulnerable to infections and tumors. AIDS is categorized as a bloodborne infectious disease and a sexually transmitted disease (STD) because it can be passed on by an exchange of bodily fluids [1]. Since its first public report in 1981 by the USCDC [2], AIDS has quickly escalated from an epidemic into a global pandemic. In 2010, the Joint United Nations Programme on HIV/AIDS (UNAIDS) estimated in its annual global report that there were 33.3 million people living with the disease at the end of 2009, including

2.6 million people that were newly infected in the year of 2009. The disease killed 1.8 million people in 2009, including 260,000 children of age 15 or younger. At present, AIDS has a prevalence of 0.8% in global adult population aged 15-49 [3]. These numbers are likely to increase further as researchers still have found no cure for AIDS after almost thirty years of hard work [4].

Human Immunodeficiency Virus (HIV) has been recognized as the causative agent of AIDS. It's a retrovirus, meaning its genetic information is carried by RNA instead of DNA. HIV uses reverse transcriptase to transcribe its RNA to DNA, which in turn integrates into the host's DNA [5]. HIV can be divided into two species, HIV-1 and HIV-2, with HIV-1 being more infective and prevalent [6, 7]. As of today, we have been able to control the disease with an antiretroviral drug cocktail regimen, by targeting multiple stages of HIV's replication cycle [8-10]. But current treatments are unable to eradicate the dormant virus integrated in the genomic DNA of the host. Moreover, HIV virus is known for its high mutation rate [11-15]. So new anti-HIV drugs and vaccines are needed to attack resistant strains, reduce side effects, improve the effectiveness of therapies and prevention methods, as well as to control treatment cost [16-20].

1.2. HIV Replication cycle

Understanding HIV's replication cycle has been key to the treatment of AIDS. Many details have been revealed about HIV and other retroviruses. As depicted in Figure 1-1

(D'Souza and Summers, 2005 [21]), a typical replication cycle of HIV-1 can be roughly divided into 9 steps:

Step 1: Attachment. The surface subunit of HIV-1's Env protein, gp120, recognizes and binds to the CD4 receptors on the surface of HIV-1's host cells (T cells, macrophages and dendritic cells) [22]. The binding to CD4 induces a conformational change of gp120 [23, 24], leading to its high-affinity binding to chemokine receptors, including CCR5 and CXCR4 [25, 26], which function as co-receptors for HIV-1 on the host cell surface.

Step 2: Fusion. Interactions of gp120 with co-receptors trigger a cascade of conformational changes of Env that result in activation of the trans-membrane subunit of Env, gp41 [27]. The free energy released when gp41 undergoes conformational change then induces the fusion of virion and cell membranes [28]. The HIV-1 capsid core containing its genetic material is then released into the cytoplasm of the host cell.

Step 3: Uncoating and Reverse Transcription. After viral entry, the HIV-1 capsid core disassembles, in which the capsid proteins (CA) are dissociated from the HIV-1 nucleoprotein complex. This event is called uncoating. Uncoating of the capsid core is accompanied with the formation of subviral particles called reverse-

transcription complexes (RTCs) and pre-integration complexes (PICs) [29, 30], in which the positive strand of viral RNA is transcribed to double-stranded DNA [5].

Step 4: Integration. The PICs containing the double-stranded proviral DNA are transported through the nuclear membrane. The proviral DNA is then inserted into the genomic DNA of the host with the help of the viral integrase [31]. The integrated proviral DNA is replicated together with the host genome but lies dormant in the latent stage of the replication cycle until being activated or induced by cellular or viral signals [32, 33].

Step 5: Transcription and Translation. Utilizing the host's transcription machinery, HIV-1 viral transcription from proviral DNA is initiated and driven by cellular transcription factors and the viral trans-activating protein Tat [34]. In the early stage, most mRNA is spliced and used to encode viral accessory proteins Tat, Rev and Nef. Full-length viral RNA is later transported to the cytoplasmic ribosome for translation with the help of Rev protein. Three viral polyproteins, Gag, Gag-Pol and Env, are synthesized in the cytoplasm and at the endoplasmic reticulum. These polyproteins are later processed by proteolytic cleavage to generate structural and functional viral proteins [35].

Step 6: Genome Recognition. After the viral Gag and Gag-Pol polyproteins are synthesized, they recognize the full length viral RNA by the interaction between

the NC domain of Gag protein and the packaging signal at the 5'-end of viral RNA. Upon recognition of the viral RNA, Gag polyproteins assemble into multimeric structures around the RNA molecule and localize to the cell membrane [36].

Step 7: Assembly. As Gag polyproteins accumulate on the inner surface of the plasma membrane, they recruit the Env polyproteins, which are transported from the ER onto the cell surface via the secretory pathway, and start the assembly process of virus particles [35]. An immature particle is first formed with a relatively stable capsid core of spherical shape [37].

Step 8: Budding. The assembly of immature virion particles induces the formation of a bud on cell membrane and eventually the departure of the bud from host cell membrane. During this process, the virion particle acquires a plasma membrane-derived lipid envelope that incorporates Env polyproteins and is tightly associated with a layer of Gag polyproteins. Budding offers a non-lytic approach for virus egress and allows the host cell to remain viable and continue to produce more progeny virions [37].

Step 9: Maturation. The freshly budded virion particle is immature and non-infectious. During and immediately after budding, the virion undergoes a dramatic morphologic change known as maturation. This step involves proteolytic cleavage of the Gag and Gag-Pol polyproteins and a large-scale rearrangement of

the cleavage products resulting in the structural transition of the virus core from a spherical to a conical shape [37-39].

1.3. Gag Polyprotein and NC (Nucleocapsid) Domain

The HIV-1 genome contains nine open reading frames (ORFs), as shown in Figure 1-2 (Freed, 1998 [36]). Three of these ORFs encode the Gag, Pol and Env polyproteins, which are later processed by viral and cellular proteases to generate individual structural proteins and enzymes [35]. The other six ORFs encode six accessory proteins that have regulatory or protective roles in the viral replication cycle [35, 40].

The 55kD Gag polyprotein is a major structural protein of HIV-1. During the maturation step of the viral replication cycle, it is proteolyzed into four proteins, MA (matrix), CA (capsid), NC (nucleocapsid) and p6, which make up the virion core containing the viral genomic RNA [35]. The Gag protein has important roles in many steps of HIV-1's replication cycle, including viral particle assembly, localization, budding, maturation, early post-entry steps and reverse transcription [36].

Many of the Gag polyprotein's functions are associated with the NC (nucleocapsid) domain or the mature NC protein [36, 41-56]. The most significant function of the NC domain or NC protein is to recognize the viral genomic RNA for correct virus packaging in a sequence-specific manner [57-61]. In the gag precursor, the NC domain is

responsible for full discrimination of genomic from non-genomic RNA via several RNA-NC interactions [62-66]. NC also interacts with viral RNA or proviral DNA in a sequence-nonspecific manner. In a HIV-1 virion particle, there are 2000 to 3000 gag precursors coating the dimeric viral RNA, equivalent to about one NC protein per every seven nucleotides [42, 67]. This Histone-like coating pattern of NC protects viral RNA or proviral DNA from nuclease digestion [46]. The NC domain or mature NC protein can destabilize sequence-nonspecific nucleic acid duplex [52], facilitate nucleic acid strand annealing and complex nucleic acid secondary structure rearrangements [52], chaperone viral RNA's conformational rearrangement [68], activate viral RNA's dimerization [69], anneal the primer tRNA onto the genomic RNA for reverse transcription [70], and facilitate efficient minus- and plus-strand transfer in reverse transcription [71]. NC protein promotes primer-specific synthesis of proviral DNA and reduces non-specific self-primed reverse transcription of viral and cellular RNAs. NC also contributes to the high mutation rate of HIV-1 by enhancing the incorporation of mutations during minus strand DNA elongation [42].

NC also interacts with various viral and cellular proteins to perform its multiple functions. NC domain is essential to the self association of Gag polyprotein. It works in concert with MA and p6 domains as the driving force behind Gag multimerization and the assembly process of HIV-1 [72, 73]. NC is directly involved in the recruitment of HIV-1 reverse transcriptase (RT) into nucleoprotein complexes. In collaboration with Vif, NC enhances RT processivity to promote synthesis of full-length cDNA [44, 74, 75]. NC is

also reported to regulate the timing of reverse transcription by inhibiting premature cDNA synthesis by RT during virion assembly [76, 77]. NC is required for incorporation of Vpr into the viral particle. Moreover, NC improves RNA recognition by Vpr through interaction with the C-terminus of Vpr [78, 79]. NC protein can stimulate the integration of HIV-1 proviral DNA by HIV-1 integrase [80, 81]. By directly binding F-actin, NC domain mediates the association of Gag precursor with the host cell cytoskeleton and plays a role in virus assembly and budding [82]. NC domain binds cellular protein Alix and cooperates with p6 domain on Gag precursor to recruit cellular proteins, such as HP68, Tsg101, EF-1 α , Staufen, VAN and many other cellular proteins, which are proved or hypothesized to participate in viral assembly, budding, morphogenesis and other steps of HIV-1 replication cycle [55, 83, 84]. The cellular protein APOBEC3G, which is shown to have an anti-retroviral activity, is encapsidated into virus-like particles through a direct interaction with NC [49, 85]. Drugs to attack the nucleocapsid have the potential to interfere with critical functions at many stages of the viral replication cycle [45, 86].

The mature HIV-1 nucleocapsid protein contains 55 amino acid residues (see Figure 1-3) and has an apparent molecular weight around 7kD, hence the name NCp7. NCp7 has a highly conserved amino acid sequence among HIV-1 strains, as shown in Figure 1-4 (Lin and Borer, unpublished data [87]), and has several distinct structural features:

- (1) NC or NCp7 contains two conserved CCHC zinc finger motifs in the form: Cys-X₂-Cys-X₄-His-X₄-Cys (X is any amino acid residue) [88-90]. This CCHC type zinc finger

motif is highly conserved among retroviral nucleocapsid proteins and is relatively rare among cellular zinc finger proteins in which CCHH or CCCC motifs are more common [91, 92]. In coordination with zinc ions in a 1:1 stoichiometry, these zinc fingers maintain a unique backbone structure and play an essential role in viral genomic RNA packaging and infectivity. The zinc fingers, especially the intact N-terminal zinc finger, are required for the binding specificity and target selectivity of NC for viral genomic RNA [93, 94]. Intact structure of the zinc fingers is required for the temporal control of reverse transcription by NC throughout the virus replication cycle [76, 77]. Mutations within the zinc fingers are usually inhibitory. Zinc finger mutations can impair intracellular HIV-1 Gag localization, virion biogenesis and maturation and infectivity [95]. Mutations that render NC incompetent for zinc binding can destroy the capacity to recognize and package genomic RNA [91, 96, 97]. Zinc finger interactions contribute to NC's nucleic acid chaperone activity by facilitating efficient minus- and plus-strand transfer and nucleic acid annealing [71, 98, 99]. The conserved aromatic and hydrophobic residues, including F16, I24, A25 of zinc finger 1 and W37, Q45, M46 of zinc finger 2, are displayed on the protein surface and form a hydrophobic core that could serve as the nucleic acid interactive site [89]. Besides interaction with viral RNA, the zinc fingers are also found to directly interact with other proteins such as Vpr [78]. The N-terminal zinc finger (F1 in Figure 1-3) is more conserved among retroviruses and is proved more essential than the C-terminal zinc finger

(F2 in Figure 1-3), whereas F2 serves an accessory role in NC's chaperoning activity [99-103].

(2) One important feature of NC domain or NCp7 is its high content of basic residues.

At neutral pH, NCp7 carries a net charge of +9. The basic residues in NC are important for virus budding. The mutation of basic residues in NC can cause a pronounced decrease in virus release from cultured cells [56]. The first zinc finger of NC is flanked on both sides by a highly conserved and basic consensus sequence, RXXRK, where X represents any amino acid residue. This motif is also present in other retroviruses and many non-retroviral RNA-binding proteins and is probably responsible for the unique hybridization properties of NC proteins, which enhance RNA dimerization, tRNA annealing to the RNA primer binding and the turnover rates of the interaction of a ribozyme with its RNA substrate [104, 105]. Basic residues in those motifs as well as the first zinc finger have been shown to be required for specific binding of NC to HIV-1 RNA. In particular, R7, R32 and K33 are the most essential basic residues for the interaction between NCp7 and viral RNA [102, 104]. There's an overall proportional relationship between the basicity of NC mutants and their non-specific RNA binding activity [104, 106]. Basic residues in NC domain are required for Gag-Gag interactions and Gag multimerization. But the specific interaction between Gag and viral RNA is more dependent on the protein tertiary structure than on the basic residues [107, 108].

- (3) The linker region between the two zinc fingers has a highly conserved short basic sequence, ²⁹RAPRKKG³⁵. Most notably, the proline residue at position 31 occurs only in the *trans* conformation and induces a bend in the otherwise flexible linker that is likely to bring the two fingers into close proximity. Mutations that could destabilize the structure of the linker region are detrimental to the infectivity of virion particles. Substitution of Pro³¹ by D-Pro³¹ causes spatial rearrangement of the linker backbone, which leads to a severe reduction of viral RNA dimerization *in vitro* [109]. The mutation P31L inhibits incorporation of Gag-Pol polyprotein and tRNA^{Lys3} into virion particles and results in a total loss of RT activity and the formation of non-infectious and immature viral particles [109, 110]. The overall structure, rather than the basicity, of the central globular domain of NC, which consists of the two zinc fingers and the linker region, has been shown to be critical to the structure and infectivity of virions [108].
- (4) Another characteristic of NC is that it contains one and only one tryptophan residue (Trp³⁷), whose side chain resides in a hydrophobic core between the two zinc motifs and is key to NC's interaction with viral RNA. Tight and saturated binding of RNA molecule onto NC can quench the natural fluorescence of Trp³⁷ by over 90%. Therefore the tryptophan fluorescence in NCp7 can be used to monitor and measure the binding of viral RNA to NC in *in vitro* assays [111].

Its multitude and significance in function and its sequence conservation and distinct structure make the NC domain or mature NC protein an attractive target for anti-HIV drug discovery. Even though there hasn't been an approved NC inhibitor as anti-HIV drug because of lack of selectivity and efficacy in drug candidates so far [112, 113], researchers have been actively working on finding an effective anti-HIV drug that targets NC domain or NCp7 [45, 86, 114-119]. Investigation of the structure-function relationship in the interaction between NCp7 and viral RNA would shed light on the rational design of anti-HIV drugs.

Mature NCp7 interacts with nucleic acids *in vitro* in a similar manner as NC domain does *in vivo* [120, 121]. Synthetic or recombinant NCp7 protein has been widely used as a model for structural and functional studies on NC and NC domain [57, 102, 104, 121-124]. Therefore in this project I used recombinant NCp7 mutant and wild type proteins over-expressed in *E. coli* as targets to investigate the interactions between them and viral RNA *in vitro*.

1.4. SL3 Packaging Signal

The HIV-1 genome contains two copies of a 9.7-kb RNA molecule in a dimeric form, which are coated with nucleocapsid proteins [125]. Besides serving as the HIV-1 genome, the full length viral RNA also functions as an mRNA to encode the Gag and Gag-Pol polyproteins [126]. Many functions of NC domain or NCp7 involve interactions between

NC and the viral RNA. NC demonstrates both specific and non-specific interactions with RNA. The non-specific interaction can be attributed largely to its high basicity, whereas the specific interaction likely results from the structure of NC domain and/or the sequence and structure of viral RNA, especially the 5'-untranslated region (5'-UTR) on the viral RNA.

The 5'-untranslated region (5'-UTR) on the viral RNA is the most conserved part of the HIV-1 genome and contains multiple distinct secondary structures, as shown in Figure 1-5 [57, 127-129], that can mediate several crucial steps in the viral replication cycle [130]. These structural elements include a *trans*-acting responsive (TAR) hairpin which binds the transcriptional *trans*-activator protein Tat to regulate viral transcription [131, 132]; a polyadenylation signal hairpin (poly(A)) which is inhibited by downstream splicing signals [133, 134]; a primer binding site (PBS) domain where the tRNA^{Lys3} anneals to the viral RNA as a primer to initiate reverse transcription [135, 136]; three stem-loops containing the dimerization initiation site (DIS or SL1) [137, 138], the splice donor site (SD or SL2) [139, 140] and the packaging signal (Ψ or SL3) domain [139, 141]; as well as the start codon of the Gag polyprotein.

The NC domain or NCp7 can interact with many RNA motifs on the 5'-UTR region [53, 69, 98]. But NC bears the highest specificity and affinity for the major packaging signal SL3 [123, 142], which is considered as NC's major RNA partner and primary recognition site on HIV-1 genome. SL3 is highly conserved in sequences among different strains of HIV-1

and well structured, which makes it of particular interest to many researchers [57, 61-66, 141, 143]. Inhibition of NC-SL3 interaction can significantly reduce the release of virus particles from infected cells[144]. Deletion of the packaging signal causes loss of HIV-1 specific RNA from the virus-like particles being released from infected cells [141]. Disruption of the stem-loop structure of SL3 significantly diminishes both viral RNA packaging and dimerization [145]. Further study of the interaction of NCp7 and SL3 could provide useful information that can facilitate rational design of anti-HIV drugs. To investigate the interaction between NCp7 and SL3, De Guzman et al. (1998) constructed a 20-nucleotide RNA molecule containing the 14-nt core sequence of SL3 (as shown in Figure 1-5 inset) [57]. For consistency and convenience, in this study I used the same 20-mer synthetic RNA molecule for affinity assays.

1.5. Interactions within the NCp7-SL3 complex

In 1998, the 3-D structure of the NCp7-SL3 complex (PDB 1A1T) was solved using heteronuclear magnetic resonance spectroscopy by the Summers lab at University of Maryland-Baltimore County, Baltimore, MD, with collaboration of the Borer lab at Syracuse University, Syracuse, NY [57]. This structure revealed many details of the interactions between NCp7 and SL3 RNA as summarized below, which served as a basis for designing the mutation sites in this study.

(1) Overall structure: The stem nucleotides of SL3 RNA form an A helix and the tetra loop bases project away from the stem and interact directly with NCp7. In NCp7, the N-terminal residues from Lys³ to Arg¹⁰ form a 3₁₀ helix that binds within the major groove of the RNA molecule. The two zinc fingers F1 and F2 are in close proximity and interact with the G³²⁰ and G³¹⁸ bases on the RNA tetra loop, respectively, as shown in Figure 1-6.

(2) Hydrophobic interactions play a significant role in maintaining the protein structure and contribute to the specific interactions between NC and SL3. Hydrophobic interactions among Phe⁶ of the 3₁₀ helix, Val¹³ and Ile²⁴ of F1 bring the 3₁₀ helix tightly against the F1 knuckle, whereas the hydrophobic interactions among Trp³⁷ of F2 and Phe¹⁶, Asn¹⁷, and Gly¹⁹ of F1 bring the two zinc fingers tightly together. G³²⁰ binds to a hydrophobic cleft formed by the side chains of Val¹³, Phe¹⁶, Ile²⁴ and Ala²⁵. G³¹⁸ interacts with the F2 finger in a similar manner, by fitting into a hydrophobic cleft formed by Trp³⁷, Gln⁴⁵ and Met⁴⁶ side chains. A³¹⁹ makes hydrophobic contacts with the beta methyl or methylene groups of Ala²⁵, Phe¹⁶ and Asn¹⁷. A zoomed-in view of the hydrophobic clefts on the surface model of the complex is shown in Figure 1-7.

(3) Hydrogen bonding is another important factor in the intra- and intermolecular interactions within NCp7-SL3 complex. Phe¹⁶ forms a hydrogen bond with Trp³⁷.

Extensive hydrogen bonding between the side chain atoms of Asn¹⁷ to Cys²⁸, Pro³¹ and Lys³³ stabilizes a single conformation of the linker region, which links the two zinc fingers in close proximity. G³²⁰ forms hydrogen bonds with the backbone atoms of Lys¹⁴, Phe¹⁶ and Ala²⁵. G³¹⁸ forms hydrogen bonds with the backbone atoms of Trp³⁷, Met⁴⁶ and Gly³⁵. A³¹⁹ forms a hydrogen bond with the side chain of Arg³². The side chain carbonyl of Asn⁵ forms a hydrogen bond with the RNA stem and its NH₂ group interacts with atoms of G³²⁰ and G³²¹.

- (4) Electrostatic interactions also contribute to maintaining the structure of the NCp7-SL3 complex. Ten out of 15 basic residues are involved in intra- or intermolecular interactions in the NC-SL3 complex. Most notably, three salt bridges are formed between Lys¹⁴-Glu²¹, Lys³³-Glu⁴² and Lys³⁸-Glu⁵¹. The salt bridges involving Lys¹⁴-Glu²¹ and Lys³⁸-Glu⁵¹ are believed to stabilize the folding of the F1 and F2 fingers, respectively. And the salt bridge between Lys³³ and Glu⁴² is thought to stabilize F2-linker interactions. Lys²⁶ and Lys⁴⁷ are responsible for anchoring the zinc fingers to the RNA through electrostatic interactions. Other basic residues are involved in non-specific electrostatic contacts with the RNA.

1.6. Previous mutational studies

The obvious significance and the unusual conservation of HIV-1 NC had invited extensive mutational analysis on the protein. Deletion and replacement analysis indicated that zinc finger F1 plays a more important role than F2 in NC's functions [99, 102, 146, 147]. The intact CCHC motif of zinc finger F1 is required for the optimal target selectivity of NCp7 [93]. The four conserved C-C-H-C residues involved in the zinc fingers were shown to be critical for the functions of NC and infectivity of HIV-1 viral particles [76, 78, 95, 101, 103, 110, 148]. Random single mutations in the zinc motifs of NC resulted in more relaxed RNA binding specificity of the protein [94]. Site-directed mutagenesis was carried out on each residue in zinc finger F1 [148]. Five mutants, including F16A and other four mutants on the C-C-H-C motif, showed no virus replication and significantly lower viral RNA content in mutant viral particles. Replacing zinc finger F1 with F2 also rendered similar adverse effects on virus replication and viral RNA content [148]. Binding assay using chemically synthesized mutants of NC protein showed that zinc finger F1 and its flanking basic amino acid residues are sufficient for specific RNA binding. Among those residues, F16 and T24 (a conserved variation of I24) and the arginine residues, as well as the C-C-H-C motif, are essential for NC's specific RNA binding [102]. Substitution of D-proline or leucine for Pro31 demonstrated reduced protein function and viral infectivity [108-110]. Mutational analysis on the linker region between the two zinc fingers indicated that the overall structure, rather than the basic nature, of the linker region is critical for virion structure and infectivity [108]. Mutational analysis of

basic residues in NCp7 suggested proportional relationship between the basicity of NC mutants and their binding affinity for RNA or infectivity of mutant viruses. Nonetheless, some basic amino acid residues, including R7, K14, R32 and K33, were proved to be more critical than others for NC's activity or the structure and infectivity of mutant viruses [104, 149].

Some mutants reported in this dissertation, such as F16A and E21A, have been previously studied by various binding or functional assays [82, 102, 146, 148]. Most *in vitro* binding assays were carried out under different conditions than studied in this thesis, and some of the findings weren't consistent with each other (studies at low ionic strength are largely irrelevant due to the dominance of non-sequence-specific interactions [124]). Our work addresses these conflicts via a systematic and thorough mutational and binding study on the effect of side chain contacts of NC residues on the interactions between NC and SL3.

1.7. Specific Aims

The goal of this project was to locate the key residues in the NCp7-SL3 interaction that are responsible for recognition and packaging of viral genomic RNA into nascent viral particles. Based on previous studies and current knowledge of HIV-1 biology, the following key hypotheses were proposed:

- (1) The RNA-nucleocapsid interaction is nearly the same in the NC-domain of the Gag precursor as in the mature NCp7 protein.
- (2) The NC-SL3 interaction is mainly defined by direct side chain contacts between the protein and RNA.
- (3) Important RNA-protein contacts that stabilize the NCp7-SL3 complex can be probed by varying the protein sequence.

These hypotheses lead to the following specific aims:

- A. Construct and express NCp7 variant proteins in which certain side chain contacts are altered. Site-directed mutagenesis was used to construct the mutants. The proteins were over-expressed in *E. coli* as recombinant proteins and used in *in vitro* affinity assays for SL3 RNA.
- B. Determine affinities of NCp7 variant proteins for SL3 RNA. Two types of affinity assays were used in this study. The primary assay is a fluorescence spectroscopy based assay, taking advantage of the natural fluorescence of Trp³⁷. The secondary assay is a direct thermodynamic assay and is used to confirm and validate the data collected from the primary assay.

- C. Map the binding surface on NCp7 for SL3 RNA by comparing the affinity of NCp7 variants with wild type, so as to locate key residues in NCp7-SL3 interaction.

The methods and results are described and discussed in the following chapters.

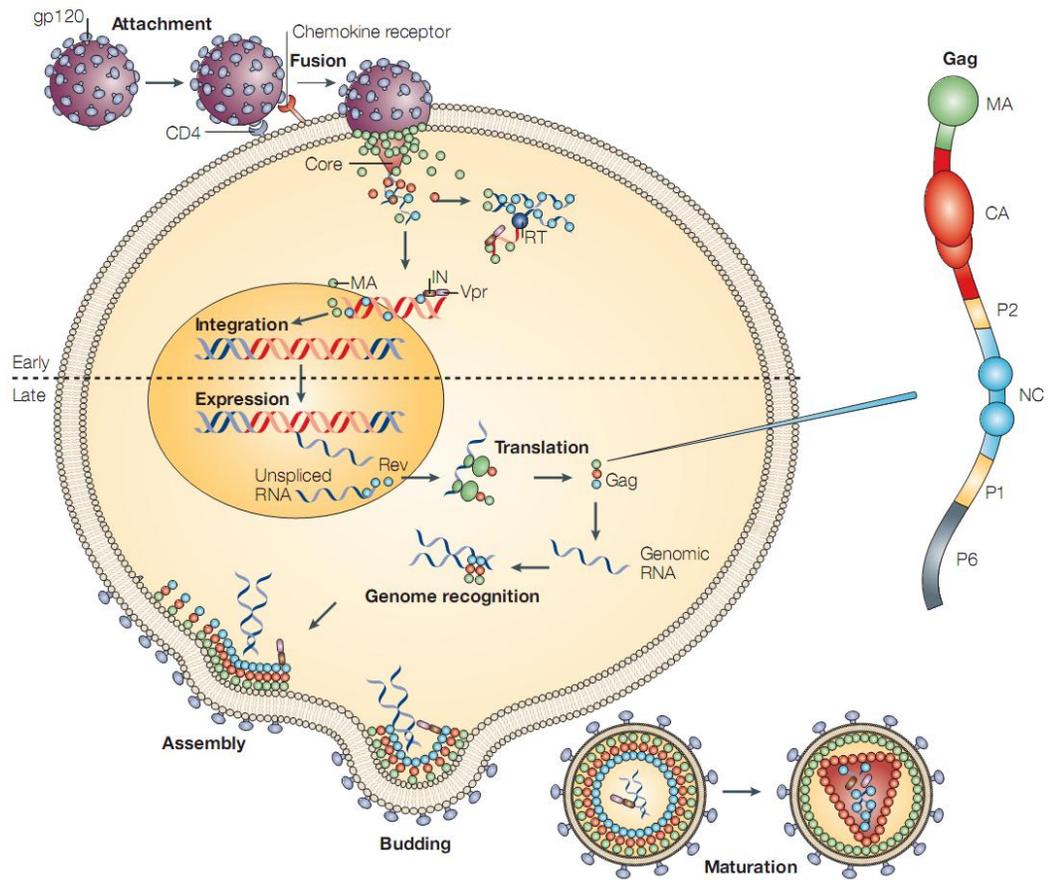
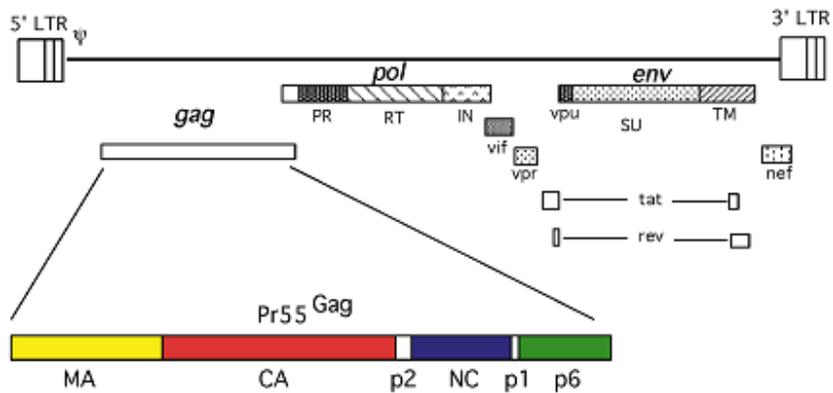


Figure 1-1. General features of HIV-1 replication cycle

Viral proteins labeled in the graph are: Gag, group-specific antigen protein; matrix (MA, green), capsid (CA, red) and nucleocapsid (NC, blue); IN, integrase; Pol, polymerase; Rev, regulator of viral protein expression; RT, reverse transcriptase; Vpr, viral propagation.

Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology, copyright 2005 (D'Souza and Summers [21]).

A HIV-1 GENOME ORGANIZATION



B HIV-1 VIRION

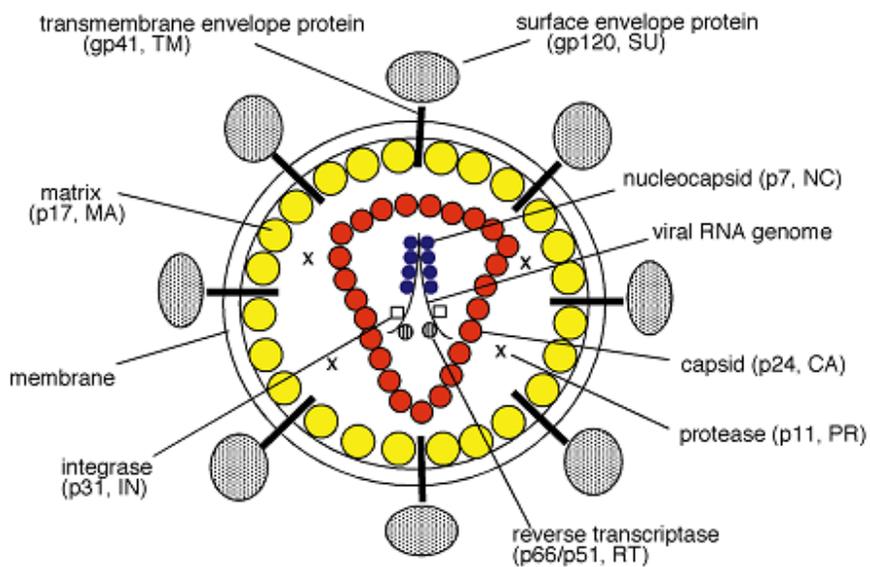


Figure 1-2. Schematic representation of the HIV-1 genome and virion

(A) The location of the HIV-1 open reading frames is indicated. The RNA packaging signal, ψ , is shown near the 5'-end of the genome. The gag open reading frame is enlarged.

(B) The HIV-1 virion, indicating the approximate location of Gag proteins, the Env glycoproteins and the pol-encoded enzymes IN, RT, and PR

Reprinted from *Virology*, 251(1), Freed EO, pages 1-15, Copyright 1998, with permission from Elsevier [36].

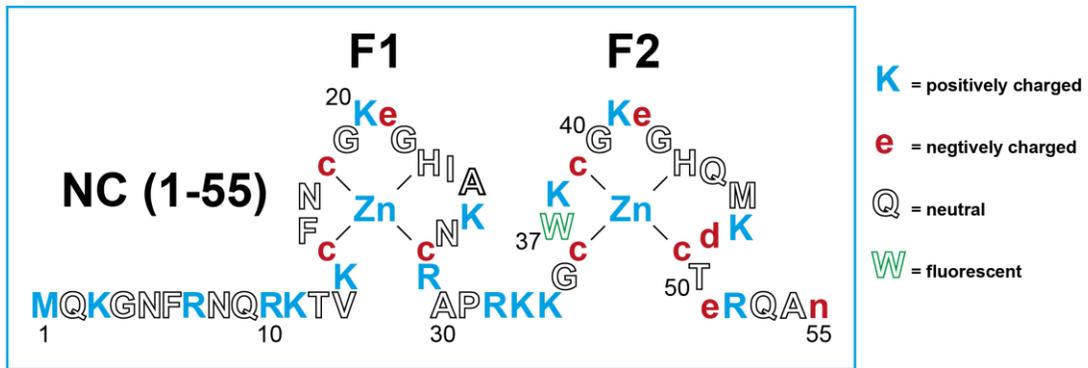


Figure 1-3. The amino acid sequence of HIV-1 NCp7 protein

Residues carrying a charge at neutral pH are shown in solid letters (positive = blue capital letter, negative = red lower case); $Zn_2 \cdot NCp7$ (1-55) carries a +9 charge at neutral pH. The green outlined letter W indicates that the residue Trp³⁷ has natural fluorescence.

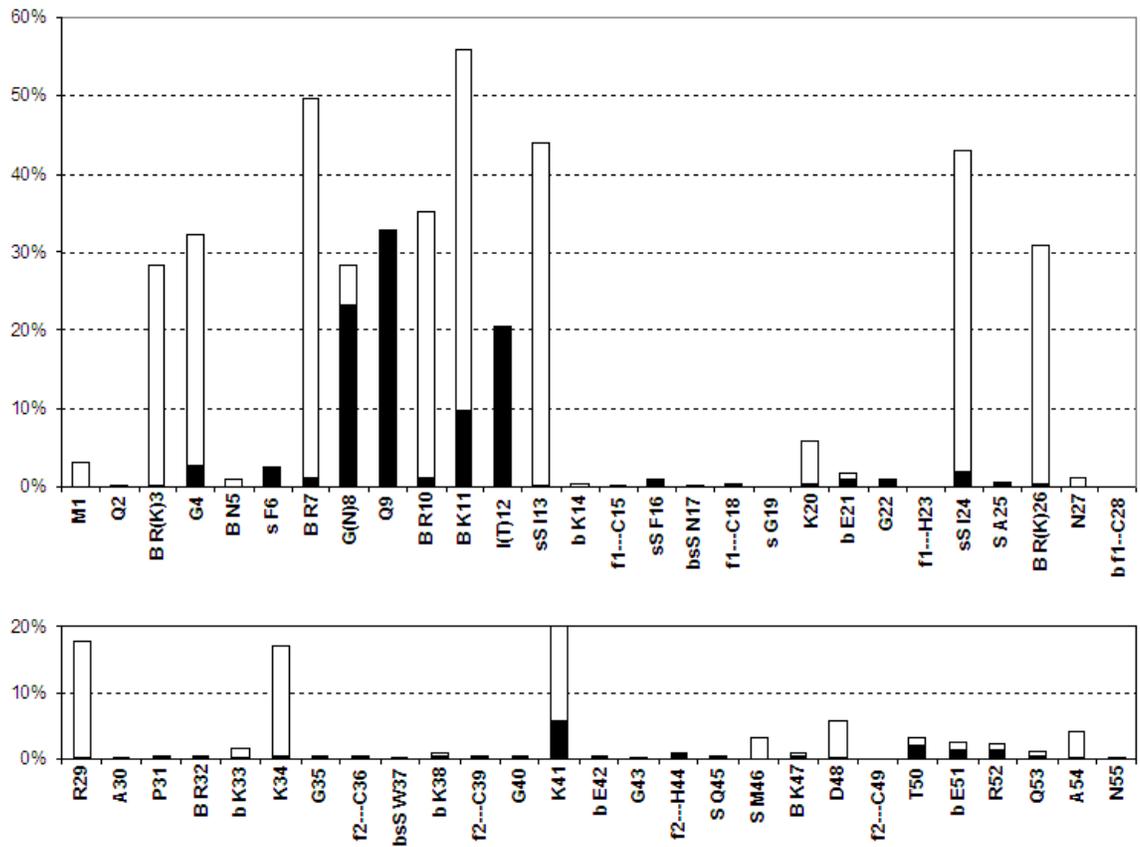


Figure 1-4. Variation in 1700 NCp7 sequences

Open bars: conservative amino acid substitutions, closed bars: non-conservative.

Notation for RNA to protein side-chain interactions in the SL3-NCp7 complex: B = H-bonds or salt-bridges, S = steric contacts. Side-chain protein to protein contacts are also noted: b = H-bonds or salt-bridges, s = steric contacts [57]. The consensus sequence is shown, with different residues in the pNL4-3 isolate used in this work noted in parentheses. CCHC in F1 and F2 are marked.

Reprinted with permission from Lin and Borer, unpublished data [87].

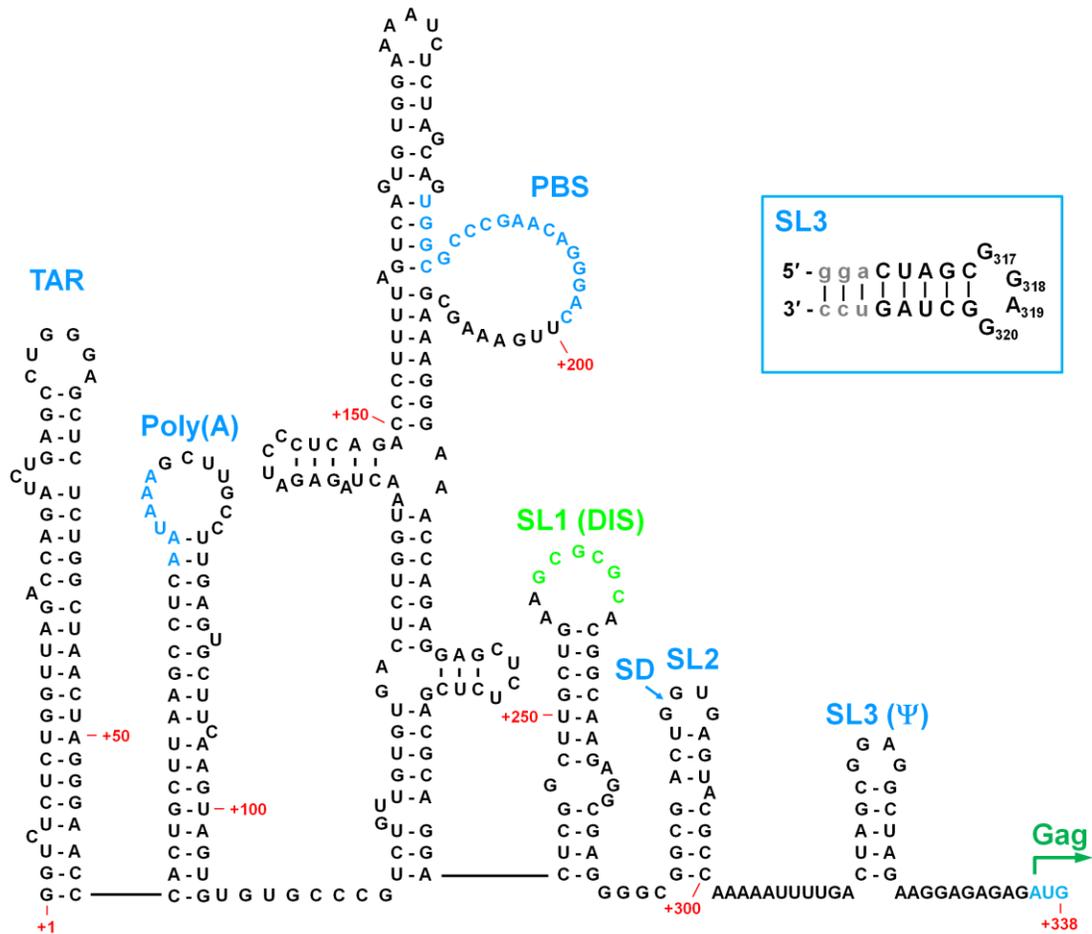


Figure 1-5. Secondary RNA structure model of HIV-1 5' un-translated region (5'-UTR)

Schematic illustration of a consensus hypothetical model of the HIV-1 5'-UTR showing multiple stem-loop structures important for virus replication (from left to right): TAR

element, the poly(A) hairpin, the PBS domain, Stem-Loop 1 (DIS), Stem-Loop 2 (splice donor, SD), Stem-Loop 3 (packaging signal, Ψ), and the *Gag* start codon, respectively.

Inset shows a 20-mer synthetic RNA molecule, which contains the core sequence of SL3

and is used in this study as a model for SL3. (Adapted from Russell et al. 2004 [127],

Paillart et al. 2004 [128], De Guzman et al. 1998 [57] and Watts et al. 2009 [129]).

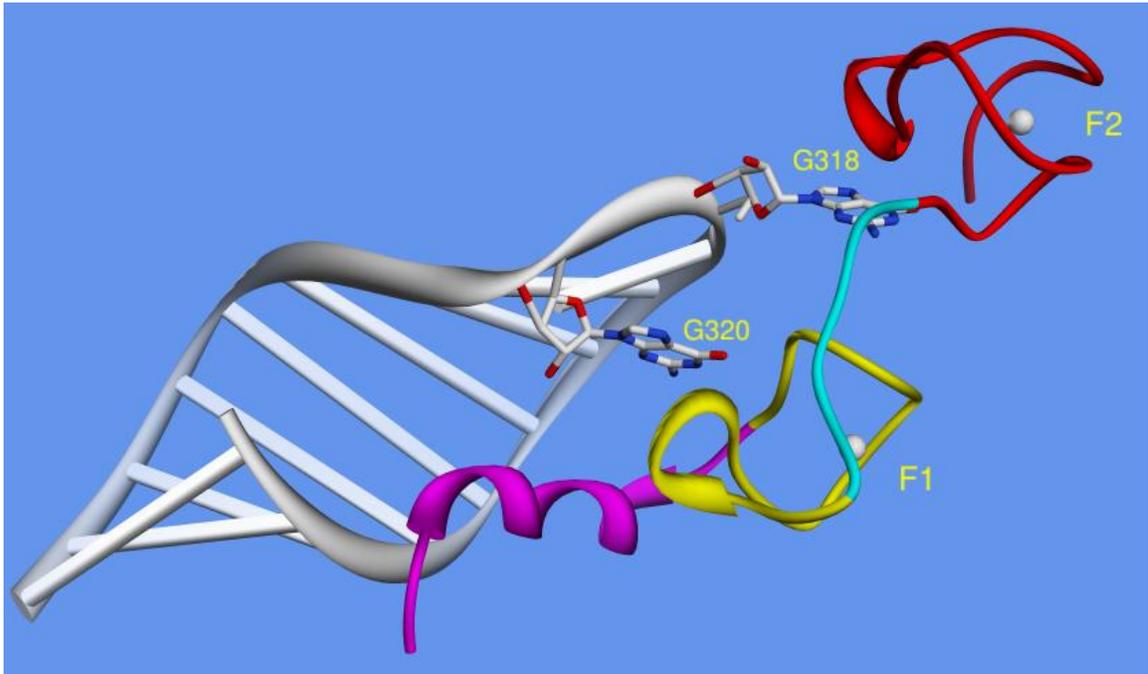


Figure 1-6. NMR structure of NCp7-SL3 complex

White ribbon: SL3 RNA; Stick model: backbone and base of G³¹⁸ and G³²⁰; Magenta ribbon: N-terminal 3-10 helix of NCp7; Yellow ribbon: N-terminal zinc finger (F1); Cyan ribbon: linker region; Red ribbon: C-terminal zinc finger (F2); White sphere: zinc. Drawn from PDB ID 1A1T using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081) [150].

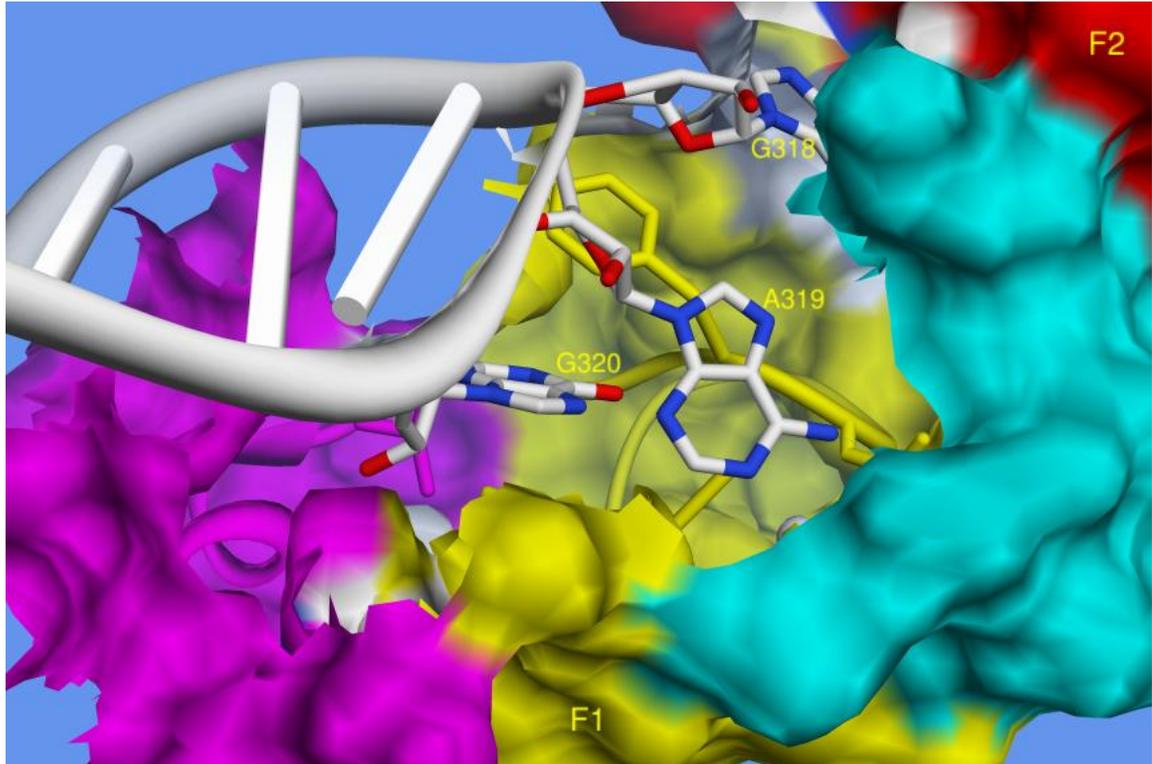


Figure 1-7. Hydrophobic cleft on NCp7-SL3 complex

White ribbon: SL3 RNA; Stick model: backbone and base of G³¹⁸, A³¹⁹ and G³²⁰; Surface model: NCp7 protein; Magenta: N-terminal 3-10 helix of NCp7; Yellow: N-terminal zinc finger (F1); Cyan: linker region; Red: C-terminal zinc finger (F2). Drawn from PDB ID 1A1T using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081) [150].

Chapter 2. Construction, Expression, Purification and Characterization of Variants of NCp7 Protein

2.1. Introduction

2.1.1. Site-directed mutagenesis

Site-directed mutagenesis is a widely used technique in molecular biology to introduce single point mutations or multiple mutations, insertions and deletions into a cloned DNA fragment by polymerase chain reaction (PCR) [151]. As shown in Figure 2-1 (Stratagene, 2005 [152]), the desired mutation is designed and incorporated into a pair of mutagenic primers. The primers are annealed to the circular dsDNA template (plasmid) and extended by a high fidelity polymerase (PfuUltra).

DNA methylation is not essential but nonetheless universal in *E. coli* [153]. Many laboratory strains of *E. coli*, including the XL1-blue and BL21(DE3)pLysS strains used in this study, encode three sequence-specific methylases: the *hsd* (host-specificity), *dam* (DNA adenine methylation), and *dcm* (DNA cytosine methylation) [153]. Foreign

plasmids transformed into such *E. coli* strains will also become methylated if they contain sequences recognized by those methylases [154, 155]. The *dam* methylase specifically methylates the adenine residue in sequence GATC on both strands of the duplex DNA. The *Dpn* I restriction enzyme cleaves on the 3' side of methylated adenine within the sequence GA*TC (* indicates methylation), which means all *dam* modified dsDNA molecules are susceptible to *Dpn* I digestion [155, 156].

After sufficient thermal cycles, the mutagenesis PCR reaction produces a large number of dsDNA molecules amplified by PfuUltra, which contain mutations introduced by the primers and do not have *dam* methylation. The amplified PCR mixture is then subjected to *Dpn* I digestion [152]. Because PCR synthesized products are not methylated, they can survive the *Dpn* I digestion and stay intact, while the parental plasmid templates containing *dam* methylation are degraded by *Dpn* I. Then the PCR synthesized DNA containing the desired mutation is transformed into *E. coli* cells for preservation and expression. Site-directed mutagenesis simplifies the process of gene manipulation and has become an important tool in protein engineering, functional genomics and structural biology.

Alanine scanning mutagenesis is probably the most common practice of site-directed mutagenesis, in which the codon for one amino acid residue is substituted with a codon for alanine [157]. Alanine was chosen as the substitution because it's the simplest chiral molecule of the 22 proteinogenic amino acids, with a chemically inert methyl group as

its side chain. By substituting alanine for other amino acid residues, the side chain contact on the mutated residue will be eliminated or greatly reduced whereas the protein backbone remains unaltered. Thus the role of the side chain contact of the mutated residue can be speculated from quantitative binding analysis of the alanine scanning mutants.

2.1.2. Over-Expression and Purification of Recombinant NCp7

Over-expression of recombinant proteins in *E. coli* has been a routine technique in modern biochemistry labs. Protocols for recombinant wild type NCp7 protein over-expression and purification have been well established and documented [111, 158, 159]. Because of the unusually high content of basic residues in NCp7, few proteins in *E. coli* bear a higher charge than NCp7 does at neutral pH. Based on this fact, a two-part ion exchange purification procedure was successfully used for the separation and purification of NCp7. Because all mutants constructed in this study have the same or even higher charge than wild type NCp7 at neutral pH, similar purification protocols can be applied to the mutant proteins with only minimal modification.

In previously established protocols for purification of NCp7 protein, reducing agents such as dithiothreitol (DTT) and beta-mercaptoethanol (BME) were used extensively in order to protect the thiol group on the cysteine residues of the recombinant protein from oxidation and forming disulfide bonds. DTT and BME are both widely used and

very effective reducing agents. But UV background interference tends to be a problem in accurate concentration determination of protein samples containing DTT or BME. The oxidized forms of both DTT and BME have high absorption around the UV₂₈₀ region. Therefore they can interfere or even mask the UV absorption by a protein sample, especially when the protein concentration is low. Air oxidation of DTT and BME makes it difficult to avoid the interference by simply using a buffer blank.

Tris(2-Carboxyethyl) phosphine Hydrochloride (TCEP·HCl) is an attractive alternative to DTT and BME [160]. It is a powerful reducing agent with a reducing potency comparable to that of DTT. It is also non-volatile, odorless and resistant to air oxidation. More importantly, it does not have absorption peak in the UV-Vis range so that it won't interfere with the measurement of protein concentration by UV absorption. All these advantages make TCEP·HCl an ideal substitution for DTT and BME in the purification of recombinant NCp7 proteins in this study. The results in this study demonstrated satisfactory and reproducible purification of NCp7 with TCEP·HCl as the protective reducing agent.

2.1.3. Protein characterization assays

SDS-PAGE was routinely used in this study to check the expression level of recombinant NCp7 proteins and the purity of the purified proteins. Besides SDS-PAGE, two other

assays were used in this study to identify and characterize the proteins before they are subjected to affinity assays with SL3 RNA.

Mass spectrometry has been a preferred technology in protein identification for many years [161]. Among different MS techniques, MALDI-TOF (Matrix-assisted laser desorption/ionization (MALDI) in combination with time-of-flight (TOF) mass analyzers) mass spectrometry is especially useful in protein and peptide analysis. MALDI-TOF mass spectrometry is applicable to a wide range of biomacromolecules and is more tolerant of common contaminants than other techniques. MALDI ionization preserves the structural integrity of ionized biomacromolecules and extends the molecular weight and polarity range of sample detection, allowing for easy sample preparation, large m/z range and enhanced sensitivity [162, 163]. MALDI-TOF mass spectrometry was mainly used in this study to determine the accurate molecular weight of the expressed and purified protein. It also further confirmed the purity of the protein sample which was checked on SDS-PAGE after purification.

Analytical ultracentrifugation is often used as a “gold standard” for monitoring aggregation status of protein samples [164]. Protein oligomerization or aggregation can affect the stoichiometry and apparent affinity of the studied protein for its ligands. So it is important to determine the aggregation status of the protein of interest before subjecting it to affinity assays. Our previous fluorescence titration model had assumed that NCp7 is in monomer form and binds to SL3 RNA in a 1:1 ratio. The titration data

fitted the model well [165]. But this assumption had never been directly proved by experiment until now.

2.2. Materials and Methods

Unless otherwise noted, all chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO), Thermo Fisher (Waltham, MA) or Thermo Pierce (Rockford, IL). All *E. coli* strains and site-directed mutagenesis kits were purchased from Stratagene (now Agilent Genomics, Santa Clara, CA). All buffers and media were prepared with analytical grade water produced from a Millipore Elix water purification system or ultrapure water produced from a Millipore Milli-Q Biocel ultrapure water system (Millipore, Billerica, MA).

Strains, plasmids and media. XL1-Blue *E. coli* strain was used for plasmid cloning and preservation. BL21-(DE3)pLysS *E. coli* strain was used for the over-expression of NCp7 wild type and mutant proteins. The pRD2 plasmid containing the coding sequence of the 55-amino acid NCp7 from HIV-1 pNL4-3 strain[159] (a generous gift from Dr. Michael Summers at University of Maryland, Baltimore County, referred to as “wild type” in the following text. See plasmid map and sequence in appendix A) was used as template for site-directed mutagenesis to construct all the NCp7 mutants characterized in this study. The recombinant protein has methionine at position one of NCp7, coded by the AUG start codon, rather than isoleucine, which occurs in wild type NCp7 from HIV-1 pNL4-3

strain. Luria-Bertan (LB) medium was prepared according to Sambrook *et al.* [166] or using pre-packed LB medium capsules (MP Biomedicals, Solon, OH). XL1-Blue strains were grown in LB medium containing Ampicillin (Amp) (100µg/ml) as a selective antibiotic. BL21-(DE3)pLysS strains were grown in presence of Amp (100µg/ml) and Chloramphenicol (Cam) (34µg/ml). For protein over-expression, the LB-Amp-Cam media were also supplemented with glucose (4g/L), MgSO₄ (1mM), ZnCl₂ (100µM), and M9 salts (NH₄Cl (1g/L), Na₂HPO₄·7H₂O (3.2g/L) and KH₂PO₄ (3g/L) [111, 159, 167]. The sugar and salt solutions were prepared as 10X, 100X or 1000X stock solutions, sterilized separately and added to the media each time before starting culture.

Plasmid preparation. A single XL1-Blue colony picked from a freshly grown LB-Amp plate was used to inoculate 5ml of LB-Amp media and grown at 37°C overnight, shaking at 200-250rpm. The overnight culture was harvested by centrifugation in a refrigerated Eppendorf centrifuge Model 5804 (Eppendorf, Hauppauge, NY) for 5min at 3000rpm. The plasmid was prepared from the harvested XL1-Blue cells using a QIAprep Spin Miniprep kit (Qiagen, Valencia, CA) and eluted with 50µl of H₂O or 1X TE (10mM Tris-HCl, 1 mM EDTA, pH 8.0) [168] (also see appendix B-2). Plasmids are stored in 1X TE at -80°C for long term storage. The purity of the plasmid miniprep was confirmed by 1% agarose gel electrophoresis and the concentration of the plasmid was determined using a Nanodrop ND-1000 UV-Vis spectrophotometer (Thermo Nanodrop, Wilmington, DE).

Site-directed mutagenesis. The Stratagene QuikChange® or QuikChange® II site-directed mutagenesis kit was used to introduce site-specific mutations at positions listed in Figure 2-2. Site-directed mutagenesis primers were designed using Stratagene QuikChange Primer Design Program (<http://www.genomics.agilent.com>) or PrimerX online tool (<http://bioinformatics.org/primerx/>) and synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). The mutagenesis PCR reactions were set up according to the manufacturer's instructions [152] (also see appendix B-3). After the PCR reactions were finished, 0.5 µl of *Dpn* I restriction enzyme (provided in the site-directed mutagenesis kit) was added into each reaction and incubated for 1 hr. The digested PCR products were run on a 1% Agarose gel to check product size and estimate DNA concentration. Then the digested PCR product was transformed into XL1-Blue *E. coli* competent cells according to the manufacturer's instructions [152] (also see appendix B-4). XL1-Blue strains transformed with plasmids can be preserved in 15% glycerol (final concentration) at -80°C for long term storage. Frozen stock of XL1-Blue strains should be renewed every six months by remaking frozen stock from fresh overnight culture (also see appendix B-6). Repeat the transformation if necessary. Plasmids containing the desired mutations were prepared and stored as described above.

Sequencing. The prepared wild type and mutant plasmids were sequenced at the Core Facility for DNA Sequencing & Flow Cytometry at State University of New York Upstate Medical University, using a universal T7 promoter primer as sequencing primer. The

sequences of mutants and wild type were aligned using the ClustalW2 online tool [169], to confirm the position and sequence of the introduced mutation.

Transformation of BL21-(DE3)pLysS competent cells. The mutated and confirmed plasmids were transformed into BL21-(DE3)pLysS *E. coli* competent cells according to the manufacturer's instruction [170] (also see appendix B-5). BL21-(DE3)pLysS strains transformed with plasmids can be preserved in 15% glycerol (final concentration) at -80°C for long term storage. Frozen stock of BL21-(DE3)pLysS strains should be renewed every six months by remaking frozen stock from fresh overnight culture (also see appendix B-6). Repeat the transformation if necessary.

Over-expression of NCp7 Variants. The wild type and mutant NCp7 strains were grown in LB media and the proteins were over-expressed using a revised protocol based on that of Lee et al. [159] and Shubsda et al. [111] (also see appendix B-7). A Single BL21-(DE3)pLysS colony was used to inoculate 5ml of supplemented LB-Amp-Cam media as starter culture and grown at 37°C overnight, shaking at 200-250rpm. The overnight starter culture was added to 1L of supplemented LB-Amp-Cam media and grown at 37°C, shaking at 200-250rpm. The culture was induced with 400µM IPTG (isopropyl-β-D-thiogalactopyranoside) when its absorbance at 600nm reached 0.6-0.8. The induced cells were grown for 4 more hours and harvested by centrifugation on an Avanti® J-E centrifuge (Beckman Coulter, Brea, CA) using the JLA-10.5 rotor at 5000rpm for 15 minutes. The cell pellets were kept frozen at -20°C before subjected to cell lysis.

Cell Lysis. Cell lysis was performed using a modified protocol based on that of Shubsda *et al.* [111] (see also appendix B-8) Cell pellets from 1L culture were re-suspended on ice in 30ml of lysis buffer (50mM Tris-HCl (pH 8.0), 10% (v/v) glycerol, 0.1M NaCl, 1 μ M ZnCl₂, 5mM TCEP-HCl (Tris(2-Carboxyethyl) phosphine Hydrochloride, 60 μ g/ml Lysozyme, 53 μ M PMSF (phenylmethylsulfonyl fluoride), 1 μ g/ml Pepstatin A, 5mM sodium deoxycholate). The mixture was left on ice for 20minutes before 4ml of B-PER II Protein Extraction Reagent (Thermo Pierce) was added. The cells were then left on ice for 10 minutes before subjected to 4x 25 seconds of sonication using a Microson XL2000 Ultrasonic Liquid Processor (Misonix, now Qsonica, Newtown, CT) at an output power of 20-25 watts. 2ml of 4% PEI (polyethyleneimine) was then added to the mixture and mixed thoroughly. The lysed cells were centrifuged at 17,000rpm for 30min at 4°C on an Avanti® J-E centrifuge using the JA-25.5 rotor and the supernatant (lysate) was collected for FPLC purification.

FPLC purification of NCp7 proteins. FPLC purification was carried out on an AKTA FPLC system (Amersham, now GE Healthcare, Piscataway, NJ). Cell lysate was loaded onto a HiPrep Q FF and a HiPrep SP FF column (GE Healthcare) connected in series and previously equilibrated with at least 200ml of buffer A (50mM Tris-HCl (pH 8.0), 10% (v/v) glycerol, 0.1M NaCl, 1 μ M ZnCl₂, 1mM TCEP-HCl). After the A₂₈₀ value returned to baseline, the Q-column was disconnected and NCp7 protein was eluted using a one-hour linear gradient from 20% to 50% buffer B (50mM Tris-HCl (pH 8.0), 10% (v/v)

glycerol, 1.0M NaCl, 1 μ M ZnCl₂, 1mM TCEP·HCl) (also see appendix B-9). Homogeneity of the eluted protein was determined by SDS-PAGE. The purified NCp7 protein fractions were then pooled and concentrated using a VIVASPIN 15ml Concentrator (5000 MWCO) (GE Healthcare). The NCp7 concentration was determined by UV₂₈₀ absorbance using a molar extinction coefficient of $\epsilon_{280} = 6050 \text{ M}^{-1} \cdot \text{cm}^{-1}$ [171, 172]. Sample purity and homogeneity was further confirmed by MALDI-TOF mass spectrometry and analytical ultracentrifugation. Purified NCp7 protein was stored at -80°C both in aliquots (typically in 100 μ l aliquots at a concentration of 50 μ M of protein in buffer A) and concentrate stocks (usually range from 0.5-2mM in buffer A).

SDS-PAGE. Purity of the eluted protein was determined by SDS-PAGE [166]. Ten microliters of each fraction was loaded and run on a 15% acrylamide/SDS Tris-Tricine gel stacked with a 5% stacking gel. The gel was stained by Coomassie blue staining. The fractions of the correct size and at least 95% purity are pooled and concentrated. Some recent batches of over-expressed and purified protein were analyzed on precast gradient gels purchased from Invitrogen (Carlsbad, CA), Bio-Rad (Hercules, CA) or Lonza (Walkersville, MD) and stained using SimpleBlue™ SafeStain (Invitrogen).

Mass Spectrometry. Purified and concentrated protein samples were dialyzed against MS buffer (5mM sodium phosphate, 0.2M NaCl, 1 μ M ZnCl₂, 1mM TCEP·HCl, pH 7.0) overnight. The concentrations of the dialyzed samples were determined by UV using MS buffer as blank. The samples were then diluted to 20 μ M in 0.1% aqueous trifluoroacetic

acid (TFA) solution. The dilute samples were then mixed in 1:1 ratio with freshly prepared matrix solution (saturated Sinapinic Acid in Acetonitrile:0.1% aqueous TFA 1:2). 1 μ l of each sample mixture was spotted onto a MALDI sample plate and was allowed to air dry. MALDI-TOF mass spectrometry was performed on an Autoflex III Smartbeam MALDI-TOF system (Bruker Daltonics, Billerica, MA) at State University of New York College of Environmental Science and Forestry, Syracuse, NY. The spectrometer was operated in reflectron mode. Triplicate measurements were taken for each protein sample.

Analytical Ultracentrifugation. Protein samples were dialyzed against NCp7 MS buffer overnight. The concentrations of the dialyzed samples were determined by UV using MS buffer as blank. Samples were then diluted in MS buffer to an absorbance of 1-1.5 at 280nm before being analyzed on a ProteomeLab™ XL-A Protein Characterization System (Beckman Coulter). 400 μ l of sample was loaded each time. Sedimentation velocity experiments were carried out at 50,000 rpm (200,000 x g) and 10°C overnight in 3-mm two-sector charcoal-filled Epon centerpieces with quartz windows. Sedimentation data were collected at 280nm and analyzed by the continuous sedimentation coefficient distribution method using the program SEDFIT [173] (Acknowledgement: the AUC sample run and data analysis were done by Dr. Anamika Patel (Laboratory of Prof. Michael S. Cosgrove, Biology department, Syracuse University)).

2.3. Results

Starting from the wild type pRD2 plasmid, sixteen mutants of NCp7 protein were made as laid out in Figure 2-2. Thirteen out of the sixteen mutants are alanine scanning mutants in which one non-alanine residue in the wild type protein sequence was mutated into an alanine. These mutated residues were chosen because either they are predicted to be involved in direct side chain contact with SL3 RNA or other parts of the protein [57] (e.g. F16A) or they are structurally rigid (e.g. P31A) or flexible (e.g. G19A). The other three mutants contain double mutations, which resulted in position switching of the two residues involved in each of the three salt bridges in the wild type NCp7 structure. The mutant proteins were then over-expressed in *E. coli* and purified by ion-exchange chromatography on FPLC.

Site-Directed Mutagenesis Primers. Designed mutagenesis primers are listed in table 2-1. The primers were designed to have a melting temperature (T_m) between 75-85°C, a GC content between 40-60% and a length between 30-55 bases [152]. The mutated codons were designed to have minimal change from the template codon sequences in the wild type and optimized for codon usage in *E. coli* [174, 175].

Sequence alignments. All 16 mutants and the wild type plasmids were sequenced and the coding sequence regions on each plasmid were aligned and shown in Figure 2-3. The sequencing results show that the desired mutations were introduced in the specific sites

successfully. As reference, the plasmid map of pET-3a (on which pRD2 was constructed [159]) was obtained from Novagen (now EMD4Biosciences, Gibbstown, NJ) [176] and shown in appendix A-1. The full sequence of pRD2 is listed in appendix A-2.

FPLC chromatogram. A typical FPLC chromatogram is shown in Figure 2-4. Except for the mutant E21A, all other NCp7 mutant proteins and wild type NCp7 were eluted off SP FF column during the gradient from around 28% -42% Buffer B, with a peak at around 35% Buffer B (the conductivity of sample at peak is around 27 mS/cm). E21A was eluted later in the gradient because of its higher net charge than wild type.

SDS-PAGE. An example of SDS-PAGE gel image is shown in Figure 2-5. The gel clearly shows that the recombinant NCp7 protein is expressed successfully and is a major protein component in the crude lysate of the induced cells. Most of the NCp7 fractions collected during the FPLC gradient elution are more than 95% pure. For the purpose of affinity assays, this level of purity is acceptable. Precast gels produce excellent resolution of NCp7 from other proteins. Representative gel images using precast gels from three different suppliers are shown in Figure 2-6, 2-7 and 2-8. Invitrogen precast gels have the best performance of the three. But in general all three kinds of precast gels produce satisfactory results with good reproducibility and great convenience.

MALDI-TOF. An example of MALDI-TOF mass spectra of purified NCp7 protein is shown in Figure 2-9. The spectrum shows that the observed mass of the highest peak (M+)

agrees well with the calculated mass of the analyzed protein. The observed molecular weight and calculated molecular weight of purified NCp7 wild type and all mutant NCp7 proteins are listed in table 2-2.

Analytical Ultracentrifugation. Two sets of AUC experiments were performed by Dr. Anamika Patel using two different batches of NCp7 wild type protein samples. Results from both experiments had only one single peak in the sedimentation velocity profile with a calculated mass around 6.5 kDa, indicating that NCp7 is in the monomeric form in the buffer conditions used in this study and is stable enough to be used in overnight experiments. An example of the AUC profile of NCp7 wild type is shown in Figure 2-10.

2.4. Discussion

Protease inhibition. In cell lysis, the *E. coli* cells were homogenized by the concerted efforts of lysozyme digestion, detergent dissolution and ultrasonic disruption. The cellular proteins released after lysis were protected from endogenous proteases by addition of protease inhibitors PMSF and Pepstatin A. Nonetheless, the protection from protease inhibitors wears off soon due to their degradation in aqueous solution. For this reason, the FPLC purification needs to be carried out on the same day of cell lysis to avoid protein degradation.

DNA contamination. Genomic DNA is a common contamination in recombinant protein purification. NCp7 can bind DNA non-specifically because of its high basicity. So DNA contamination is especially a problem for NCp7 purification. Polyethyleneimine (PEI) is a basic cationic polymer which is positively charged at neutral pH. It can bind to negatively charged macromolecules like nucleic acid and acidic protein molecules and precipitates together with them. By addition of PEI and subsequent centrifugation, genomic DNA can be removed from the lysate [177].

FPLC purification. The FPLC purification of NCp7 proteins involves two parts: removal of negatively charged impurities by the Q FF column and separation of positively charged NCp7 protein by the SP FF column. The purpose of flowing the cell lysate through the Q FF column before SP FF column is to reduce the burden on the SP FF column and increase the yield and purity of the separated NCp7 protein. SP FF column alone can be used to separate NCp7 protein from the crude cell lysate. But the existence of negatively charged cellular proteins can compete with the resin in the SP column for NCp7 protein, and therefore reduce the binding capacity of the SP column. Moreover, the negatively charged proteins might be co-eluted with NCp7 protein as a contaminant. So it is necessary to include the Q FF column in the FPLC purification procedure.

Zinc concentration in media and buffers. The zinc concentration in the LB media used in NCp7 over-expression was 100 μ M. But in the following purification and analysis process, the zinc concentration used in all the buffers was reduced to 1 μ M. This was because of

the instability of TCEP·HCl observed in presence of high concentration of zinc ions (data not shown). The high concentration of zinc in the culture was to make sure that there was enough zinc to be incorporated into the protein for the zinc finger structure to correctly fold. Results from following affinity assays showed that NCp7 wild type protein has similar K_d for SL3 RNA compared to reported K_d values of NCp7 wild type for SL3 RNA in the presence of higher zinc concentration (100 μ M), which indicates that NCp7 protein maintains its zinc finger structure in the presence of lower zinc concentration in the buffer (1 μ M).

MALDI-TOF peaks. Before we had ready access to the MALDI-TOF mass spectrometer and worked out the optimal conditions for measuring NCp7 variants on MALDI, the most reliable method for protein identification available to us was Sanger sequencing of the plasmid being used for protein over-expression, which is indirect and often affected by sequencing noise. SDS-PAGE gels give a good indication of the purity of protein but don't make accurate measurement of the molecular weight of protein and cannot distinguish mutant proteins that are similar in size. Therefore MALDI-TOF and other mass spectrometry techniques are powerful tools in protein chemistry. In fact, several mutants which had exhibited unexpected binding patterns in preliminary experiments were later found to contain unwanted mutations after being examined by MALDI-TOF. For example, the original F6A mutant showed very weak binding affinity for SL3 RNA and was later confirmed to be a double mutant, which contains both F6A and F16A mutations. An early stop codon was introduced in the original P31A mutant because of a

shift in the reading frame, which resulted in a shorter translated product (31aa, mw. 3565). Those mutants were later re-constructed from scratch and confirmed to be the correct ones by MALDI-TOF.

In the MALDI-TOF mass spectra of the NCp7 wild type and mutant proteins, the molecular ion peak (M^+) is the most abundant peak. In some cases, small peaks with higher mass were observed. These peaks can be attributed to the zinc-bound protein molecules. The most frequently observed small peak has a mass around $M+65$, which can be attributed to the protein molecules coordinated with one zinc ion. Less frequently, a peak around $M+130$ can be observed, indicating the NCp7 molecules coordinated with two zinc ions. The low abundance of these peaks is probably because the ionization condition of MALDI is too harsh for the zinc-bound protein to keep the zinc ions upon the impact of laser. Using a milder ionization method such as electrospray ionization (ESI) might help improve the abundance of zinc bound protein peaks.

Table 2-1. Site-directed mutagenesis primers for construction of NCp7 mutants

Primer		Primer Sequence	Length	T _m
N5A	Fw	5'-GGAGATATACATATGCAGAAAGGCC <u>GCA</u> TTTAGGAACCAAAGAAAGACTGTT-3'	51	78.5°C
	Rv	5'-AACAGTCTTCTTTGGTTCCTAAA <u>TGC</u> GCCTTCTGCATATGTATATCTCC-3'	51	78.5°C
F6A	Fw	5'-GAGATATACATATGCAGAAAGGCAAT <u>GCT</u> AGGAACCAAAGAAAGACTGTAA-3'	52	78.1°C
	Rv	5'-TTAACAGTCTTCTTTGGTTCCT <u>TGC</u> ATTGCCTTCTGCATATGTATATCTC-3'	52	78.1°C
V13A	Fw	5'-TTTTAGGAACCAAAGAAAGACT <u>GCT</u> AAGTGTTTCAATTGTGGCAAAG-3'	47	79.0°C
	Rv	5'-CTTTGCCACAATTGAAACACTT <u>AGC</u> AGTCTTCTTTGGTTCCTAAA-3'	47	79.0°C
F16A	Fw	5'-CCAAAGAAAGACTGTTAAGTGT <u>GCC</u> AATTGTGGCAAAGAAGGGCAC-3'	46	79.4°C
	Rv	5'-GTGCCCTTCTTTGCCACAATT <u>GCC</u> ACACTTAACAGTCTTCTTTGG-3'	46	79.4°C
N17A	Fw	5'-CAAAGAAAGACTGTTAAGTGTTC <u>GCA</u> TGTGGCAAAGAAGGGCACATAG-3'	49	79.2°C
	Rv	5'-CTATGTGCCCTTCTTTGCCACA <u>TGC</u> GAAACACTTAACAGTCTTCTTTG-3'	49	79.2°C
G19A	Fw	5'-TGTTAAGTGTTCATTGT <u>GCC</u> AAAGAAGGGCACATAGC-3'	39	78.4°C
	Rv	5'-GCTATGTGCCCTTCTTT <u>GCC</u> ACAATTGAAACACTTAACA-3'	39	78.4°C
E21A	Fw	5'-GTTTCAATTGTGGCAA <u>GCA</u> AGGCACATAGCCAAA-3'	35	78.1°C
	Rv	5'-TTTGGCTATGTGCC <u>TGC</u> TTTGGCACAATTGAAAC-3'	35	78.1°C
G22A	Fw	5'-TTCAATTGTGGCAAAGAA <u>GCG</u> CACATAGCCAAAATTGC-3'	39	78.4°C
	Rv	5'-GCAATTTTGGCTATGT <u>GCG</u> TCTTTGCCACAATTGAA-3'	39	78.4°C
I24A	Fw	5'-TGGCAAAGAAGGGCAC <u>GCA</u> GCCAAAAATTGCAGGG-3'	35	78.8°C
	Rv	5'-CCCTGCAATTTTGGC <u>TGC</u> GTGCCCTTCTTTGCCA-3'	35	78.8°C
P31A	Fw	5'-CAAAAATTGCAGGGCC <u>GCT</u> AGGAAAAGGGCTGTT-3'	35	79.3°C
	Rv	5'-AACAGCCCTTTTCTCT <u>AGC</u> GGCCCTGCAATTTTG-3'	35	79.3°C
G40A	Fw	5'-GGGCTGTTGAAATGT <u>GCA</u> AAGGAAGGACACCAA-3'	35	79.3°C
	Rv	5'-TTTGGTGTCTTCTTT <u>TGC</u> ACATTTCCAACAGCCC-3'	35	79.3°C
Q45A	Fw	5'-TGTGAAAGGAAGGACAC <u>GCA</u> ATGAAAGATTGTACTGAGAG-3'	41	78.2°C
	Rv	5'-CTCTCAGTACAATCTTTCAT <u>TGC</u> GTGTCTTCTTTCCACA-3'	41	78.2°C
M46A	Fw	5'-GTGAAAGGAAGGACACCA <u>GCG</u> AAAGATTGTACTGAGAGAC-3'	42	78.2°C
	Rv	5'-GTCTCTCAGTACAATCTTT <u>GCG</u> TGGTGTCTTCTTTCCAC-3'	42	78.2°C
K14E	Fw	5'-CCAAAGAAAGACTGTT <u>GAG</u> TGTTTCAATTGTGGC-3'	34	75.6°C
	Rv	5'-GCCACAATTGAAACA <u>CTC</u> AACAGTCTTCTTTGG-3'	34	75.6°C
K14E/ E21K	Fw	5'-GTTTCAATTGTGGCAA <u>AAA</u> AGGCACATAGCCAA-3'	34	75.6°C
	Rv	5'-TTGGCTATGTGCC <u>TTT</u> TTTGGCACAATTGAAAC-3'	34	75.6°C
K33E	Fw	5'-TGCAGGGCCCTAGG <u>GAA</u> AAGGGCTGTGGAA-3'	32	81.6°C
	Rv	5'-TTCCAACAGCCCTT <u>TTT</u> CCTAGGGCCCTGCA-3'	32	81.6°C
K33E/ E42K	Fw	5'-GCTGTTGAAATGTGAAAG <u>AAA</u> AGGACACCAAATGAAAGA-3'	40	78.5°C
	Rv	5'-TCTTTCATTTGGTGTCC <u>TTT</u> CTTTCACATTTCCAACAGC-3'	40	78.5°C
K38E	Fw	5'-GAAAAGGGCTGTGG <u>GAA</u> TGTGAAAGGAAG-3'	32	76.5°C
	Rv	5'-CTTCTTTCCACA <u>TTT</u> CCAACAGCCCTTTTTC-3'	32	76.5°C
K38E/ E51K	Fw	5'-CACCAAATGAAAGATTGTACT <u>AAG</u> AGACAGGCTAATTGAATTC-3'	43	77.8°C
	Rv	5'-GAATTCATTAGCCTGTCT <u>CTT</u> AGTACAATCTTTCATTTGGTG-3'	43	77.8°C

Table 2-2. Observed and calculated molecular weight of purified NCp7 proteins

#	Name	Observed M_r	Calculated M_r	Number of Measurements
1	Wild Type	6369.7 ± 0.7	6369.4	11
2	N5A	6327.1 ± 0.4	6326.4	7
3	F6A	6293.0 ± 0.8	6293.3	6
4	V13A	6340.5 ± 2.0	6341.3	8
5	F16A	6293.2 ± 0.9	6293.3	9
6	N17A	6327.2 ± 1.3	6326.4	9
7	G19A	6383.4 ± 2.0	6383.4	8
8	E21A	6311.0 ± 1.9	6311.3	8
9	G22A	6385.0 ± 0.7	6383.4	6
10	I24A	6328.4 ± 0.4	6327.3	7
11	P31A	6344.3 ± 0.1	6343.3	6
12	G40A	6384.6 ± 0.5	6383.4	7
13	Q45A	6313.7 ± 0.6	6312.3	7
14	M46A	6310.3 ± 0.5	6309.3	4
15	K14E-E21K	6365.4 ± 0.2	6369.4	3
16	K33E-E42K	6369.4 ± 1.2	6369.4	4
17	K38E-E51K	6368.2 ± 0.1	6369.4	3

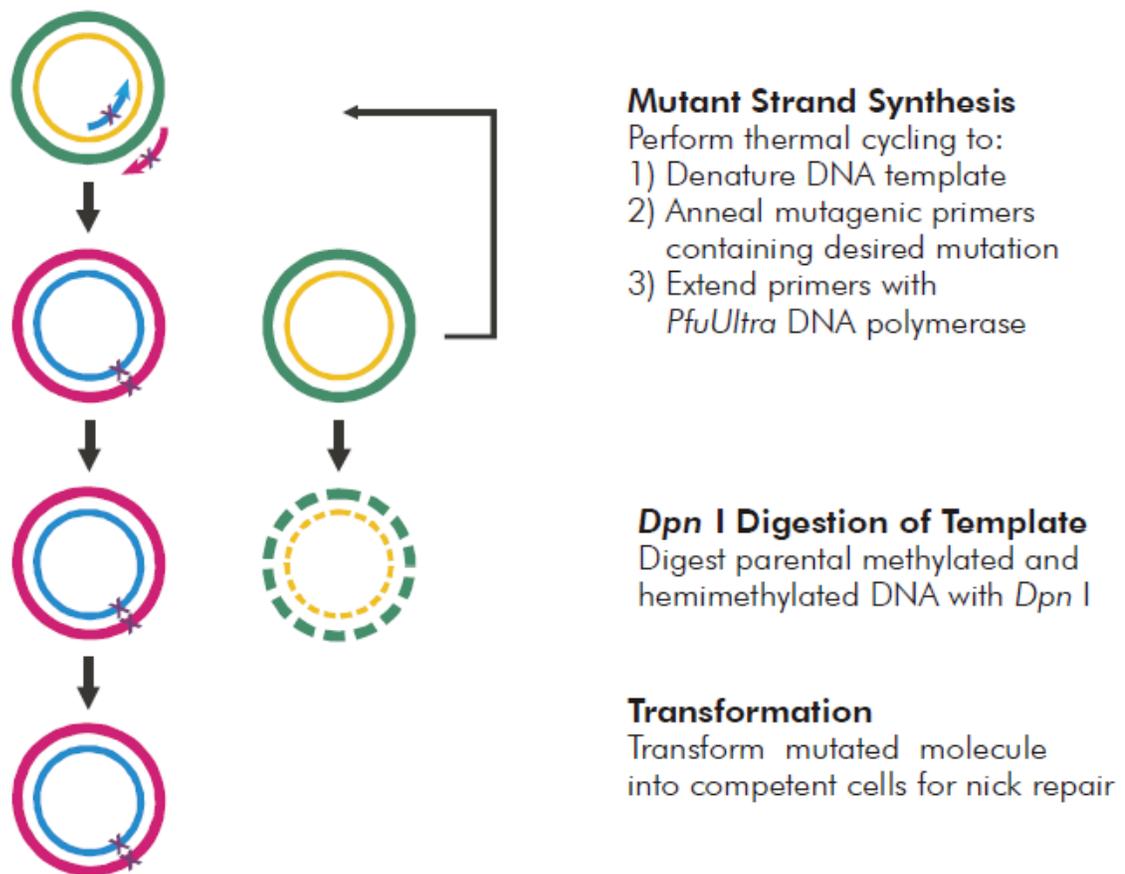


Figure 2-1. Schematic procedure of site-directed mutagenesis

Circles indicate circular double stranded DNA templates and the nicked PCR amplified products. Arc arrows indicate the mutagenesis PCR primers. Crosses indicate the mutations being introduced via the primers. Dashed circles indicate the digested DNA templates. © Agilent Technologies, Inc. 2005 [152]. Reproduced with permission, courtesy of Agilent Technologies, Inc.

	1	10	20	30	40	50
WT	M	Q	K	G	N	F
	R	N	Q	R	K	T
	V	K	C	F	N	C
	G	K	E	G	H	I
	A	K	N	C	R	A
	P	R	K	K	G	C
	W	K	C	G	K	E
	G	H	Q	M	K	D
	C	T	E	R	Q	A
	N					
N5A	-	-	-	-	-	-
	-	A	-	-	-	-
F6A	-	-	-	-	-	-
	-	-	A	-	-	-
V13A	-	-	-	-	-	-
	-	-	-	A	-	-
F16A	-	-	-	-	-	-
	-	-	-	-	A	-
N17A	-	-	-	-	-	-
	-	-	-	-	-	A
G19A	-	-	-	-	-	-
	-	-	-	-	-	A
E21A	-	-	-	-	-	-
	-	-	-	-	-	A
G22A	-	-	-	-	-	-
	-	-	-	-	-	A
I24A	-	-	-	-	-	-
	-	-	-	-	-	A
P31A	-	-	-	-	-	-
	-	-	-	-	-	A
G40A	-	-	-	-	-	-
	-	-	-	-	-	A
Q45A	-	-	-	-	-	-
	-	-	-	-	-	A
M46A	-	-	-	-	-	-
	-	-	-	-	-	A
K14E-E21K	-	-	-	-	-	-
	-	-	E	-	-	K
K33E-E42K	-	-	-	-	-	-
	-	-	-	-	E	-
	-	-	-	-	-	K
K38E-E51K	-	-	-	-	-	-
	-	-	-	-	E	-
	-	-	-	-	-	K

Figure 2-2. HIV-1 NCp7 variants characterized in this study.

The amino acid sequence of wild-type NCp7 (pNL4-3 strain) is given at the top with residue numbers labeled, followed by mutant sequences. The name of each mutant is listed on the left. Identical residues to that of wild-type are indicated by dashes.

	M	Q	K	G	N	F	R	N	Q	R	K	T	V	K	C	F	N	C	G	K
WT	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
N5A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
F6A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
V13A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
F16A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
N17A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
G19A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
E21A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
G22A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
I24A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
P31A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
G40A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
Q45A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
M46A	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
K14E-E21K	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
K33E-E42K	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
K38E-E51K	ATGCAGAAAGGCAATTTTAGGAACCAAGAAAGACTGTTAAGTGTTC	CAATTGTGGCAA	60																	
	*****	*****	*****																	
	E	G	H	I	A	K	N	C	R	A	P	R	K	K	G	C	W	K	C	G
WT	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
N5A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
F6A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
V13A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
F16A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
N17A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
G19A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
E21A	<u>GCA</u> GGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
G22A	<u>GAAGCG</u> CACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
I24A	GAAGGGCAC <u>GCA</u> GCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
P31A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
G40A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
Q45A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
M46A	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
K14E-E21K	<u>AAA</u> GGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
K33E-E42K	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
K38E-E51K	GAAGGGCACATAGCCAAAAATTGCAGGGGCCCTAGGAAAAAGGGCTGT	TGGAAATGTGGA	120																	
	**	****	*****																	
	K	E	G	H	Q	M	K	D	C	T	E	R	Q	A	N	X				
WT	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
N5A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
F6A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
V13A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
F16A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
N17A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
G19A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
E21A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
G22A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
I24A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
P31A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
G40A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
Q45A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
M46A	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
K14E-E21K	AAGGAAGGACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
K33E-E42K	AAG <u>AAA</u> GACACCAATGAAAGATTGTA	CTGAGAGACAGGCTAATTAG	168																	
K38E-E51K	AAGGAAGGACACCAATGAAAGATTGTA	CT <u>AAG</u> GACAGGCTAATTAG	168																	
	***	*****	*																	

Figure 2-3. Sequence alignment of NCp7 wild type and mutants

The name of each mutant is listed on the left. The amino acid sequence of the wild type is shown on top of the alignment. Mutated codons are indicated by underscored bold red font. X indicates the stop codon.

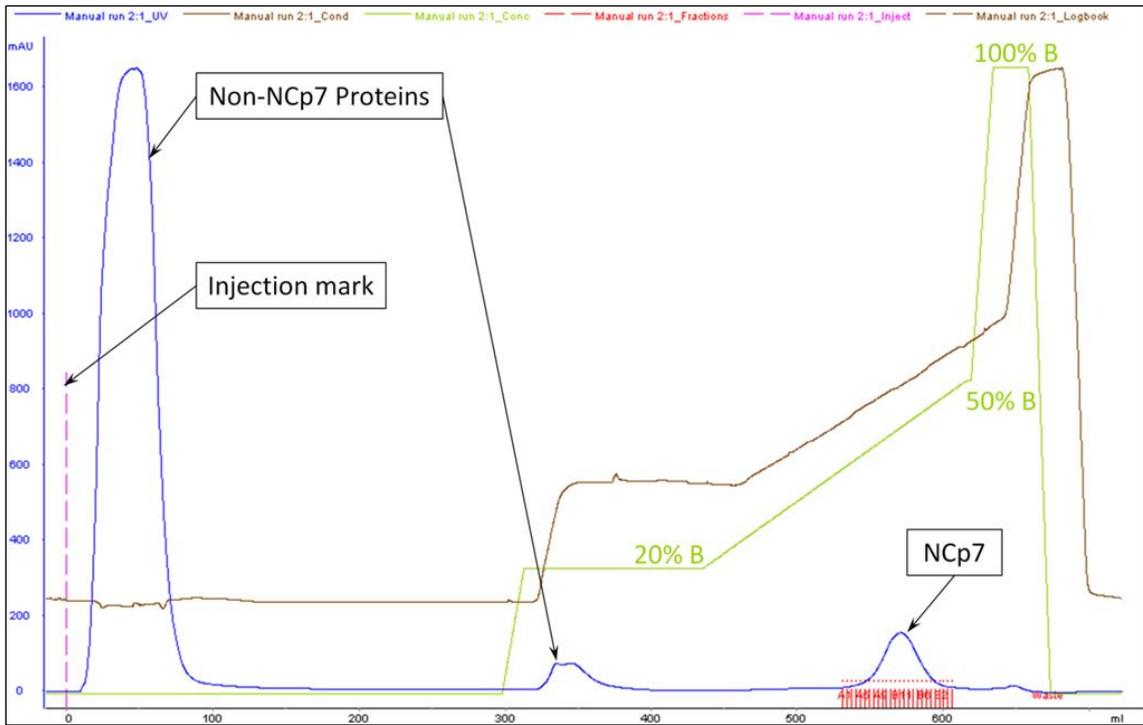


Figure 2-4. Typical FPLC chromatogram of NCp7 purification

Blue curve: UV absorption at 280nm; Brown curve: conductivity of sample; Green curve: buffer gradient; Pink dashed line: Injection mark; Red tick marks: collected fractions.

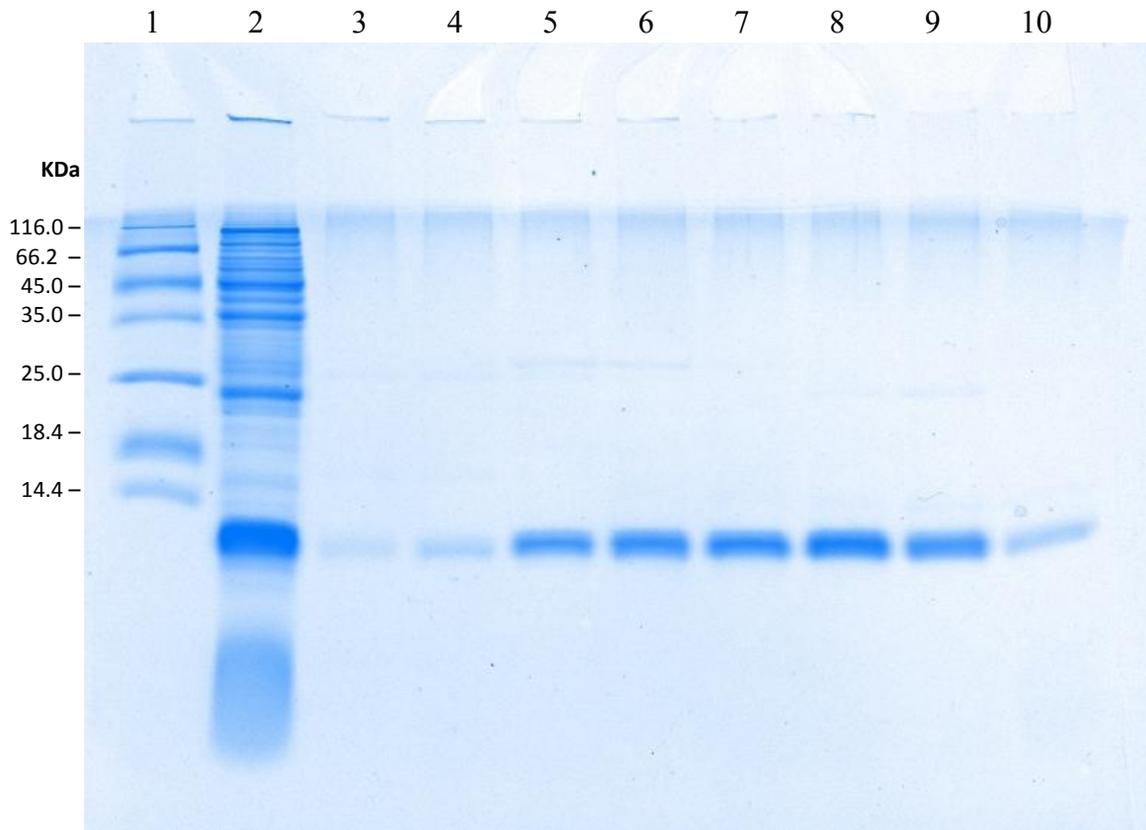


Figure 2-5. SDS-PAGE gel image of NCp7 protein sample (Tris-Tricine)

Lane 1: Protein mass ladder; lane 2: crude lysate of induced cells; lane 3-10: NCp7 sample fractions collected during FPLC gradient elution. Stacking gel: 5%. Separating gel: 15%. Gel thickness: 1.5mm. Running buffer: Tris-Tricine discontinuous buffer (see appendix B-1). Run condition: 160V, 45 minutes.

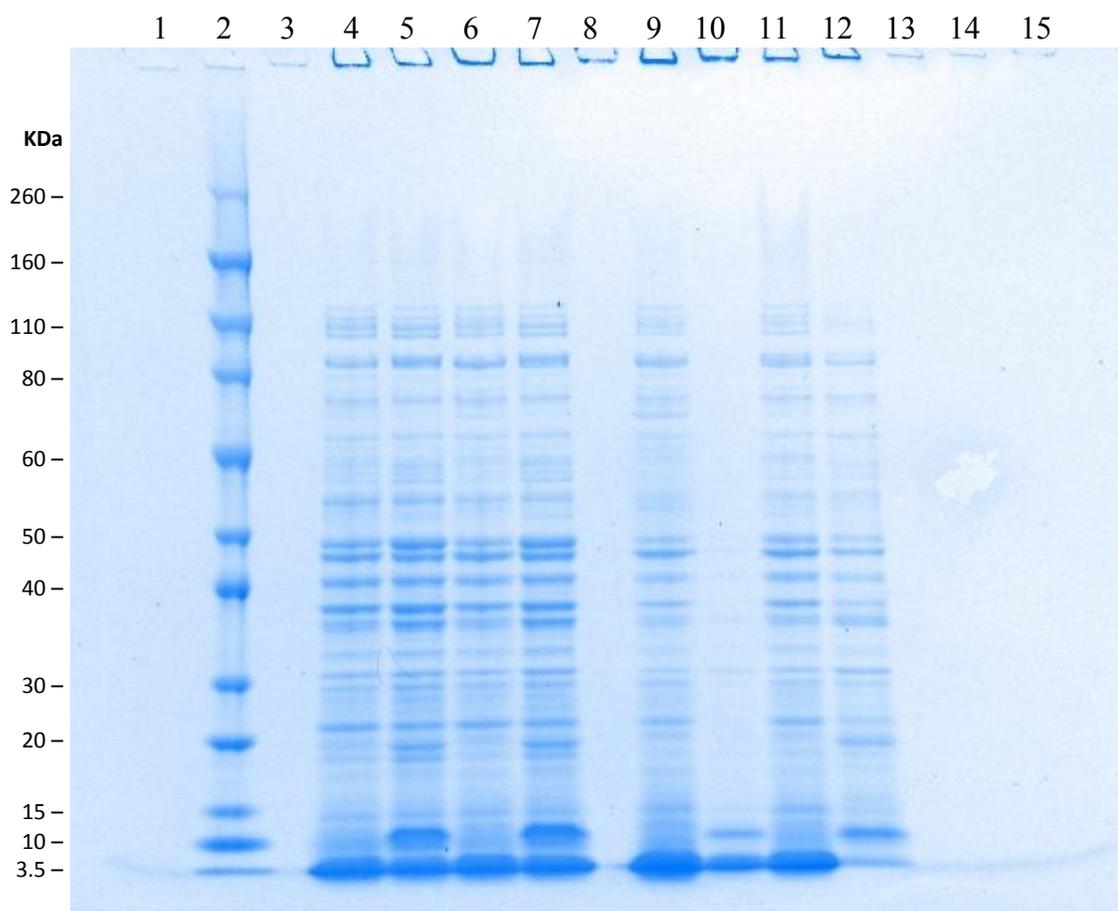


Figure 2-6. SDS-PAGE gel image of NCp7 protein sample (Invitrogen Precast Gel)

Lane 1: blank; lane 2: protein mass ladder; lane 3: blank; lane 4, 6, 9, 11: crude lysate of un-induced cells; lane 5, 7, 10, 12: crude lysate of induced cells; lane 8 and 13-15: blank.

Gel: Invitrogen NuPAGE® Novex® 4-12% Bis-Tris Precast Gel. Gel thickness: 1.0mm.

Running buffer: 1X NuPAGE® MOPS SDS Running Buffer (Invitrogen). Run condition:

200V, 50 minutes.

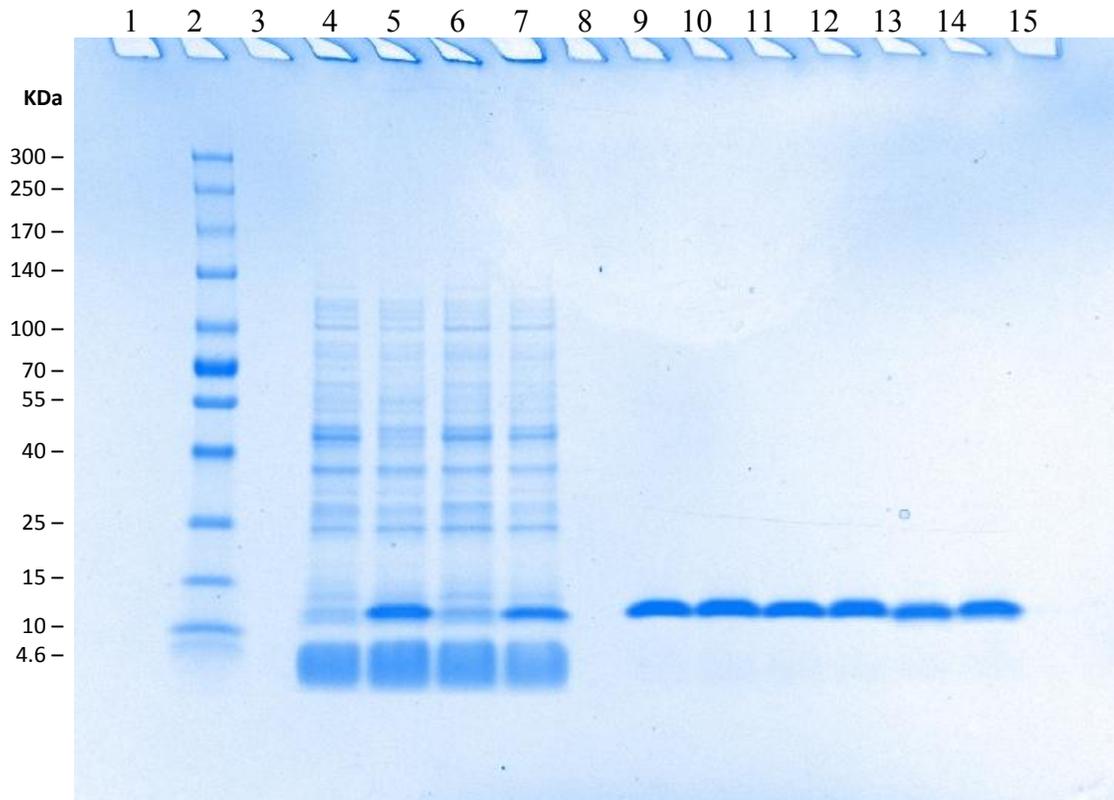


Figure 2-7. SDS-PAGE gel image of NCp7 protein sample (Bio-Rad Precast Gel)

Lane 1: blank; lane 2: protein mass ladder; lane 3: blank; lane 4 and 6: crude lysate of un-induced cells; lane 5 and 7: crude lysate of induced cells; lane 8: blank; lane 9-14: purified protein samples; lane 15: blank. Gel: Bio-Rad 4-20% Mini-Protean® TGX™ Precast Gel. Running buffer: 1X Tris-Glycine buffer (Bio-Rad). Run condition: 100V, 90 minutes.

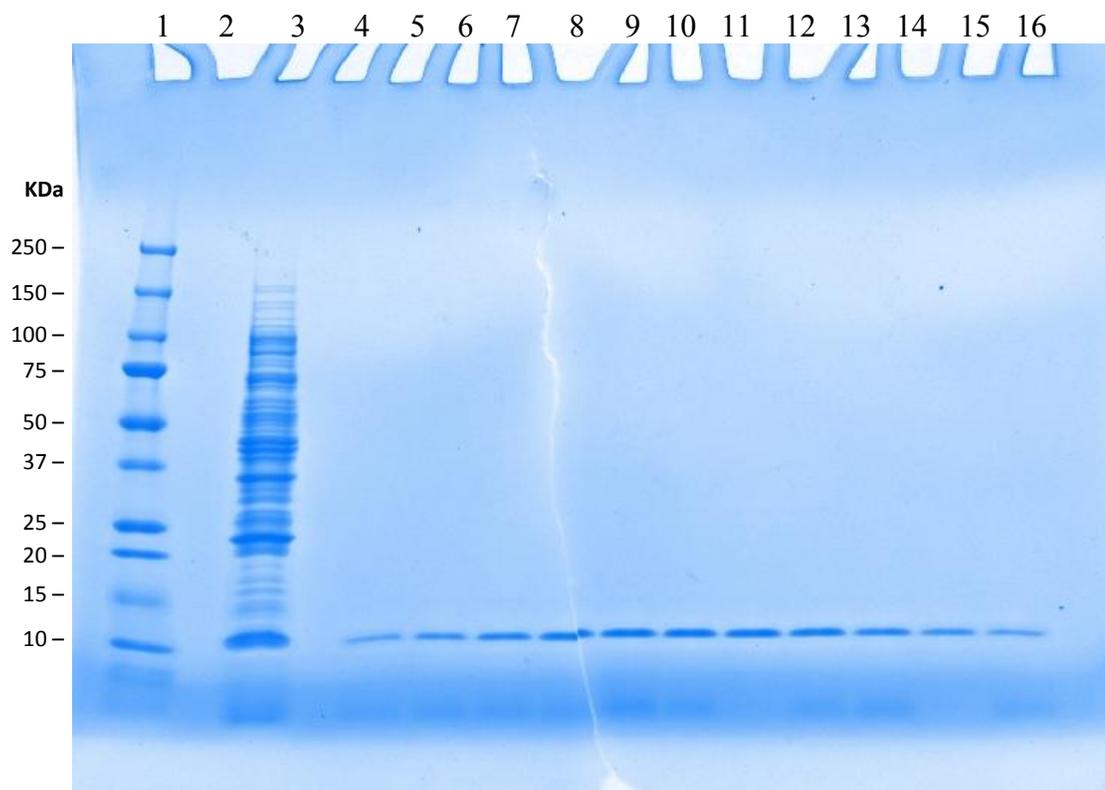


Figure 2-8. SDS-PAGE gel image of NCp7 protein sample (Lonza Precast Gel)

Lane 1: protein mass ladder; lane 2: blank; lane 3: crude lysate of induced cells; lane 4: blank; lane 5-15: NCp7 sample fractions collected during FPLC gradient elution; lane 16: blank. Gel: Lonza PAGEr® 4-20%T-G Precast Gel. Running buffer: 1X Tris-Glycine buffer (Bio-Rad). Run condition: 125V, 6 minutes.

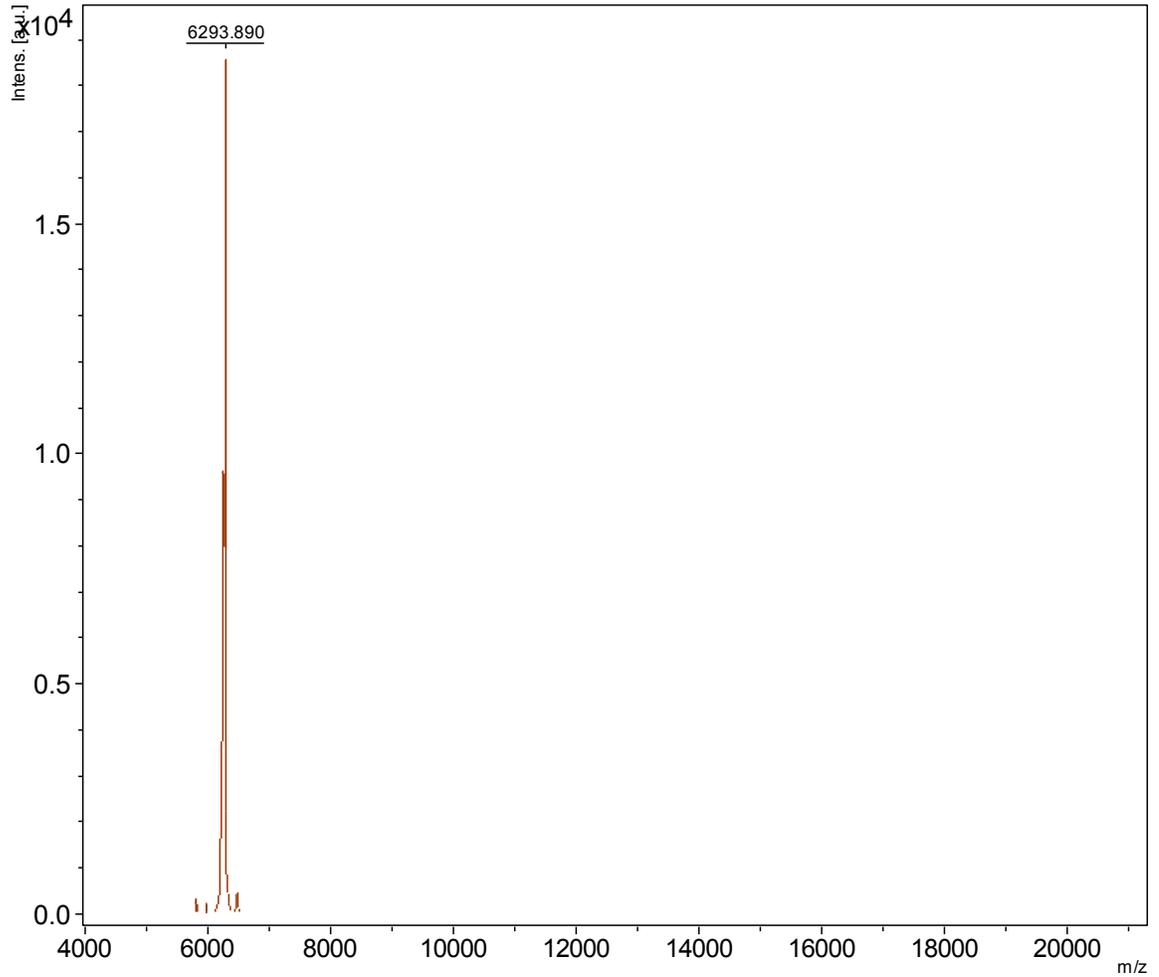


Figure 2-9. Sample MALDI-TOF spectrum of NCp7 protein variants

Sample MALDI-TOF mass spectrum acquired on a Bruker Daltonics Autoflex III

Smartbeam MALDI-TOF system. Analyzed protein sample shown in this spectrum is F16A.

Spectrum was acquired in reflectron mode and recorded in the mass range between

4000 and 22500 m/z.

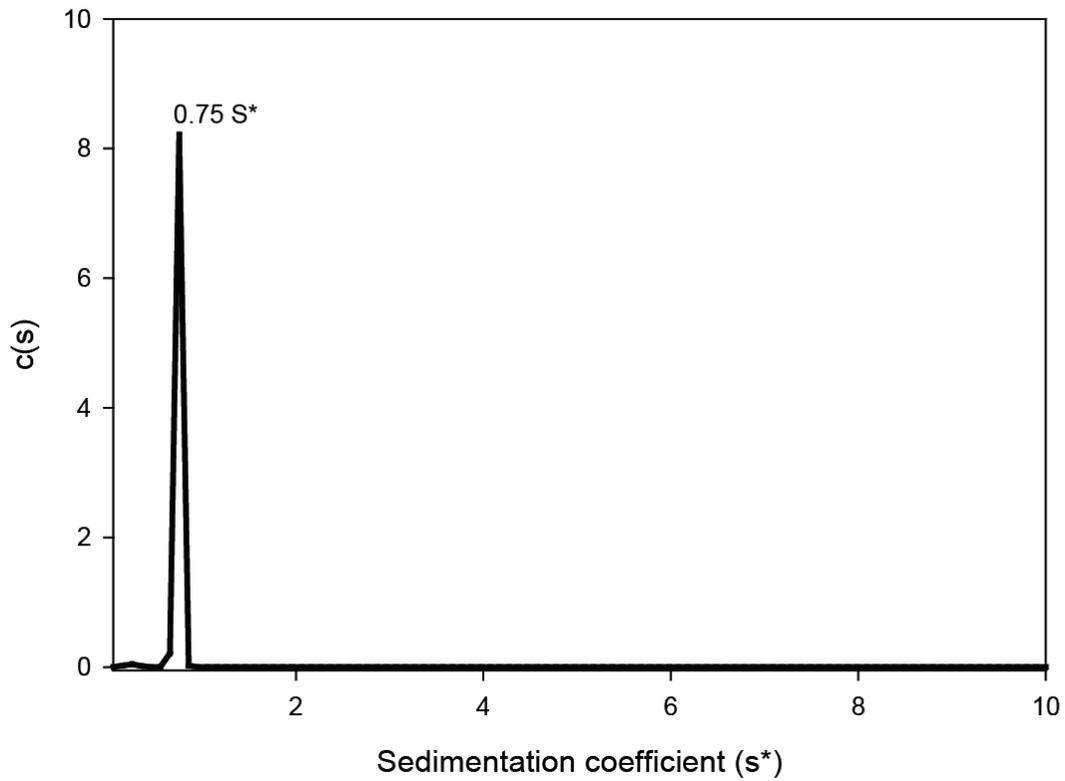


Figure 2-10. Sedimentation velocity profile for NCp7 wild type

Diffusion-free sedimentation coefficient distributions ($c(s)$) derived from sedimentation velocity data of NCp7 in MS buffer. Sedimentation data were collected at 280nm, centrifuging at 50,000rpm and 10°C overnight. Sedimentation velocity profile for NCp7 wild type was fitted to a distribution of Lamm equation solutions using the SEDFIT software [173].

Chapter 3. Quantitative Binding

Analysis of NCp7 Variant Proteins with SL3 RNA

3.1. Introduction

After sufficient amounts (>1 mg) of pure NCp7 wild type and mutant proteins were acquired and identified, the proteins were subjected to quantitative binding analysis to determine their affinity for SL3 RNA. Two kinds of affinity assays, tryptophan fluorescence titration and isothermal titration calorimetry (ITC), were used in this study to measure the affinity of the proteins for HIV-1 SL3 RNA, as described below.

3.1.1. Tryptophan fluorescence titration

Tryptophan is the primary contributor of the intrinsic fluorescence of proteins [178]. The fluorescence emission of tryptophan is always dominant over that of other fluorescent residues such as tyrosine and phenylalanine. Fluorescence of phenylalanine is much weaker than that of tryptophan and tyrosine and can only be observed in the absence of those two stronger fluorescent residues. Tryptophan fluorescence is very sensitive to its microenvironment, which makes it especially useful as an intrinsic fluorescence probe in

gathering information about protein structures and conformational transitions. Changes of tryptophan fluorescence upon such transitions as protein unfolding, ligand binding or protein-protein association allow for acquisition of thermodynamic and kinetic information about these processes.

As mentioned in Chapter 1, the NCp7 protein contains one and only one tryptophan residue (Trp³⁷), two phenylalanine residues (Phe⁶ and Phe¹⁶) and no tyrosine residue. Therefore the W³⁷ residue accounts for most of the intrinsic fluorescence of NCp7 protein. The NMR structure of NCp7-SL3 complex revealed that the G³¹⁸ base of SL3 RNA stacks tightly on the side chain of Trp³⁷ upon binding [57]. This kind of stacking effectively quenches the fluorescence of tryptophan by nearly 100%, which is the underlying mechanism of the fluorescence titration assay used in this study. The assay can be conveniently conducted with 0.6 nmoles (~ 3.8 µg) of NCp7 protein and 2 nmoles (~ 13 µg) of SL3 RNA.

In the tryptophan fluorescence titration assay, SL3 RNA was added in small aliquots into a protein solution under constant stirring in a quartz cuvette. After each addition, the solution is excited at 290nm and emission is recorded at 350nm. NMR structures showed that except for the zinc fingers, NCp7 is largely disordered in the absence of RNA [159], and the Trp³⁷ residue is almost fully exposed to the solvent. Therefore an emission maximum at around 350nm is expected in the fluorescence emission spectrum of the NCp7 protein [179, 180]. By monitoring fluorescence emission at 350nm, the

highest sensitivity and signal-to-noise ratio can be achieved in the titration. Upon increasing concentration of SL3, the fluorescence emission was observed to drop following a quadratic curve with the 1:1 molar ratio trend line and the x axis as its tangent lines, and can be fitted using a simple 1:1 binding model of NCp7 and SL3 RNA [165].

3.1.2. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a thermodynamic technique for probing interactions between biomolecules. It directly measures the heat change upon binding of two or more species of molecules. This method is label-free and does not require the sample to have optical characteristics that change in response to a structural change, such as in fluorescence or UV spectrophotometry. ITC also doesn't require sample immobilization as in surface plasma resonance (SPR) and doesn't have a limitation on molecular weight. It can simultaneously determine multiple binding parameters such as the binding constant (K), enthalpy (ΔH), entropy (ΔS) and the binding stoichiometry (n). All of these make ITC a powerful and universal method to study biomolecular interactions [181]. A typical ITC assay requires 30 nmoles ($\sim 190 \mu\text{g}$) of NCp7 protein and 10 nmoles ($\sim 65 \mu\text{g}$) of SL3 RNA.

A schematic representation of isothermal titration calorimetry is shown in Figure 3-1 (Freyer *et al.* 2008 [182]). In an ITC instrument (a calorimeter), a pair of identical coin

shaped cells made of highly efficient thermal conducting alloy are enclosed in an adiabatic shield. Both cells are filled with liquid during experiments. During a typical ITC experiment, the reference cell is filled with water and the sample cell is filled with one reactant (sample) in titration buffer. A spinning syringe is filled with the other reactant (titrant) in the same buffer and used for injecting and subsequent mixing of the two reactants. The reference cell is maintained at a user-defined constant temperature by applying a small constant power to it. The temperature difference between the sample and reference cells (ΔT) is kept at a constant value by increase or decrease of power being applied to the sample cell, as appropriate, using a feedback circuit. When precisely known aliquots of titrant are titrated into the sample cell through the syringe, heat gets released or absorbed, resulting in the increase or decrease of the sample cell's temperature. This change activates the feedback circuit, causing the power being applied to the sample well to decrease or increase, in order to keep ΔT constant. This change in the power is recorded and used to calculate the thermodynamic parameters using preset or user-defined fitting models in Origin software [183].

ITC has its limitations, as well. It usually requires significantly higher amounts of sample than other binding assays do. The reason for that is to generate large enough heat change to be accurately measured by the calorimeter. The concentration of the sample is usually set between 1 and 1000 times of the estimated or expected K_d of the sample for the titrant [184]. The concentration of the titrant is set so that the final molar ratio

of titrant to sample is above 2 or higher, ensuring the sample to be saturated by titrant toward the end of titration.

3.2. Materials and Methods

SL3 RNA. Desalted SL3 RNA (rGrGrArCrUrArGrCrGrGrArGrGrCrUrArGrUrCrC) was purchased from Integrated DNA Technologies, Inc. (Coralville, IA) and used without further purification. SL3 RNA was reconstituted in NCp7 titration buffer (5mM sodium phosphate, 0.2M NaCl, 1 μ M ZnCl₂, 1mM TCEP·HCl, 0.1% PEG 8000, pH 7.0). The concentration of SL3 RNA was determined by UV₂₆₀ absorbance using a molar extinction coefficient of $\epsilon_{260} = 160,000 \text{ M}^{-1}\cdot\text{cm}^{-1}$ as described previously [111]. Purity of SL3 RNA was examined by ion exchange HPLC using a DNAPac[®] PA200 4X250mm column (Dionex, Sunnyvale, CA) on an Agilent 1200 HPLC system. SL3 was heat denatured at 95°C and snap cooled on ice before being loaded onto HPLC. Ten microliters of ~30 μ M SL3 RNA was injected onto the PA200 column and eluted with a gradient from 0-65% B over 10 minutes (solvent A=0.01M Tris-HClO₄, 20% acetonitrile, pH 8.0; solvent B=0.01M Tris-HClO₄, 20% acetonitrile, 0.33M NaClO₄, pH 8.0).

Tryptophan Fluorescence Titration Assay. NCp7 wild type and mutant proteins were titrated with SL3 RNA using conditions similar to those described previously [111]. NCp7 wild type protein was titrated every time as a positive control along with the mutant proteins. Fluorescence emission was measured at 350nm with a QuantaMaster QM-

4/2005SE fluorometer (Photon Technology International, Birmingham, NJ) in the laboratory of Prof. Bruce Hudson at Chemistry Department, Syracuse University, Syracuse, NY, using a 290-nm excitation wavelength, a 4-nm excitation band-pass and a 1.5-nm emission band pass. The fluorescence reading was acquired 2 minutes after each oligo aliquot using Felix 5.1 software. The titration curves were fitted to a model assuming 1:1 stoichiometry for the molar ratio of NCp7 to bound SL3 RNA as described by Paoletti et al. [165]. The titration curves were fitted to the following equation:

$$(I - I_{\infty}) / I_0 = \left\{ -(R_t - P_t + K_d) + \left[(R_t - P_t + K_d)^2 + 4P_t K_d \right]^{1/2} \right\} / 2P_t \quad (3-1)$$

where R_t and P_t are the total RNA and protein concentrations, respectively, I_0 is the intensity at $R_t = 0$, and I_{∞} is the limiting intensity at saturation. In this study I_{∞} was set to zero, leaving K_d as the only adjustable fitting parameter. The NCp7 concentration was fixed at 0.3 μM , and titrations were extended to $R_t/P_t \approx 3.0$. K_d values were fitted using the above model in Origin 7.0 software (OriginLab, Northampton, MA).

Isothermal Titration Calorimetry. Isothermal Titration Calorimetry (ITC) experiments were performed on a Microcal VP-ITC Micro Calorimeter (Microcal, now GE Healthcare) at State University of New York Upstate Medical University, Syracuse, NY, in the laboratory of Prof. Stephan Wilkens. Origin 7.0 software was used for data collection and analysis. The sample cell has a volume of 1.4 ml and the total available injection volume in the syringe is 300 μl . The sample and titrant were degassed for 5 minutes before being loaded into the sample cell and syringe. Each time around 2ml of 5 μM SL3 RNA dissolved in NCp7 titration buffer was used to load the sample cell and the

overflow was removed by a syringe. About 500 μ l of NCp7 protein at a concentration around 60 μ M was needed for filling the syringe in each titration. Binding reactions were carried out at 30°C. Reference power was set to 5 μ cal/sec. The first injection volume was set to 2 μ L, 4sec, followed by 29 injections of 10 μ l, 20sec. The last injection volume was set to 8 μ L, 16sec. Interval between each injection was set to 240 sec. Syringe stirring speed was set at 295 rpm. The feedback gain was set at high and equilibrium set at fast, auto. The above-mentioned concentrations of sample and titrant were chosen so that the titration would reach saturation at least 5-10 data points before the last injection. A baseline was drawn by linear extrapolation using the data points collected under saturation conditions and subtracted from the whole data set to correct for the heat of dilution [185, 186]. Binding isotherms were then fitted to a one-binding-site model using the supplied manufacturer's ITC add-on for Origin 7.0 software, with K_d and the binding stoichiometry, n , as adjustable parameters.

3.3. Results

SL3 RNA purity. The HPLC chromatogram of SL3 RNA is shown in Figure 3-2. The RNA sample shows no sign of degradation and the purity is higher than 90% as judged by integrals of the peaks in the HPLC chromatogram. This level of purity is acceptable for the purpose of the binding affinity assays used in this study. Therefore the SL3 RNA was used without further purification.

Tryptophan Fluorescence Titration Assay. Seven of the seventeen protein-SL3 titration curves are shown in Figure 3-3 covering the wide range of K_d values of the wild-type and mutant NCp7 proteins for SL3. All seventeen K_d values calculated from Trp titration data are compiled in Table 3-1 as well as in Appendix C-1, ranging from 9 nM to 720 nM. Pictorial summaries are displayed in Figures 3-4, 3-5 and 3-7. All individual Trp titration curves are shown in Appendix C-2. The K_d value for binding of the wild type NCp7 with SL3 was calculated to be 29 nM, which agreed with our previous experimental data [124, 165] and served as a good positive control for the mutant proteins. Dissociation constants were calculated by non-linear regression to get the best fit titration curves, assuming a 1:1 complex. No evidence for other stoichiometries was indicated by the data at 0.2 M NaCl [111, 124]. The binding isotherms for low-affinity complexes (e.g. F16A) were not extended to saturation because the nature of the complex may change in the presence of a large excess of RNA [165]. I_∞ was set at zero because in the range of the K_d values for these titration curves, very little effect of the residual fluorescence was observed [165].

Isothermal Titration Calorimetry. Due to limitations of material and time, only 8 mutant NCp7 proteins and NCp7 wild type protein and were examined using ITC assay. A set of ITC titration curves for all 9 proteins are shown in Figure 3-6. The K_d values and other binding parameters calculated from ITC experiments are compiled in table 3-2 as well as in Appendix C-1. All individual ITC titration curves are shown in Appendix C-3. The K_d values calculated from the ITC assays (termed $K_{d(ITC)}$) were compared with the ones

calculated from the tryptophan fluorescence titration assay (termed $K_{d_{\text{Trp}}}$) in Figure 3-5. Taking into account that ITC experiments were carried out at 30°C, the K_d values calculated from the ITC data were converted to equivalents at room temperature (22°C) using the van't Hoff equation as shown in equations (3-2), assuming the ΔH^0 is constant over the limited temperature range of eight degrees. The K_d values calculated and converted from ITC data were generally larger than their counterparts calculated from tryptophan fluorescence titration data. But the overall trend of the changes in $K_{d_{\text{ITC}}}$ values agrees well with the trend observed in $K_{d_{\text{Trp}}}$ values, as demonstrated in Figure 3-5. The average number of binding sites determined by ITC also suggested that NC-SL3 binding follows a 1:1 binding model. These observations confirmed the validity of the tryptophan fluorescence titration as a good affinity assay for NCp7 proteins and SL3.

$$\begin{aligned}
 -RT \ln K_a &= \Delta G^0 = \Delta H^0 - T\Delta S^0 \\
 \ln K_a &= \frac{\Delta H^0 - T\Delta S^0}{-RT} = \frac{\Delta H^0}{-RT} + \frac{\Delta S^0}{R} \\
 \ln\left(\frac{1}{K_d}\right) &= \frac{\Delta H^0}{-RT} + \frac{\Delta S^0}{R} \\
 \ln K_d &= \frac{\Delta H^0}{RT} - \frac{\Delta S^0}{R} \tag{3-2} \\
 \ln K_{d1} - \ln K_{d2} &= \frac{\Delta H^0}{RT_1} - \frac{\Delta H^0}{RT_2} \\
 \ln\left(\frac{K_{d1}}{K_{d2}}\right) &= \frac{\Delta H^0}{R} \left(\frac{1}{T_1} - \frac{1}{T_2}\right)
 \end{aligned}$$

3.4. Discussion

For the convenience of discussion, the mutants are divided into groups based on their K_d values calculated from the tryptophan titration data. The factor (F) by which K_d change was calculated and is listed in Table 3-1 for each mutant protein by dividing its K_d value by the K_d value of wild type NCp7 (29nM). Mutants with a change factor less than 2 are considered as mutants with little or no change in affinity. Those with a change factor between 2 to 10 are considered to have moderate change in affinity. Those with an F value greater than 10 are considered as mutants with significant change in affinity. The observed changes from the fluorescence titrations are summarized in Figure 3-4 and 3-5 and will be further discussed in the following section. A set of ITC data for the wild-type and eight mutants are displayed in Figure 3-6.

3.4.1. Single mutants

Mutants having little change in affinity. Most of the single mutants have a K_d value close to that of wild type NCp7. These mutants include N5A, F6A, V13A, N17A, G19A, E21A, G22A, P31A and G40A.

In the case of F6A, this observation is not surprising because Phe⁶ does not participate in any RNA-protein or protein-protein side chain contacts and therefore may not contribute much in stabilizing the structure of the protein or protein-RNA complex.

Considering the position and structure of Gly¹⁹, Gly²² and Gly⁴⁰ (each reside in similar positions in the two zinc finger motifs), we hypothesized that substitution of alanine for those residues might give rise to unwanted side chain contacts between NC and RNA or within the NC protein, or to a loss in flexibility in the zinc fingers, which would result in loss of affinity of NCp7 for SL3 RNA. The observations however indicated the opposite, that substitutions of alanine for glycine on these positions do not cause unfavorable effects.

Pro³¹ itself does not have any RNA-protein or protein-protein side chain contact. But given its unique structural rigidity and high conservation, we expected that the P31A mutation might affect its affinity for SL3 RNA. In fact, it's reported that Pro³¹ helps determine the spatial proximity of the two zinc fingers [109]. Substitution of Pro³¹ would result in significant impairment of viral RNA dimerization, virion formation and infectivity [108, 109]. So the observation that P31A only shows a slight change in affinity for SL3 RNA was surprising. This observation indicates that the affinity of NCp7 for SL3 RNA doesn't require specific spatial orientation of the zinc fingers.

Val¹³ does participate in NCp7-SL3 interaction by forming a hydrophobic cleft, which interacts with G³¹⁷ of SL3 RNA, with neighboring residues Phe¹⁶, Ile²⁴ and Ala²⁵. When NCp7 is bound to SL3 RNA, Val¹³ also contributes to intra-protein hydrophobic interactions between the 3₁₀ helix and zinc finger 1 [57]. Molecular dynamics study suggests that Val¹³ resides within a hydrophobic core with Phe¹⁶, Ile²⁴, Ala²⁵, Trp³⁷, Gln⁴⁵

and Met⁴⁶ [53, 89]. Substitution of Val¹³ with alanine shortened the side chain but still maintained similar polarity and hydrophobicity with those of valine. The observation that V13A mutation resulted in a smaller change in affinity than the other mutants in the same hydrophobic core suggests that Val¹³ contributes little to the RNA binding function of NCp7, except by maintaining the hydrophobic core. The fact that Val¹³ is among the residues with a higher frequency of natural variation with hydrophobic residues (see Figure 1-4) also suggests that the specific side chain of Val¹³ is less important to maintaining the hydrophobic core than some of the other residues.

Asn⁵ and Asn¹⁷, especially Asn⁵, were thought to play a more significant role in NCp7's RNA recognition. Asn⁵ is part of the 3₁₀ helix that binds the major groove of SL3 RNA and is also the only NC residue that forms specific hydrogen-bonds with the RNA stem. Moreover, Asn⁵ is one of the most conserved residues in NCp7. Asn¹⁷ is part of the F1 knuckle and makes hydrophobic contacts with the SL3-tetraloop. It also stabilizes the protein structure through multiple hydrogen-bonding and hydrophobic interactions [57]. The observation that both N5A and N17A displayed no significant change in affinity for SL3 was unexpected. This may indicate that either these two residues are not essential in NCp7-SL3 binding or that conformational changes in the mutants could compensate for the loss of H-bonds that should be caused by these mutations.

Another interesting mutant in this group is E21A. Glu²¹ forms a salt bridge with Lys¹⁴, which appears to stabilize the folding of the F1 knuckle [57]. Mutating Glu²¹ into Alanine

effectively destroys the above mentioned salt bridge. One would expect such a change might result in a noticeable loss in the mutant protein's affinity for SL3 RNA. However, the affinity of E21A for SL3 RNA turned out to be even higher than that of wild-type NCp7. The K_d of E21A is calculated to be 9nM vs. 29nM for the NCp7 wild type. One explanation could be that although the mutation from Glu²¹ to alanine destroyed the salt bridge between Glu²¹ to Lys¹⁴, it also resulted in an increase of the mutated protein's net charge from +9 to +10, which helps the protein to attract the negatively charged RNA. Also the zinc finger's structure is stabilized by many forces including the salt bridge. While the loss of the salt bridge may destabilize the structure to some extent, other forces such as hydrophobic interaction could be enough to keep the structure of the zinc finger intact. Also the bound RNA might play a role to keep the protein correctly folded. Moreover, nucleic acid binding studies suggest that binding affinity of NC for RNA is strongly dependent on the net charge of the protein [104, 187, 188].

Mutants with moderately change in affinity. I24A, Q45A and M46A show a moderately lower affinity with K_d values ranging from 2 to 10 times of that of wild-type NCp7. As shown in Figure 3-7, these residues occupy positions on the same side of the two fingers, with Ile²⁴ and Gln⁴⁵ in equivalent positions.

Ile²⁴, Gln⁴⁵ and Met⁴⁶ are members of the hydrophobic core described above. They all show certain steric interactions with the RNA molecule in the NMR structure [57]. The

loss of affinity caused by these mutations indicates that these interactions may contribute more to the stabilization of the complex than the residues in the first group.

Mutant with significant change in affinity. F16A has a much larger K_d value than wild-type NCp7 (approximately 25-fold). This observation also agrees with previously reported northwestern blot (meaning protein-RNA blotting) data and molecular dynamics analysis [148, 189]. Phe¹⁶ forms a hydrophobic cleft with several other amino acid residues and closely interacts with Trp³⁷, for which fluorescence is monitored in the binding assay to determine the affinity of the binding. The mutation from Phe¹⁶ to alanine could result in relocation of Trp³⁷ in the protein-RNA complex so that the binding of SL3 may no longer quench its fluorescence efficiently.

3.4.2. Salt-bridge switching double mutants

Two of the three double mutants, K33E-E42K and K38E-E51K, do not produce large changes in their affinity for SL3 RNA. But the other double mutant, K14E-E21K, which contains mutations within the F1 knuckle, showed a moderate change in its affinity for SL3. The K_d value of this double mutant for SL3 was calculated to be 140 nM. Compared to the other two double mutants, this observation indicated that structure of F1 knuckle is more susceptible to mutations and/or more crucial to NC's affinity for SL3 RNA. The tryptophan titration curve of K14E-E21K doesn't fit well with the model described by equation (3-1), as shown in Figure 3-8. One explanation could be that this mutant might

bind to SL3 RNA differently than the wild type NCp7 does. Its interaction with SL3 RNA might involve a tight binding phase and a weak binding phase, giving rise to the unique shape of the titration curve.

The structure of NCp7-SL3 complex may offer a clearer explanation of the differences in the affinity of the three salt-bridge switching mutants for SL3. As shown in Figure 3-9, the salt bridge between Lys³³ and Glu⁴² and the one between Lys³⁸ and Glu⁵¹ are both directly exposed to solvent, while the one between Lys¹⁴ and Glu²¹ is partially buried inside the complex, especially residue Lys¹⁴. Salt bridges exposed to the solvent may not contribute to stabilizing protein structures because of the interference of solvent. But semi-buried salt bridges like the one between Lys¹⁴ and Glu²¹ can be a strong stabilizing force [190]. The moderate change of this double mutant's affinity indicates that even though the switched residues are still in close proximity, their ability to form salt bridges might be modified.

3.4.3. Comparison of Trp titration and ITC titration results

The K_d values calculated from ITC titration data don't match exactly with those calculated from Trp titration data for the same mutant. But overall they show the same trend. The same criterion (the change factor of K_d value, F) used to categorize the mutants based on their Trp titration data can be used on their ITC titration data and the mutants can be grouped exactly the same way (see Table 3-1, 3-2 and Figure 3-5). This

shows that ITC is a good alternative assay for measuring NCp7-SL3 RNA binding isotherms.

Some of the differences in the K_d values calculated from the two methods may be attributed to the different nature of the assays. Tryptophan fluorescence titration relies on the direct interaction of Trp³⁷ and the bases in the bound RNA, while ITC experiments directly measure the heat change upon binding, which is not dependent on the structure of protein or RNA. The ITC experiment can be more accurate than Trp titration, but this is only true at much higher concentrations than are practical for NC-SL3 titrations. The ITC data, especially for the weakly bound complexes with F16A, I24A, K14E-E21K, are noisy or have unusual shape (See Figure 3-6). This causes more uncertainty in K_d . Trp titration has the advantage of being a convenient and straightforward assay with much higher sensitivity than ITC. The observation that ITC data follow the same trends for Trp titration data also endorses the validity of Trp titration assay for measuring the affinity of NCp7 for SL3 RNA.

The order of titration in the two assays is also different in this work. In tryptophan fluorescence titrations, SL3 RNA is being titrated into NCp7 protein because it's a well established and convenient way of titration and lots of titration data had already been collected before the ITC assay came into play. Our lab has recently worked out the conditions and fitting model for the AddPro approach of tryptophan fluorescence titration, in which NCp7 protein aliquots were added into a large excess of SL3 RNA

solution. Excellent agreement in K_d values was found at the nearly physiological salt concentrations used in the present work [124].

While in ITC titrations, NCp7 protein is loaded into syringe and titrated into SL3 RNA solution in the cell. The amount of titrant (syringe sample) in ITC experiments needs to be at least two to three times of that of sample (cell sample). Considering the cost efficiency, it was reasonable to use SL3 RNA as the cell sample which requires smaller amount of the two binding partners, because SL3 RNA is the most expensive material needed in this assay.

3.4.4. ITC data analysis

Origin software defines the "ligand" as the species with only one site and calls the species with n greater than 1.0 the "macromolecule", irrespective of their molecular weights [191]. In theory, when two binding partners follow a 1:1 binding stoichiometry and have only one binding site, as in the case of NCp7 and SL3, it shouldn't matter which one is defined as "ligand" and which one is loaded into syringe, because the system would be symmetrical. But in reality, systematic errors and other factors would prevent one from acquiring perfect data that produces an exact $n=1$ stoichiometry. So it's important to be consistent in the ITC experiment setup and data analysis for the same set of samples. In this study, the SL3 RNA was defined as "ligand" and was loaded into the cell in all the experiments.

One should note that the order of titration in ITC experiments, i.e. whether the “ligand” is in the syringe or in the cell, needs to be specified under the “ITC” menu in the Origin software before data analysis. The default option in the software is the “ligand-in-syringe” mode. The second option is the “ligand-in-cell” mode. It has been observed that the same set of data can produce similar shape of fitted curve but very different binding parameters. The binding stoichiometry n calculated in one mode is normally the reciprocal of the n value calculated from the same set of data in the other mode. But the K_d and ΔH values are less straightforward, thus much more difficult, to convert between the two modes. So it’s necessary to choose the appropriate mode for the data being analyzed. All the ITC data in this study were fitted in the “ligand-in-cell” mode because SL3 was the loaded in the cell in all experiments.

Before fitting the ITC titration data, it’s necessary to subtract from the data set the baseline or background heat signal, which is largely due to the heat change generated from dilution of the injected titrant. There are two approaches to establish the baseline for ITC data analysis [186]. One is to run a separate background titration by injecting the same titrant into buffer only. The other is to include extra data points at the end of titration when the sample in the cell has already been saturated by the titrant. The heat change in this phase is mainly because of dilution of the titrant. I have tried both approaches of background correction and found that overall the two approaches demonstrated good agreement with each other. When there is discrepancy in data

analyzed with different approaches, the ones corrected with terminal injection data generate a better fit (see Appendix C-3).

3.5. Conclusion

Overall, most of the mutations did not result in significant loss of affinity for SL3 RNA in the mutant NC proteins. This indicates that the side chain contacts of those involved residues in wild type NCp7 either contribute little to its stability or can be compensated by new contacts in the mutants. On the other hand, conservation of many of these residues could arise from NC's interactions with other proteins or in other aspects of its many roles in HIV infection.

The F16A mutant exhibits a profound loss of affinity for SL3 RNA and is confirmed by both fluorescence titration and ITC assay. This observation agrees with previously reported NMR structure and molecular dynamics analysis [57, 189]. All of these indicate that Phe¹⁶ is a key site in the NCp7-SL3 interaction and warrant further structural and functional study focusing on Phe¹⁶. Other interesting mutants include I24A and K14E-E21K.

Tolerance to alanine mutagenesis mutation in some structurally important sites, such as Asn⁵, Gly¹⁹, Pro³¹ and Gly⁴⁰, indicates the key interactions between NCp7 and SL3 might be confined in a small locale on protein surface and is not sensitive to the overall

protein structure. Yet the high conservation of those residues suggests that they have important roles in other functions of NCp7, if not in the NCp7-SL3 interactions.

Fluorescence titration and ITC are complementary methods, both of which have their own flaws and advantages. Both assays make reliable and reproducible measurements of the K_d value. Both methods have good agreement with each other's data and have confirmed an $n=1$ stoichiometry for NCp7-SL3 binding, which is shown in the complexes used in previous NMR study [57].

The Trp titration assay is robust, sensitive and straightforward. It also doesn't require large amount of sample. But the Trp titration assay requires that the protein has intrinsic fluorescence and the fluorescent residue has to be directly involved in the binding site, which limit the application of this assay for other proteins and one NCp7 mutant of great interest, W37A. Also because Trp titration measures the fluorescence signal of protein, it can distinguish complexes with R_1P_1 (complexes consist of one molecule of RNA and one molecule of protein) from R_1P_2 (complexes consist of one molecule of RNA and two molecule of protein), R_1P_3 , etc., but not R_1P_1 from R_2P_1 , R_3P_1 , etc. (See Athavale et al., 2010 [124]).

ITC can distinguish those different stoichiometries of complexes, because it measures the heat change of the binding reaction, which is independent from species and optical characteristics. The nature of ITC assay makes it a universal and direct method for

measuring parameters in a binding reaction. But ITC experiments require a much larger amount of sample and much longer time to finish than Trp titrations. And ITC experiments are more susceptible to background noises, thus more difficult to be optimized with the experimental conditions. The data analysis of ITC is also trickier than Trp titration. Different titration mode can affect fitted data as mentioned earlier. There are also more variables that cannot be fixed when fitting ITC data, such as the binding stoichiometry n . This can produce more accurate fit of data but on the other hand also makes the fitted data more difficult to interpret.

3.6. Suggestions for future work

The discoveries in this study shed light on some promising projects for future work. Since ITC has been confirmed as a good alternative assay for measuring K_d value of NCp7-SL3 interactions, the mutant W37A, which was not able to be measured by the Trp titration assay, can now be subjected to ITC assay. One would expect to see K_d changes in W37A similar to the level of F16A, if not greater, given its key position in the binding site and direct involvement in the binding as demonstrated in the Trp titration assays.

There are other interesting mutants that can be constructed and subjected to either Trp titration or ITC assay. These include G43A, for its similar position in F2 finger as Gly²² in F1 finger; K14A, K33A, K38A, E42A, E51A and combinations of these mutations, to

further investigate the stabilizing effect of salt bridges for the zinc finger structures; F16W, W37F and F16W-W37F, to investigate the involvement of Phe¹⁶ and Trp³⁷ in the interaction between NCp7 and SL3.

Some interesting mutants such as N5A and F16A should also be studied with NMR. Solving the structures of these mutant proteins in complex with SL3 RNA can lead to better understanding of the mechanism underlying the NCp7-SL3 recognition and interactions and further decipher the roles of NCp7 in HIV-1 biology.

Mutants with moderate to significant K_d changes, such as F16A, I24A and K14E-E21K, should be further studied by functional assays such as viral infectivity assay in human cell lines to determine the relationship of the affinity for SL3 RNA and the viability of mutant viruses.

The findings of this project will also benefit drug screening for anti-HIV drugs or aptamers. Prior work done by Dr. Chris DeCiantis has identified about 20 drug leads from over 2000 compounds in the NCI diversity set [192]. Because the similarity of F16A and NCp7 WT in sequence and basicity, we can use F16A to eliminate the non-specific hits from those drug leads or from an aptamer library.

Table 3-1. Trp titration results of NCp7 wild type and mutant proteins

#	Mutant Name	K _d _{Trp} (nM)				m ^b
		Avg.	SD	SD%	F ^a	
1	WT	29	± 4.5	± 15%	<u>1.0</u>	8
2	N5A	29	± 5.6	± 19%	<u>1.0</u>	5
3	F6A	33	± 3.3	± 10%	<u>1.1</u>	3
4	V13A	31	± 6.9	± 22%	<u>1.1</u>	3
5	F16A	720	± 26	± 4%	<u>25</u>	3
6	N17A	32	± 1.4	± 4%	<u>1.1</u>	3
7	G19A	45	± 2.6	± 6%	<u>1.5</u>	3
8	E21A	8.6	± 2.3	± 26%	<u>0.3</u>	3
9	G22A	45	± 6.7	± 15%	<u>1.6</u>	3
10	I24A	170	± 29	± 17%	<u>5.8</u>	5
11	P31A	23	± 5.8	± 25%	<u>0.8</u>	3
12	G40A	33	± 6.5	± 20%	<u>1.1</u>	3
13	Q45A	73	± 14	± 20%	<u>2.5</u>	4
14	M46A	62	± 10	± 16%	<u>2.1</u>	3
15	K14E-E21K	140	± 5.1	± 4%	<u>4.8</u>	7
16	K33E-E42K	40	± 11	± 27%	<u>1.4</u>	3
17	K38E-E51K	33	± 5.1	± 15%	<u>1.1</u>	3

a. F is the change factor for the K_d of a mutant NC compared with that of wild type

$$\text{NC. } F = (\text{K}_d \text{ of mutant}) / (\text{K}_d \text{ of WT}).$$

b. m is the number of measurements.

Table 3-2. ITC titration results of NCp7 wild type and mutant proteins

#	Mutant Name	K _d _{ITC} (nM)					N ^b		ΔH (kcal/mol)		m ^c
		30°C	RT ^a	SD	SD%	F	Avg.	SD	Avg.	SD	
1	WT	69	42	± 5.0	± 12%	<u>1.0</u>	1.1	± 0.034	-11E+03	± 600	4
2	N5A	124	80	± 26	± 33%	<u>1.9</u>	1.1	± 0.082	-9.8E+03	± 1500	4
3	F6A	97	58	± 23	± 40%	<u>1.4</u>	0.87	± 0.041	-12E+03	± 690	3
4	F16A	1790	1200	± 540	± 45%	<u>28</u>	1.4	± 0.13	-9.0E+03	± 2500	3
5	I24A	483	310	± 130	± 42%	<u>7.3</u>	1.4	± 0.18	-9.0E+03	± 910	4
6	P31A	81	46	± 5.6	± 12%	<u>1.1</u>	1.1	± 0.044	-13E+03	± 320	3
7	Q45A	177	120	± 12	± 10%	<u>2.8</u>	1.2	± 0.087	-8.2E+03	± 360	3
8	K14E-E21K	195	150	± 14	± 9%	<u>3.5</u>	0.79	± 0.021	-6.3E+03	± 430	3
9	K33E-E42K	75	43	± 4.1	± 9%	<u>1.0</u>	1.1	± 0.049	-12E+03	± 1100	3

a. K_d values measured at 30°C were converted to equivalents at room temperature (22°C) using van't Hoff equation. The K_d values at 22°C are rounded to two significant digits.

b. N is the average number of binding sites determined by ITC.

c. m is the number of measurements.

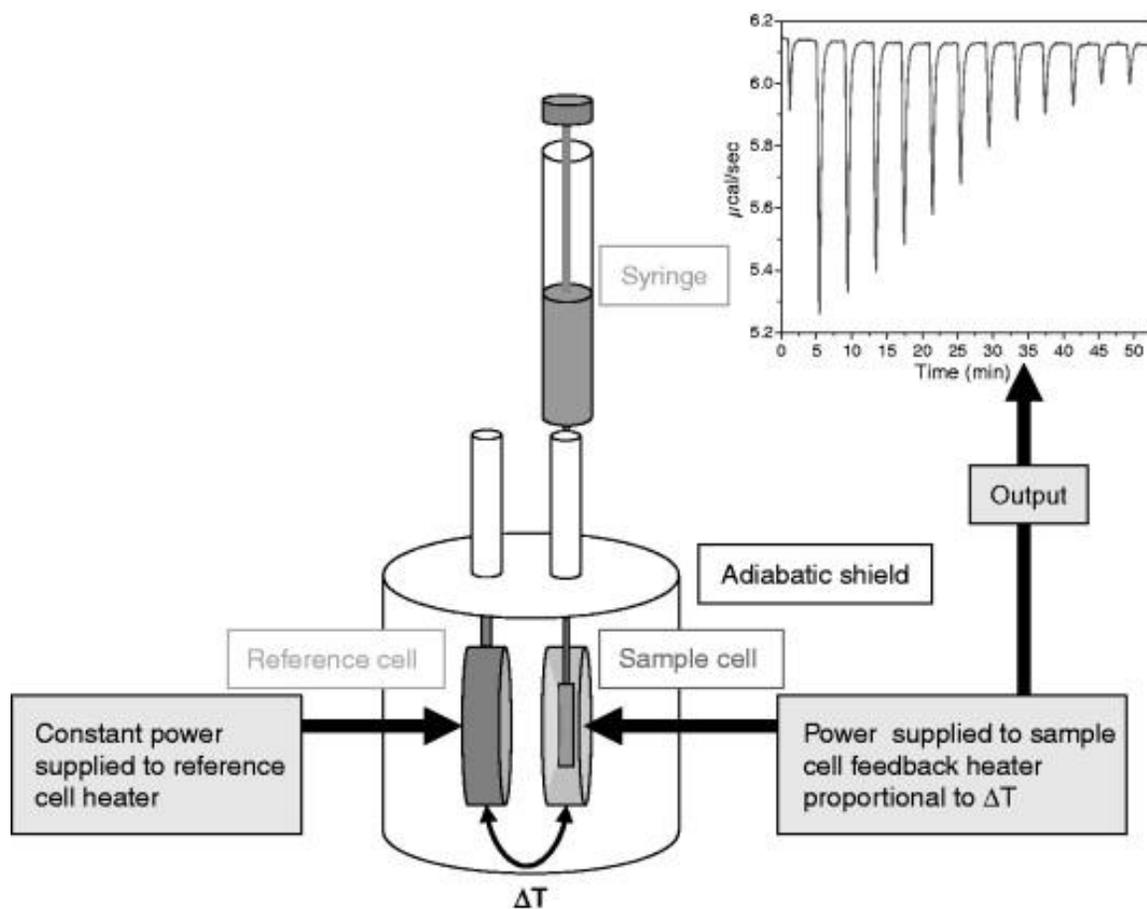


Figure 3-1. Schematic representation of isothermal titration calorimetry (ITC)

Major features of ITC instrument such as the reference and sample cells, sample injection syringe and the adiabatic shield are noted in the figure. This schematic diagram shows how the power applied by the instrument to maintain constant temperature between the reference and sample cells is measured resulting in the instrument signal.

Reprinted from *Methods in Cell Biology*, vol. 84, Freyer MW & Lewis EA, Pages 79-113,

Copyright 2008 [182], with permission from Elsevier.

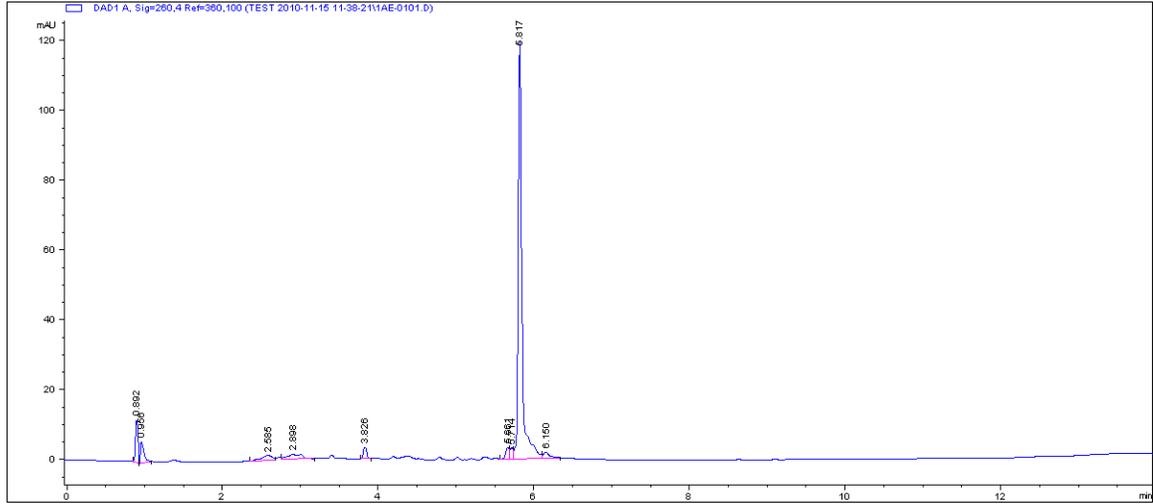


Figure 3-2. HPLC chromatogram of SL3 RNA

SL3 RNA sample purchased from IDT was run on an ion-exchange column on Agilent 1200 HPLC system. UV absorption signal was monitored and recorded at 260nm.

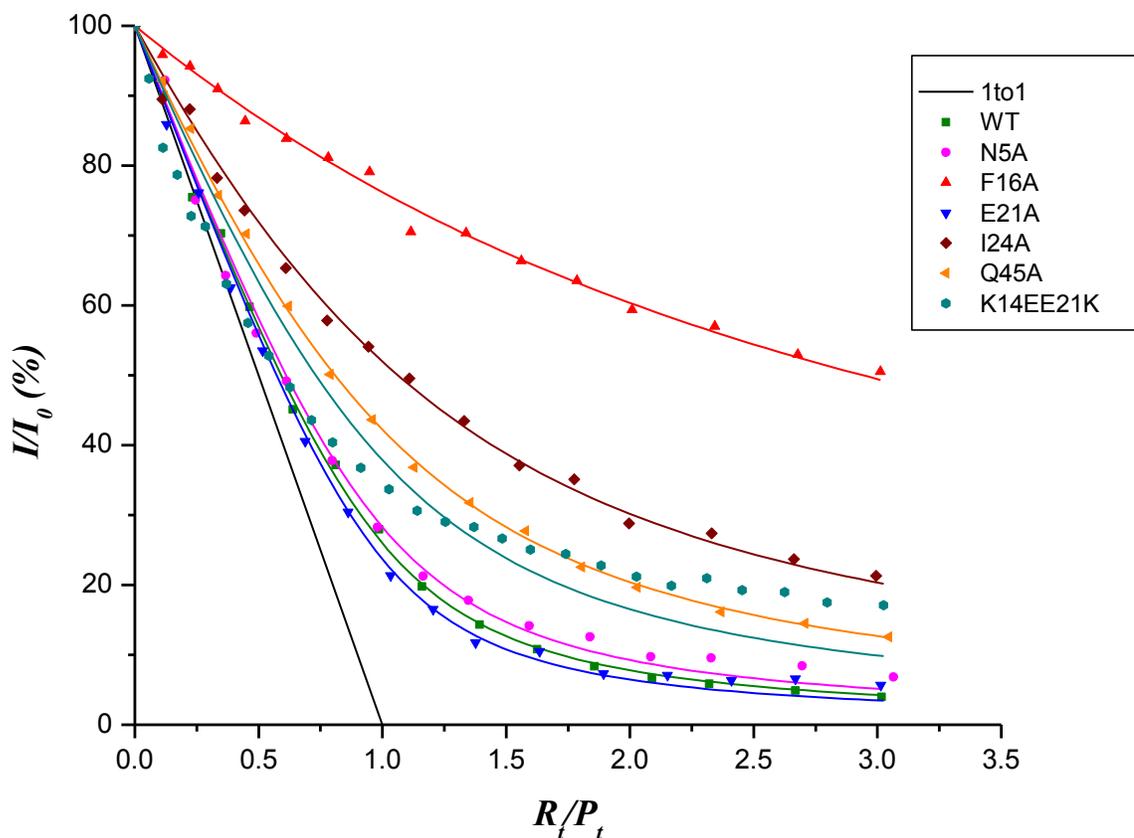


Figure 3-3. Tryptophan fluorescence titration curves of NCp7 wild type and mutant proteins with SL3 RNA

Compilation of tryptophan fluorescence titration curves of NCp7 wild type and mutant proteins with SL3 RNA. Only 7 sample curves are shown for clear presentation. Molar ratio of total RNA to total protein, R_t/P_t , increases from left to right. Colored curves are fitted curves using the data points of the same color coding. Data point sets are as following: green square, NCp7 wild type; pink circle, N5A; red triangle, F16A; blue triangle, E21A; brown diamond, I24A; orange triangle, Q45A; teal circle, K14E-E21K. The black line is the trend line for a 1:1 complex with an infinite binding constant (K_d^{-1}). See all calculated K_d values in table 3-1.

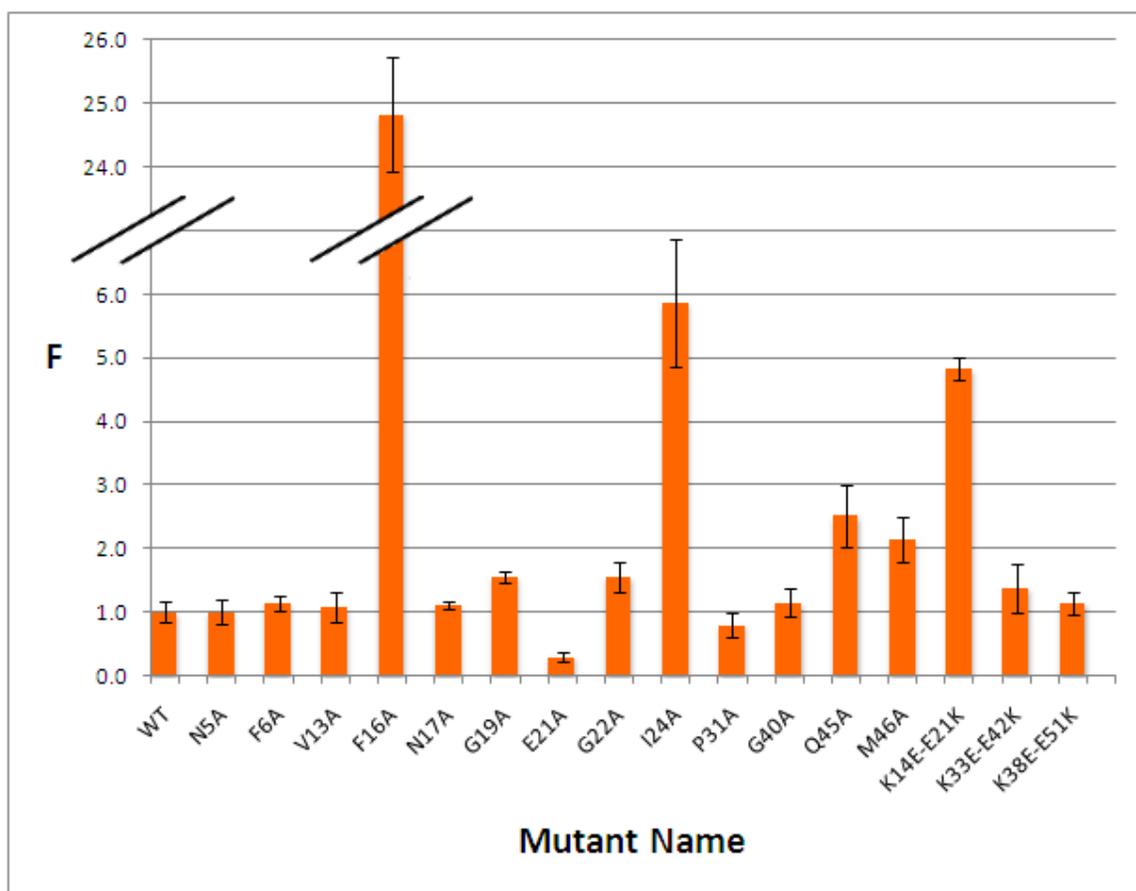


Figure 3-4. The effect of site-directed mutagenesis on the K_d values of NCp7 mutant proteins for SL3 RNA

K_d values of NCp7 mutants for SL3 RNA calculated from the tryptophan titration data were taken from table 3-1 and shown in the above bar chart as change relative to the K_d value of wild type NCp7. Y axis is the change factor F. $F = (K_d \text{ of mutant}) / (K_d \text{ of WT})$. The three mutants with highest K_d values are: F16A, I24A and K14E-E21K.

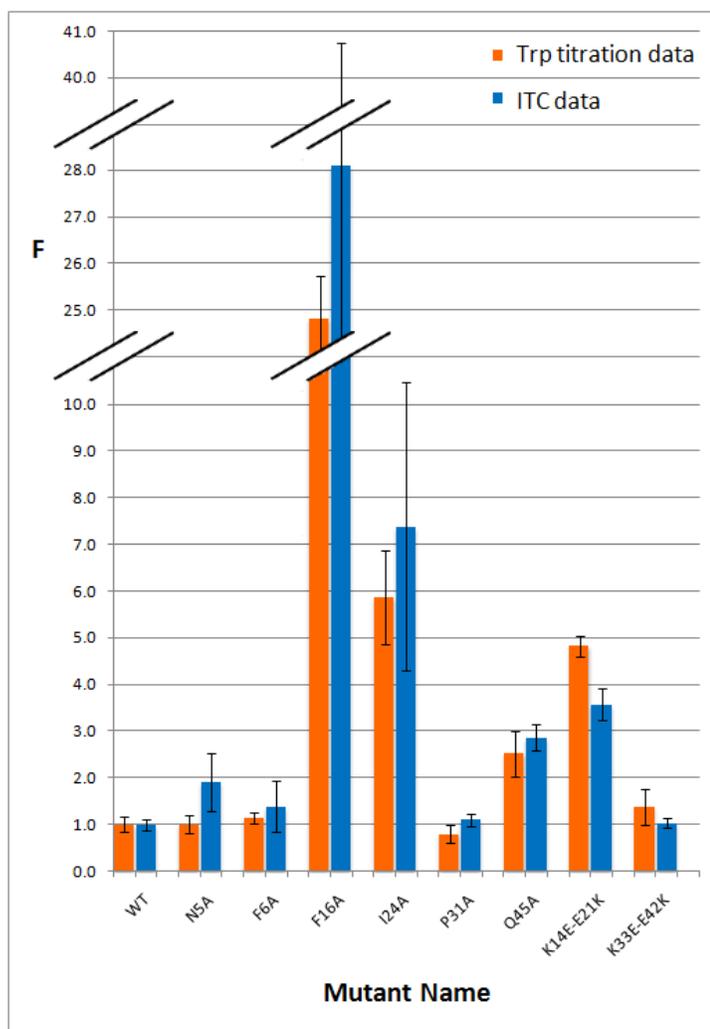
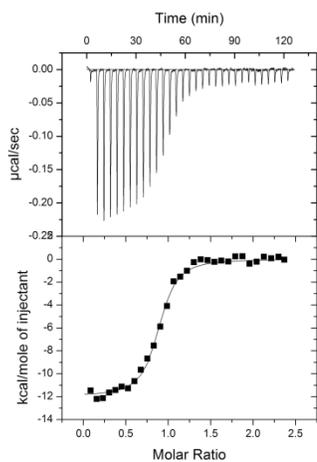


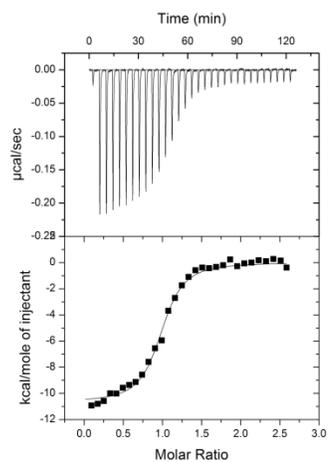
Figure 3-5. Comparison of the Trp Titration and ITC Data

K_d values of NCp7 mutants for SL3 RNA calculated from the tryptophan titration data (orange bar) and ITC data (blue bar) were taken from table 3-1 and 3-2 and shown in the above bar chart as change relative to the K_d value of wild type NCp7. Y axis is the change factor F. $F = (K_d \text{ of mutant}) / (K_d \text{ of WT})$.

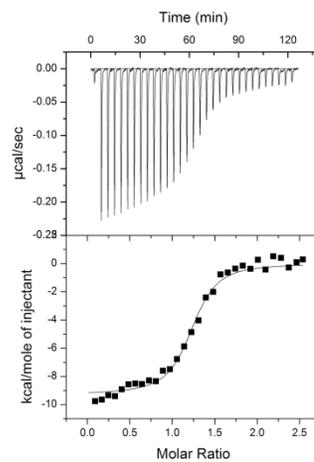
WT



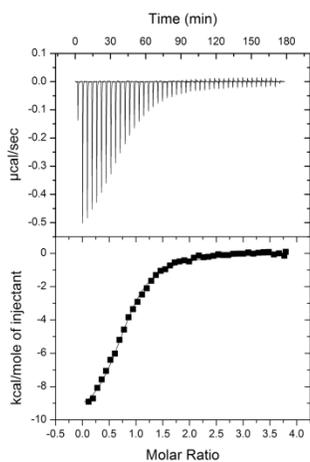
N5A



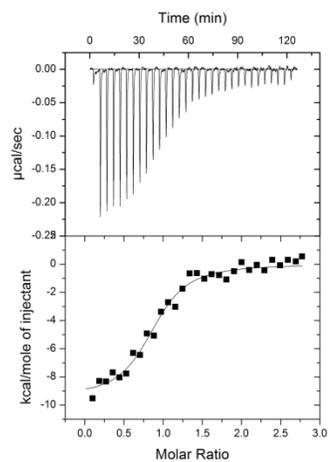
F6A



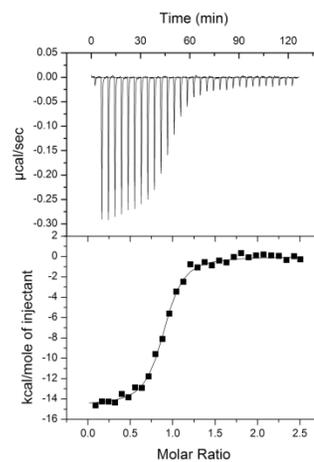
F16A



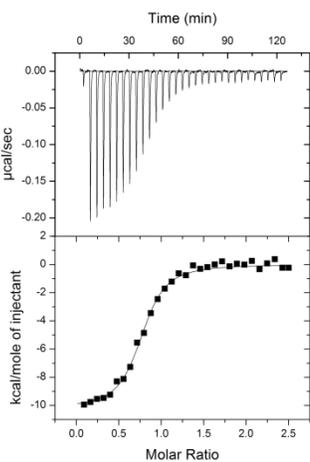
I24A



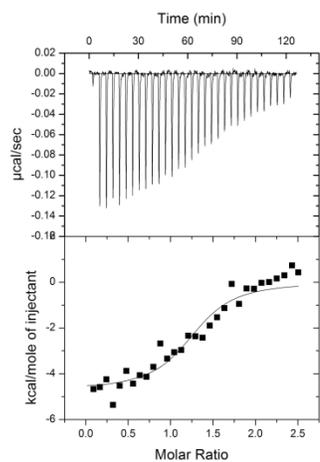
P31A



Q45A



K14E-E21K



K33E-E42K

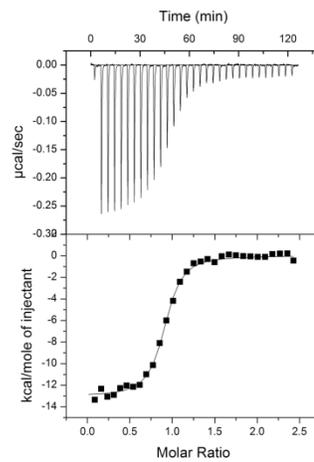


Figure 3-6. Representative ITC titration curves of NCp7 WT and mutant proteins
Note that the vertical axis scales are not constant.

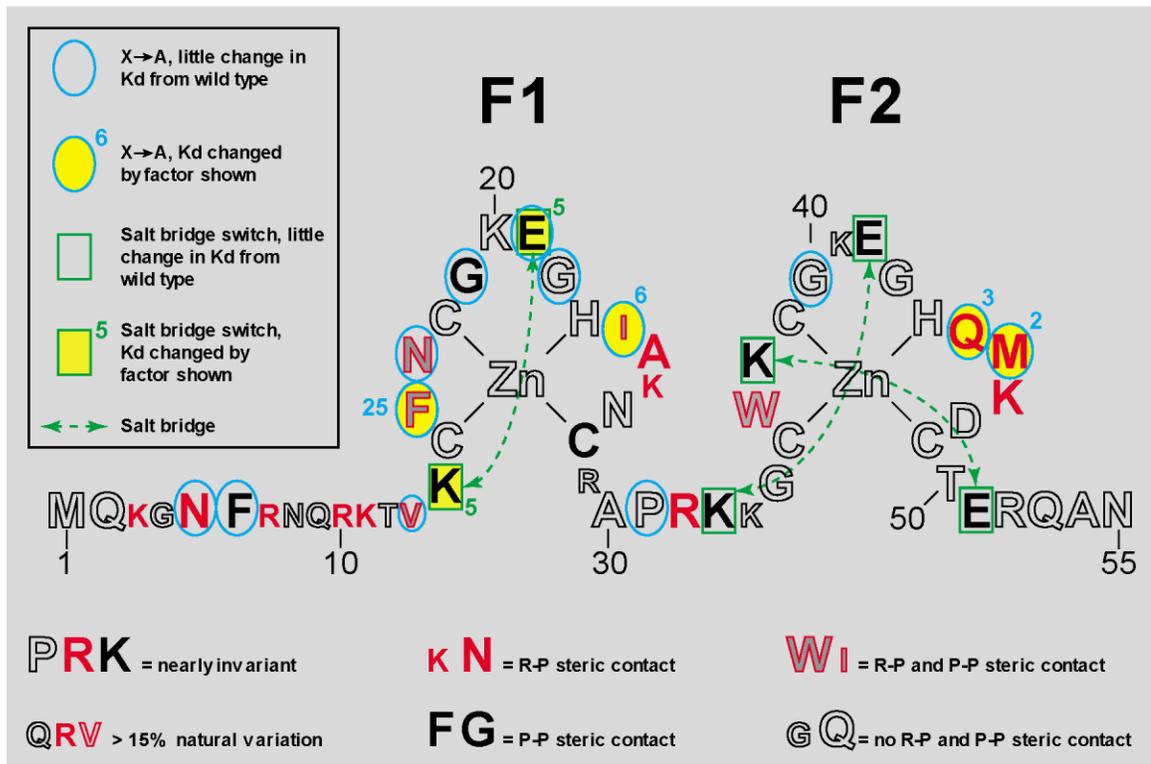


Figure 3-7. Comparison of the changes in NCp7 protein's affinity for SL3 RNA by alanine mutations on different residues

The highly conserved residues are shown in larger fonts and the residues in smaller fonts are more tolerant to variations. The red solid letters indicate residues with RNA-Protein steric contact. Black solid letters indicate residues with protein to protein steric contact. The red hollow letters are the residues with both protein-to-protein and RNA-to-protein steric contacts. The black hollow letters are the ones with no steric contacts at all. The green boxes mark the salt bridge switch mutants. The blue ovals indicate alanine scanning mutants. Yellow background is used to highlight the mutants with moderate to significant change in affinity, with the K_d change factor (F) shown next to the residue(s) (See Table 3-1).

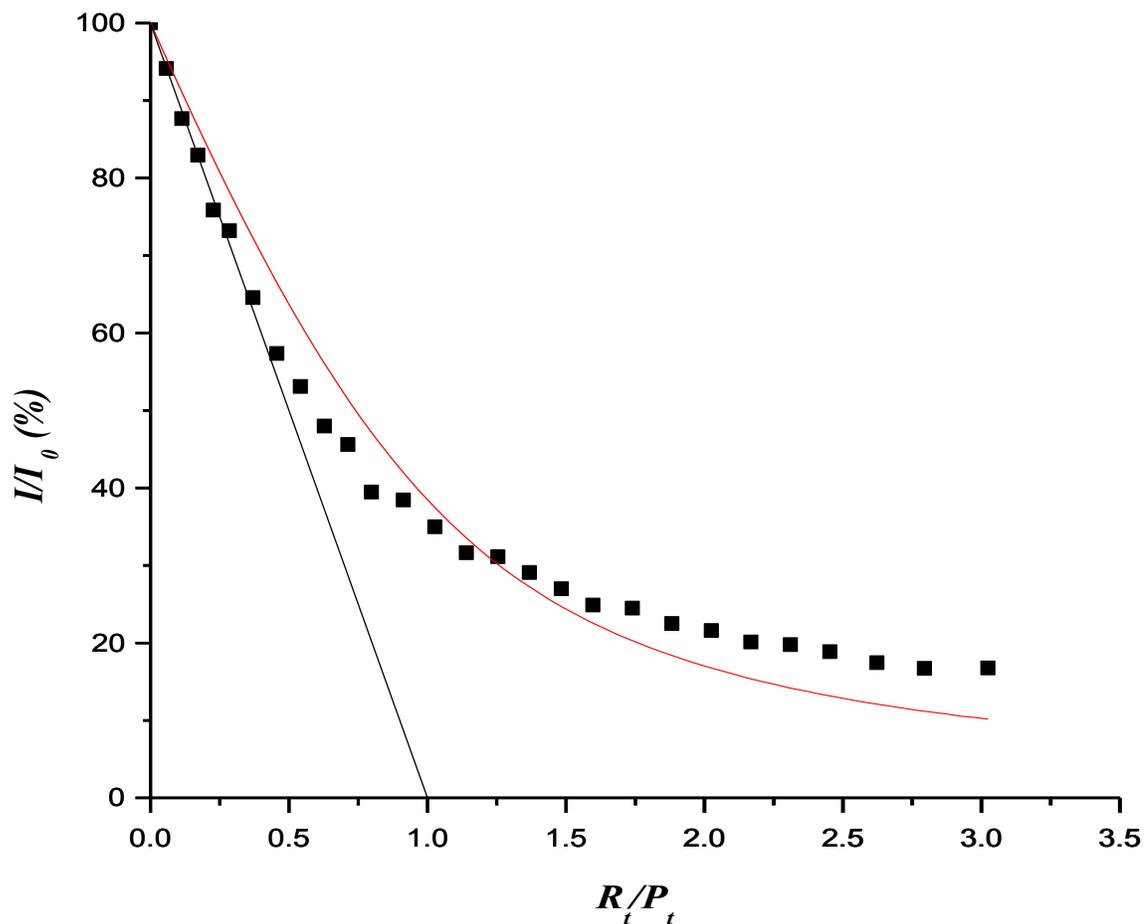


Figure 3-8. Tryptophan fluorescence titration curve of K14E-E21K for SL3 RNA

K14E-E21K titrated with SL3 RNA. The fitted titration curve does not provide a good fit for the data points, indicating a different binding model might play a role in K14E-E21K's binding with SL3 RNA.

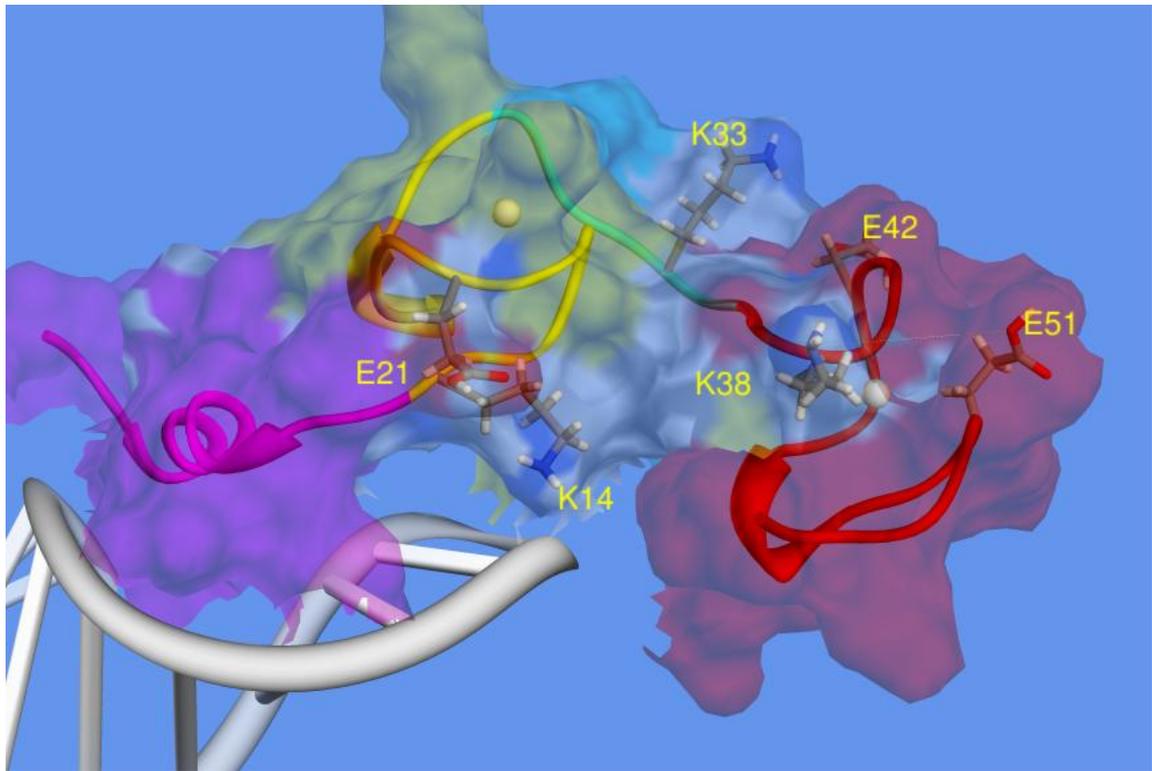


Figure 3-9. Positions of the three salt bridges in NC-SL3 complex

White ribbon: SL3 RNA; Colored ribbon: NCp7 protein; Stick model: side chain atoms of Lys¹⁴, Glu²¹, Lys³³, Lys³⁸, Glu⁴² and Glu⁵¹; Surface model: NCp7 protein; Magenta: N-terminal 3-10 helix of NCp7; Yellow: N-terminal zinc finger (F1); Cyan: linker region; Red: C-terminal zinc finger (F2). Surface is shown in 60% transparency. Drawn from PDB ID 1A1T using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081) [150].

Part II. Methods Development for Aptamer Discovery

Chapter 4. Introduction

For decades after the elucidation of the genetic code [193] and the articulation of the central dogma of molecular biology [194], nucleic acids had been envisioned mainly as carriers of genetic information through their sequences. Nonetheless, extensive studies on the relationship between the structure and function of nucleic acids had indicated that the structure of nucleic acids is equally, if not more, important than their sequence. The stem-loop structure has been shown to be critical for the recognition of HIV-1 packaging signal by NCp7 [145]. Single stranded nucleic acids can carry out various biological functions such as enzymatic activities and specific target binding [195-198]. The secondary and tertiary structure of a DNA or RNA molecule can be affected by the sequence of the molecule, by the ionic strength of the buffer or by the ligand bound to it. Because of the vast diversity in the secondary and tertiary structure of nucleic acids, researchers have begun to explore a new application of nucleic acids as, other than the genetic material, affinity ligands.

4.1. Aptamers

Aptamers (termed from the Latin *aptus*, meaning “to fit” [199]) are synthetic RNA or ssDNA molecules with stable three-dimensional structures and high affinity and specificity for certain target molecules. Since first described in 1990 [199, 200], aptamers have been developed over the past two decades to be a powerful tool for

drug discovery, biosensors, diagnosis and therapy [201-219]. They are now widely used in place of antibodies, especially in diagnostic applications [220].

Aptamers can be composed of naturally occurring nucleotide bases or chemically synthesized or modified derivatives. By incorporation of unnatural nucleotides or post-selection modifications, aptamer molecules can withstand hydrolysis or enzymatic degradation. The length of aptamers ranges from 15 nt to about 100 nt long. After modification, the size of aptamer molecules can range from 6kDa to 40kDa, most commonly be seen around 10kDa [220]. Because of the small size of aptamers, they usually penetrate or diffuse in biological tissues better than antibodies do. There also has been no evidence of aptamers' immunogenicity [221].

Aptamers usually have well-defined three-dimensional structures, especially when complexed with target molecules. Their structures result from a combination of Watson-Crick and non-canonical interactions. Common structural motifs seen in aptamers include G-quartets [222], non-canonical base pairs, in particular, purine-purine base pairs like G-G base pairs [223], and the GNRA tetra-loop (N is any nucleotide and R is a purine) [224, 225]. RNA aptamers usually can adopt a greater variety of complex tertiary structures than DNA aptamers can. This is because RNA molecules are more lenient in allowing non-Watson-Crick base pairs and have great rotational freedom in the backbone of their non-base-paired regions[226]. In addition, the 2'-hydroxyl groups on

the riboses of RNA molecules are potential hydrogen bond donors in intermolecular interactions [227].

The affinity of aptamers for their targets can vary a lot from target to target. They usually have affinities for small molecules in the micro-molar range, nano-molar level affinities for nucleic acid binding molecules and affinities in the nano-molar to subnano-molar range for proteins. Targets that have high affinity aptamers also tend to have measurable non-specific affinity for nucleic acids [220].

Nonetheless, aptamers can show significant specificity for their target molecules. Aptamers can distinguish different enzymes with similar activity, enantiomers of small molecules or minor sequence variants of macromolecules with several orders of magnitude difference in K_d values between the two similar species of molecules being distinguished by the aptamers. The targets of aptamers can be as small as metal ions or as big as large glycoproteins. The targets of aptamers are also chemically diverse, including organic dyes, neutral disaccharides, aminoglycoside antibiotics, dopamine, etc. [220].

To sum it up, aptamers are interesting because they are:

- 1) Synthetic molecules. The manufacturing cost of aptamers therefore is considerably lower than that of antibodies, which requires much longer time and

much more expensive animals to produce the desired antibodies. They are also easy to make, modify and label.

- 2) Structurally diverse and stable. The diverse structure allows for selection of aptamers for a large variety of target molecules, making aptamers good candidates in a wide range of applications. Assisted with unnatural nucleotides and chemical modifications, aptamers can have a much longer shelf life than antibodies.
- 3) Ligands with high affinity and specificity for their targets. Comparable affinities and specificities as antibodies make aptamers good substitutes for antibodies. Low immunogenicity and smaller size also make them attractive as therapeutic agents.

4.2. SELEX

Aptamers are usually selected from a large pool of random DNA or RNA sequences by an iterative *in vitro* process termed SELEX (systematic evolution of ligands by exponential enrichment) [200]. The name “SELEX” was coined two decades ago, but the history of *in vitro* selection for RNA ligands goes back much further.

In 1974, Sol Spiegelman and his colleagues reported an *in vitro* evolutionary method to select an RNA ligand that has specificity for Q β phage RNA polymerase [228]. This experiment and other work by Spiegelman and his co-workers are considered as

pioneering work to SELEX and aptamer science[197]. These experiments and SELEX share the same rationale. These methods consisted of rounds of replication and selection procedures and enabled selection of specific RNA ligands from a random pool of RNA sequences, which is in essence an *in vitro* Darwinian evolutionary process.

A schematic representation of SELEX procedure is shown in Figure 4-1 [229]. SELEX is a cyclic combinatorial procedure. Each round of SELEX involves three major steps:

1. Binding. Target molecules are added to and incubated with a pool of DNA/RNA sequences which have a completely random sequence region with flanking defined primer binding sites required in the following PCR amplification. The starting DNA/RNA pool usually contains 10^{13} - 10^{15} molecules and each one is a unique sequence. By mixing this DNA/RNA library with target molecules, potential aptamer molecules will bind tightly to the target molecules, while most sequences will remain unbound or loosely bound to the target molecules.
2. Partitioning. In this step, the bound and unbound species of DNA/RNA sequences are partitioned to remove the sequences with no or low affinities for the target. This is a critical step in the SELEX procedure. Partitioning can be achieved by affinity chromatography, magnetic beads, electrophoresis or microfluidics. After partitioning, most unbound or loosely bound DNA/RNA sequences will be washed off the target molecules and removed from the library.

3. Regeneration by amplification. The bound DNA/RNA molecules are eluted off the target molecules and used as PCR template to amplify and regenerate a new library which contains only the surviving members of the original library. But each sequence remaining in the new library now has multiple copies. This newly generated library is then used as starting library for the next round of SELEX.

A single round of selection is not sufficient to recover a tight binding aptamer sequence from a large random pool. Usually it takes multiple (10 to 15) rounds of SELEX runs before a successful aptamer sequence is enriched to the extent that it accounts for a significant portion of the final DNA/RNA library. At that point, a small number (1-100) of survivors will be cloned and sequenced to identify the potential aptamer sequence.

SELEX has its advantages. It is an *in vitro* process, which makes it easy for optimization and manipulation. The complexity of the starting pool is much greater than other combinatorial methods. And the selection does not require any prior knowledge about the structure of the target or the library.

However, SELEX also has its limitations. The multiple cycles of amplification and partitioning could make SELEX time consuming, labor intensive and require substantial amounts of purified target. A typical SELEX process may take several weeks before an aptamer sequence is identified. Repetitive PCR amplifications tend to produce a higher background in identifying the final aptamer sequences, making it difficult to tell a tight

binding aptamer from sequences with moderate affinities. More importantly, PCR amplifications might produce cycling artifacts, in which case some sequences are favored because they are easier to amplify, not because they have higher affinity for the target molecules.

4.3. Acyclic identification of aptamers

Many of the limitations of SELEX are intrinsic to repetitive cycling. To overcome these limitations, our lab proposed and validated a method for acyclic identification of aptamers (AIA) [230]. A schematic comparison between SELEX and AIA is shown in Figure 4-2 (McPike, M.P., 2011 [231]). This method eliminates the need for cycling by utilizing guided library design and the large sequencing capacity of next generation sequencing (NGS) technology. The overall screening time for AIA is shortened to less than a week compared with several weeks for SELEX. One successful AIA run can turn out millions of sequences while SELEX usually produces less than 100 sequences each time.

A typical AIA procedure includes the following steps:

1. Binding. Similar to SELEX, a random library of 10^{14} molecules is used to incubate with the target molecules in this step. The key difference is the design of the library. Even though an AIA library can contain a number of molecules similar to

SELEX libraries, the variable region of the AIA library is usually shorter (22nt or less) than that of SELEX libraries (often 40nt or more). This design ensures that the complexity of AIA library is lower than that of SELEX library so that each unique sequence in the AIA library is represented multiple times in the library. This kind of over-representation is essential for the acyclic selection to be able to obtain a meaningful signal-to-noise ratio in the final sequencing results.

2. Partitioning. This step is basically the same as the partitioning step in SELEX, where unbound and loosely bound sequences are separated from the tight binding sequences and removed from the library by stringent washing.
3. Sequencing and counting. After elution from the target, the tightly bound molecules are directly subjected to sequencing without going through iterative steps of amplification and partitioning. The partitioned library is linked to two universal adapters and sequenced on a next generation sequencer such as an Illumina GAII sequencer. The final sequences are analyzed using a Perl script and the good reads (sequences with intact flanking sequences) are counted. The count number of each sequence should be proportional to its affinity for the target molecules. So the sequence with highest count number is also in theory the sequence with highest affinity for the target.

As described so far, AIA overcomes the main bottleneck in SELEX by eliminating cyclic enrichment. But it has its own limitation of low complexity in the library. However, for libraries with high complexity and low representation, a pre-selection step can be added

before the AIA procedure to increase the sample representation level to a suitable level for AIA experiments [232]. Twenty years of research on SELEX can also be used to guide the design of AIA libraries to focus on regions of sequence and fold-space that are especially productive for aptamer discovery. AIA can be used to further refine the structure and sequence of aptamers selected using SELEX.

AIA also has great multiplexing potential. Multiplexing is limited with SELEX because an especially high affinity aptamer for one target can take over the evolving pool. While in AIA, multiple targets can be screened in parallel and aptamers with different level of affinities can be selected at the same time because the aptamers with lower affinity are not suppressed by evolutionary selection.

4.4. Next-generation sequencing (NGS) technologies

The success of AIA relies in large part on the commercial realization of next generation sequencing (NGS) technologies. Ever since next-generation sequencing instruments became commercially available in 2004, these fast-paced technologies have already made a significant impact in many fields of biological sciences, such as genome-wide association studies, transcriptome analysis, synthetic biology and systems biology.

Next-generation sequencing technologies generate fragment libraries by ligating platform-specific linkers to DNA fragments, amplify single strands of this library and

perform sequencing reactions on amplified strands. Next-generation sequencing technologies are highly streamlined, high throughput and high yield. Each NGS run can generate from several hundred thousand to tens of millions of reads [233].

Most imaging equipment used in modern sequencing systems cannot detect single fluorescent events, therefore in many cases templates need to be amplified before they can be sequenced. Based on the strategies used for template preparation, current commercially available NGS technologies can be divided into two major categories, which I'll briefly describe below [234]:

4.4.1. Bridge amplification based NGS technology (Illumina)

Illumina (San Diego, CA) is currently the market leader and the most widely used platform in the NGS industry. The Illumina technology generates DNA clusters by bridge amplification as shown in figure 4-3-a (adapted from Metzker, M.L. 2010) [234]. Each cluster contains thousands of copies of one single DNA fragment, ensuring sufficient signal intensity for detection during sequencing. There are 8 lanes on an Illumina sample slide. Each lane can produce 10-40 million sample reads. The clusters are sequenced using the sequence-by-synthesis approach.

4.4.2. Emulsion PCR based NGS technology (454/SOLiD/Polonator)

Emulsion PCR is commonly used for preparing sequencing templates in several types of NGS technologies (Roche/454, Life/APG (SOLiD) and Polonator). Emulsion PCR was originally designed to analyze rare gene mutations in a large pool of gene templates [235, 236]. In emulsion PCR, each emulsion droplet contains only one template, as shown in figure 4-3-b (adapted from Metzker, M.L. 2010) [234]. After amplification, each droplet contains thousands of copies of the single starting template sequence. Thus high signal to noise ratio can be achieved. Millions of droplets can be formed in one emulsion and make high throughput screening possible. The homogeneity and multiplexing nature of emulsion PCR makes it an ideal method for aptamer screening. The amplified sequences on the emulsion beads are sequenced using either the sequence-by-ligation or pyrosequencing approach, depending on the platform of choice.

4.5. Potential of NGS in aptamer discovery

A typical sequencing run on an Illumina GAII sequencer can generate over 100 million sample reads per slide and similar output can be expected from an ABI SOLiD sequencer. We prefer to get 5-15 million reads from each partitioned AIA library sample. This huge difference between the sequencing capability and the multitude of sample necessitates multiplexing of samples.

Besides multiplexing of samples, there's another way that we can utilize the full capacity of next generation sequencing technology. I call it the sequence-and-screen approach.

Compared to this approach, the traditional SELEX and AIA method can both be attributed as screen-and-sequence approaches. In screen-and-sequence approaches, the initial library is partitioned by incubating with target molecules, the partitioned library is subjected to sequencing and the resulting sequences are aligned and counted to generate the final aptamer candidate sequences. Whereas in sequence-and-screen approaches, the initial library is directly subjected to next generation sequencing to get the full coverage of every possible sequence in the library. The sequenced library should have formed small clusters of single sequences on the slide that was used in earlier sequencing. We propose that these clusters can be detected and screened by ds-DNA binding fluorescent dyes, fluorescently labeled complementary DNA probes or fluorescently labeled targets to generate hits in the clusters. The sequence of those hit clusters would then be tracked and pulled out of the sequence file and tested for affinities and specificities for the probes used.

The advantages of sequence-and-screen approach are as below:

1. There is no partitioning step, which could save time and resources.
2. Target molecules are not immobilized on the slide except by binding to aptamers, which are attached to the surface. This might be beneficial for the target to retain its natural structure or activity.
3. Multiplexing of target molecules is easy and convenient, parallel or subsequent screening is possible.

4. The library will get full coverage after screening. Effect of PCR artifacts would be minimal.

4.6. Specific aims

When the idea of the sequence-and-screen approach was first brought up by Dr. Mark McPike and I around 2007, next-generation sequencing technologies and instruments were much more expensive than they are as of today. Without easy access to a high capacity next-generation sequencer, we decided to create our own prototype sequence-and-screen platform based on known principles of commercialized next generation sequencers at that time. What we were trying to accomplish are listed as below:

- A. Establish prototype sample preparation platforms similar in principle as commercialized next-generation sequencers.
- B. Develop sensitive and efficient detection and screening methods on the established prototype platforms.
- C. Develop multiplexed screening protocols using the above-mentioned platforms and methods.

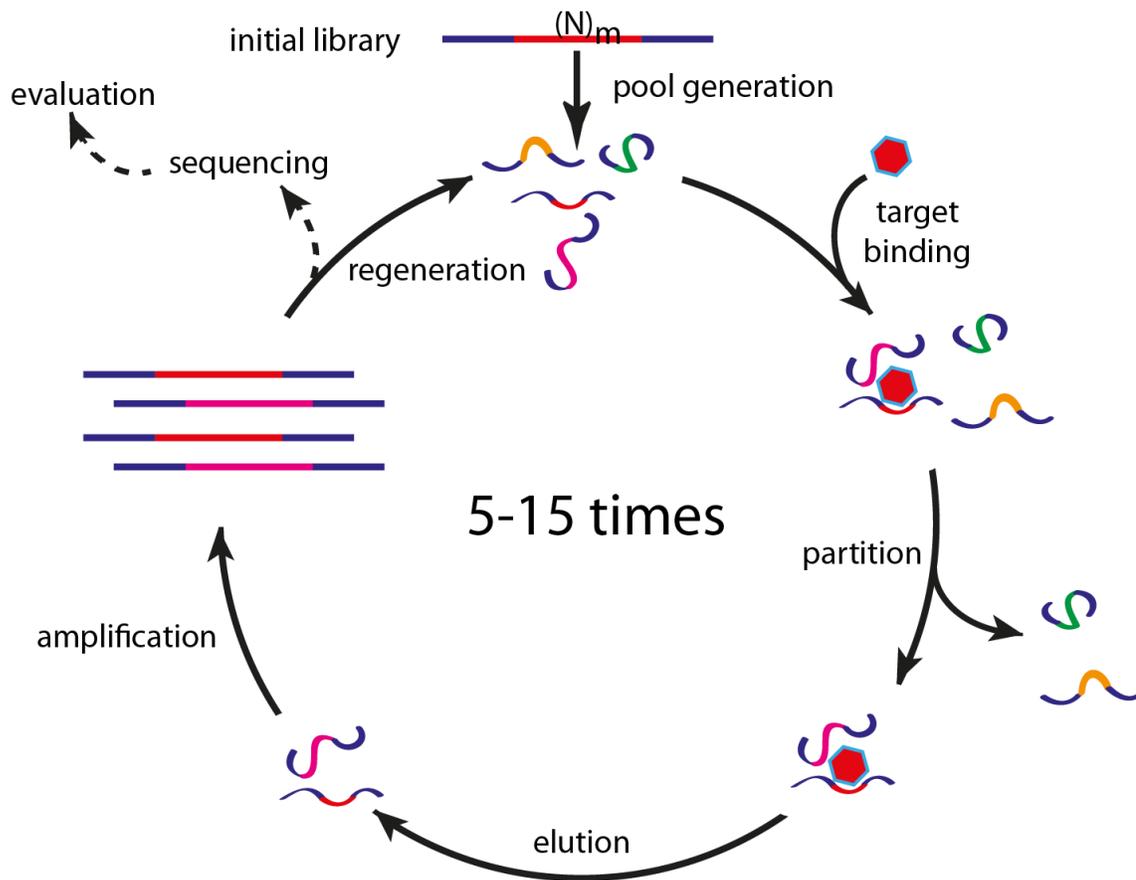


Figure 4-1. Schematic SELEX (Systematic Evolution of Ligands by Exponential Enrichment) procedure

Adapted from <http://www.molgen.mpg.de/~nabt/background.html> [229]. See description in text.

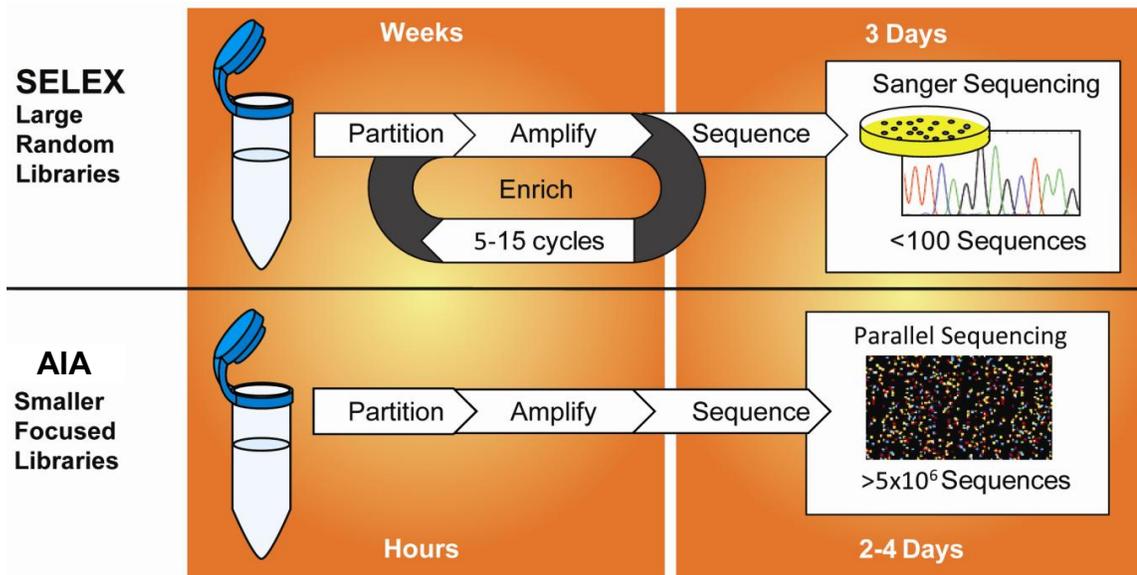


Figure 4-2. Comparison of SELEX and AIA

Reproduced from McPike, M.P. 2011 [231] with permission. See description in text.

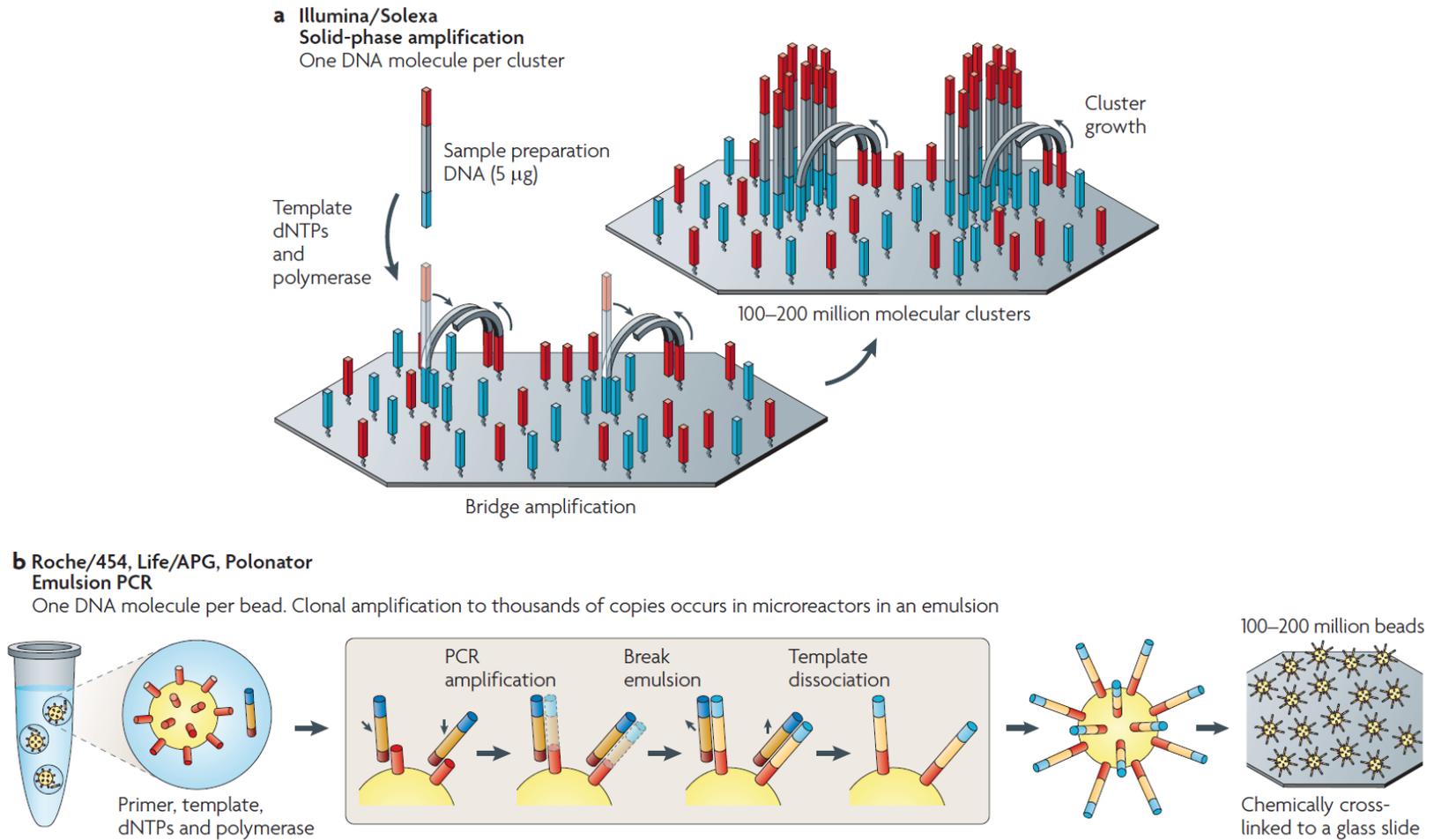


Figure 4-3. Template preparation strategies in NGS

Adapted by permission from Macmillan Publishers Ltd: Nature Review Genetics [234], copyright 2010. See description in text.

Chapter 5. Experimental Design and Preliminary Data

5.1. Experimental design

At the time these preliminary experiments began, our lab was planning to purchase a high-throughput DNA sequencer. Under consideration were the Illumina/Solexa GAIIx (using bridge amplification) and the Life Technologies (formerly Applied Biosystems) SOLiD or Dover Systems Polonator (both the SOLiD and Polonator use emulsion PCR for amplification). We decided to investigate strategies to utilize both bridge amplification and emulsion PCR for aptamer discovery.

In 1992, Bock et al. reported a 15-nt long DNA sequence (5'-GGTTGGTGTGGTTGG-3') for human alpha thrombin, which was later named thrombin binding aptamer (TBA) [201]. Thrombin has high affinity for TBA ($K_d \approx 100$ nM) but not for a similar sequence (5'-GGTGGTTGTTGTGGT-3') (TBA scrambled, or TBA-S) ($K_d > 1$ μ M) [237]. Because of this distinction, TBA and TBA scrambled sequences have often been used with thrombin as a model system in aptamer application development experiments to examine the validity and efficiency of the applications being developed [238-240]. In this project, I also chose

to use TBA and TBA scrambled sequences with thrombin in a set of proof-of-concept experiments for the proved robustness and convenience of this model system.

5.1.1. Bridge amplification

In this proof-of-concept experiment, I tried to create DNA clusters consisting of either positive control sequence (TBA sequence) or negative control sequence (TBA scrambled sequence) by bridge amplification on a glass slide, as depicted in figure 5-1. A lawn of a universal primer is attached onto a glass slide by tight binding between the streptavidin coating on the slide and the dual biotin label on the 5'-end of the primer. Also attached onto the slide is a template sequence with a stem-loop structure. The stem region of the template contains an identical sequence to the universal primer and its complementary sequence. The loop region contains the target sequence, in this case, either TBA or TBA scrambled sequence. Sequences and secondary structures of the two templates are shown in figures 5-2 and 5-3. The template sequences are attached to the slide in a small ratio to the primer sequences. Then bridge amplification reaction is carried out on the surface of the slide. If the amplification is successful, clusters containing the target sequence and its complimentary sequence should be formed on a background of the primer lawn, which would look like bacteria colonies on a LB-plate. These colonies could be detected by different probes such as DNA stain, labeled complementary DNA probe or labeled protein affinity ligand, as represented in figure 5-4.

5.1.2. Emulsion PCR

At the beginning of emulsion PCR, each emulsion bead is in an isolated compartment containing no more than one copy of the starting template sequence. Therefore each amplified bead should be equivalent to a DNA cluster containing a single DNA sequence and its complementary sequence. These beads are magnetic and can be harvested using a magnetic separator. The harvested beads are attached onto a glass slide and then subjected to detection and screening.

To successfully attach the emulsion PCR beads onto the glass slide, a capping oligo and a bridging oligo sequence are adopted according to the Polonator chemistry which is also based on emulsion PCR [241]. The capping oligo has 5'-phosphate and 3'-amine modification. The bridging oligo is partly complementary to the 3'-end of amplified sequence and partly complementary to the capping oligo. The capping oligo is ligated by T4 DNA ligase onto the 3'-end of the amplified sequences on emulsion PCR beads via splinted ligation guided by the bridging oligo [242]. Two set of capping oligos and bridging oligos were used to attach both control beads and sample beads onto the slide. The processed beads are then covalently attached onto a glass slide pretreated with aminosilane. A schematic setup for emulsion PCR process is shown in figure 5-5. Uniformly dispersed beads containing the successfully amplified sequences can then be detected and screened in a similar manner as shown in figure 5-4.

5.2. Methods and Materials

Primers and templates. The primers and templates used in this study are listed in table 5-1. All oligos were purchased from Integrated DNA Technologies, Inc. (Coralville, IA) in desalted form and used without further purification.

5.2.1. Methods for bridge amplification based detection

Sample coating. Dissolve biotinylated primer and template oligos in Micro Spotting Solution Plus (ArrayIt, Sunnyvale, CA) at designed final concentration. Mix the primer and template oligos in designed ratio. Attach a 25- μ l frame-seal incubation chamber (Bio-Rad) onto a SuperStreptavidin Substrates slide (ArrayIt, Sunnyvale, CA). Spot mixed DNA samples inside the chamber using a micropipette. Hybridize the sample at room temperature for 30 minutes.

Slide Blocking. Block the slide with 25 μ l SuperStreptavidin blocking buffer (ArrayIt, Sunnyvale, CA). Incubate the slide at room temperature for 1 hour. Wash the slide twice in 1x PBST (NaCl 8g/L, KCl 0.2g/L, Na₂HPO₄ 1.44g/L, KH₂PO₄ 0.24g/L, Tween-20 0.05%, MgCl₂ 5mM, pH 7.4) for 5 minutes each. Spin dry the slide.

Bridge Amplification. Prepare PCR mix 25 μ l: Turbo Pfu 2 μ l (2.5U/ μ l), dNTP 0.25mM each, 1x Pfu buffer. Fill the chamber with the PCR mix. Seal the chamber with a plastic

cover slip. Run PCR in DNA Engine Dyad Thermocycler or Thermo PX2 Thermocycler. PCR cycles were set up as the following: 92°C, 1 minute; (92°C, 30 seconds; 55°C, 30 seconds; 72°C, 30 seconds) x 40 cycles; 72°C, 1 minute; 4°C hold.

Sample staining. After amplification, remove the cover slip and rinse the slide twice with 1X PBST for 5 minutes each time. Stain the slide with 20 µl of Sybr555 DNA stain (Invitrogen) in 27ml of 1X PBST at room temperature for 5 minutes. Wash the slide twice with 1X PBST for 5 minutes each time. Spin dry the slide (optional).

Imaging. Stained slides were scanned in a GenePix 4400A microarray scanner (Axon™ Instruments, Union City, CA, USA). Images were taken using GENEPIX PRO 5.0 software (Axon).

5.2.2. Methods for emulsion PCR based detection

Bead loading. Re-suspend Dynabeads MyOne streptavidin C1 beads (10mg/ml) in stock vial by vortexing. Take 100 µl of suspended beads and transfer into a 1.5 ml eppendorf microcentrifuge tube. Add 100 µl of Bind and Wash buffer (5mM Tris-HCl, 0.5mM EDTA, 1.0M NaCl, pH 7.5). Mix by vortexing and remove all liquid on a magnetic particle concentrator (Dynal MPC-S magnet). Wash 2 more times in 200 µl Bind and Wash buffer, and re-suspend in 198 µl Bind and Wash buffer. Add 2 µl of 1mM forward primer with 5' dual-biotin label. Mix well by pipetting and incubate the mixture at room

temperature on a Dynabeads rotary mixer for 20 minutes. Remove all liquid, and wash twice in 200 μ l of Bind and Wash buffer, then once in 200 μ l of 1X TE (10mM Tris-HCl, 1 mM EDTA, pH 8.0). Remove all liquid and re-suspend beads in 200 μ l of 1X TE. These beads should now be around 5×10^9 beads/ml.

Emulsion PCR. Prepare the aqueous and oil phase according to the protocols established by Church lab at Harvard [241, 243] (also described in Appendix D-3), scale up or down when necessary. Mix oil and aqueous phase, vortex well for 5 minutes, distribute into PCR tubes or 96-well plate for approximately 50 μ l per tube/well. PCR amplify for 120 cycles.

Breaking emulsion. Add 100 μ l isopropanol to each tube or well containing the amplified PCR mixture, mix well by pipetting. Pool the mixture into a 15 ml polypropylene falcon tube, vortex for 1 minute. Spin down the beads for 30 seconds at 4000 rpm. Remove supernatant by decanting. Add 2 ml isopropanol and re-suspend the pellet by pipetting. Add 1 ml of isopropanol, vortex for 30 seconds. Spin down the beads for 30 seconds at 4000 rpm. Remove supernatant as much as possible. Re-suspend the pellet in 3 ml of NXS (10mM Tris pH 7.5, 1mM EDTA pH 8.0, 100mM NaCl, 1% (v/v) Triton X-100, 1% (w/v) SDS) and mix well by vortexing. Split the suspension into two 1.5ml siliconized microcentrifuge tubes, spin down the beads for 30 seconds at 3000 rpm. Remove the supernatant on a magnet. Wash the pellet in 1 ml of NXS once and then in 1 ml of 1X TE twice. Re-suspend the beads in 500 μ l 0.1 M NaOH. Incubate the suspension at room

temperature on a Dynabeads rotary mixer for 5 minutes. Wash the beads once with 500 μ l of 0.1 M NaOH, then twice with 500 μ l of 1X TE. Re-suspend and combine the beads in 30 μ l of 1X TE.

Bead capping. Prepare the hybridization mixture by adding 1.8 μ l 1mM bridging oligo 1 and 1.8 μ l 1mM bridging oligo 2 into 90 μ l 6X SSPE with 0.01% Triton X-100. Re-suspend amplified emulsion PCR beads in this solution and mix well by vortexing. Incubate the mixture at 56°C for 10 minutes. Mix well again and then incubate the mixture at room temperature on a Dynabeads rotary mixer for 10 more minutes. Wash the beads twice with 200 μ l of 1X TE. Prepare ligation mix by mixing the following: 84 μ l dH₂O; 10 μ l 10x T4 Ligase Buffer; 2 μ l 1mM Capping oligo 1; 2 μ l 1mM Capping oligo 2; 2 μ l T4 DNA ligase (3U/ μ l). Remove all liquid from beads and re-suspend them in the ligation mix solution. Incubate the mixture at room temperature on a Dynabeads rotary mixer for 1 hour. Remove ligation mixture solution on a magnet. Wash beads once with 200 μ l of 1X TE, once with 200 μ l of NX (10mM Tris pH 7.5, 1mM EDTA pH 8.0, 100mM NaCl, 1% (v/v) Triton X-100), then twice with 200 μ l of 1X TE. Transfer the beads to a new tube, wash once with 200 μ l of 1X TE, then re-suspend the beads in 30 μ l of 1X TE.

Aminosilane treatment of glass slide. Load slides (Erie Scientific, Portsmouth, New Hampshire) into a glass rack. Wash the slides by immersing the loaded rack in solution of 0.5% Triton X-100 in dH₂O and manually agitating the rack for ~1 minute. Drain all the liquid and wash slides thoroughly under running water to remove all detergent, then

rinse slides with dH₂O. Drain all the liquid and rinse the slides in 100% EtOH, then let air dry. Use glass pipettes to prepare 2% aminopropyl triethoxysilane (Pierce) solution in dry acetone in a glass beaker. Immerse the rack with completely dry (and ethanol-free) slides in the silane solution for 30 seconds. Manually agitate the rack. Remove and immediately immerse in fresh dry acetone, rinse for 1 minute. Remove rack and blow-dry partially with compressed air (to remove the bulk of the acetone). Let the slides completely air dry and store them under vacuum at room temperature.

Bead attachment. Wash beads three times in 50 µl 1x PBS (NaCl 8g/L, KCl 0.2g/L, Na₂HPO₄ 1.44g/L, KH₂PO₄ 0.24g/L, pH 7.4). Quickly re-suspend beads in 35 µl fresh BS3 (Bis(Sulfosuccinimidyl) suberate) solution in 1x PBS (2.86 mg BS3 in 1 ml 1x PBS). Add the beads onto the aminosilane treated center circle on a Teflon coated glass slide, apply the cover slip. Incubate the slide at room temperature for 1 hour. Wash the slide twice with 1x TE for 5 minutes each time to remove any unattached beads.

Staining the beads. Add 35 µl of 6x SSPE (0.9M NaCl, 60mM NaH₂PO₄, 60mM EDTA, pH 7.4) with 0.01% Triton X-100 to each circle, cover with a cover slip. Incubate at 56°C for 10 minutes. Wash immediately with 1x TE twice for 2 minutes on labquake. Add 35 µl of 6.7 µM 6-FAM labeled probe dissolved in 6x SSPE w/ 0.01% Triton X-100 to each slide, cover with a cover slip. Incubate at 56°C for 2 minutes, then 40°C for 10 minutes and 25°C for 10 minutes. Wash twice with 1x TE for 5 minutes each time on a labquake. Spin dry the slide. Add 25ul 1x self seal reagent (Bio-Rad) in 1X TE, cover with a cover slip. In

the case of thrombin staining, stain the slide overnight at room temperature using Cy3 or Cy5 labeled thrombin in 1X PBST (thrombin concentration 500nM). Keep the slide covered throughout the staining procedure. Wash the stained slide with 1X PBST twice and twice with 1X PBS for 10 minutes each time. Add 25 μ l of 1x self-seal reagent in 1X PBS, cover with a cover slip.

Visualization. Stained slides were visualized using epifluorescence microscopy on a Nikon TE2000 microscope (Nikon, Tokyo, Japan) equipped with a CCD camera, a 100X, 1.4 numerical aperture objective and standard FITC-Texas Red dichroic filter sets. Snap images were taken by the CCD camera using SPOT software (Diagnostic Instruments, Sterling Heights, MI), in both bright field mode and fluorescence mode. Bright field images and the fluorescent images were superimposed using Adobe Photoshop software.

5.3. Preliminary Results

5.3.1. Bridge amplification

Figure 5-6 shows the preliminary result from bridge amplification approach. 0.5 μ M template (TBA 55mer) was spotted on the slide as a positive control. The initial sample spotting solution contained 0.5 μ M primer (BA primer) and 0.01 pM template (TBA 55mer). The negative control spotting solution contained 0.5 μ M primer (BA primer) and

1 pM template (TBA 55mer) but no polymerase was added in the PCR mix for negative control. The slides were subjected to bridge amplification and subsequent staining with Sybr555 DNA stain.

Both positive control and sample spots were stained by Sybr555. Negative control spots weren't visible. Sample spots were much brighter than the positive control spots. This could be considered as a sign of successful amplification. But the bridge amplification slides suffered from uneven staining and high background problems. The staining results were not consistent and at times the staining results for controls and samples were not distinguishable. I also tried staining with Cy5-labeled thrombin. The staining results weren't satisfactory with very high background issues (data not shown).

I also used an open loop template design (TBA and TBA-S 63mer as templates and BA fwd and rev primers in table 5-1) as shown in figure 5-7 and tried to repeat *in situ* polony PCR in gel pads as Mitra et al. reported in 1999 [244]. The open loop template design was to make the template more accessible to primers and thus make the amplification more efficient. The polony PCR protocol was adopted also to increase PCR efficiency. The gel pad was supposed to hold the PCR mixture within and create a solution-like environment for the reaction while holding the amplified DNA clusters in place. However, both approaches did not help solve the high background problem. In fact, the background was worse in the gel pads (data not shown). The possible reason

might be that the gel monomers and/or APS and TEMED inhibited the enzyme activity and the probes were difficult to diffuse into or to be washed out of the gel pad.

5.3.2. Emulsion PCR

One set of the initial emulsion PCR experiment results is shown in Figure 5-8. In this experiment, the results show that the control beads (no template added) and sample beads (using TBA 89mer as template) were both attached to the silane-treated glass slide. But only the sample slide was stained by 6-FAM labeled TBA probe. Successful amplification on the sample beads was indicated by the staining result. Logically, the next step was to see whether the sample beads could be stained by fluorophore-labeled thrombin.

As can be seen in Figure 5-9, the emulsion PCR beads were stained with Cy3-labeled thrombin. But to our disappointment, both control slide and sample slide were stained by Cy3-thrombin. And there weren't distinct differences between the fluorescence signal intensity of the two slides, which suggested non-specific binding between thrombin and the control beads.

To reduce this kind of non-specific binding, 10mg/ml casein was added into the block solution and 2mg/ml casein was added into the staining solution. And one more set of negative control (where no forward primers were loaded onto the beads) was

examined. But the staining result still show signs of non-specific binding of thrombin to the control beads, as shown in figure 5-10. It suggested that casein did not provide sufficient blocking in this experimental setup.

To determine the source of the non-specific binding, another set of control experiments was carried out. Beads loaded with forward primers but not stained were used as negative control 1. Unloaded beads were used as negative control 2. Positive control beads are loaded with a 5'-dual biotin labeled DNA 35mer consisting 20 T's on the 5'-end and the 15mer TBA sequence on the 3'-end (T20TBA). Both negative control 2 and the positive control beads were stained with Cy5-labeled thrombin in eppendorf tubes for overnight instead of 30 minutes on slide. The staining results showed clear distinction between the negative and positive control beads, as shown in Figure 5-11. This suggests that the non-specific binding was not between thrombin and the bead surface. Also the beads or the primers did not have background fluorescence. The positive control had very satisfactory staining result. The contributing factors might be the overnight staining in tubes and the TBA sequence being at the far end of the attached oligos.

To examine the effect of staining conditions, a set of experiments similar to that shown in Figure 5-9, except that the beads were stained overnight in tubes instead of 30 minutes on slide. The results showed that the fluorescence signal on sample beads did

improve compared with the control beads, as shown in Figure 5-12. But the improved signals were not as bright as the positive sample in Figure 5-11.

To test the hypothesis that the TBA sequence binds thrombin better on the free end of the sequence than in the middle, a new emulsion PCR template (Temp-Alol) was designed as shown in Table 5-1 and Figure 5-13. A restriction enzyme recognition site was designed into the template to the 5'-end of the TBA region. After amplification, the beads were digested with the restriction enzyme Alol to expose the TBA sequences so as to allow easy access for thrombin binding. The control (no template) and sample beads were digested and stained with 6-FAM labeled TBA probe (Figure 5-14) or Cy5-labeled thrombin (Figure 5-15) in parallel. The results showed that Alol digestion increased the Cy5 signal but decreases the 6-FAM signal. And non-specific binding between thrombin and the control beads are still noticeable.

5.4. Discussion

Amplification efficiency. It has been reported that PCR amplifications on a solid surface has much lower efficiency than solution-phase PCR reactions [245, 246], which might explain the considerably greater difficulties in establishing a prototype bridge amplification platform than an emulsion PCR platform. The close proximity of the clustered complementary DNA strands makes them not able to freely move inside the reaction chamber and results in low efficiency of bridge amplification. The low

amplification efficiency made it difficult to distinguish between samples and controls, therefore the high background issues. The amplicon length also has a negative correlation with the efficiency of the emulsion PCR [247]. The in-gel polony PCR approach was supposed to improve the amplification efficiency and help with polony formation [244]. But the less-than-satisfactory results called for more optimization before an ideal condition for bridge amplification could be worked out.

Staining efficiency. Staining of the amplicons also works better in solution phase than on a solid surface, as demonstrated in Figure 5-11 and previously stained images. This is probably because the probes, especially the protein probes, have less steric hindrance staining the amplicons floating in solution than staining the ones being attached onto a planar solid surface. Also, the on-slide staining for emulsion PCR beads was performed in a small chamber, which usually has smaller volume than tubes and is less efficient for the staining solution and/or the PCR mix to reach the beads. When comparing different probes, the performance can be listed from the best to the worst as: thrombin (Cy5) > thrombin (Cy3) > TBA probe (6-FAM) > Sybr555. The sybr555 DNA dye is a non-specific DNA dye and prefers double stranded DNA over single stranded DNA. The clusters formed by bridge amplification have only partial double stranded stem region near the surface of the slide, so the staining efficiency was not good in the case of Sybr555. The TBA probe and thrombin are both specific probes. The fact that thrombin stained better than the TBA probe indicated that the TBA sequence might have higher affinity for

thrombin than for the complementary TBA probe sequence. The structural stability of a G-quadruplex makes it easier to bind thrombin than the TBA probe sequence.

Background issue. In the whole slide staining and washing steps, the slides were stained in a plastic tube filled with appropriate solutions and incubated on a lab rocker. It's observed that slides stained and washed in this manner tend to have weaker signals on both ends of the slides and a streaky pattern in background can often be seen. Using a larger tray and enough solutions to cover the slide throughout the time of incubation on the lab rocker might improve this issue of uneven staining. The catch is that more probes would be needed as the volume would be larger in this staining setup. But that should be a small price to pay. Other factors contribute to high or uneven backgrounds include insufficient blocking, air bubbles in the staining chamber, and slides drying out before wash. Non-specific binding could result from residual components from the PCR mixture, which contribute to the high background as well as false positive signal. It was also observed that Cy3- or Cy5-labeled thrombin has reduced affinity for TBA sequence, which indicates possible structure change or denaturation of thrombin due to the fluorophore labeling [248]. This fact might also contribute to the staining/background issues.

5.5. Conclusions and suggestions for future work

5.5.1. Bridge amplification

The bridge amplification platform did not work as planned, mostly because of the low amplification efficiency and staining issues. Initial data did show signs of successful amplification but the results weren't consistent.

With the recent acquirement of an Illumina GAII sequencer in our lab, the need to create a prototype bridge amplification platform is not as significant as before. We are planning to work on screening protocols using actual Illumina sample slides, which could be an extension of this project.

5.5.2. Emulsion PCR

The emulsion PCR platform can be considered as partly accomplished. There was clear sign that template sequences were successfully amplified on magnetic beads using emulsion PCR. The results were reproducible and the detection methods worked well.

But there are still issues with non-specific binding and detection sensitivity. One should be able to improve or solve these issues given more time and resources. But without access to an NGS sequencer based on emulsion PCR, this project currently remains at the prototype stage.

Table 5-1. Template and primer sequences used in aptamer discovery method development

	Name	Length	modification	Sequence (5'->3')
Bridge Amp	TBA	15	-	GGTTGGTGTGGTTGG
	TBA Scrambled	15	-	GGTGGTTGTTGTGGT
	TBA 55mer	55	5'-dual biotin	CGGTTGCGTTTACTGCCGTCGGTTGGTGTGGTTGGACGGCAGTAAACGCAACCG
	TBA-S 55mer	55	5'-dual biotin	CGGTTGCGTTTACTGCCGTCGGTGGTTGTTGTGGTGACGGCAGTAAACGCAACCG
	BA Primer	20	5'-dual biotin	CGGTTGCGTTTACTGCCGTC
	TBA 63mer	63	5'-dual biotin	AGAAGGAGAAGGAAAGGGAAAGGGGGTTGGTGTGGTTGGGGTTTGGGTTGGTTTGGGTTGGTG
	TBA-S 63mer	63	5'-dual biotin	AGAAGGAGAAGGAAAGGGAAAGGGGGTTGGTGTGGTTGGGGTTTGGGTTGGTTTGGGTTGGTG
	BA Fwd Primer	24	5'-dual biotin	AGAAGGAGAAGGAAAGGGAAAGGG
	BA Rev Primer	24	5'-dual biotin	CACCAACCCAAACCAACCCAAACC
	TBA 89mer	89	-	AACTGCCCCGGGTTCCCTCATTCTCTACGCCAACCACACCAACCCGTATCACCGACTGCCATA GAGAGGAAAGCGGAGGCGTAGTGGTT
	Fwd primer	29	5'-dual biotin	CCACTACGCCTCCGCTTTCCTCTCTATGG
	Rev primer	21	-	CTGCCCCGGGTTCCCTCATTCT
	Bridging oligo 1	24	-	GTGAGCTTCGTCTGCCCCGGGTTCC
	Bridging oligo 2	28	-	GTGAGCTTCGTCCATAGAGAGGAAAGCG
Emulsion PCR	Capping oligo 1	10	5'-phosphate, 3'-amine	CGAAGCTCAC
	Capping oligo 2	11	5'-phosphate, 3'-amine	ACGAAGCTCAC
	TBA Probe	12	5'-6-FAM	CCAACCACACCA
	T20TBA	35	5'- dual biotin	TTTTTTTTTTTTTTTTTTTTTTTTTTGGTTGGTGTGGTTGG
	Temp-Alol	77	-	AAGTCAGCTATGCCGCTACGGACCCGGGGTTCTCAGATCCCAACCACACCAACCTCCAAACAC CTCCGCATCACCTT
	Rev primer Alol	21	-	GTCAGCTATGCCGCTACGGAC
	Fwd primer Alol	21	5'-biotin	GGTGATGCGGAGGTGTTTGGGA

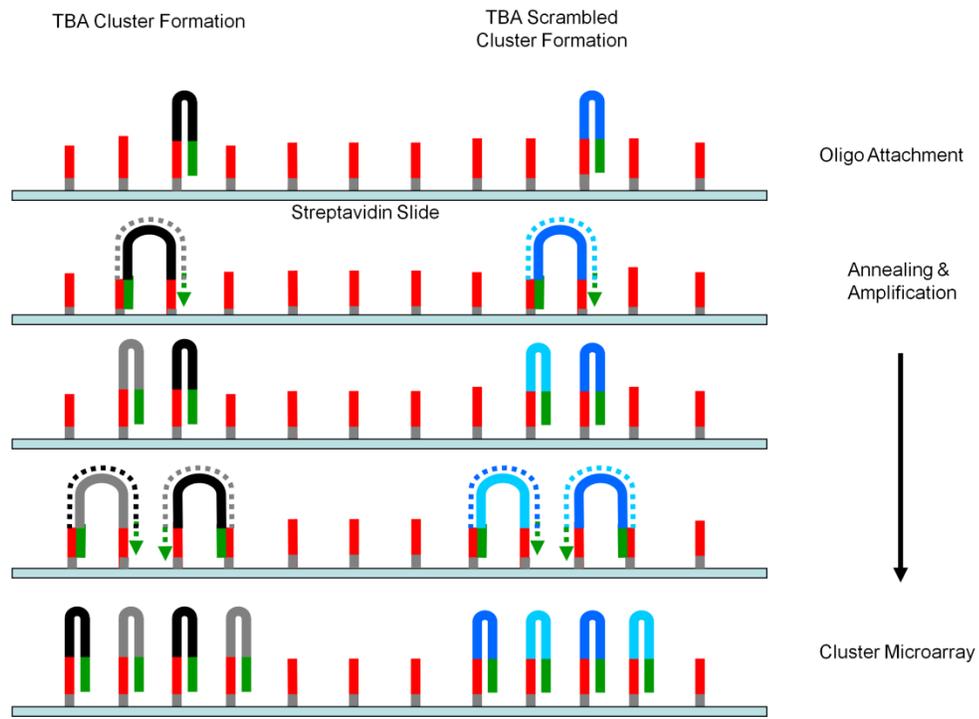


Figure 5-1. Formation of DNA clusters by bridge amplification

Red: universal primer; green: complementary region to the universal primer; Black: TBA sequence; gray: complementary sequence of TBA; blue: TBA scrambled sequence; light blue: complementary sequence of TBA scrambled; dotted arrow indicates the bridge amplification reaction and its extension direction.

p1t22gif by D. Stewart and M. Zuker
© 2006 Washington University

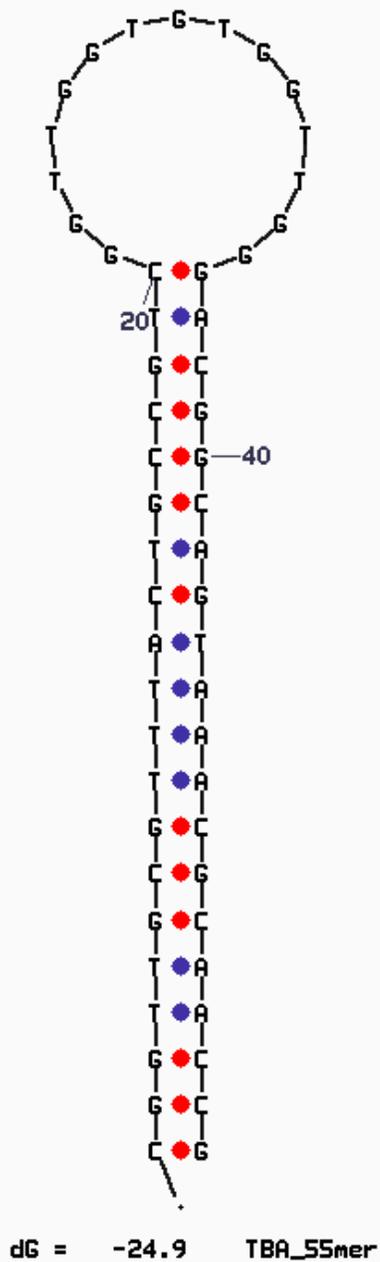


Figure 5-2. TBA 55mer

Drawn with mfold web server online tool at <http://mfold.rna.albany.edu/?q=mfold> [249].

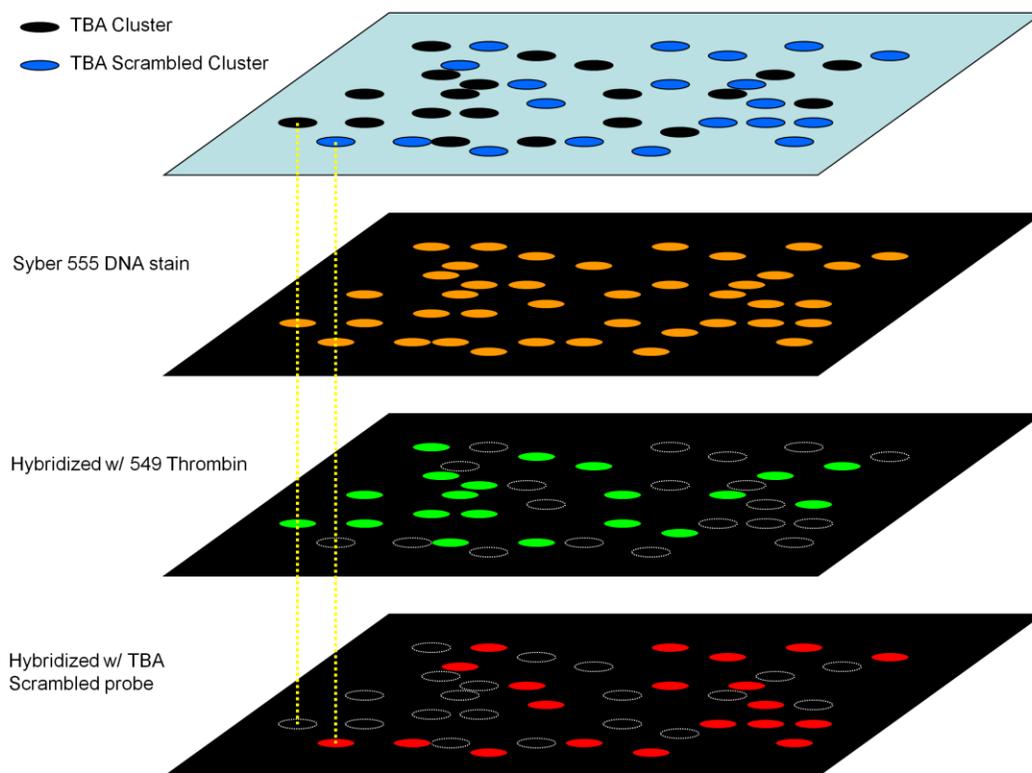


Figure 5-4. Detection of DNA clusters after bridge amplification

Black: TBA cluster; blue: TBA scrambled cluster; gold: DNA clusters stained with Sybr555 DNA stain; green: DNA clusters stained with labeled thrombin; red: DNA clusters stained with labeled TBA scrambled probe.

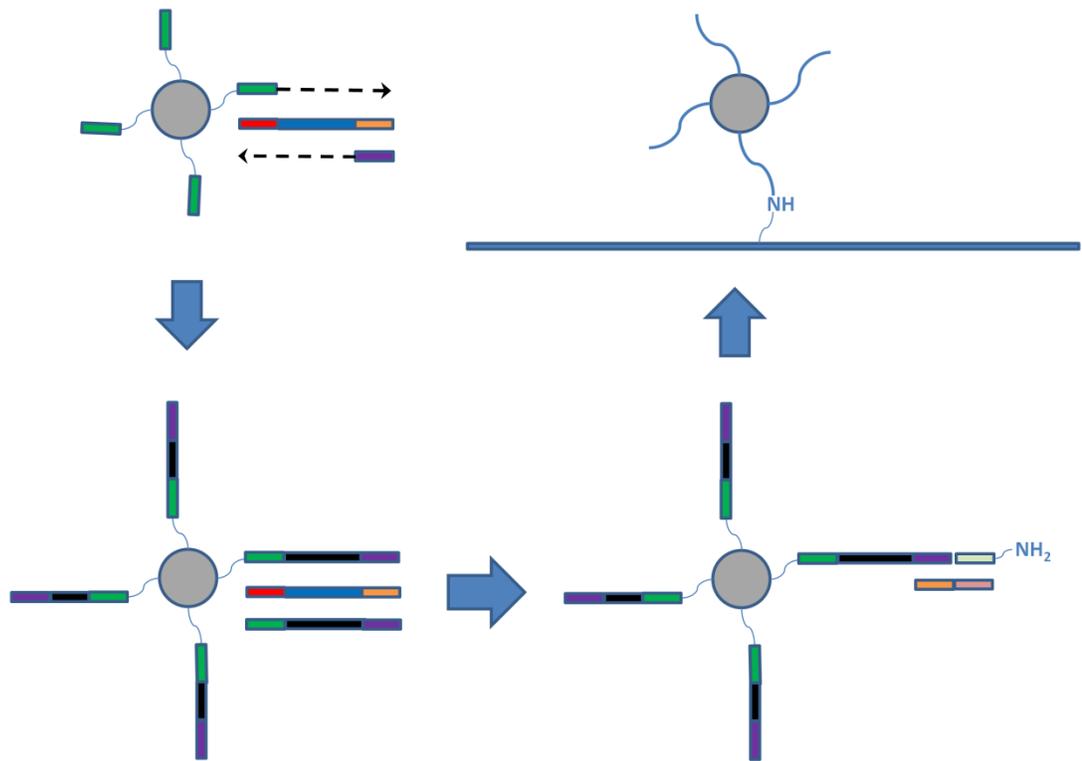


Figure 5-5. Schematic setup of emulsion PCR

Grey circle: magnetic emulsion beads; green bar: forward primer; purple bar: reverse primer; the three-part bar: template; light green bar: capping oligo; the two-part bar: bridging oligo; arrows: extension direction.

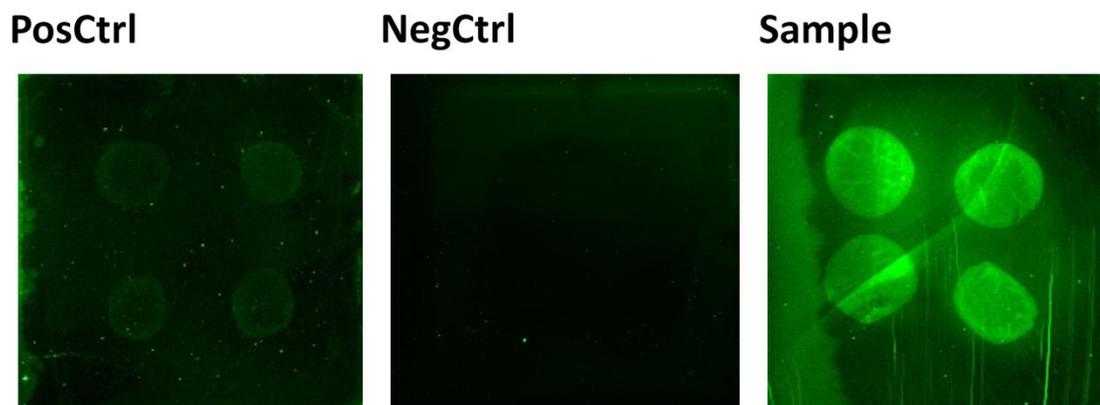


Figure 5-6. TBA clusters formed by bridge amplification

Positive control: 0.5 μ M template (TBA 55mer); Negative control: 0.5 μ M primer (BA primer) + 1 pM template (TBA 55mer), no polymerase; Sample: 0.5 μ M primer (BA primer) + 0.01 pM template (TBA 55mer). Controls and sample were stained with Sybr555 DNA stain for 5 minutes. All chambers went through the same thermo cycling and staining procedure.

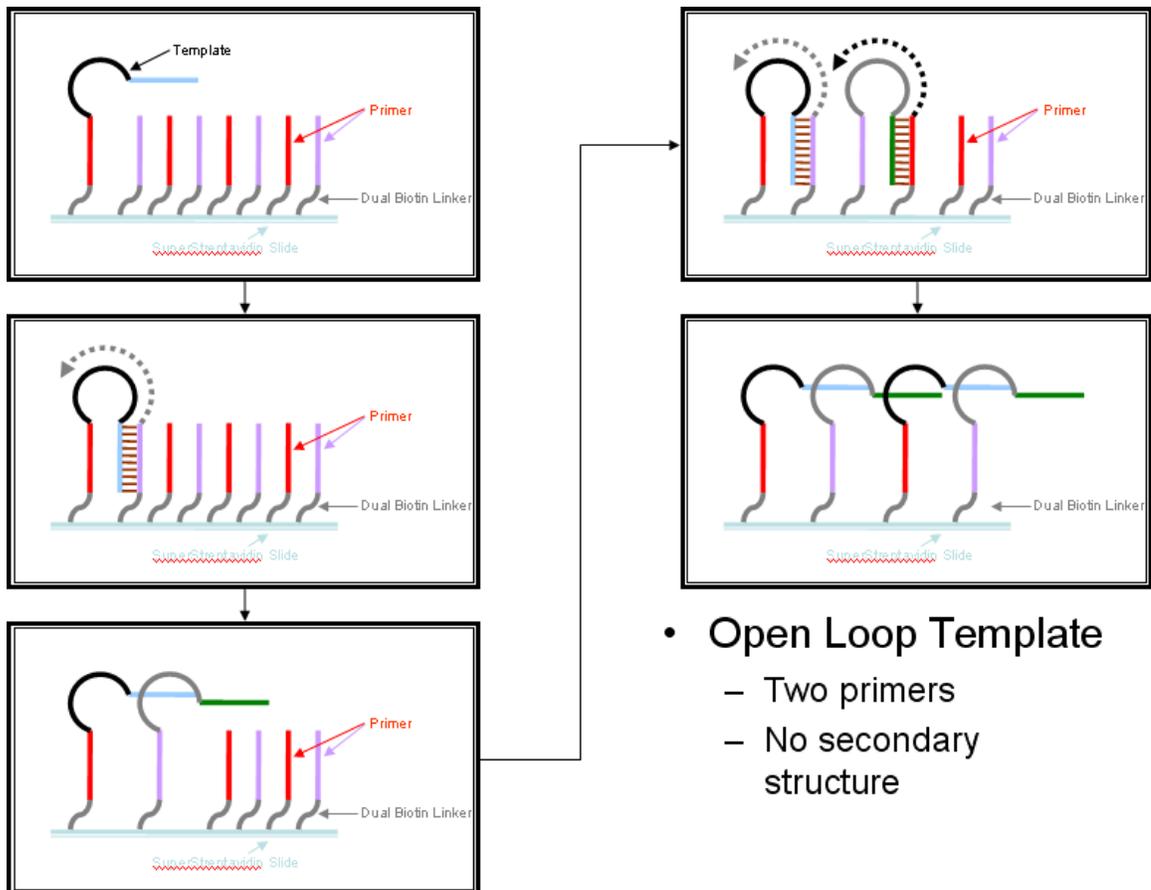


Figure 5-7. Open loop template design

Two set of non-complementary primers are attached on the solid surface. The 5'-region and 3'-region of the template match the two primers, respectively. The template does not have stable secondary structure by design.

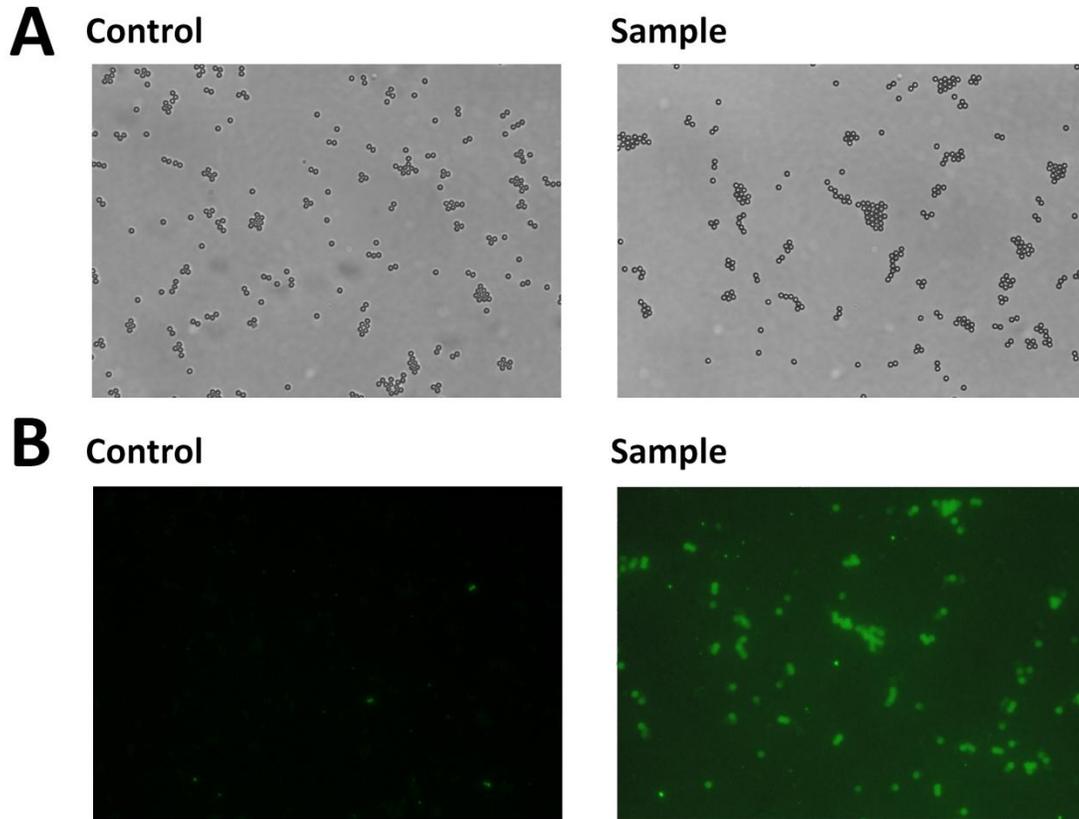


Figure 5-8. Detection of emulsion PCR beads by TBA probe (6-FAM)

Panel A: bright field images. Panel B: fluorescent images. Control beads did not contain template in the PCR mixture. Both control beads and sample beads went through the same thermo cycling and were stained with 6-FAM labeled TBA probe.

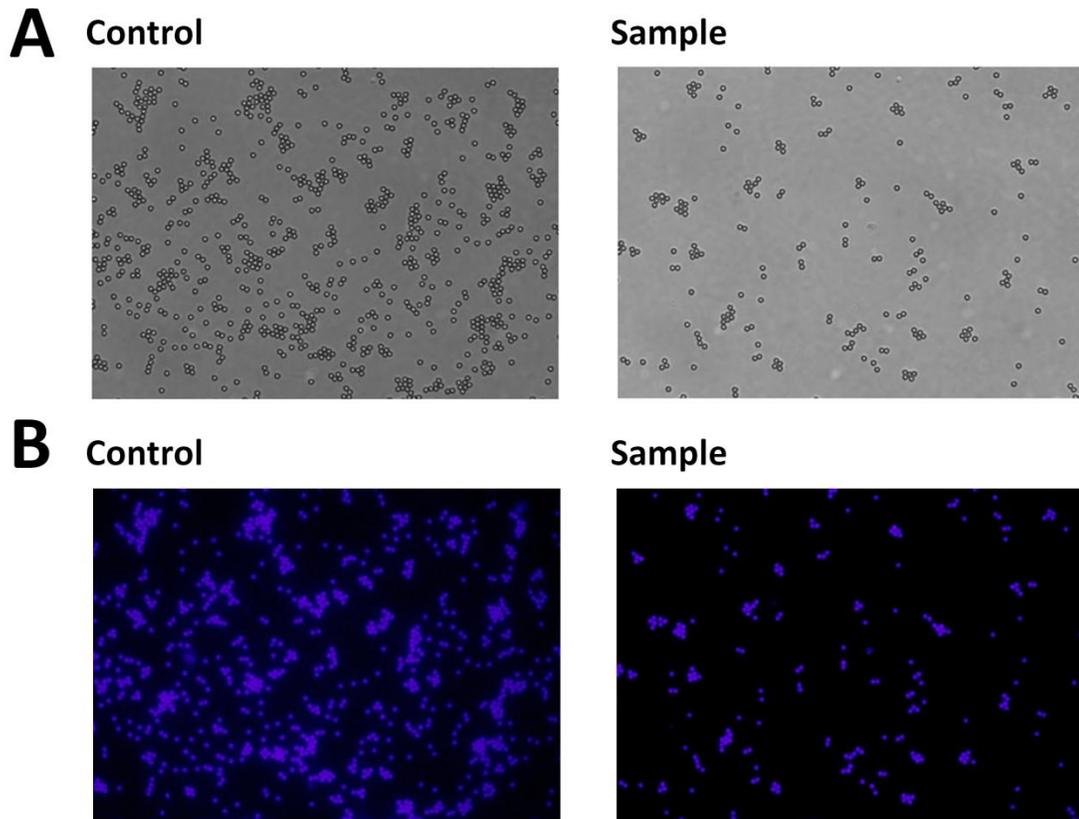


Figure 5-9. Detection of emulsion PCR beads by Thrombin (Cy3)

Panel A: bright field images. Panel B: fluorescent images. Control beads did not contain template in the PCR mixture. Both control beads and sample beads went through the same thermo cycling and were stained by thrombin labeled with Cy3.

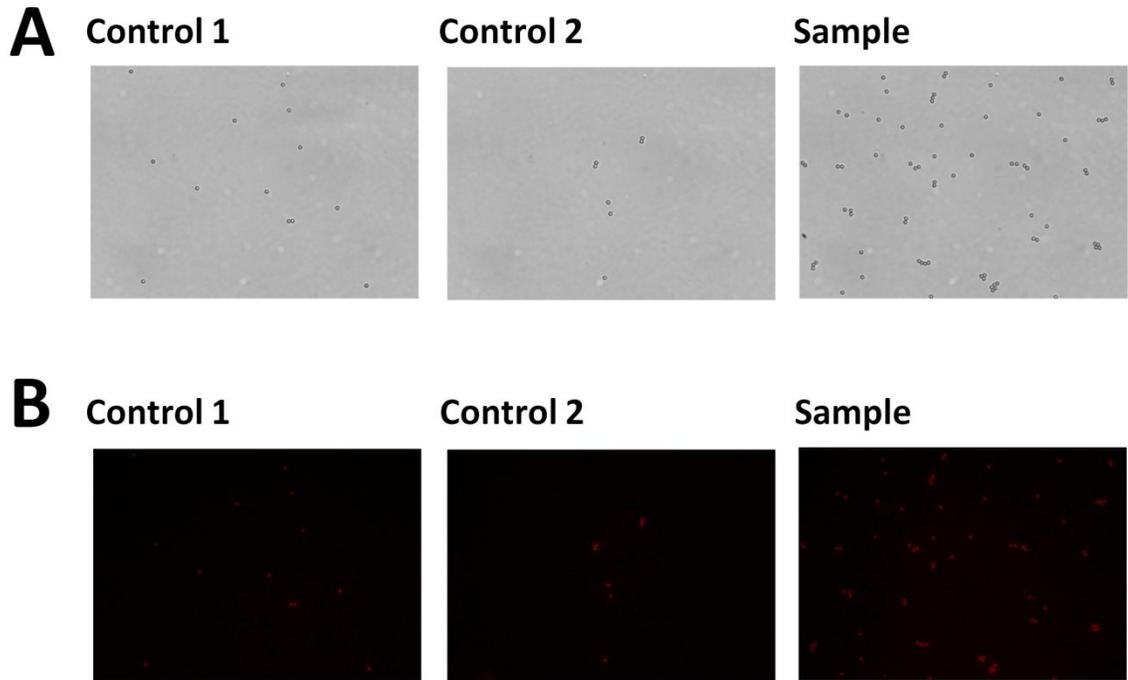


Figure 5-10. Detection of emulsion PCR beads by Thrombin (Cy5)

Panel A: bright field images. Panel B: fluorescent images. Control 1: no template; control 2: no forward primers. Both control beads and sample beads were blocked in the presence of 10mg/ml casein for 1 hour and went through the same thermo cycling. All beads were stained with Cy5-labeled thrombin in the presence of 2mg/ml casein in the staining solution.

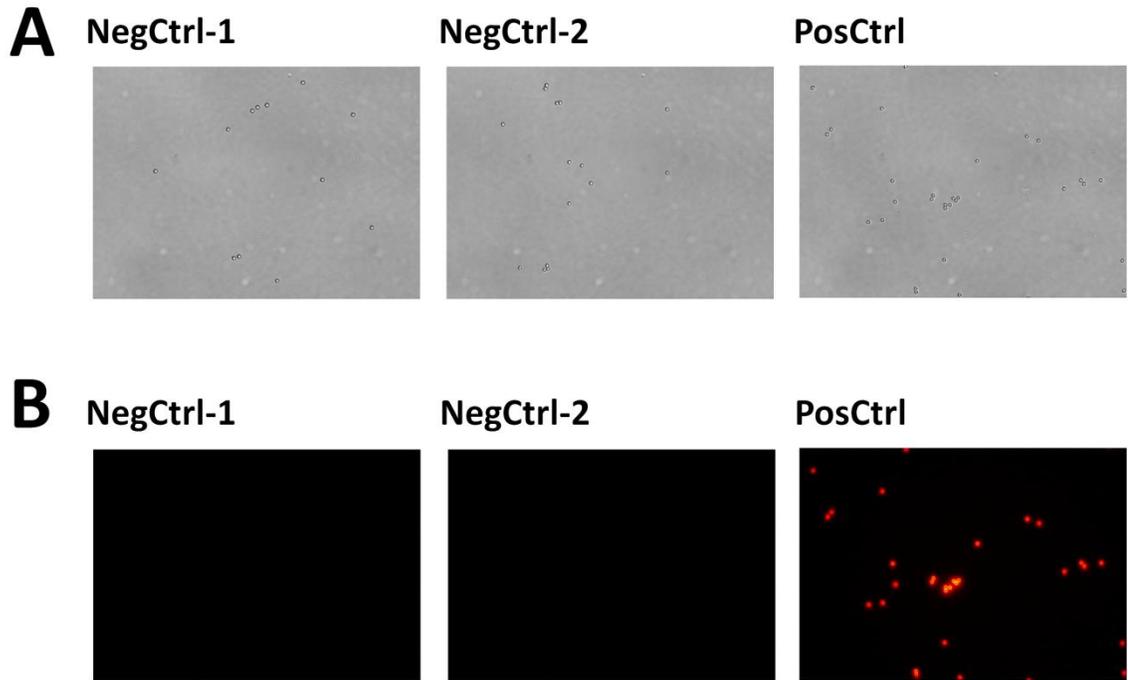


Figure 5-11. Detection of control beads by Thrombin (Cy5)

Panel A: bright field images. Panel B: fluorescent images. Negative control 1: beads loaded with forward primers but not stained; negative control 2: unloaded beads; positive control: beads loaded with T20TBA. Both negative control 2 and positive control were stained with Cy5-labeled thrombin overnight in eppendorf tubes.

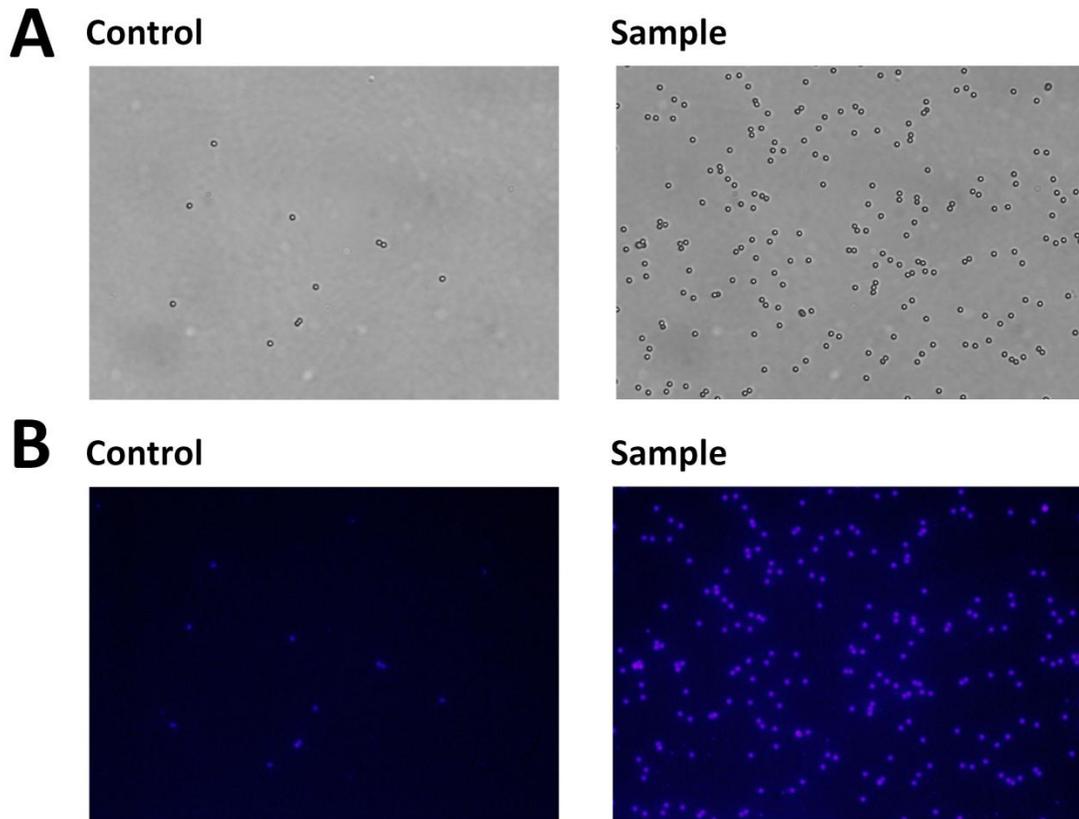


Figure 5-12. Detection of emulsion PCR beads by Thrombin (Cy3)

Panel A: bright field images. Panel B: fluorescent images. Control beads did not contain template in the PCR mixture. Both control beads and sample beads went through the same thermo cycling and were stained by thrombin labeled with Cy3 overnight in eppendorf tubes.

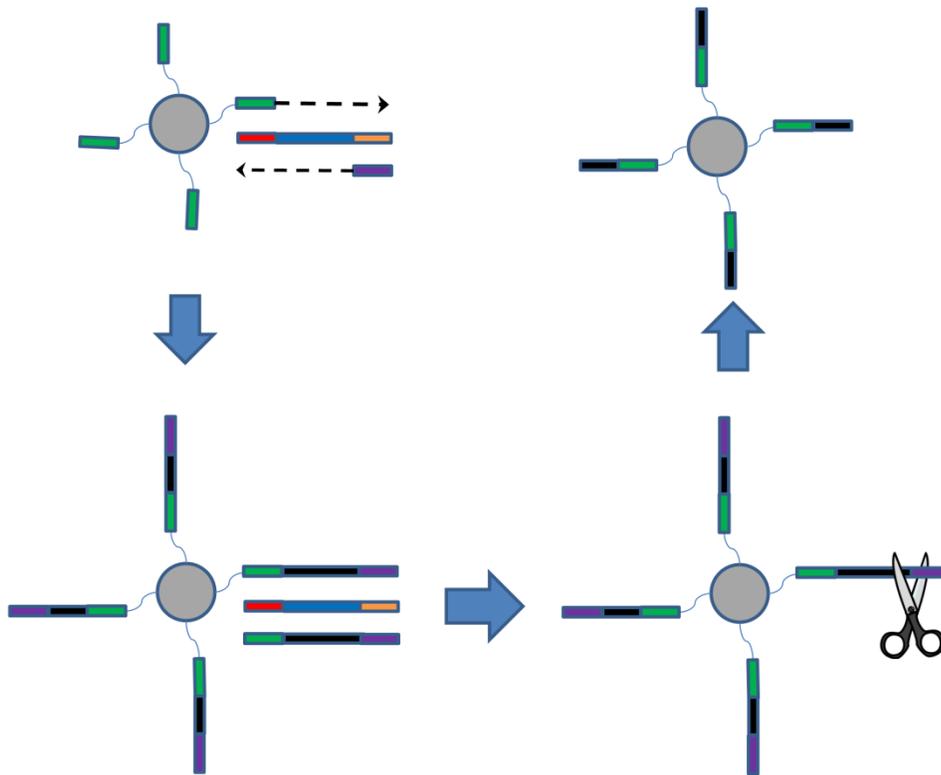


Figure 5-13. New emulsion PCR template design

Gray circle: magnetic emulsion beads; scissor: restriction enzyme (AloI); green bar: forward primer; purple bar: reverse primer; the three-part bar: template; arrows: extension direction.

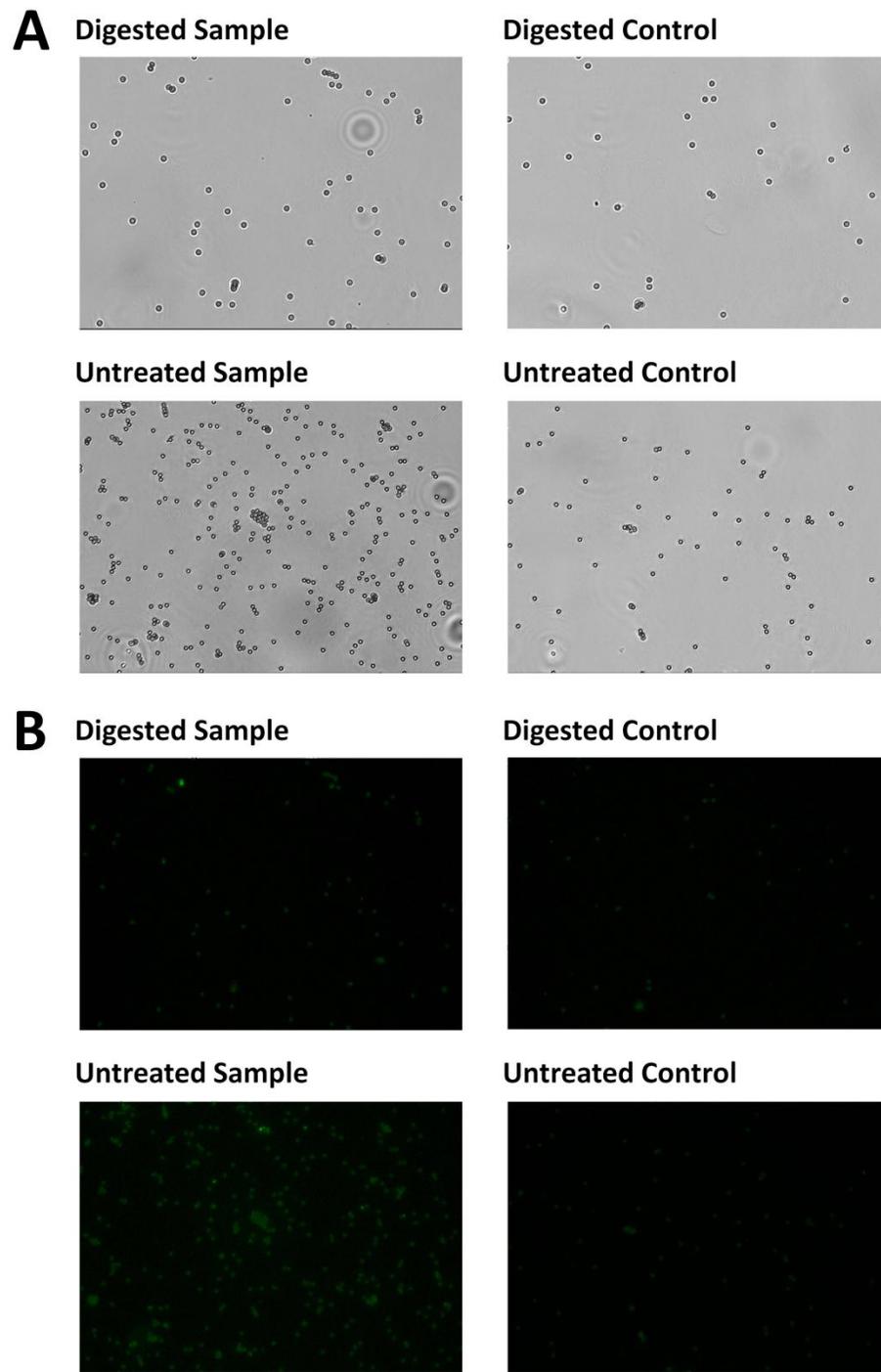


Figure 5-14. Detection of emulsion PCR beads by TBA probe (6-FAM)

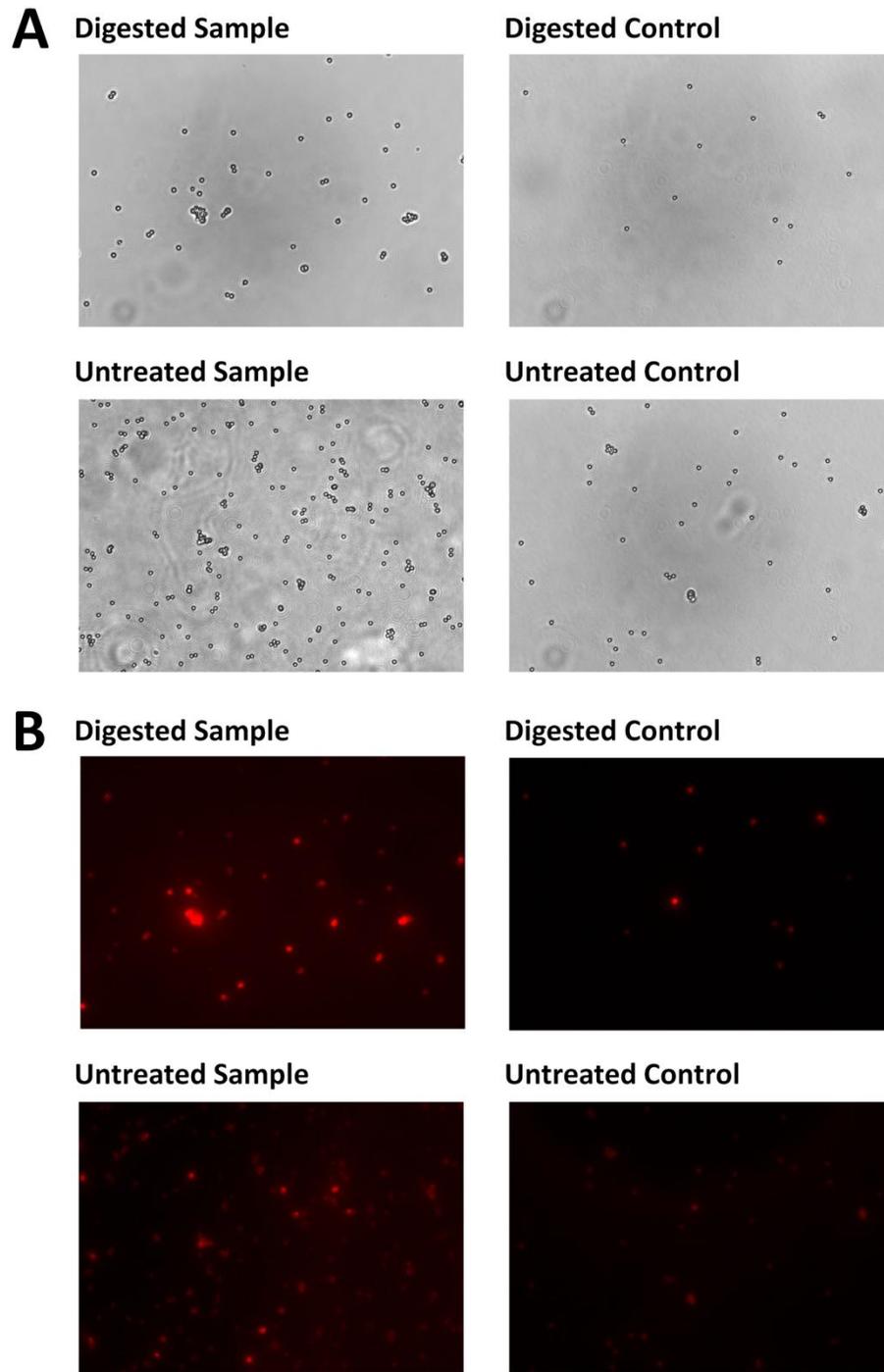


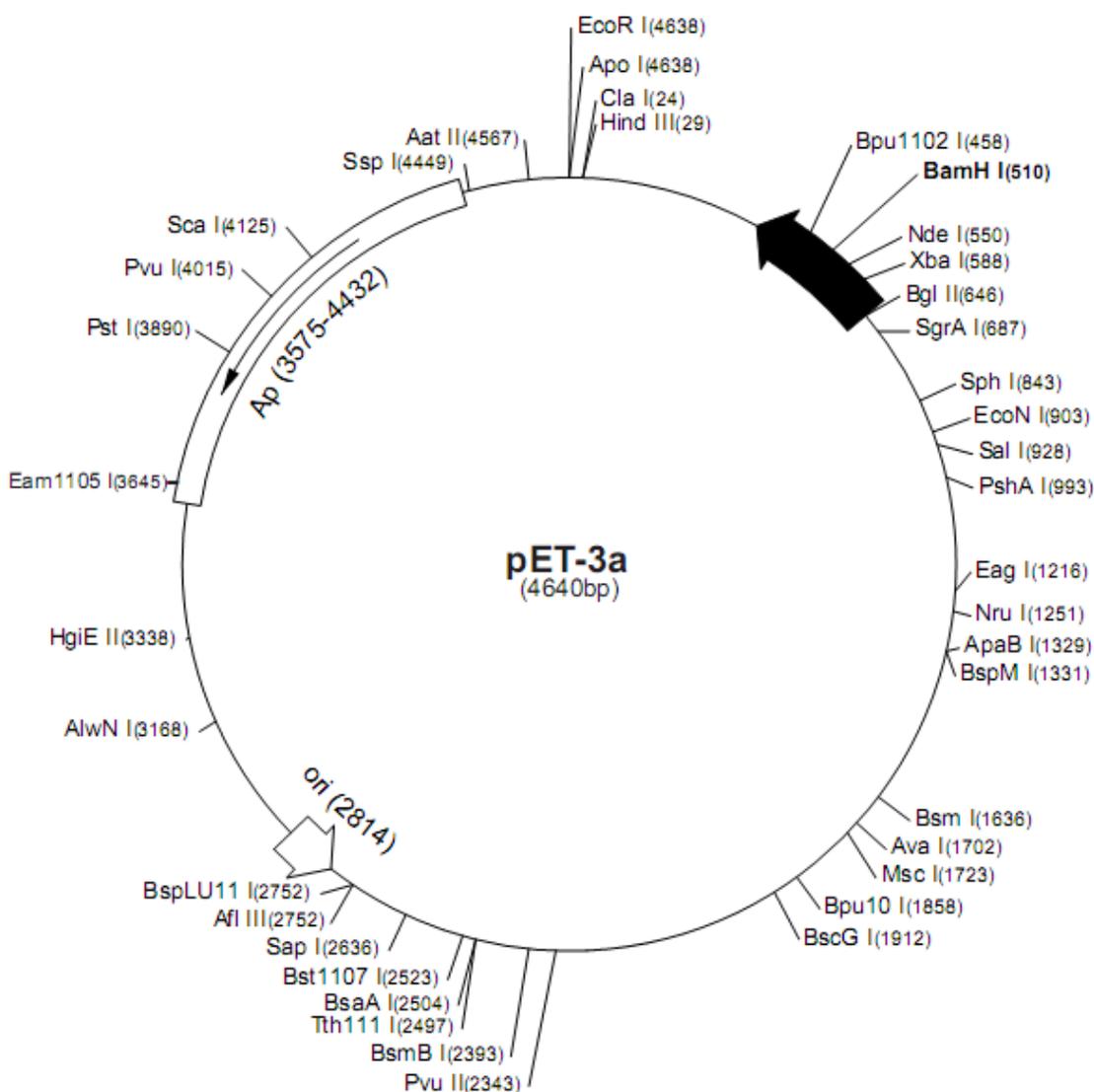
Figure 5-15. Detection of emulsion PCR beads by Thrombin (Cy5)

Appendices

Appendix A. NCp7 Site-Mutagenesis

Sequence information

A-1. Map of pET-3a plasmid



A-2. Full sequence of the pRD2 construct (from HIV-1 pNL4-3 strain)

> pRD2

```
TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAAT
AATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTG
TTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT
GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTAT
TCCCTTTTTTTCGCGGCATTTTGCCTTCCTGTTTTTGTCTACCCAGAAACGCTGGTGAAAGT
AAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAG
CGGTAAGATCCTTGAGAGTTTTTCGCCCCGAAGAACGTTTTTCCAATGATGAGCACTTTTAA
AGTTCGCTATGTGGCGCGGTATTATCCCGTGTGACGCCGGGCAAGAGCAACTCGGTGCG
CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCT
TACGGATGGCATGACAGTAAGAGAATTATGCAGTGTGCCATAACCATGAGTGATAACAC
TGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCA
CAACATGGGGGATCATGTAACCTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCAT
ACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACCT
ATTAACCTGGCGAACTACTTACTCTAGCTTCCCGCAACAATTAATAGACTGGATGGAGGC
GGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTTATTGCTGA
TAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGG
TAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACG
AAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCA
AGTTTACTCATATATACTTTAGATTGATTTAAAACCTCATTTTTAAATTTAAAAGGATCTA
GGTGAAGATCCTTTTTGATAATCTCATGACCAAAAATCCCTTAACGTGAGTTTTTCGTTCCA
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTTGTTTGCCGGA
TCAAGAGCTACCAACTCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATACAAA
TACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCC
TACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTG
TCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAAGGCGCAGCGGTTCGGCTGAAC
GGGGGTTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC
GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTG
GTATCTTTATAGTCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATG
CTCGTCAGGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCT
GGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGA
TAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCG
CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCA
TCTGTGCGGTATTTACACCCGCATATATGGTGCACCTCTCAGTACAATCTGCTCTGATGCC
GCATAGTTAAGCCAGTATACTACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCC
GACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTT
ACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTACCGTCATCAC
CGAAACGCGCGAGGCAGCTGCGGTAAGCTCATCAGCGTGGTTCGTGAAGCGATTACAGA
TGCTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGC
TTCTGATAAAGCGGGCCATGTTAAGGGCGTTTTTTTCCCTGTTTGGTCACTGATGCCTCCG
TGTAAGGGGGATTTCTGTTTCATGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCA
CGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAAC
TGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCG
TTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGA
ACATAATGGTGCAGGGCGCTGACTTCCGCGTTTTCCAGACTTTACGAAACACGGAAACCGA
AGACCATTTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTTCGCTTCACGTTT
```

GCTCGCGTATCGGTGATTTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGG
TCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCAACGCTGCCCGAGA
TGCGCCGCGTGC GGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGG
TTTGCGCATTACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATC
CGTTAGCGAGGTGCCGCGGCTTCCATTACAGGTCGAGGTGGCCCGGCTCCATGCACCGCG
ACGCAACGCGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCGTT
CCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGCGGTCCAGTGATCGAAGT
TAGGCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCGTCATCTACCTG
CCTGGACAGCATGGCCTGCAACGCGGGCATCCCAGATGCCGCGGAAGCGAGAAGAATCAT
AATGGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCAGCGCGTC
GGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGT
GACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCAT
CGTCGCGCTCCAGCGAAAGCGGTCTTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTG
TCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGGCGGACGATAGTCATGCCCCG
CGCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGACGCTCTCC
CTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCGGTTGAGCACCGC
CGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCC
TGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCAGATCTTC
CCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCC
GGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCCGCAAATTAATACGACTC
ACTATAGGGAGACCACAACGGTTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGA

NdeI

GATATACATATGCAGAAAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTC AAT
TGTGGCAAAGAAGGGCACATAGCCAAAATTGCAGGGCCCCTAGGAAAAGGGCTGTTGG
AAATGTGAAAGGAAGGACACCAAATGAAAGATTGTA CTGAGAGACAGGCTAATTAGTTA

BamHI

GGGAGGATCCGGCTGCTAACAAAGCCC GAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTG
AGCAATAACTAGCATAACCCCTTGGGGCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGA
AAGGAGGA ACTATATCCGGATATCCACAGGACGGGTGTGGTGC CCATGATCGCGTAGTCG
ATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTTCGGACAGT
GCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCA
TAGT GACTGGCGATGCTGTTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGC
ATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCA
TTGTTAGATTTTCATACACGGTGCC T GACTGCGTTAGCAATTTAACTGTGATAAACTACCG
CATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA

Appendix B. Recipes and Protocols

Used in NCp7 Mutagenesis

B-1. Recipes for medium, buffers and solutions

LB Medium

Per liter:

Tryptone	10g
Yeast extract	5g
NaCl	10g

Or use the premixed capsules (MP Biomedicals, Solon, OH), 4 capsules per liter

Sterilize by autoclaving

Store media at room temperature (RT)

Antibiotics

Chloramphenicol: 34mg/ml in ethanol, filtered, store at -20°C.

Ampicillin: 100mg/ml in ddH₂O, filtered, store at -20°C.

1X TE

Tris-HCl 10mM
EDTA 1 mM
pH 8.0

10x M9 Salts

Per liter:

NH ₄ Cl	10g
KH ₂ PO ₄	30g
Na ₂ HPO ₄ ·7H ₂ O	32g

Filtered sterile, store at RT

ZnCl₂ (200mM)

Weigh 1.363g of ZnCl₂, dissolve in 50ml ddH₂O, filtered, store at 4°C

MgSO₄ (1M)

Weigh 6.02g of MgSO₄, dissolve in 50ml ddH₂O, filtered, store at 4°C

40% Glucose (w/v)

Weigh 80g of Glucose, dissolve in 200ml ddH₂O, filtered, store at 4°C

10mg/ml Lysozyme

Weigh 10mg Lysozyme, dissolve in 1ml filtered ddH₂O. Prepare fresh solution each time immediately before use.

1mg/ml Pepstatin A

Pepstatin A comes in 5mg/bottle package (Sigma P4265). Dissolve the content in 5ml of methanol. Store at 4°C

10mM PMSF(Phenylmethylsulfonyl Fluoride)

Weigh 87.1mg PMSF, dissolve in 50ml of ethanol, filtered, store at 4°C

1% Deoxycholate

Weigh 0.5g of sodium deoxycholate monohydrate, dissolve in 50ml ddH₂O, filtered, store at 4°C

1M TCEP·HCl (Tris(2-Carboxyethyl) phosphine Hydrochloride)

Weigh 5.733g of TCEP·HCl, dissolve in 20ml ddH₂O, filtered, make 1ml aliquots, store at -20°C

4% PEI (polyethyleneimine)

Take 4ml of 50 % (w/v in water) Poly(ethyleneimine) solution (Sigma P3143), add ddH₂O to a total volume of 50ml, filtered, store at 4°C

2M NaCl

Weigh 116.88g of NaCl, dissolve in 1 L of ddH₂O, filtered, store at 4°C

1M NaOH

Weigh 40g of NaOH, dissolve in 1 L of ddH₂O, filtered, store at 4°C

NCp7 FPLC Buffers

Buffer A (1 Liter)

Tris-Base	6.06g	(final 50mM)
Glycerol	100ml	(final 10%)
NaCl	5.8g	(final 100mM)
200mM ZnCl ₂	500μl	(final 100μM)
1M TCEP·HCl	1ml	(final 1mM)

Adjust to pH8.0 with HCl

Filtered sterile, store at 4°C

Buffer B (1 Liter)

Tris-Base	6.06g	(final 50mM)
Glycerol	100ml	(final 10%)
NaCl	58.4g	(final 1M)
200mM ZnCl ₂	500μl	(final 100μM)
1M TCEP·HCl	1ml	(final 1mM)

Adjust to pH8.0 with HCl

Filtered sterile, store at 4°C

Note: The FPLC buffers were made and stored without TCEP·HCl. TCEP·HCl was only added immediately prior to use.

10X TBE

Per liter:

Tris-Base	108g	(final 890mM)
Boric Acid	55g	(final 890mM)
0.5M EDTA pH8.0	40ml	(final 20mM)

Add ddH₂O to 1 liter, filter and store at RT.

Note: Do not adjust pH. pH of the stock solution should be around 8.3. Filtering is optional and can delay precipitation.

SDS PAGE Gel

15% Separating Gel (10ml)

H ₂ O	3.55ml
40% Gel concentrate	3.75ml
1.5M Tris (pH 8.8)	2.5ml
10% SDS	0.1ml
10% Ammonium Persulfate	0.1ml
TEMED	4µl

5% Stacking Gel (4ml)

H ₂ O	2.92ml
40% Gel concentrate	0.5ml
1.5M Tris (pH 8.8)	0.5ml
10% SDS	0.04ml
10% Ammonium Persulfate	0.04ml
TEMED	4 μ l

Note: This recipe is enough for making one 7mm x 8mm x 1.5mm gel or two 7mm x 8mm x 0.75mm gel (regular gel size for Bio-Rad mini-protean tetra cell system). Scale up or down if necessary.

SDS PAGE Running Buffer

Anode Buffer (10X): 2.0M Tris HCl (pH8.9)

Dissolve 121.1 g of Tris base in 400ml ddH₂O, adjust pH to 8.9 with HCl, and add ddH₂O to final volume of 500ml. Store at RT.

Cathode Buffer (10X): 1.0M Tris HCl, 1.0M Tricine, SDS (pH8.25)

Dissolve 60.55 g of Tris base, 89.58g of Tricine and 5g of SDS in 400ml ddH₂O, adjust pH to 8.25 with HCl, and add ddH₂O to final volume of 500ml. Store at RT.

Note: Cathode Buffer goes to upper chamber; Anode Buffer goes to the bottom chamber.

SDS 2X Loading Buffer

1M Tris-HCl pH 6.8	1 ml
10% SDS	4 ml
Glycerol	2 ml
Bromophenol blue	0.02 g
1M DTT	2 ml

Adjust volume to 10 ml, divide into 1 ml aliquots, and store at -20°C

Gel Staining and Destaining Solution

Coomassie Brilliant Blue staining solution

PhastGel Blue R	1 tablet
Methanol	450ml
ddH ₂ O	450ml
Glacial acetic acid	100ml

Filter and store at RT

Destaining solution

Methanol	450ml
ddH ₂ O	450ml
Glacial acetic acid	100ml

Store at RT

10X NCp7 MS Buffer*

Per liter:

NaH ₂ PO ₄	6g	(final 50mM)
NaCl	116.8g	(final 2M)
200mM ZnCl ₂	50μl	(final 10μM)

Adjust pH to 7.0 with HCl. Filter and store at RT.

*Dilute with ddH₂O to 1X and add TCEP·HCl to 1mM final concentration before use.

10X NCp7 Titration Buffer*

Per liter:

NaH ₂ PO ₄	6g	(final 50mM)
NaCl	116.8g	(final 2M)
200mM ZnCl ₂	50μl	(final 10μM)
PEG 8000	10g	(final 1%)

Adjust pH to 7.0 with HCl. Filter and store at RT.

*Dilute with ddH₂O to 1X and add TCEP·HCl to 1mM final concentration before use.

B-2. Protocol for plasmid preparation

Materials:

- XL1-Blue strain or other endA⁻ strain containing the needed plasmid
- LB-Amp medium
- QIAprep Spin Miniprep kit

Procedure:

1. Streak the frozen culture of XL1-Blue cells onto a LB agar plate containing 100µg/ml Ampicillin. Incubate the streaked plate at 37°C overnight.
2. Pick a single colony from the overnight grown plate to inoculate 5ml of LB-Amp medium. Incubate the liquid culture at 37°C overnight, shaking at 200~250rpm.
3. Harvest the overnight culture by spinning the cells for 5 minutes at 3000rpm in a tabletop centrifuge. Remove all the supernatant.
4. Re-suspend pelleted bacterial cells in 250 µl of Buffer P1 and transfer the suspension to a microcentrifuge tube.
5. Add 250 µl of Buffer P2 and mix thoroughly by inverting the tube 4–6 times.
6. Add 350 µl of Buffer N3 and mix immediately and thoroughly by inverting the tube 4–6 times.

7. Centrifuge for 10 minutes at 13,000 rpm (~17,900 x g) in a table-top microcentrifuge.
8. Apply the supernatant (from step 7) to the QIAprep spin column by decanting or pipetting.
9. Centrifuge for 1 minute. Discard the flow-through.
10. Wash QIAprep spin column by adding 0.75 ml of Buffer PE and centrifuging for 1 minute.
11. Discard the flow-through, and centrifuge for an additional 1 minute to remove residual wash buffer.
12. To elute DNA, place the QIAprep column in a clean 1.5 ml microcentrifuge tube.
Add 50 μ l of nuclease-free water to the center of each QIAprep spin column, let stand for 1 minute, and centrifuge for 1 minute.
13. Check the purity and integrity of the miniprep by agarose gel electrophoresis.
Measure the concentration of the prepared plasmid DNA by UV absorption at 260nm.

B-3. Protocol for site-directed mutagenesis

Materials:

- Stratagene® QuikChange® II Site-Directed Mutagenesis Kit (Stored at -20°C)
- Plasmid DNA template: dilute to 10ng/μl
- Primers: dilute to 10μM (100~150ng/μl)

Procedure:

1. Prepare PCR mix as below:

10× reaction buffer	2.5 μl
dsDNA template (10ng/μl)*	0.5 μl
Forward primer (10 μM)	0.5 μl
Reverse primer (10 μM)	0.5 μl
dNTP mix (10mM)	0.5 μl
PfuUltra polymerase (2.5 U/μl)	0.5 μl
Nuclease free H ₂ O	20 μl
<hr/>	
Final volume:	25 μl

* Nuclease free H₂O was added instead of DNA template as a negative control.

2. Set up and run mutagenesis PCR reactions as below:

- 1) 95°C 30 seconds
- 2) 95°C 30 seconds
- 3) 55°C 1 minute
- 4) 68°C 6 minutes
- 5) go to step 2) for a total of 16 cycles
- 6) hold at 4°C

3. Add 0.5µl of *DpnI* restriction enzyme into each PCR reaction. Mix well and incubate the mixture at 37°C for 1 hour.

4. Check product size and integrity on agarose gel. Load 5 µl of each digested PCR reaction per well on a 1% agarose gel. Use 1X TBE as running buffer. Run the gel at 160V for 40 minutes. Stain with ethidium bromide and take gel image on a Kodak Gel Imager equipped with UV transilluminator.

5. Transform mutated DNA product into XL1-Blue competent cells (See appendix B-4 for detailed protocol).

B-4. Protocol for transformation into XL1-Blue competent cells

Materials:

- XL-1 Blue competent cells(Stratagene), stored at -80°C
- Digested PCR mix
- LB+0.4% glucose
- LB/AMP plates

Procedure:

1. Estimate concentration of PCR product based on the agarose gel. (UV absorption at this time is not reliable because of interference by dNTPs and free primers.)
2. Use ~5ng of DNA (DNA volume should not exceed 10% of reaction volume, in this case, 2µl) for each transformation. Add the DNA into a 200µl PCR tube, let it chill on ice. Use 1 µl of pUC18 control DNA (0.1ng/µl, included in the kit) as positive control. Use 1 µl of nuclease free H₂O as negative control.
3. Thaw a tube of XL-1 Blue competent cells on ice. Add 20 µl of competent cells into each tube.
4. Incubate the mixture on ice for 30 minutes. In the mean time, warm up a tube of LB+0.4% glucose at 42°C.

5. Heat-shock the cell mixture at 42°C for 45 seconds on a pre-warmed thermal cycler.
6. Immediately put tubes back on ice, incubate for 2 minutes.
7. Add 100 µl of pre-warmed LB+0.4% glucose (42°C), gently mix.
8. Pool the tubes in a 15ml falcon tube, shake at 250rpm in a 37°C incubator for 1 hour. Warm up a few LB/AMP plates during this time.
9. Spread 50 µl of the recovered cells onto each LB/AMP plate. Grow plates at 37°C overnight.
10. Pick single colony from plate, inoculate a 5-ml LB/AMP liquid culture, grow at 37°C overnight, shaking at 220~250rpm.
11. Make permanent glycerol storage culture (see appendix B-6). Store the frozen stocks at -80°C.
12. Do plasmid miniprep using QIAprep Spin Miniprep kit. Elute plasmid with 50 µl of nuclease free H₂O.
13. Measure plasmid concentration by UV absorption. Use 200-800ng of each plasmid as Sanger sequencing template. Store the rest of plasmid at -80°C.

B-5. Protocol for transformation into BL21 competent cells

Materials:

- BL21-(DE3)pLysS competent cells (Stratagene), stored at -80°C
- Mutant plasmid DNA, dilute to 10ng/μl
- LB+0.4% glucose
- LB/AMP/Chl plates

Procedure:

1. Add 1 μl of plasmid (10 ng) to a 200 μl PCR tube, let it chill on ice. Use 1 μl of pUC18 control DNA (0.1 ng/μl) as positive control. Use 1 μl of nuclease free H₂O as negative control.
2. Thaw a tube of BL21 competent cells on ice. Add 20ul of BL21 competent cells to each tube.
3. Incubate the mixture on ice for 30 minutes. In the mean time, warm up a tube of LB+0.4% glucose at 42°C.
4. Heat-shock the cell mixture at 42°C for 20 seconds on a pre-warmed thermal cycler.
5. Immediately put tubes back on ice, incubate for 2 minutes.
6. Add 100 μl of pre-warmed LB+0.4% glucose (42°C), gently mix.

7. Pool the tubes in a 15ml falcon tube, shake at 250rpm in a 37°C incubator for 1 hour. Warm up a few LB/AMP/Chl plates during this time.
8. Spread 50 µl of the recovered cells onto each LB/AMP/Chl plate. Grow plates at 37°C overnight.
9. Pick single colony from plate, inoculate a 5-ml LB/AMP/Chl liquid culture supplemented with Zinc, M9 salts and glucose (see recipe for starter culture in appendix B-5), grow at 37°C overnight, shaking at 220~250rpm.
10. Make permanent glycerol storage culture (see appendix B-6). Store the frozen culture at -80°C.

B-6. Protocol for making or renewing frozen stock of *E. coli* strains

Materials:

- Fresh LB plates with healthy colonies of *E. coli* cells (plated with freshly transformed cells or streaked from previous frozen stock)
- LB medium
- Antibiotics of choice/Glucose/M9 salts
- 50% glycerol (sterilized by autoclave)
- Dry ice

Procedure:

1. Prepare 5 ml of fresh LB media, add appropriate antibiotic(s) and nutrients at desired concentration.
2. Pick a single colony from the LB plate, inoculate the above 5-ml LB media with this colony, mix well and grow cells overnight.
3. Aliquot 150 μ l of 50% glycerol to microcentrifuge tubes or cryo-vials.
4. Add 350 μ l of fresh overnight culture to each tube containing glycerol, mix well.
5. Immediately put tubes on dry ice. Transfer tubes into -80°C freezer after the mixture is completely frozen.

B-7. Protocol for NCp7 protein over-expression

Materials:

- BL21-(DE3)pLysS cells containing the needed plasmid
- LB-Amp-Chl medium
- Supplements: 10X M9 salts, 200mM ZnCl₂, 1M MgSO₄, 40% Glucose
- 1M IPTG

Procedure:

1. Streak the frozen culture of BL21-(DE3)pLysS cells onto a LB-Amp-Chl plate.
Incubate the streaked plate at 37°C overnight.
2. Pick a single colony from the overnight grown plate to inoculate 5ml of starter culture as described below:

Starter culture media

LB media	4.45ml	
10X M9 salts*	500µl	(final conc. 1X)
34mg/ml Chloramphenicol	5µl	(final conc. 34µg/ml)
100mg/ml Ampicillin	5µl	(final conc. 100µg/ml)
200mM ZnCl ₂	2.5µl	(final conc. 100µM)
1M MgSO ₄	5µl	(final conc. 1mM)

40% Glucose	50 μ l	(final conc. 0.4%)
<hr/>		
Final Volume:	5ml	

*See appendix B-1 for recipe.

- Grow the inoculated starter culture overnight at 37°C, shaking at 200-250rpm.
- Transfer the overnight starter culture to 1 liter of fresh media as described below:

NCp7 expression media (1 Liter)

LB media	890ml	
10X M9 salts*	100ml	(final conc. 1X)
34mg/ml Chloramphenicol	1ml	(final conc. 34 μ g/ml)
100mg/ml Ampicillin	1ml	(final conc. 100 μ g/ml)
200mM ZnCl ₂	0.5ml	(final conc. 100 μ M)
1M MgSO ₄	1ml	(final conc. 1mM)
40% Glucose	10ml	(final conc. 0.4%)
<hr/>		
Final Volume:	1L	

*See appendix B-1 for recipe.

- Incubate the 1L culture at 37°C, shaking at 200-250rpm. Monitor cell density by measuring the culture's absorbance at 600nm every half hour until OD600 reaches 0.6-0.8, which usually takes 4-5 hours after inoculation.
- Induce the culture with 400 μ l of 1M IPTG (final conc. 400 μ M) when the culture's OD600 is between 0.6-0.8. Incubate the culture at 37°C, shaking at 200-250rpm, for 4 more hours after induction.

7. Harvest cells by centrifuging the culture on an Avanti® J-E centrifuge (Beckman Coulter) using the JLA-10.5 rotor, spinning at 5000rpm for 15 minutes. Dump the clear supernatant and keep the cell pellet.
8. Use the cell pellet for cell lysis or freeze the cell pellet at -20°C.

B-8. Protocol for cell lysis

Materials:

- Cell pellets from over-expression (see appendix B-6)
- Lysis buffer (see procedure below for recipe)
- B-PER II Protein Extraction Reagent (Thermo Pierce)
- 4% PEI (polyethyleneimine)

Procedure:

1. Re-suspend fresh or frozen cell pellet in ~30ml of ice-cold lysis buffer as described here:

NCp7 Lysis Buffer (for pellets from 1L culture)

NCp7 FPLC Buffer A w/o TCEP*	30ml	
10mg/ml freshly made Lysozyme	200µl	(final 60µg/ml)
1mg/ml Pepstatin A	30µl	(final 0.9µg/ml)
10mM PMSF	172µl	(final 53µM)
1% Deoxycholate	2.1ml	(final 5mM)
1M TCEP	150µl	(final 5mM)

*See appendix B-1 for recipe.

2. Transfer the re-suspended cells into a pre-chilled 50ml conical tube and left on ice for 20 minutes.
3. Add 4ml of B-PER Protein Extraction Reagents (Thermo Pierce), mix well and leave the lysate on ice for additional 10 minutes.
4. Further disrupt the cells using a Microson XL2000 Ultrasonic Liquid Processor (Misonix, now Qsonica) for 4 times, 25 seconds each time, at an output power setting of 20-25 watts, with 1-minute intervals in between pulses.
5. Add 2ml of 4% PEI (Sigma-Aldrich), mix well and transfer the lysate into a pre-chilled 50ml Oakridge tube.
6. Spin down the lysed cells on an Avanti® J-E centrifuge (Beckman Coulter) using the JA-25.5 rotor, spinning at 17,000rpm for 30 minutes at 4°C.
7. Transfer supernatant into a clean conical tube, filter with 0.45µm syringe filter when necessary. Load the collected lysate onto FPLC immediately for purification.

B-9. Protocol for FPLC purification of NCp7

Materials:

- Fresh cell lysate (see appendix B-7)
- NCp7 FPLC buffer A and B (see appendix B-1)
- SP FF and Q FF FPLC columns

Procedure:

1. Wash and equilibrate a 20ml HiPrep Q FF column (GE Healthcare) and a 20ml HiPrep SP FF column (GE Healthcare) connected in series with at least 200ml of NCp7 FPLC Buffer A, at a flow rate of 3ml/min.
2. Set flow rate to 2ml/min and load cell lysate onto the Q FF column (a strong anion exchange column).
3. After loading, set the flow rate back to 3ml/min. When the flow through peak is washed off and the UV absorbance decreases to ~15mAU, quickly take the Q FF column off the system and keep washing the SP FF column until the UV absorbance reaches a stable base line.
4. Start a 5-min linear gradient from 0% to 20% NCp7 FPLC Buffer B. Wash SP FF column with 20% NCp7 FPLC Buffer B until the UV absorbance reaches a stable base line again.

5. Start a one-hour linear gradient from 20% to 50% NCp7 FPLC Buffer B. Collect peak fractions for further analysis.
6. Start a 5-min linear gradient from 50% to 100% NCp7 FPLC Buffer B. Wash the column with 100% NCp7 FPLC Buffer B until the UV absorbance reaches a stable baseline.
7. Start a 5-min linear gradient from 100% to 0% NCp7 FPLC Buffer B. Wash the column with 100% NCp7 FPLC Buffer A until the UV absorbance reaches a stable baseline.
8. Clean the columns separately with 2M NaCl solution and ddH₂O before the next FPLC run. Clean and store the columns appropriately following the manufacturer's protocol (also see appendix B-10).

B-10. Protocols for cleaning and storing FPLC columns

Materials:

- FPLC columns
- ddH₂O
- 2M NaCl
- 1M NaOH
- 70% Ethanol
- 20% Ethanol or 20% Ethanol with 0.2 M sodium acetate

Procedure:

Regular cleaning

1. Wash the column with 2 column volumes of 2M NaCl at a flow rate of 5 ml/min after each run to elute material still bound to the column, followed by 2 column volumes of ddH₂O.
2. Re-equilibrate the column before use with at least 5 column volumes of start buffer or until the UV base-line and pH/conductivity values are stable.

More rigorous cleaning

1. Reverse the flow direction and run at a flow rate of 5 ml/min at room temperature the following sequence of solutions.
 - 1) 4 column volumes of a 2 M NaCl solution (removes ionically bound proteins from the column) followed by 3 column volumes of ddH₂O.
 - 2) 4 column volumes of a 1 M NaOH solution (removes precipitated proteins, hydrophobically bound proteins and lipoproteins from the column) followed by 3 column volumes of ddH₂O.
 - 3) 4 column volumes of 70% ethanol (removes proteins, lipoproteins and lipids that are strongly hydrophobically bound to the column) followed by 3 column volumes of ddH₂O.
2. Re-equilibrate the column before use with at least 5 column volumes of start buffer in the normal flow direction.

Storage

If the column is to be stored for more than two days after use, clean the column according to the procedure described above. Then equilibrate with at least 5 column volumes of 20% ethanol (for Q FF column) or 20% ethanol with 0.2 M sodium acetate (for SP FF column) at a flow rate of 5 ml/min. Cap the outlet tubings and store the columns at 4°C.

B-11. Protocol for FPLC column packing

Materials:

- SP FF resin or Q FF resin, 25ml/bottle
- 20% Ethanol
- XK 16 column
- ddH₂O
- FPLC buffer

Procedure:

1. Disassemble the column following the column's manual, use the right tools.
Remember to loosen the adjusting knob before pulling out the plungers.
2. Clean the parts thoroughly. Autoclave the chromatographic glass tube without the thermostat jacket and tubings.
3. Re-assemble the column. Put an end piece to the bottom.
4. Add 20% ethanol into column, use a syringe to pull ethanol through the tubing to get rid of air bubbles. Let the ethanol dripping into a clean beaker.
5. Shake the resin well, pour down the column, let it settle, add more till appropriate height.
6. Draw ethanol through the top end piece to get rid of air bubbles.

7. Insert top end piece at an angle to avoid trapping air bubbles. Assemble the column.
8. Let the resin settle overnight.
9. Adjust adaptor until it touches the surface of resin. Tighten the O-ring to lock the adaptor in position.
10. Connect the column to FPLC, wash column with 3 column volumes of ddH₂O then 3 column volumes of buffer A at regular flow rate. Re-adjust adaptor position if necessary.

Appendix C. NC-SL3 Titration Data

C-1. Compilation of NCp7-SL3 Trp titration data

#	Mutant	Titration Date	K _d (nM)	R ²	Endpoint	ITC Date	K _d (nM)	Fitting Error	N	Memo	ΔH ⁰		
1	WT	10/29/2010	23.72	0.999	4.54%	10/18/2010	63.69	11%	1.09		-1.162E+04		
		11/7/2010	21.81	0.999	3.21%	10/20/2010	65.79	15%	1.14		-1.081E+04		
		11/17/2010	32.98	0.999	5.47%	10/28/2010	65.36	21%	1.17		-1.030E+04		
		11/18/2010	34.06	0.999	5.56%	11/17/2010	80.97	13%	1.15		-1.041E+04		
		11/19/2010	31.13	0.998	5.88%								
		11/22/2010	32.39	0.999	5.09%								
		11/23/2010	27.32	0.999	3.96%								
		11/29/2010	29.34	0.999	5.82%								
				Avg. K _d =	29	M=	8	Avg. K _d =	69	Avg. N =	1.1	Avg. ΔH ⁰ =	-1.079E+04
				SD=	4.5	F=	1.0	SD=	8.1	SD=	0.034	SD=	598.2474404
				SD%=	15%			SD%=	12%	SD%=	3%	SD%=	6%
								K _d ^{RT} =	42			ln(K1/K2)=	0.4857564
								SD=	5.0			M=	4
								SD%=	12%			F=	1.0
		Analyzed w/ separate blank titration											

						10/18/2010	79.37	14%	0.928		-1.325E+04		
						10/20/2010	71.94	14%	0.886		-1.279E+04		
						10/28/2010	33.67	48%	0.835		-1.094E+04		
						11/17/2010	90.09	11%	0.879		-1.246E+04		
						Avg. K _d =	69	Avg. N =	0.88	Avg. ΔH ⁰ =	-1.236E+04		
						SD=	25	SD=	0.038	SD=	1000.566506		
						SD%=	36%	SD%=	4%	SD%=	8%		
						K _d ^{RT} =	39			ln(K1/K2)=	0.5566944		
						SD=	14			M=	4		
						SD%=	36%			F=	1.0		
2	N5A		10/20/2010	33.01	0.999	4.75%	10/28/2010	72.99	16%	0.993		-1.077E+04	
			10/29/2010	23.66	0.997	5.86%	11/11/2010	169.78	31%	1.18		-7.615E+03	
			11/9/2010	22.75	0.993	6.81%	11/15/2010	138.12	14%	1.04		-1.042E+04	
			11/15/2010	33.45	0.995	6.84%	11/17/2010	116.55	13%	1.03		-1.034E+04	
			11/18/2010	33.86	0.994	7.57%							
				Avg. K _d =	29	M=	5	Avg. K _d =	124	Avg. N =	1.1	Avg. ΔH ⁰ =	-9.786E+03
				SD=	5.6	F=	1.0	SD=	41	SD=	0.082	SD=	1459.494062
				SD%=	19%			SD%=	33%	SD%=	8%	SD%=	15%
								K _d ^{RT} =	80			ln(K1/K2)=	0.4407727
								SD=	26			M=	4
								SD%=	33%			F=	1.9

3	F6A	7/26/2010	29.63	0.999	5.81%	10/22/2010	70.92	20%	0.884		-1.097E+04	
		10/12/2010	32.73	0.987	10.60%	10/28/2010	78.74	18%	0.824		-1.125E+04	
		11/15/2010	36.17	0.988	10.26%	11/15/2010	142.25	21%	0.902		-1.228E+04	
		Avg. K _d =	33	M=	3	Avg. K _d =	97	Avg. N =	0.87	Avg. ΔH ⁰ =	-1.150E+04	
		SD=	3.3	F=	1.1	SD=	39	SD=	0.041	SD=	689.8550572	
		SD%=	10%			SD%=	40%	SD%=	5%	SD%=	6%	
						K _d ^{RT} =	58			ln(K1/K2)=	0.51796	
						SD=	23			M=	3	
				SD%=	40%			F=	1.4			
4	V13A	11/7/2010	38.79	0.992	9.60%							
		6/8/2011	25.32	0.969	8.95%							
		6/9/2011	29.13	0.976	8.65%							
		Avg. K _d =	31	M=	3							
		SD=	6.9	F=	1.1							
		SD%=	22%									
5	F16A	10/12/2010	686.35	0.986	50.00%	9/29/2010	1701	7%	1.26	12 uM	-1.051E+04	
		11/19/2010	732.30	0.994	50.51%	10/4/2010	2639	7%	1.33	19 uM	-1.036E+04	
		11/23/2010	730.75	0.987	50.22%	11/19/2010	1031	27%	1.51	5uM	-6.121E+03	
		Avg. K _d =	720	M=	3	Avg. K _d =	1790	Avg. N =	1.4	Avg. ΔH ⁰ =	-8.997E+03	
		SD=	26	F=	25	SD=	808	SD=	0.13	SD=	2491.818011	
		SD%=	4%			SD%=	45%	SD%=	9%	SD%=	28%	

7	G19A	11/3/2010	41.78	0.999	7.32%							
		11/23/2010	44.87	0.999	7.67%							
		6/8/2011	46.96	0.997	9.55%							
		Avg. K _d =	45	M=	3							
		SD=	2.6	F=	1.5							
		SD%=	6%									
8	E21A	11/23/2010	10.25	0.984	5.90%							
		6/8/2011	6.04	0.968	5.16%							
		6/9/2011	9.55	0.980	5.92%							
		Avg. K _d =	8.6	M=	3							
		SD=	2.3	F=	0.3							
		SD%=	26%									
9	G22A	11/23/2010	48.84	0.988	11.37%							
		6/8/2011	37.68	0.969	9.89%							
		6/9/2011	49.72	0.952	14.42%							
		Avg. K _d =	45	M=	3							
		SD=	6.7	F=	1.6							
		SD%=	15%									
10	I24A	10/30/2010	141.42	0.997	16.89%	10/30/2010	359.7	20%	1.19		-1.000E+04	
		11/4/2010	140.32	0.999	16.71%	11/4/2010	366.3	22%	1.13		-8.347E+03	
		11/9/2010	168.53	0.996	21.32%	11/11/2010	787.4	23%	1.49		-1.037E+04	
		11/12/2010	161.24	0.998	21.50%	11/15/2010	416.7	23%	1.32		-9.117E+03	

		11/15/2010	211.09	0.999	23.29%									
		Avg. K _d =	170	M=	5	Avg. K _d =	483	Avg. N =	1.3	Avg. ΔH ⁰ =	-9.459E+03			
		SD=	29	F=	5.8	SD=	205	SD=	0.16	SD=	908.4985782			
		SD%=	17%			SD%=	42%	SD%=	12%	SD%=	10%			
						K _d ^{RT} =	310			ln(K1/K2)=	0.42601084			
						SD=	130			M=	4			
						SD%=	42%			F=	7.3			
11	P31A	10/18/2010	17.02	0.999	3.31%	10/18/2010	69.93	9%	1.06		-1.304E+04			
		10/30/2010	28.69	1.000	4.82%	10/18/2010	87.72	9%	1.13		-1.240E+04			
		6/9/2011	22.97	0.998	6.02%	10/30/2010	86.21	10%	1.14		-1.279E+04			
		Avg. K _d =	23	M=	3	Avg. K _d =	81	Avg. N =	1.1	Avg. ΔH ⁰ =	-1.274E+04			
		SD=	5.8	F=	0.8	SD=	10	SD=	0.044	SD=	322.5419869			
		SD%=	25%			SD%=	12%	SD%=	4%	SD%=	3%			
								K _d ^{RT} =	46			ln(K1/K2)=	0.573959733	
								SD=	5.6			M=	3	
								SD%=	12%			F=	1.1	
				Analyzed w/ separate blank titration										
		10/18/2010	74.63	8%	0.95						-1.399E+04			
		10/18/2010	93.46	9%	0.89						-1.446E+04			
		10/30/2010	84.03	10%	0.88						-1.486E+04			

						Avg. K_d =	84	Avg. N =	0.9	Avg. ΔH^0 =	-1.444E+04	
						SD=	9	SD=	0.038	SD=	435.4690957	
						SD%=	11%	SD%=	4%	SD%=	3%	
						K_d^{RT} =	44			$\ln(K1/K2)$ =	0.650227467	
						SD=	4.9			M=	3	
						SD%=	11%			F=	1.1	
12	G40A	11/3/2010	30.03	0.996	5.57%							
		11/23/2010	39.97	0.998	6.70%							
		6/9/2011	27.74	0.995	7.06%							
		Avg. K_d =	33	M=	3							
		SD=	6.5	F=	1.1							
		SD%=	20%									
13	Q45A	10/13/2010	69.80	1.000	10.17%	11/1/2010	163.93	29%	1.18		-8.250E+03	
		11/1/2010	58.18	0.995	10.25%	11/4/2010	169.78	21%	1.16		-8.463E+03	
		11/4/2010	71.02	0.999	10.59%	11/22/2010	198.02	11%	1.32		-7.753E+03	
		11/22/2010	92.65	0.999	12.54%							
		Avg. K_d =	73	M=	4	Avg. K_d =	177	Avg. N =	1.2	Avg. ΔH^0 =	-8.155E+03	
		SD=	14	F=	2.5	SD=	18	SD=	0.087	SD=	364.3437022	
		SD%=	20%			SD%=	10%	SD%=	7%	SD%=	4%	
						K_d^{RT} =	120			$\ln(K1/K2)$ =	0.367316213	

16	K33E-E42K	11/1/2010	30.17	0.999	6.40%	11/1/2010	68.49	10%	1.04		-1.356E+04	
		11/18/2010	51.36	0.999	8.49%	11/17/2010	75.19	10%	1.12		-1.158E+04	
		11/22/2010	38.84	0.998	8.47%	11/22/2010	82.64	12%	1.13		-1.192E+04	
		Avg. K_d =	40	M=	3	Avg. K_d =	75	Avg. N. =	1.1	Avg. ΔH^0 =	-1.235E+04	
		SD=	11	F=	1.4	SD=	7.1	SD=	0.049	SD=	1058.741391	
		SD%=	27%			SD%=	9%	SD%=	4%	SD%=	9%	
						K_d^{RT} =	43			$\ln(K1/K2)$ =	0.556394133	
						SD=	4.1			M=	3	
				SD%=	9%			F=	1.0			
17	K38E-E51K	11/7/2010	27.80	0.950	10.80%							
		11/23/2010	37.96	0.963	11.34%							
		6/9/2011	33.72	0.938	12.17%							
		Avg. K_d =	33	M=	3							
		SD=	5.1	F=	1.1							
		SD%=	15%									

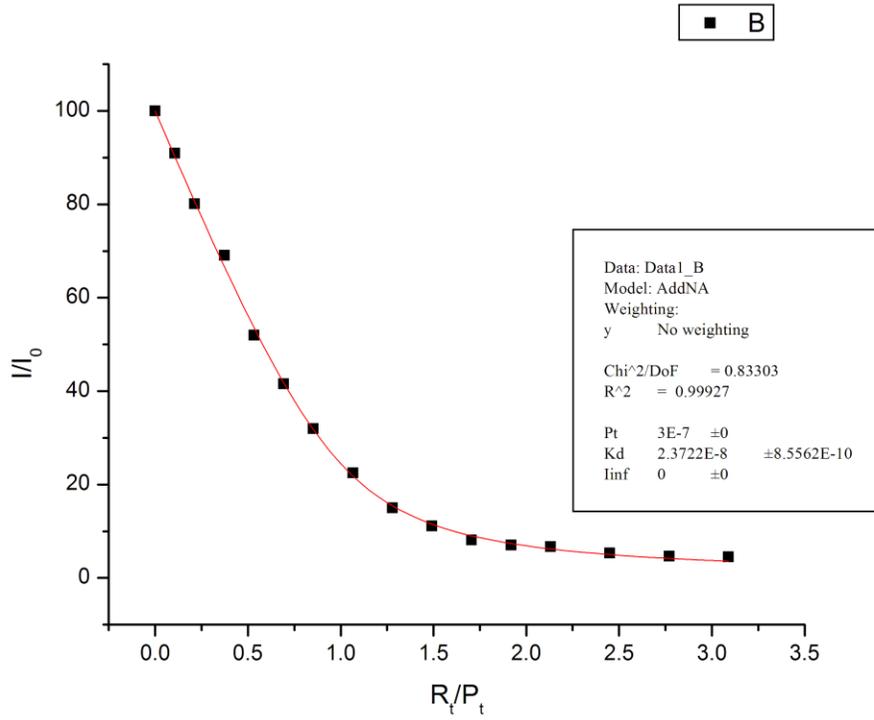
*M = number of measurements

F = Change factor of K_d

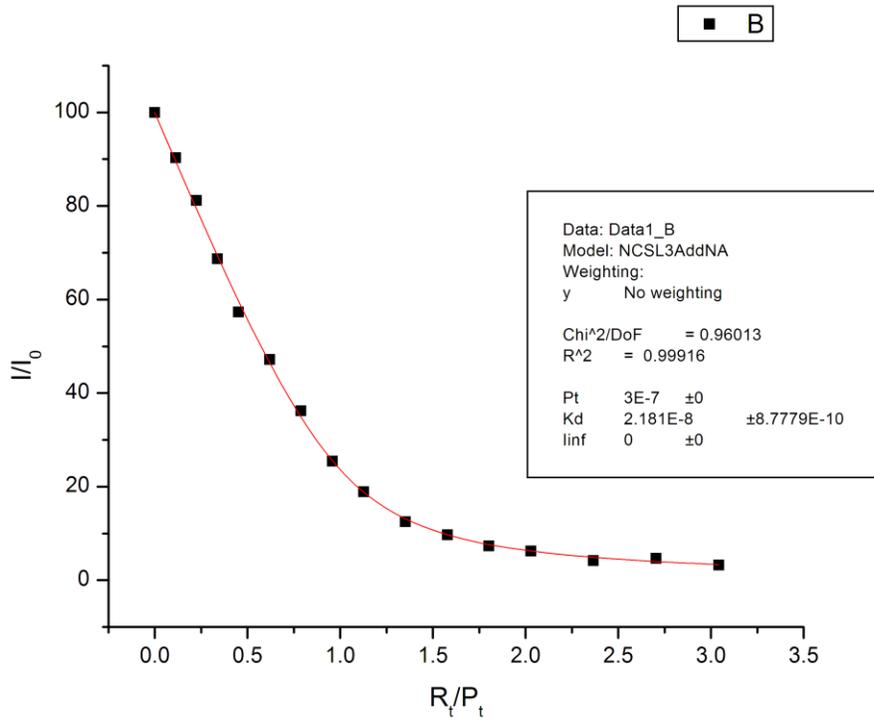
C-2. Individual NCp7-SL3 Trp Titration Curves

Individual curves start from the next page.

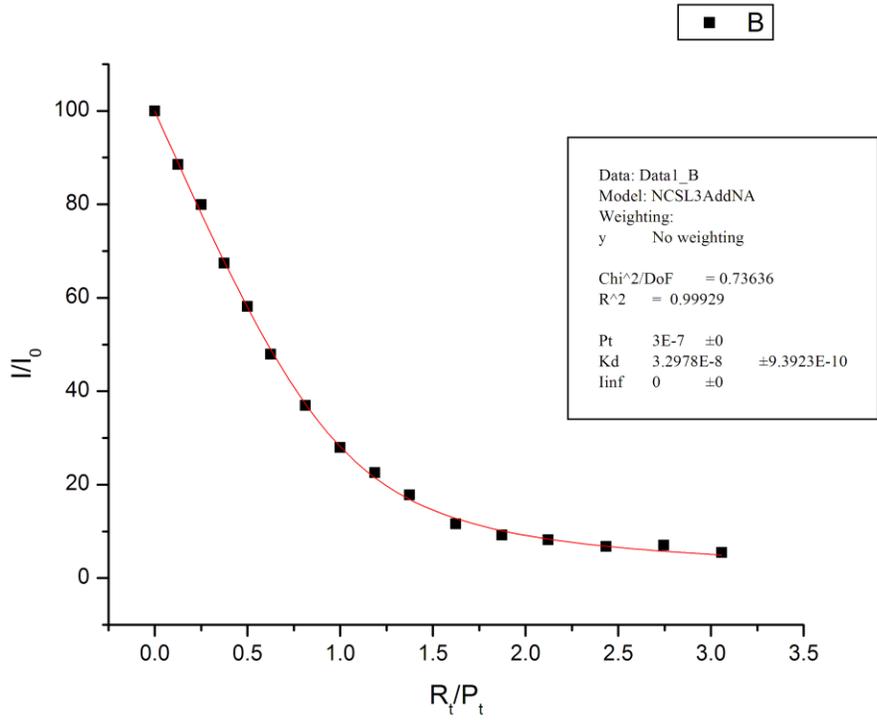
10/29/2010 WT



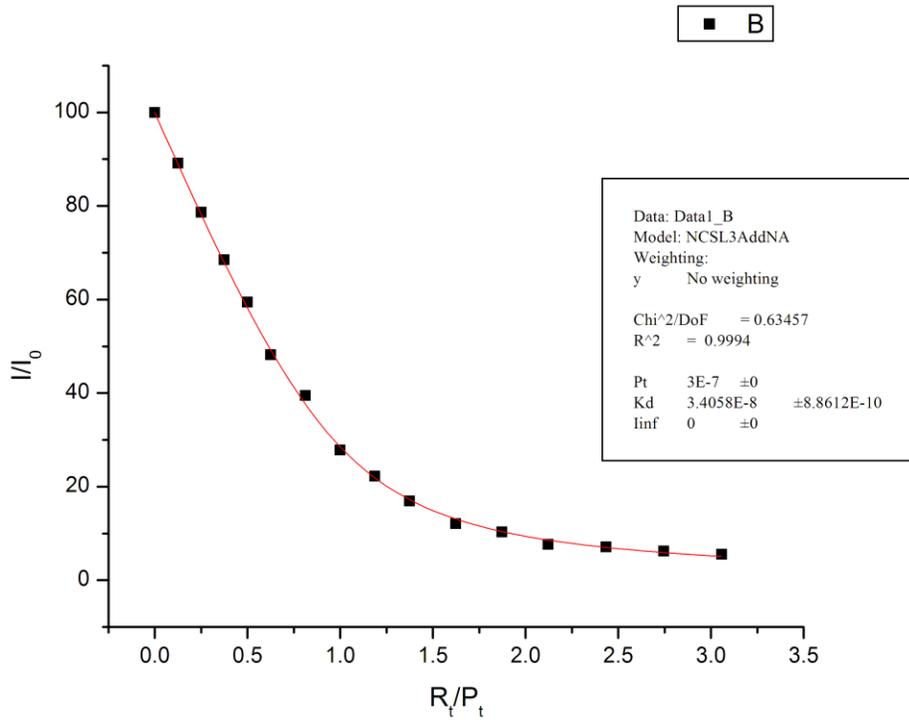
11/7/2010 WT



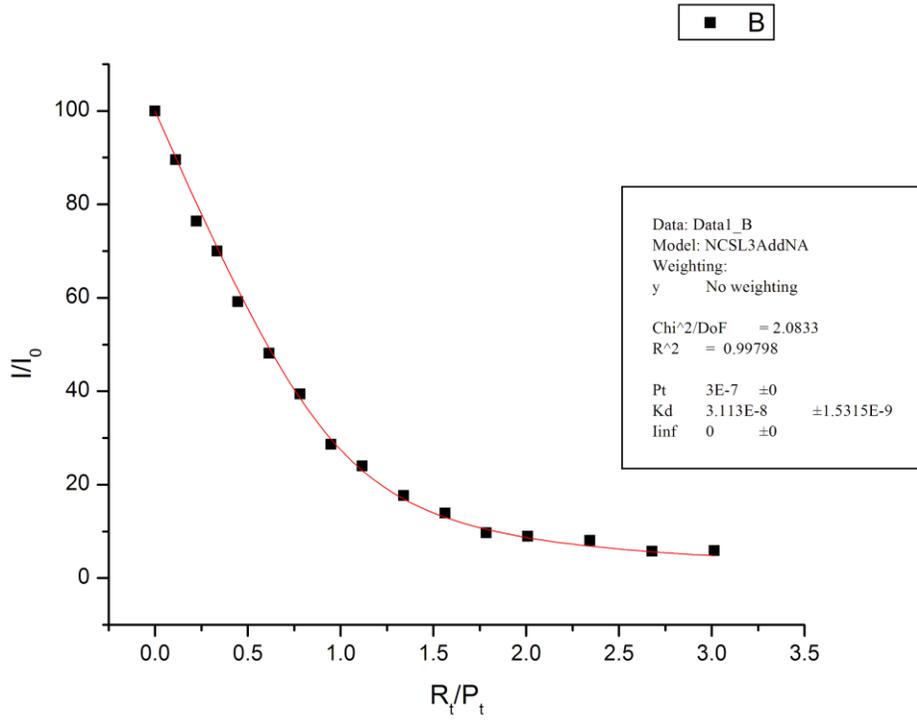
11/17/2010 WT



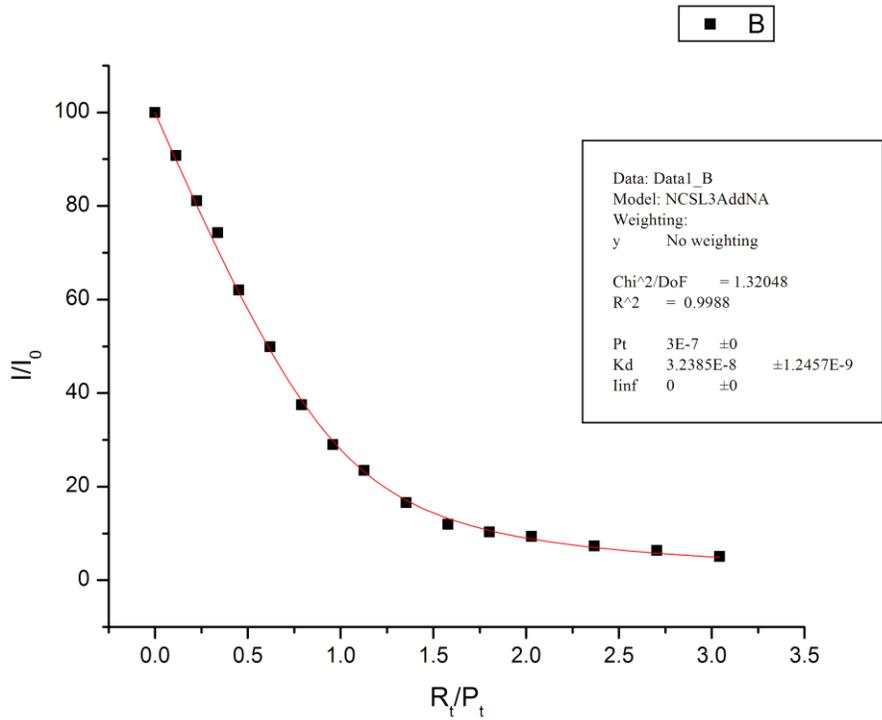
11/18/2010 WT



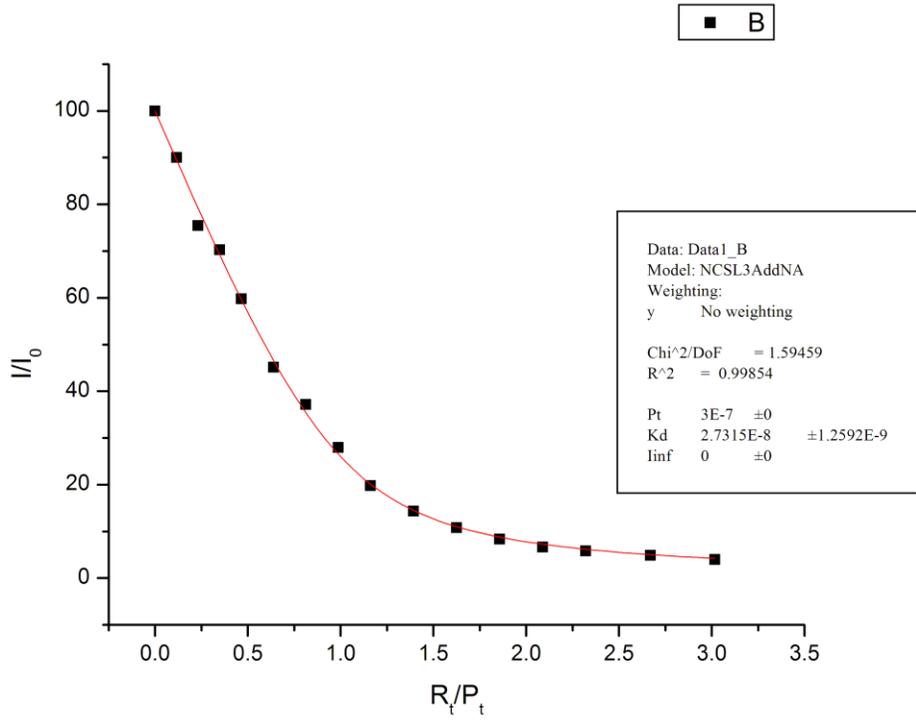
11/19/2010 WT



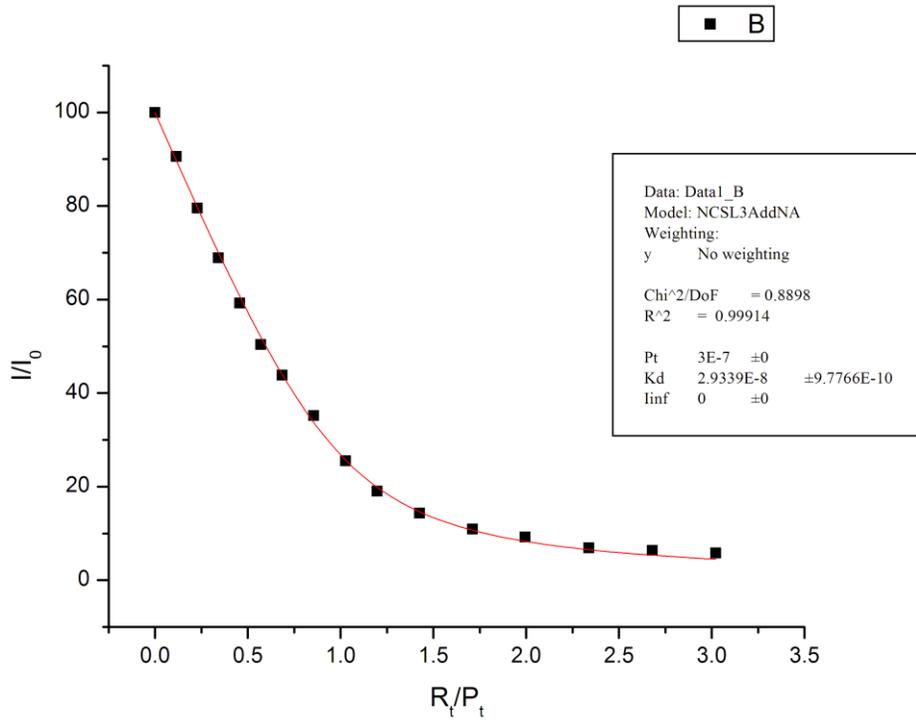
11/22/2010 WT



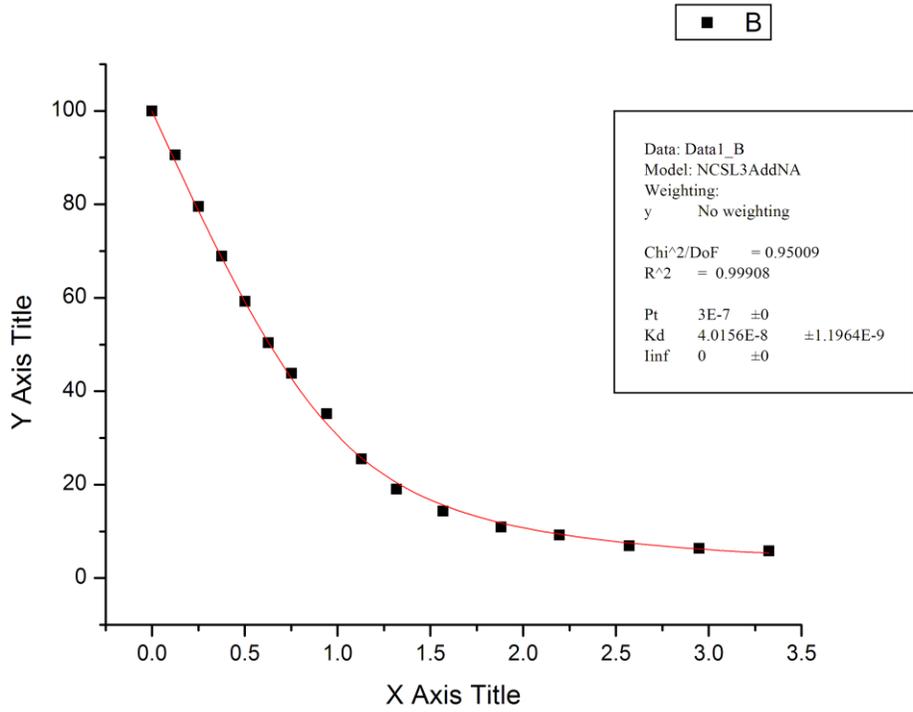
11/23/2010 WT



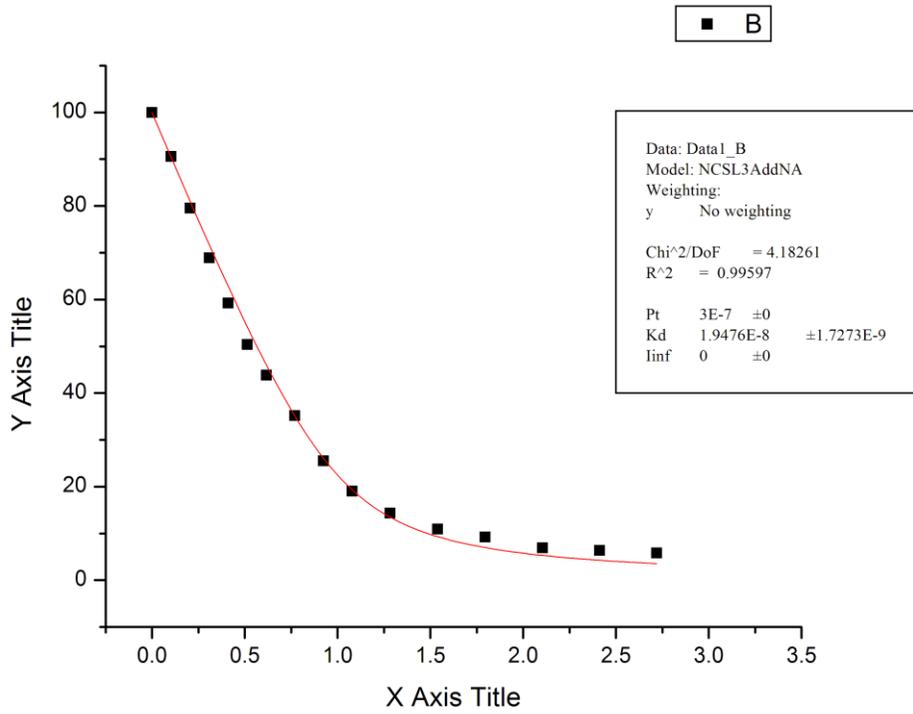
11/29/2010 WT



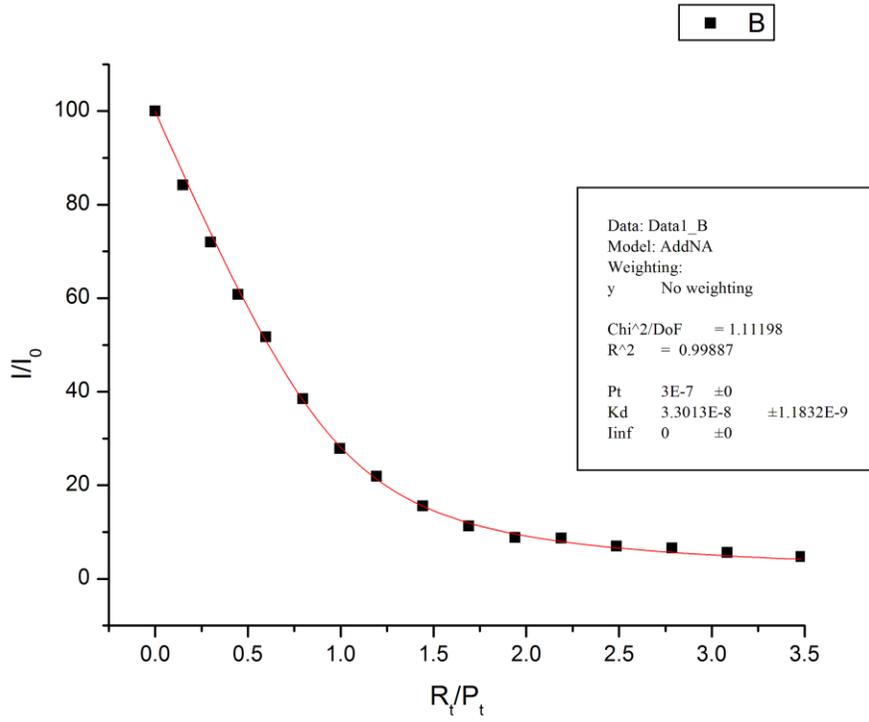
11/29/2010 WT (SL3+10%)



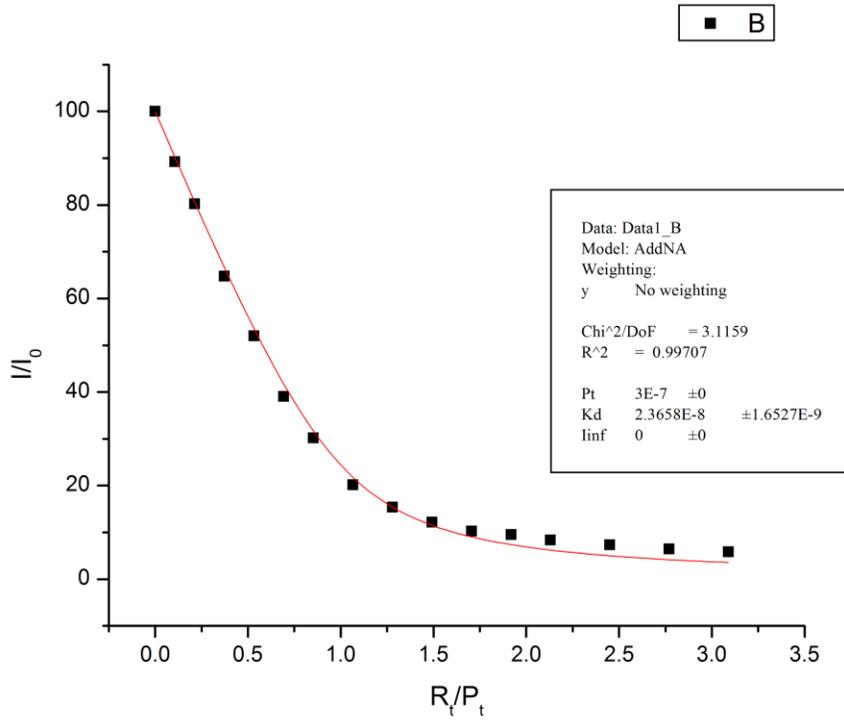
11/29/2010 WT (SL3-10%)



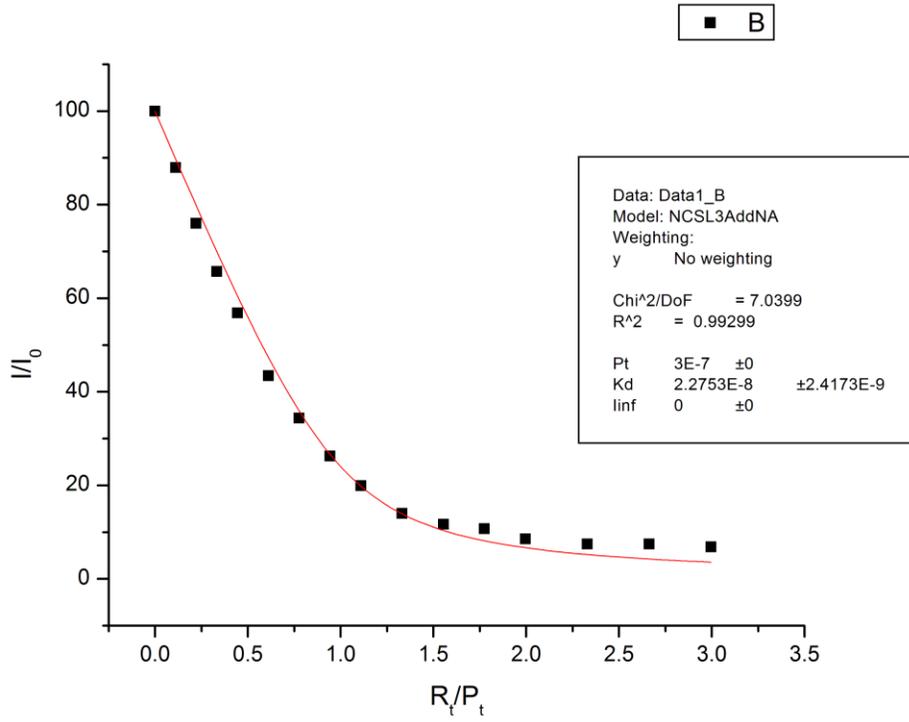
10/20/2010 N5A



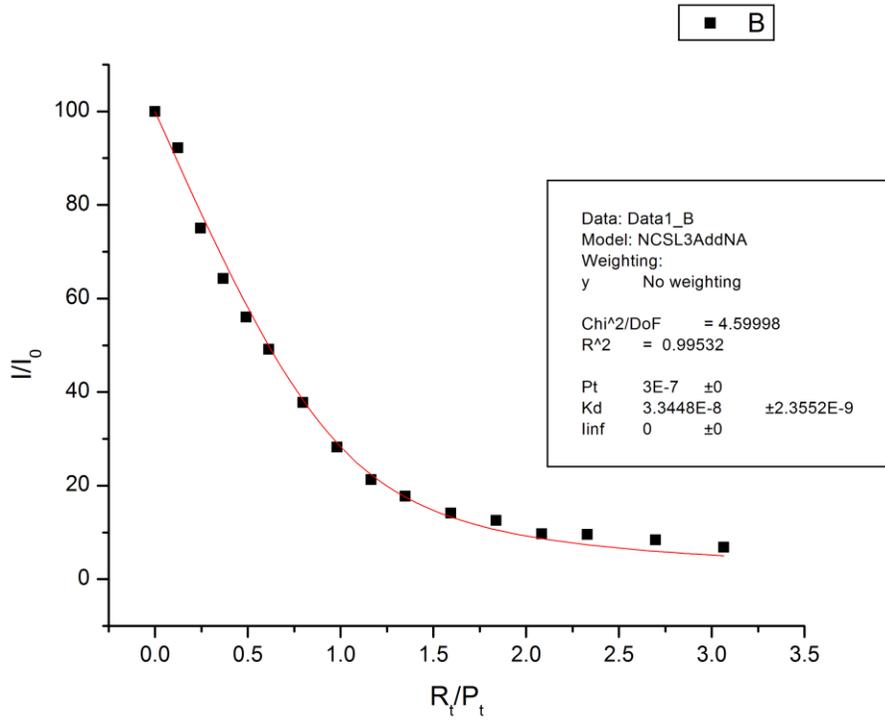
10/29/2010 N5A



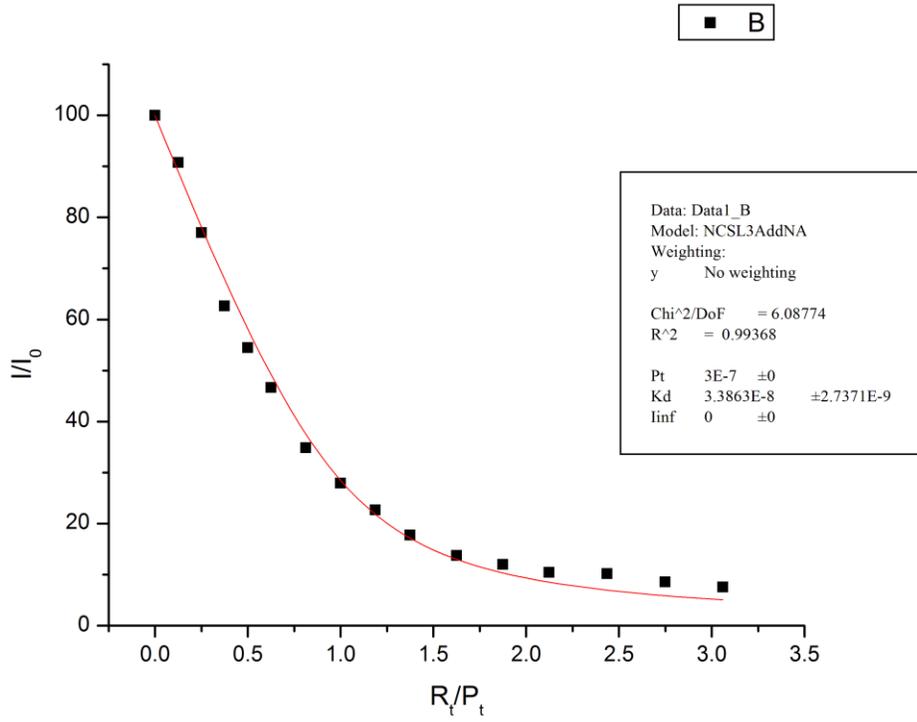
11/9/2010 N5A



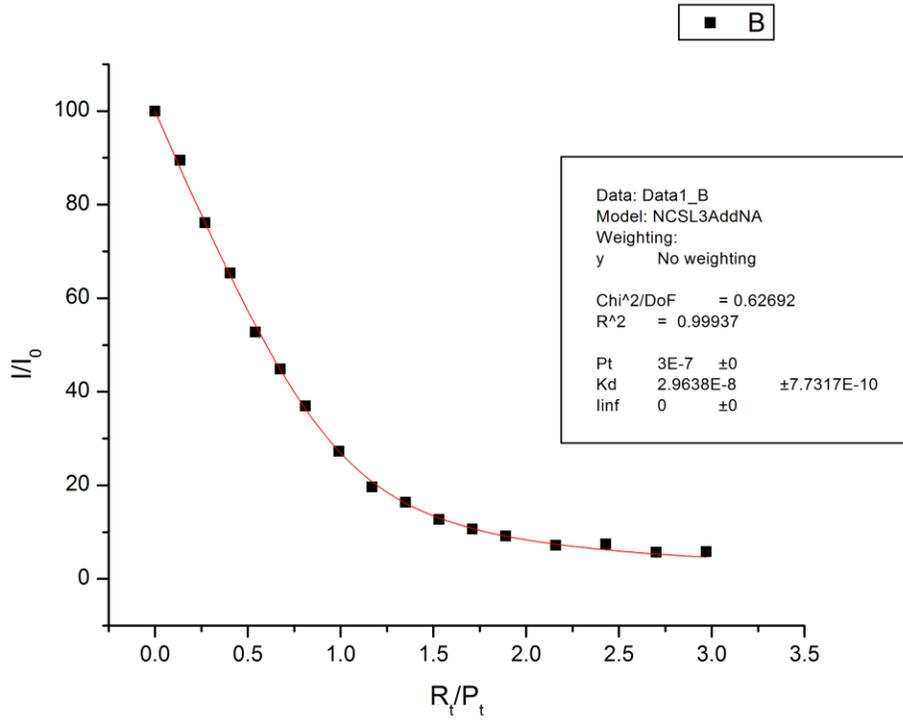
11/15/2010 N5A



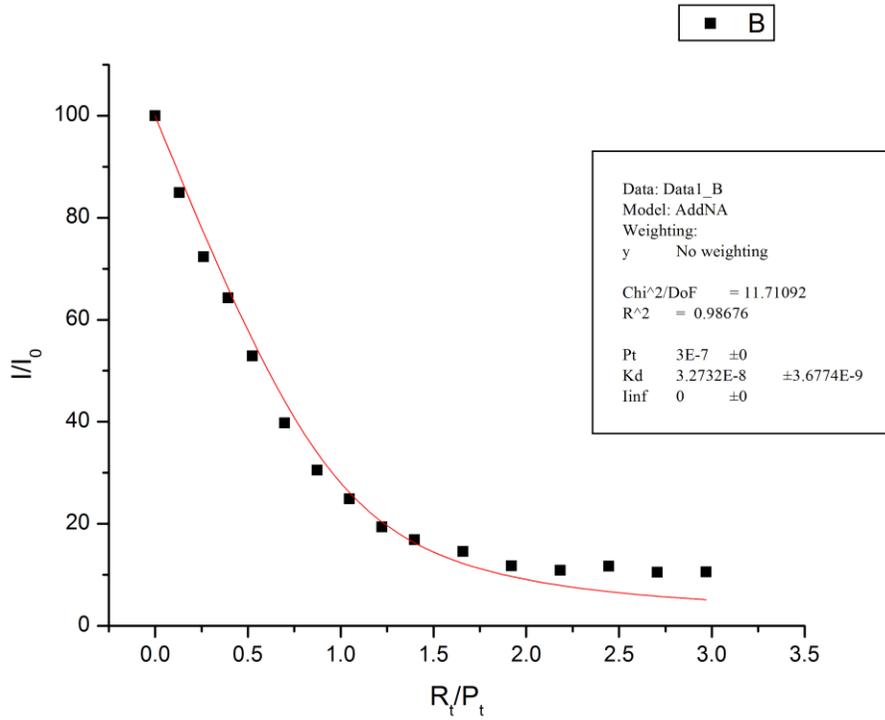
11/18/2010 N5A

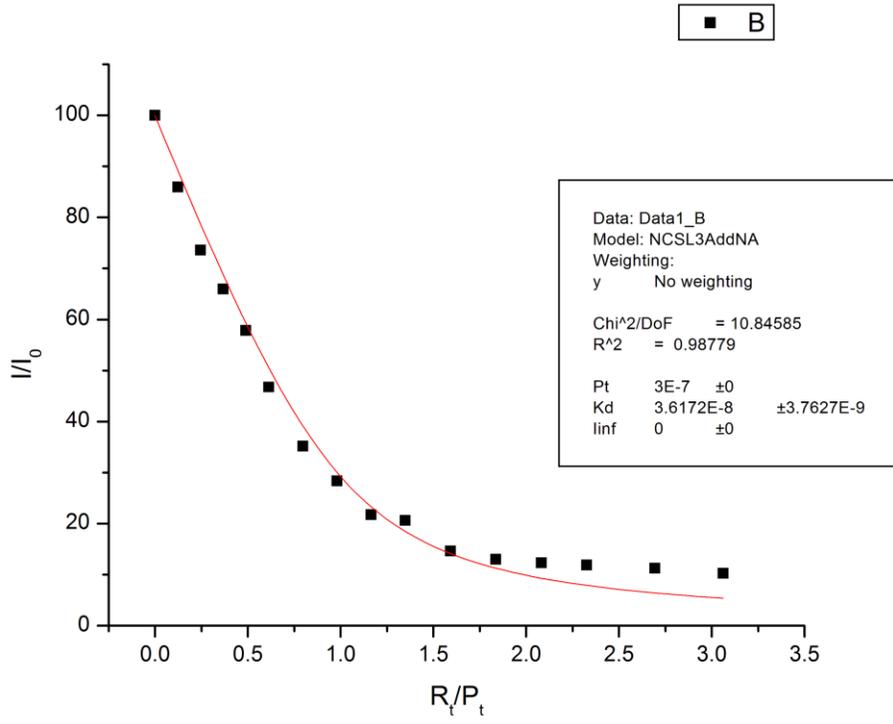


7/26/2010 F6A

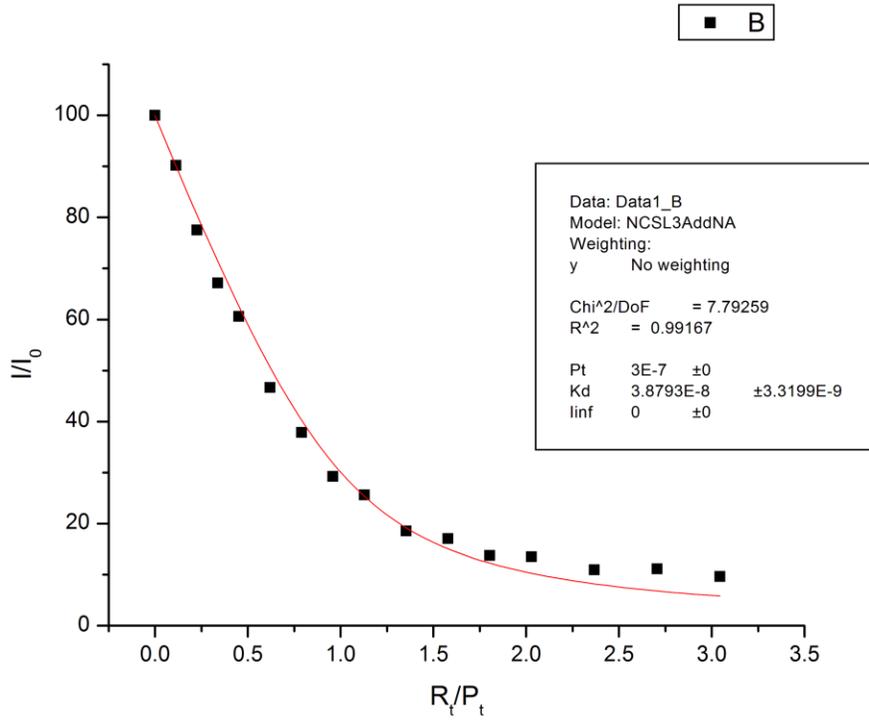


10/12/2010 F6A

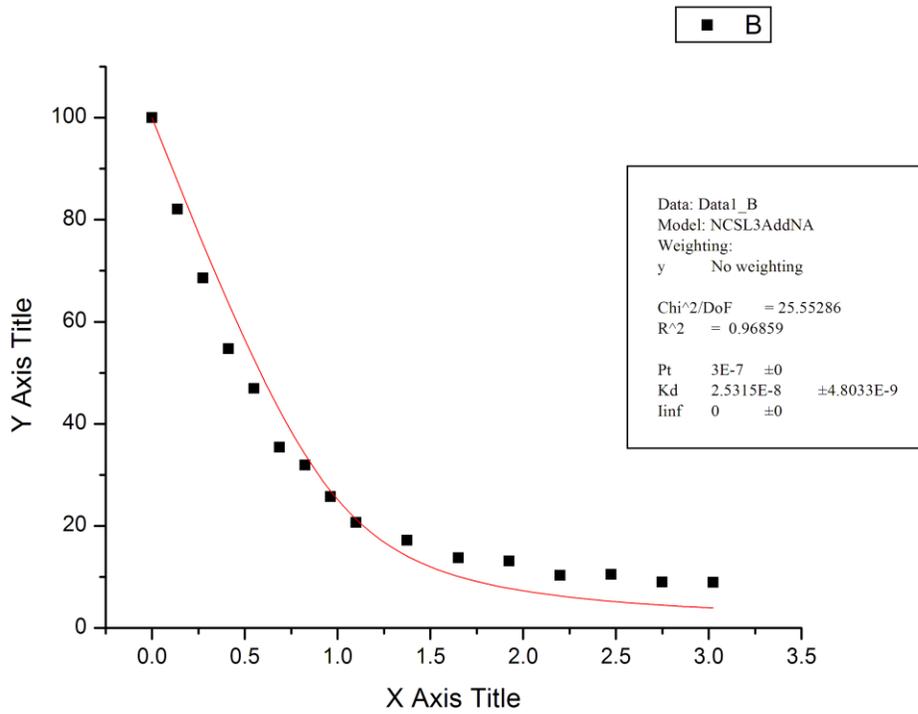




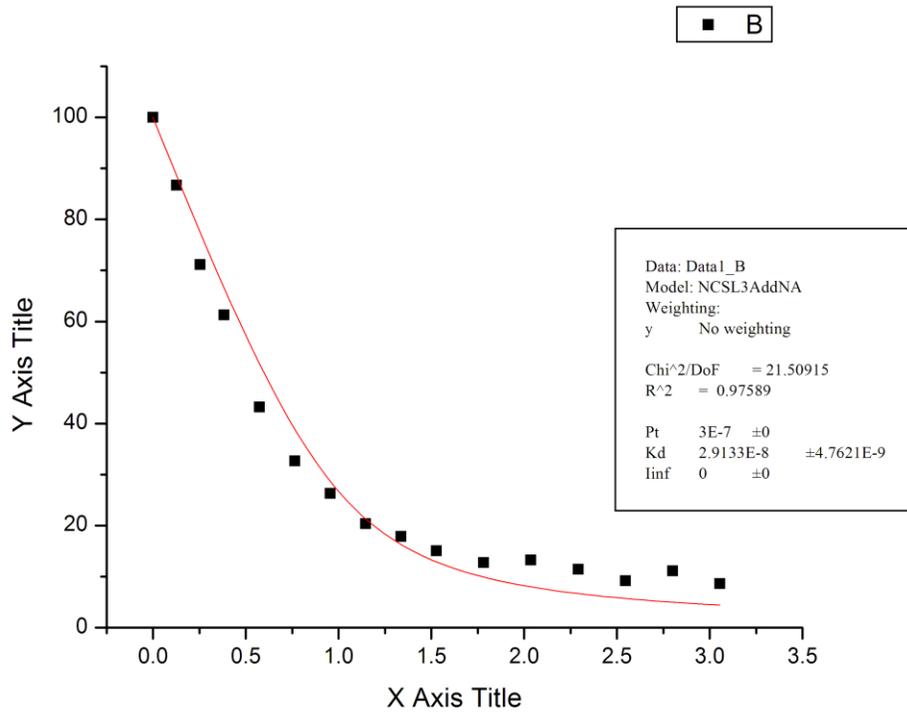
11/7/2010 V13A



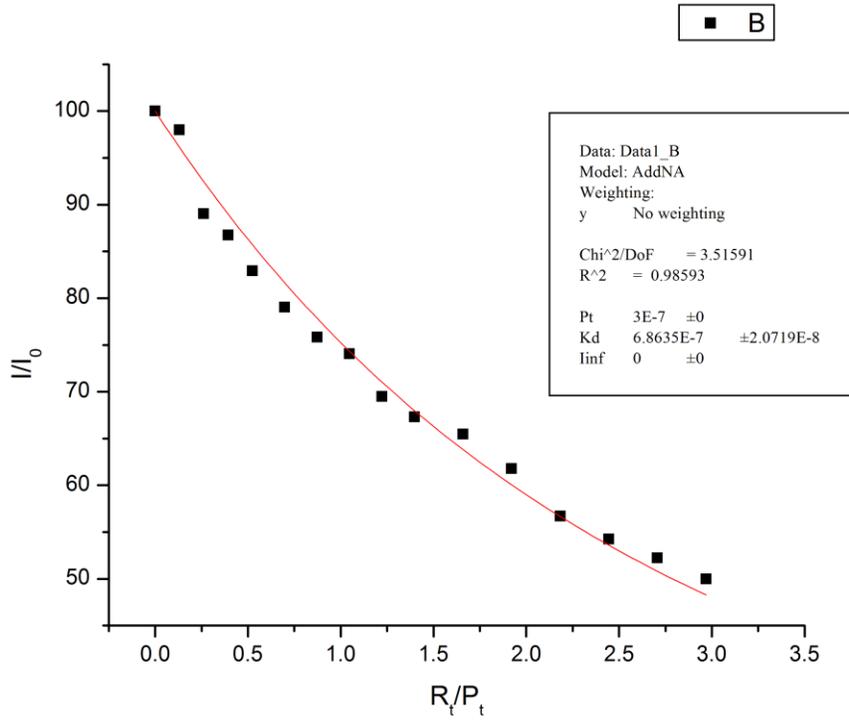
6/8/2011 V13A



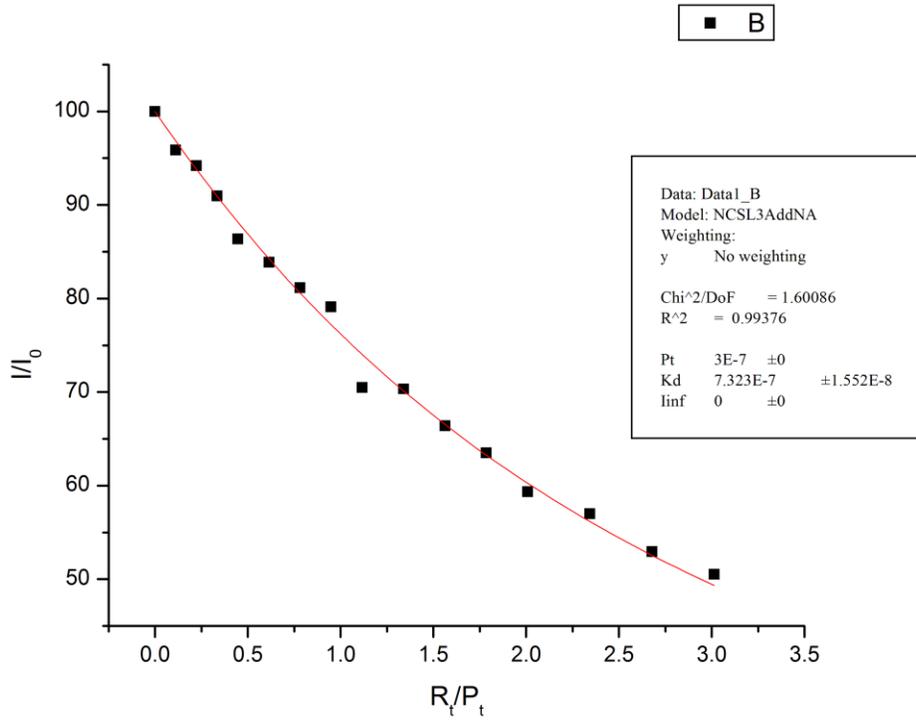
6/9/2011 V13A



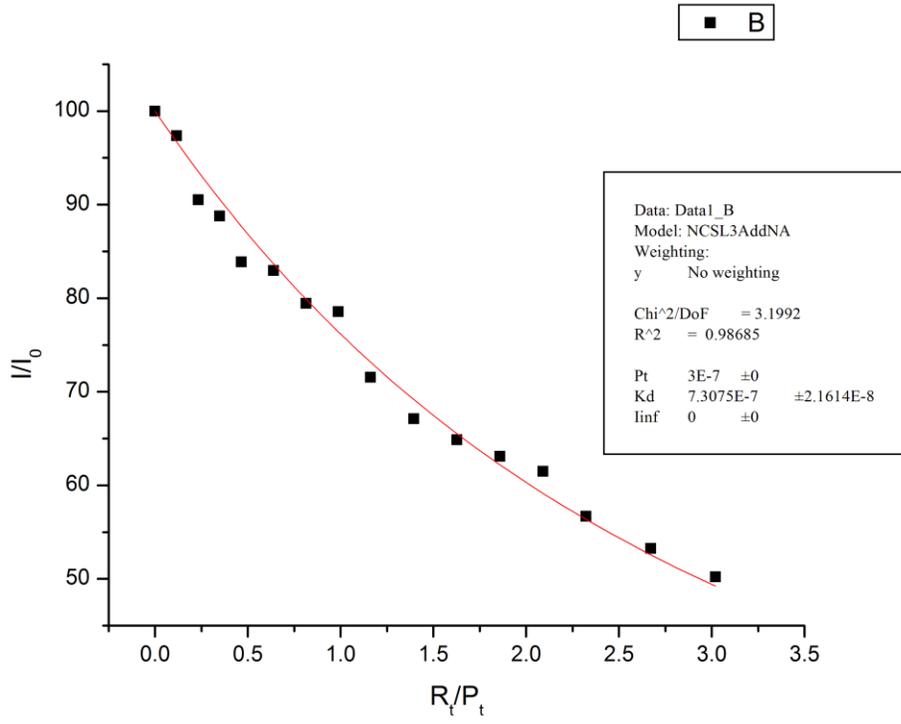
10/12/2010 F16A



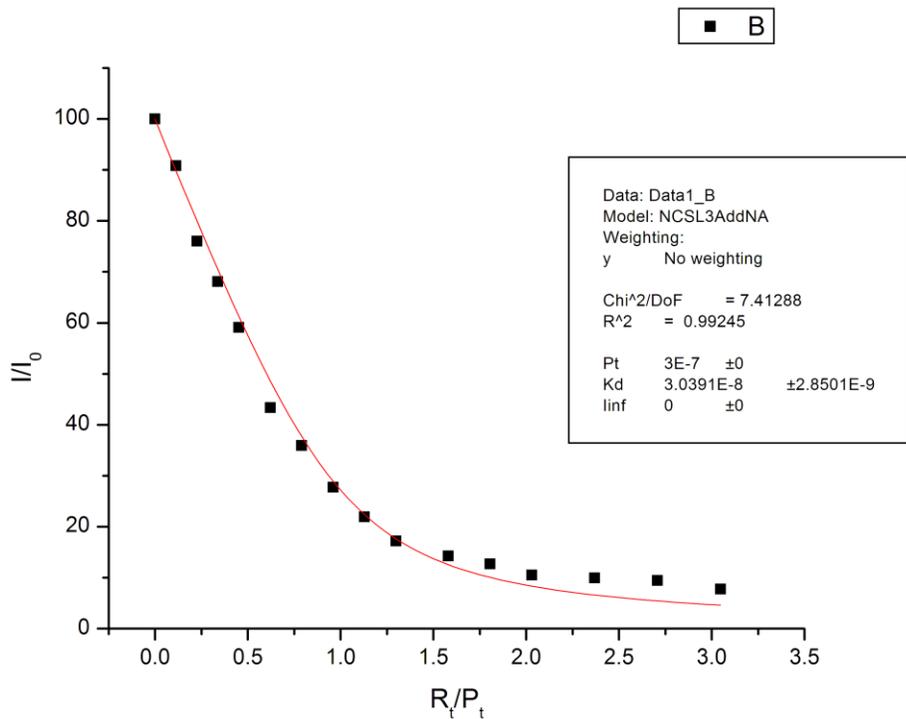
11/19/2010 F16A



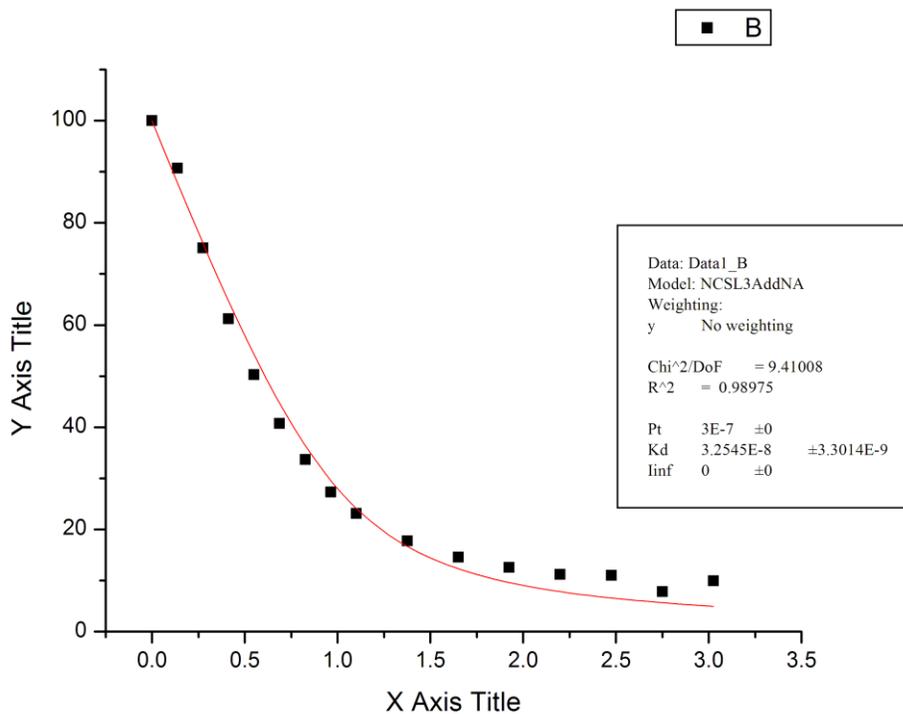
11/23/2010 F16A



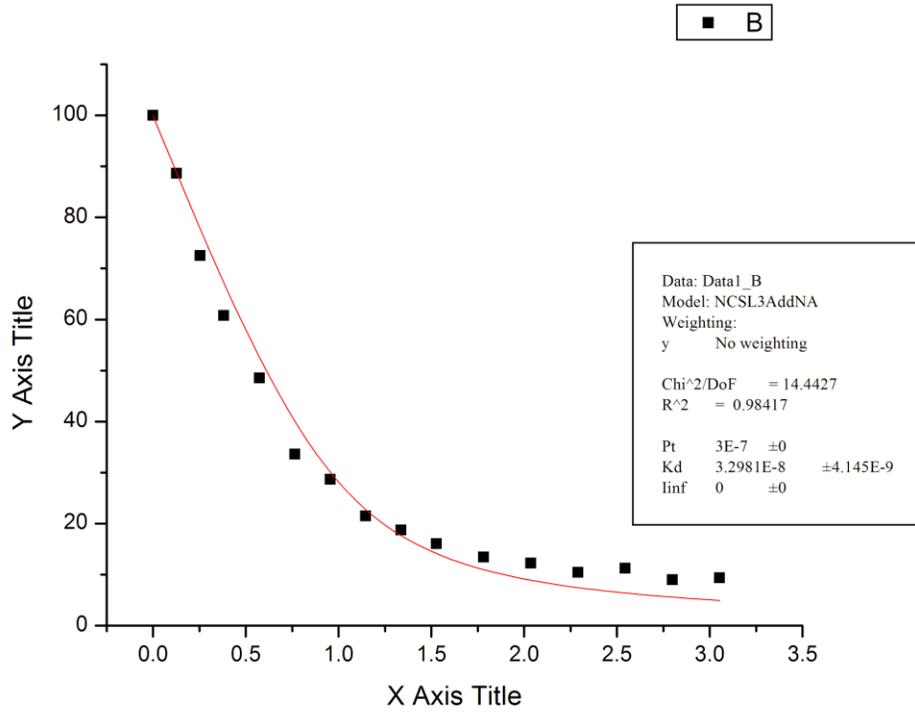
11/7/2010 N17A



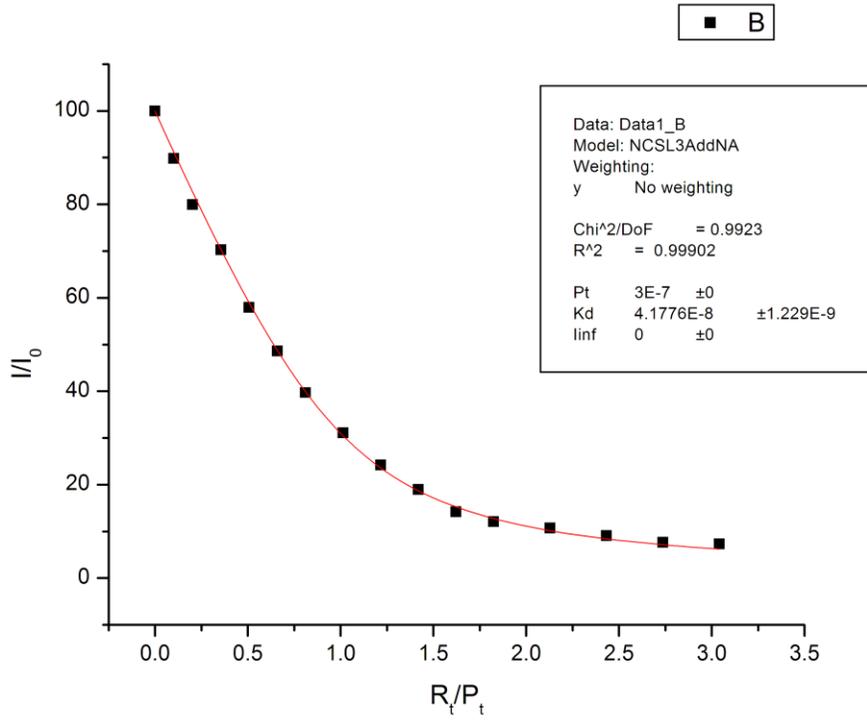
6/8/2011 N17A



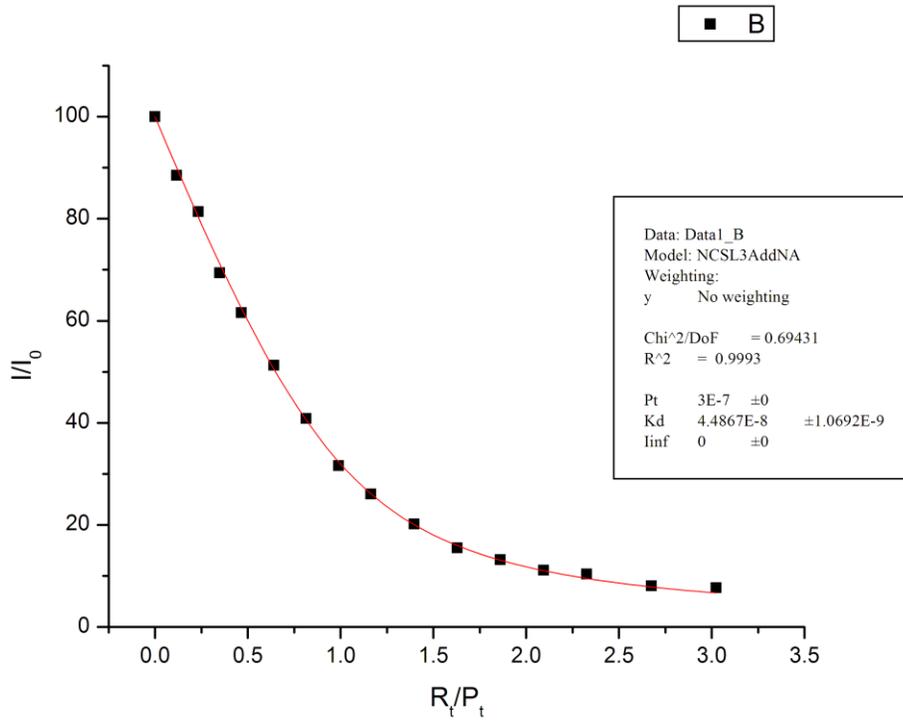
6/9/2011 N17A



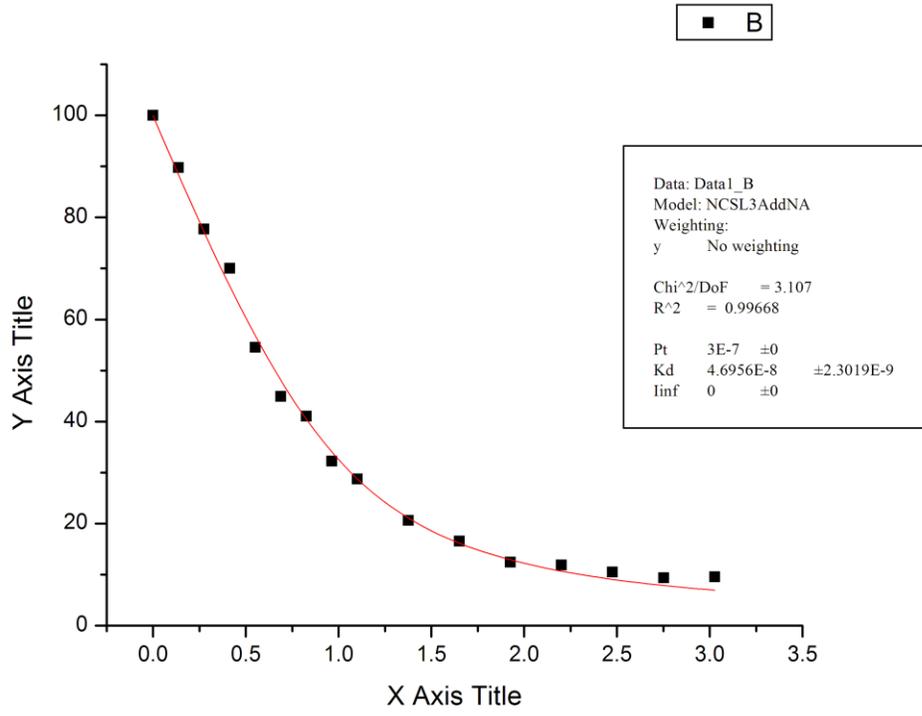
11/3/2010 G19A



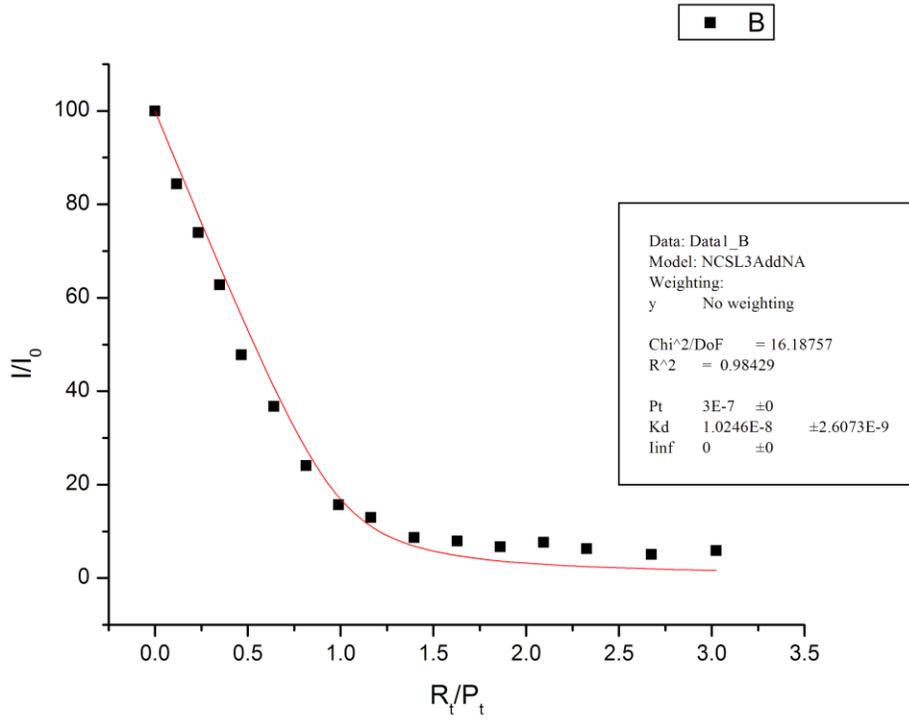
11/23/2010 G19A



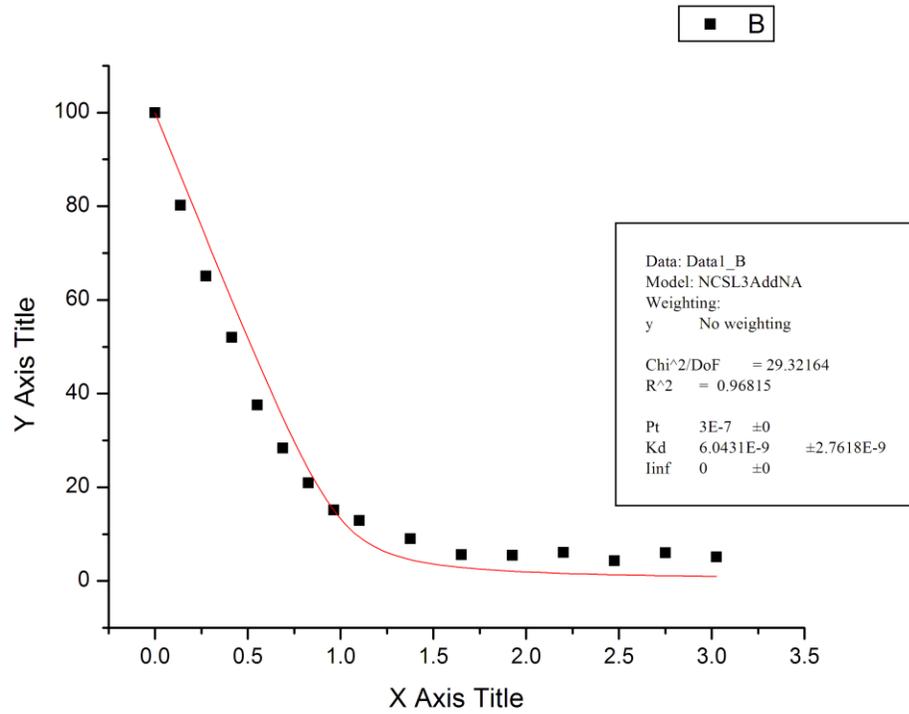
6/8/2011 G19A



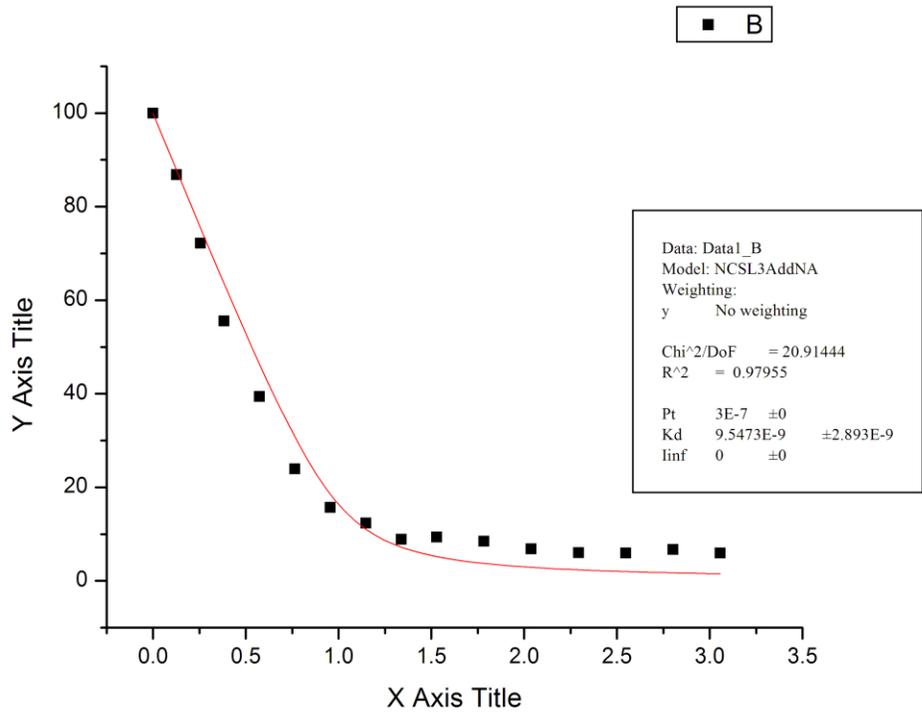
11/23/2010 E21A



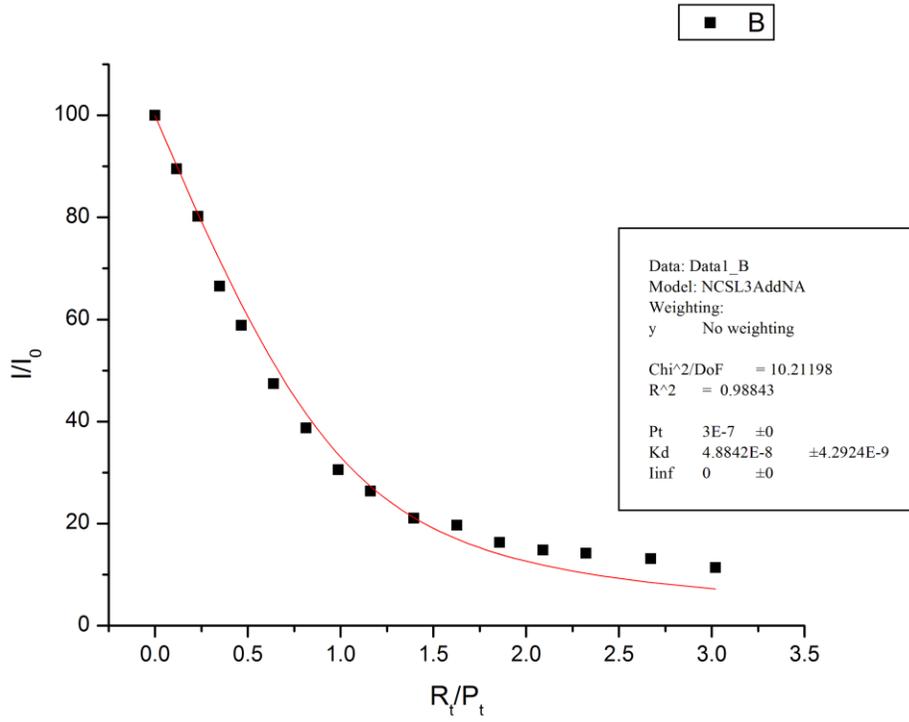
6/8/2011 E21A



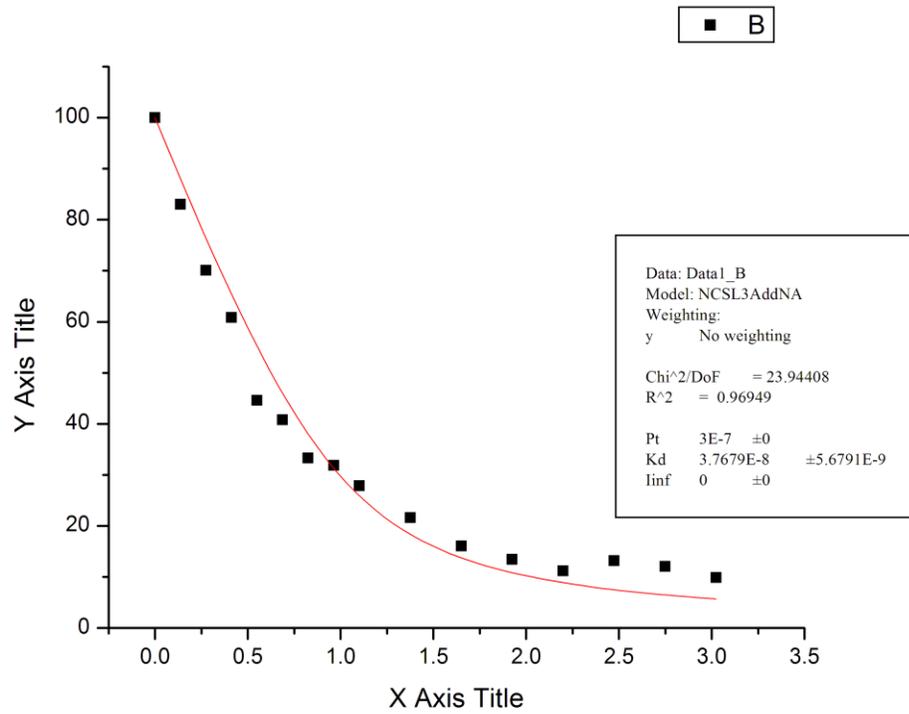
6/9/2011 E21A



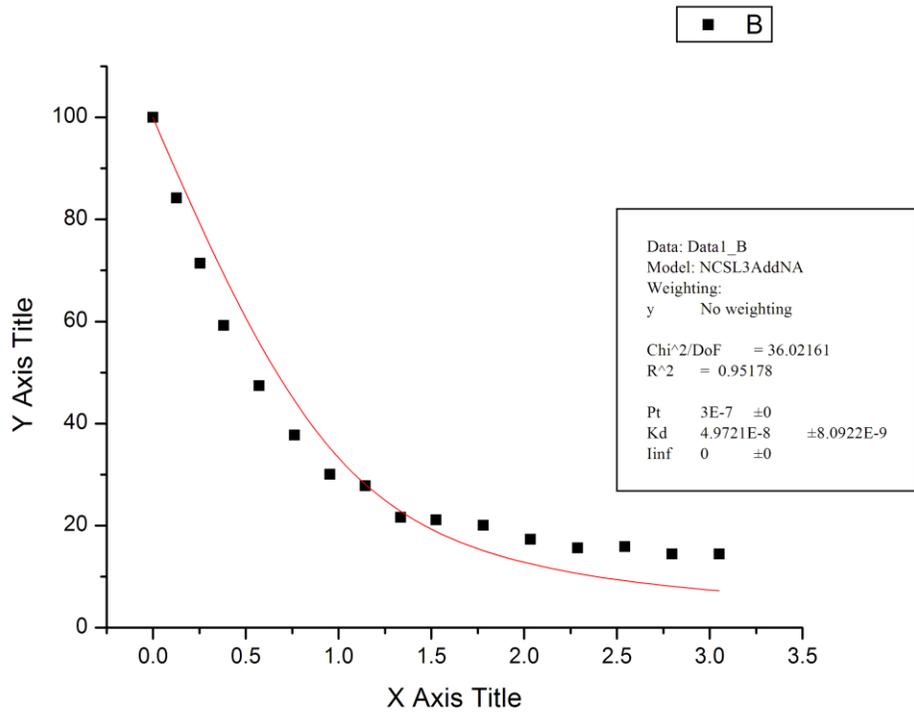
11/23/2010 G22A



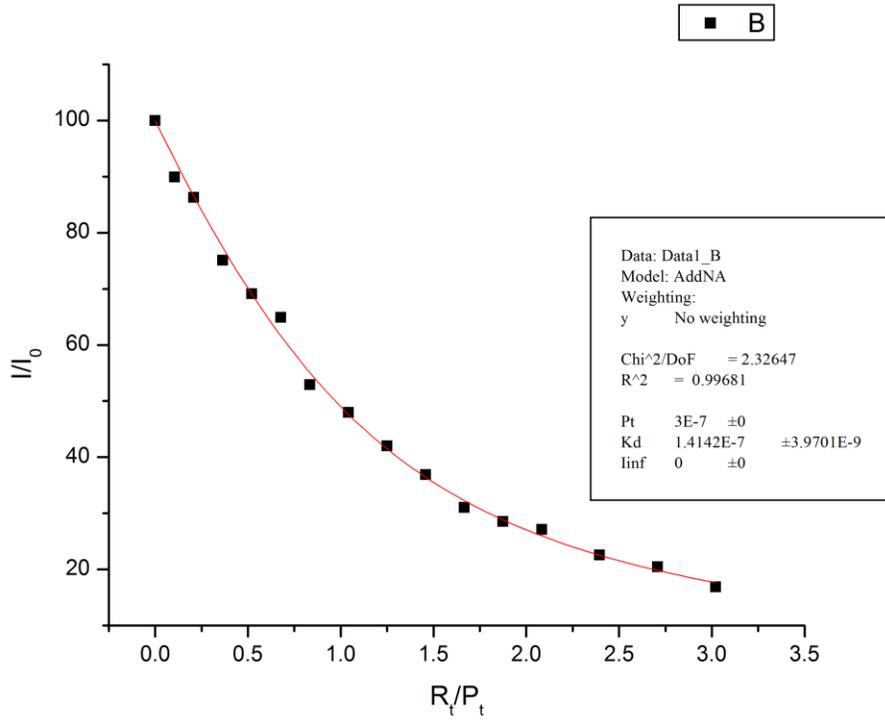
6/8/2011 G22A



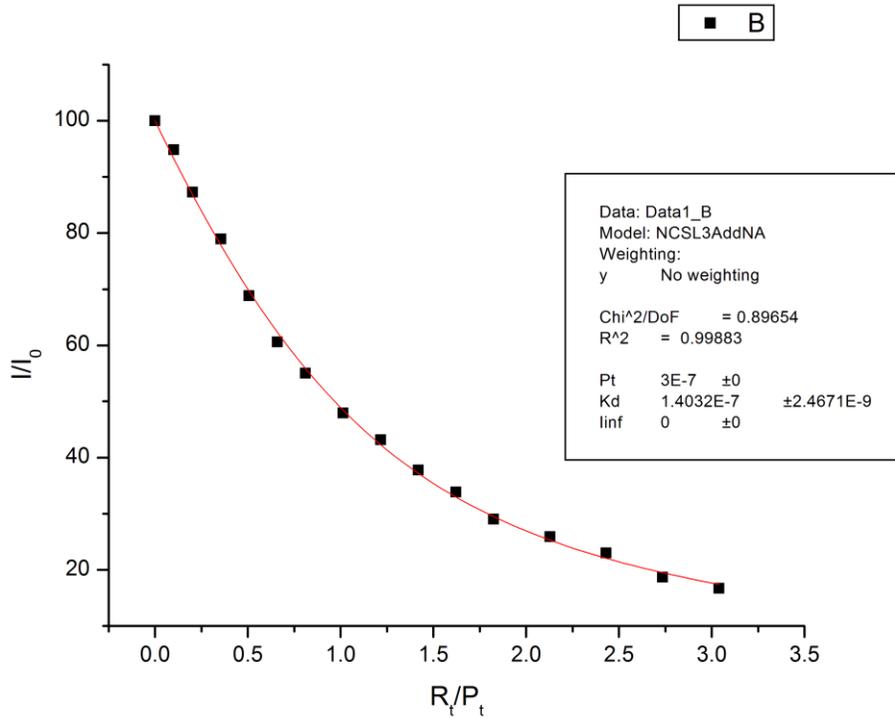
6/9/2011 G22A



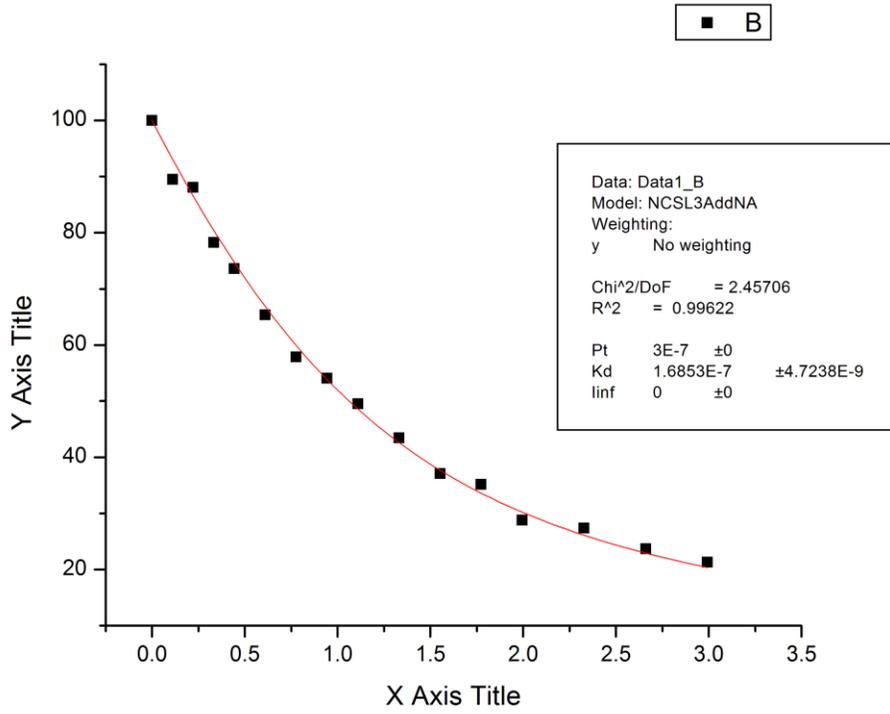
10/30/2010 I24A



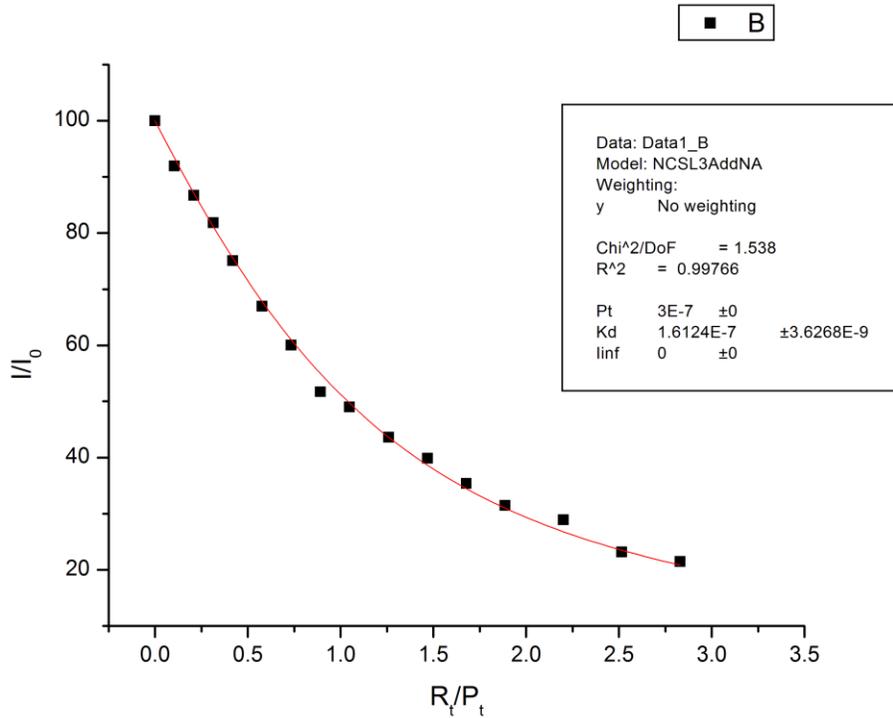
11/4/2010 I24A



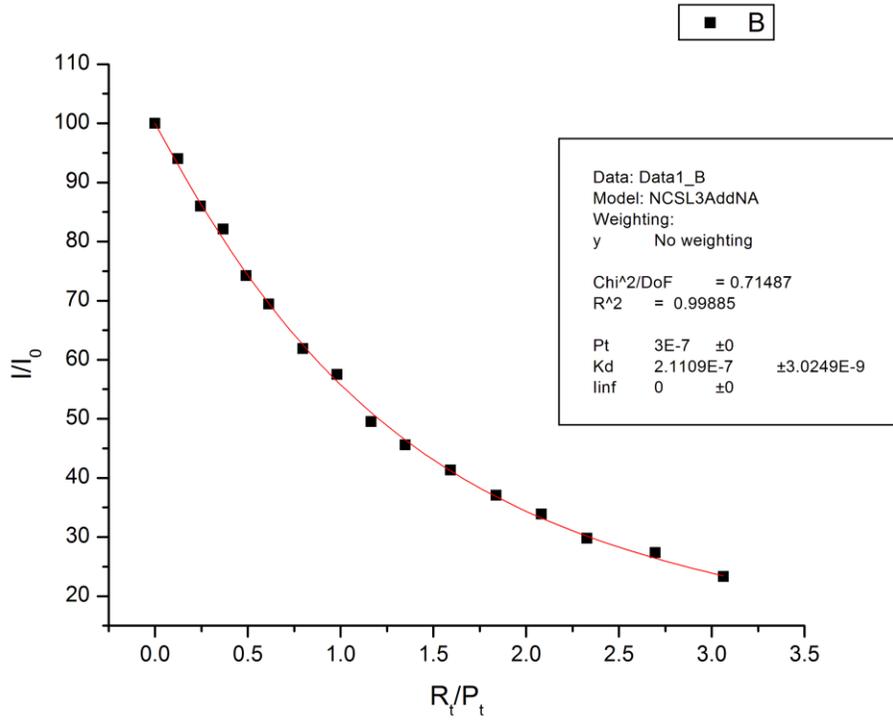
11/9/2010 I24A



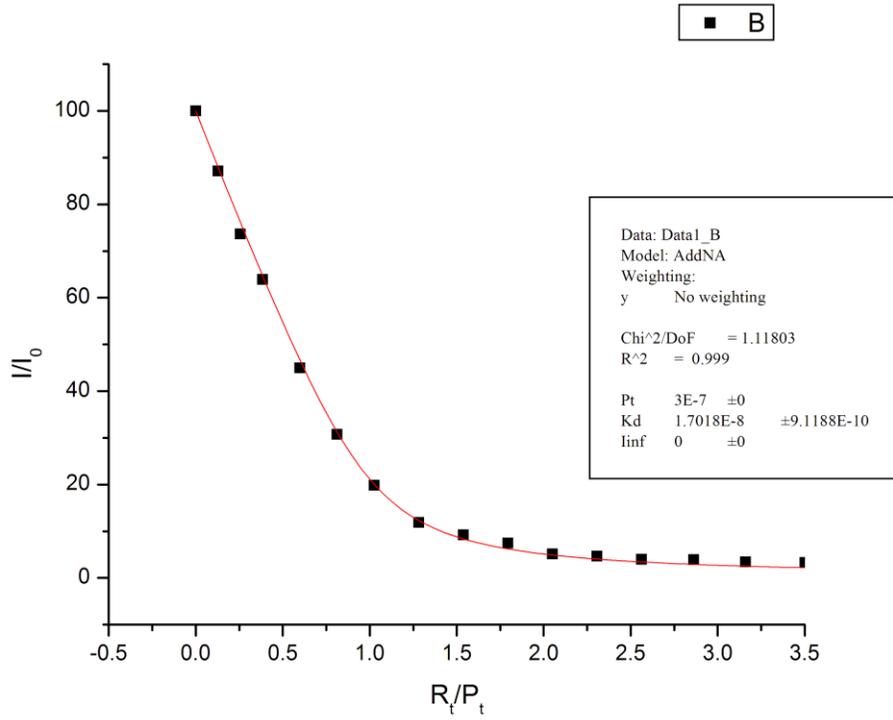
11/12/2010 I24A



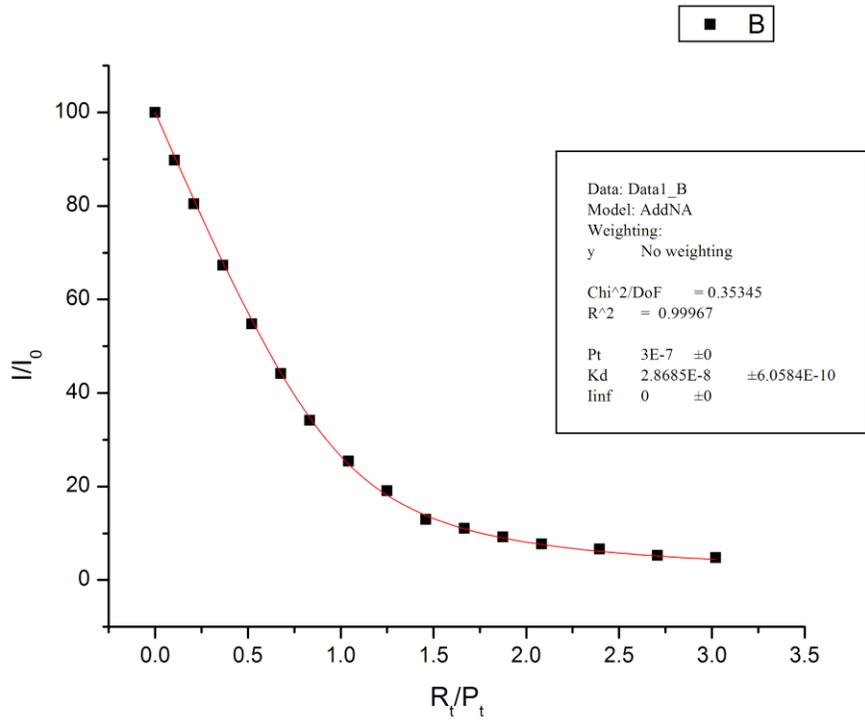
11/15/2010 I24A



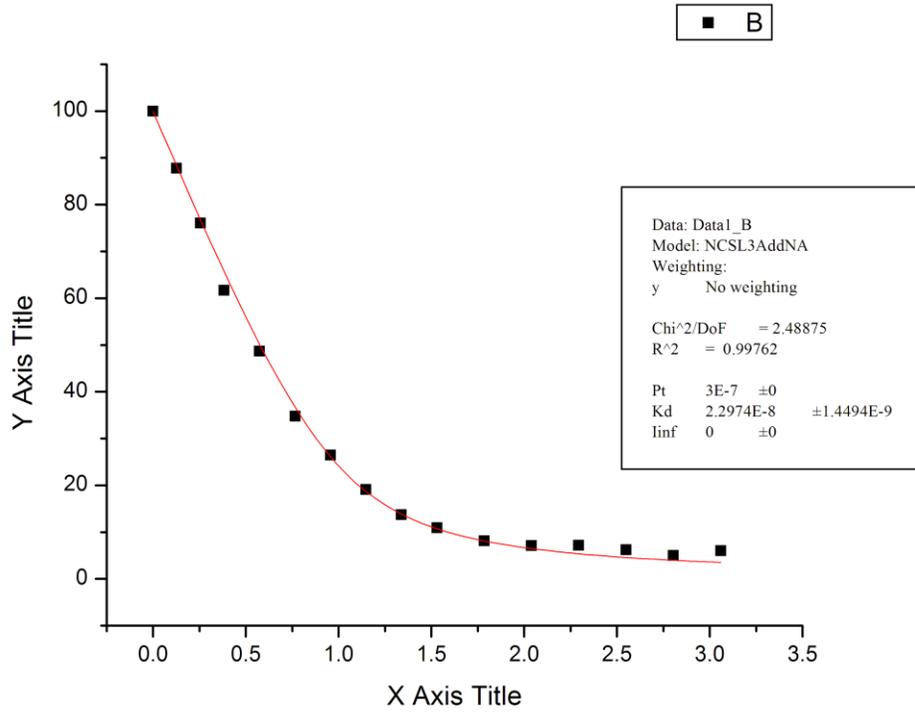
10/18/2010 P31A



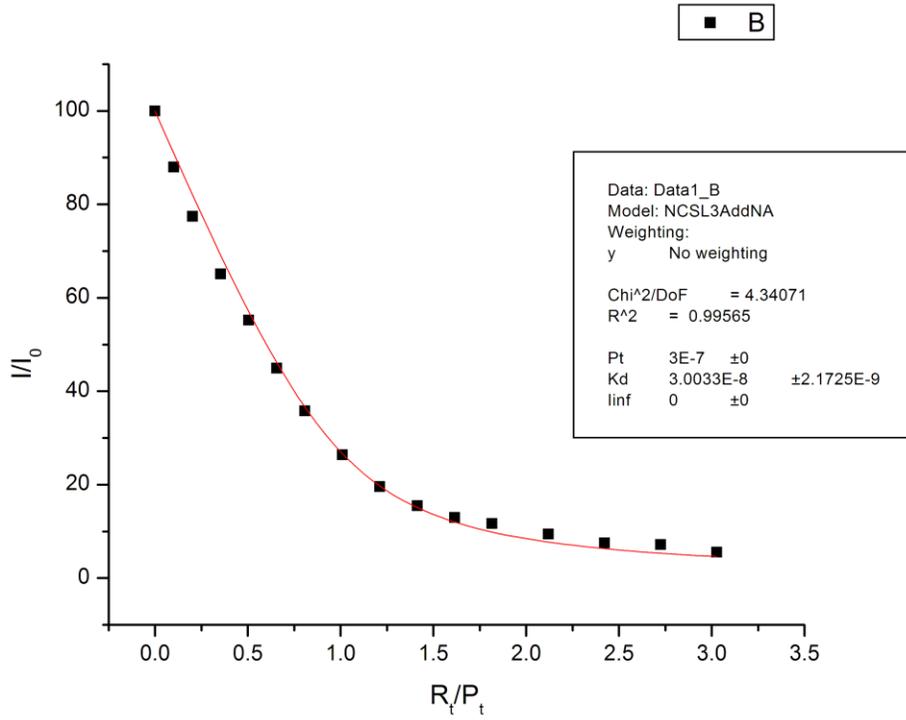
10/30/2010 P31A



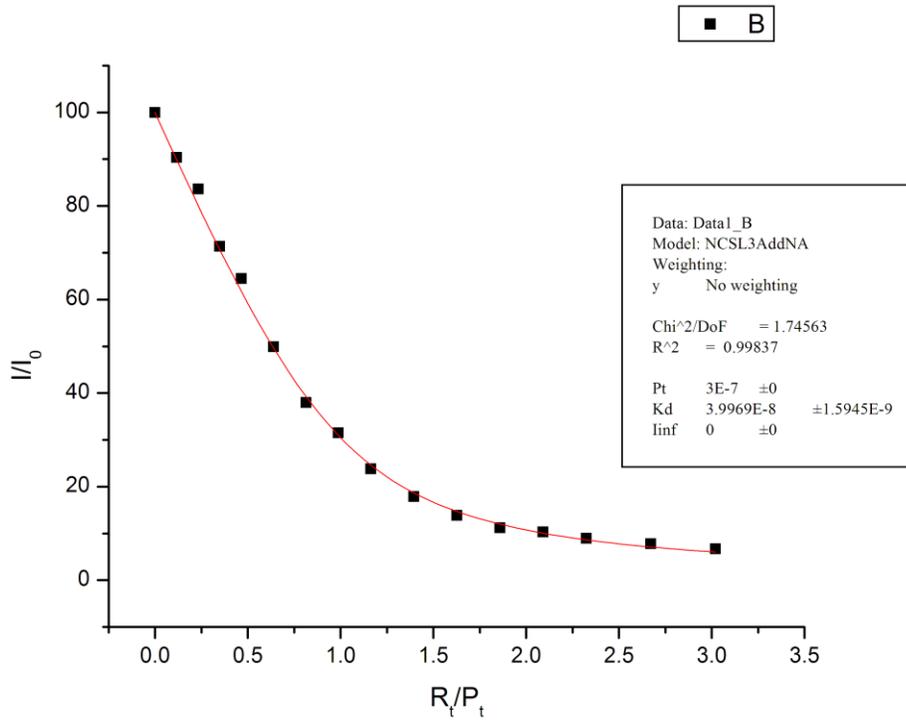
6/9/2011 P31A



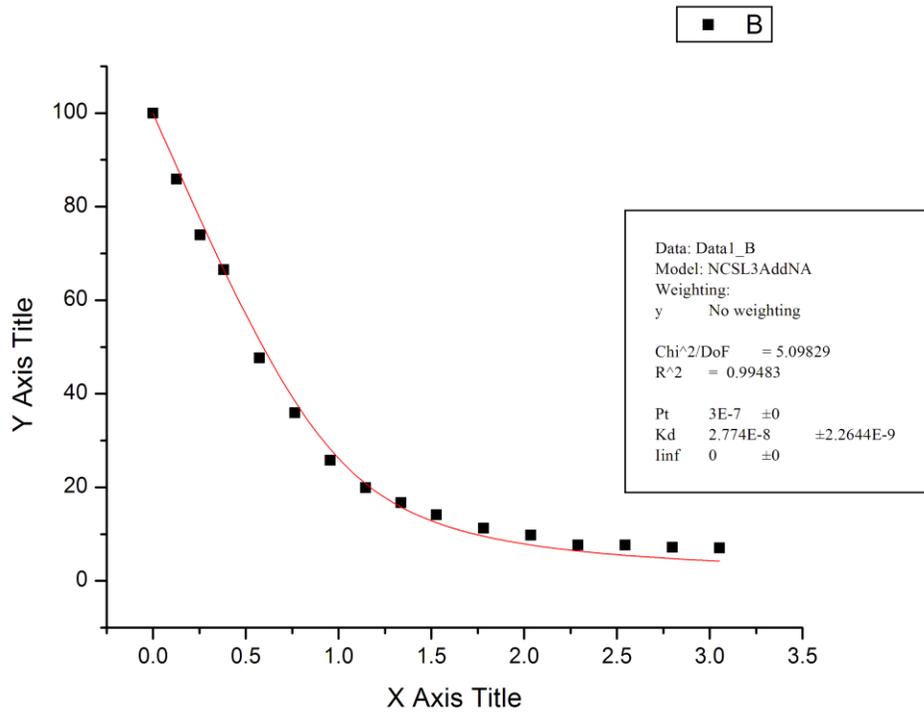
11/3/2010 G40A



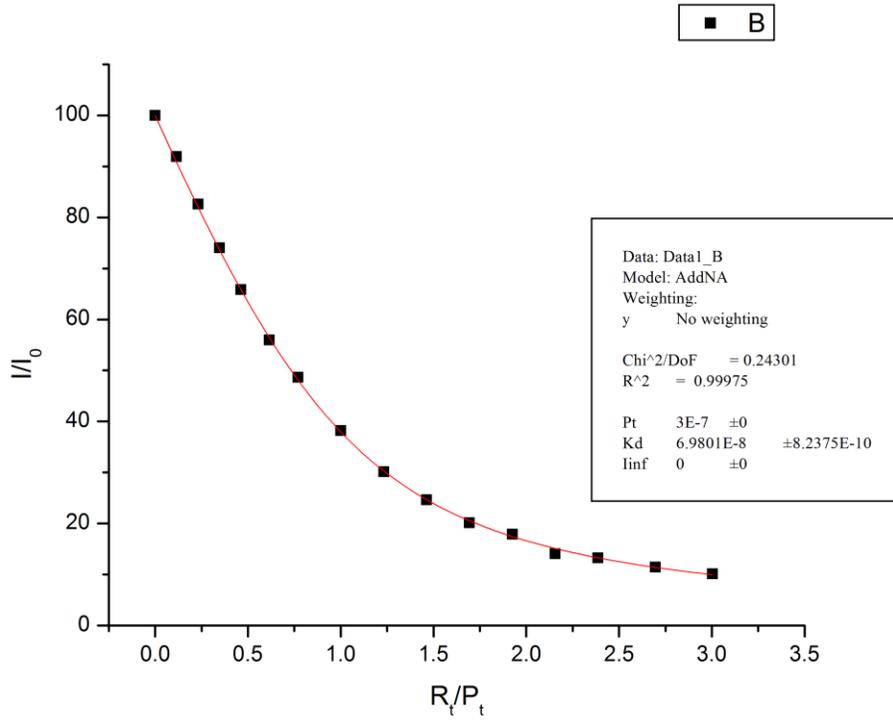
11/23/2010 G40A



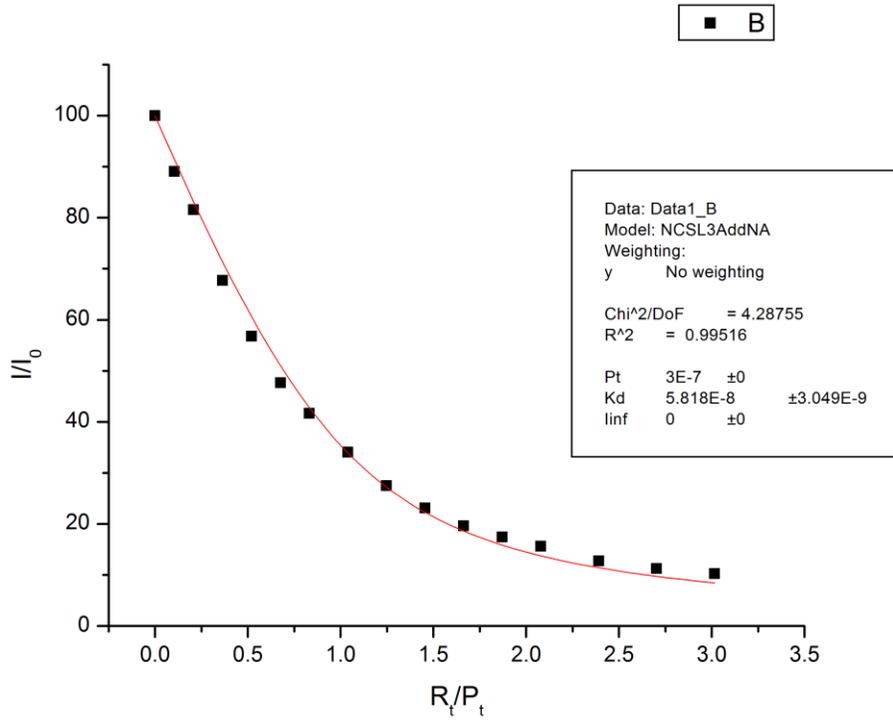
6/9/2011 G40A



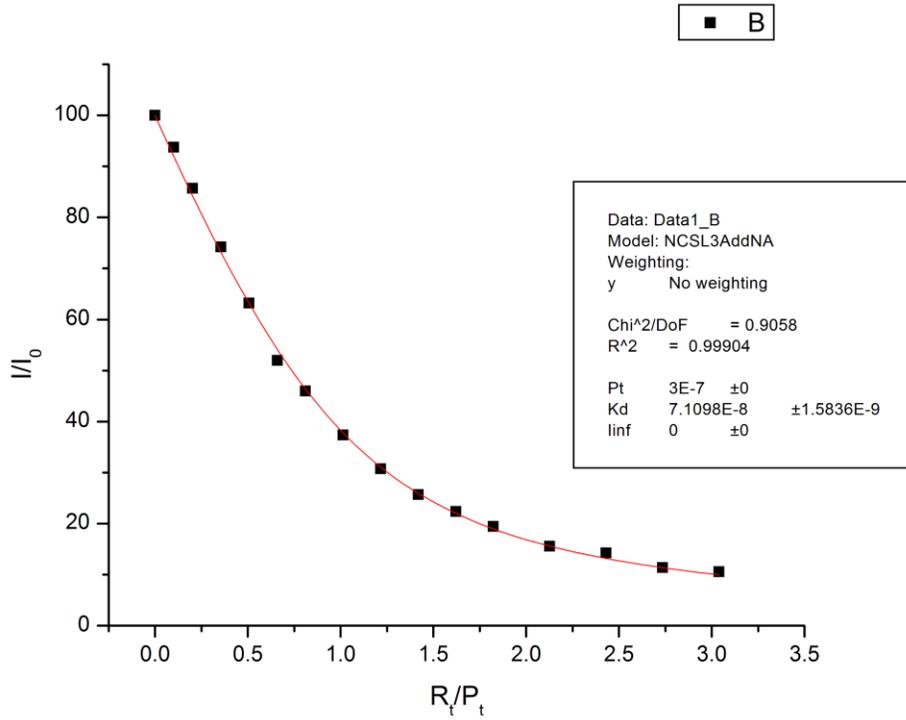
10/13/2010 Q45A



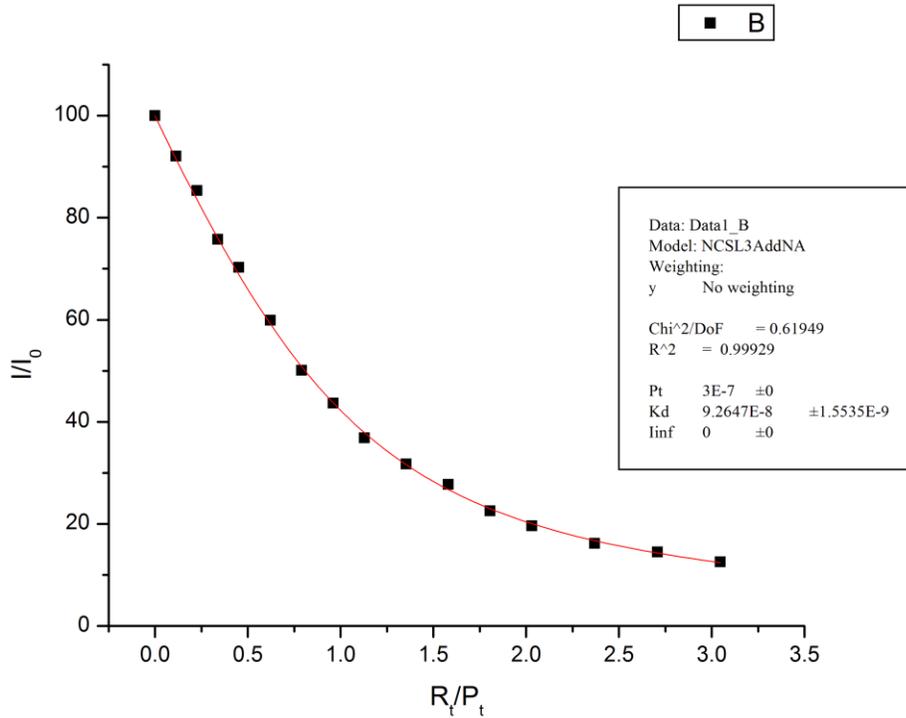
11/1/2010 Q45A



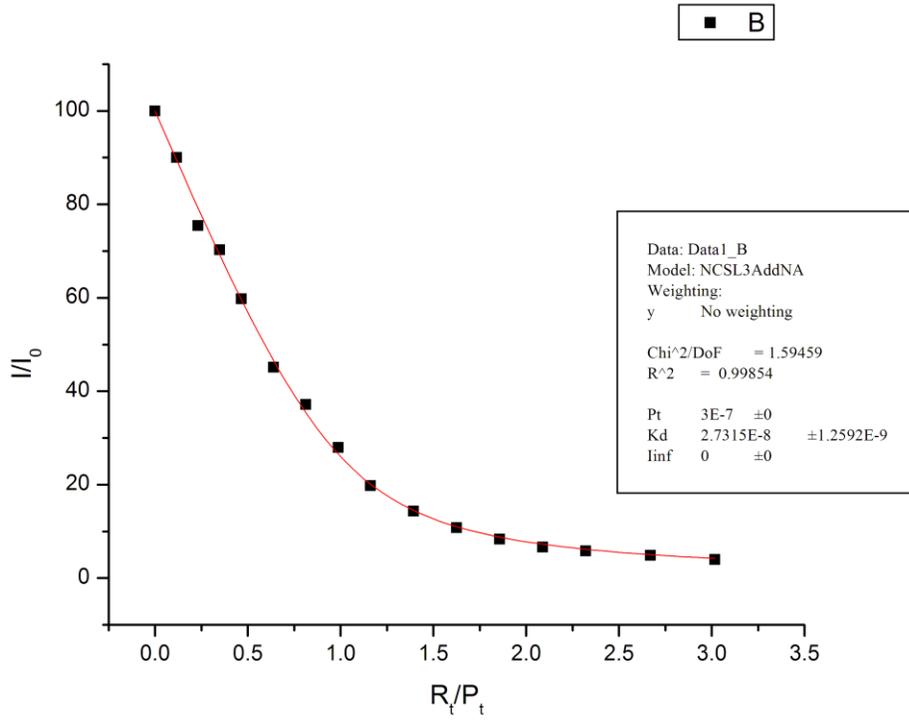
11/4/2010 Q45A



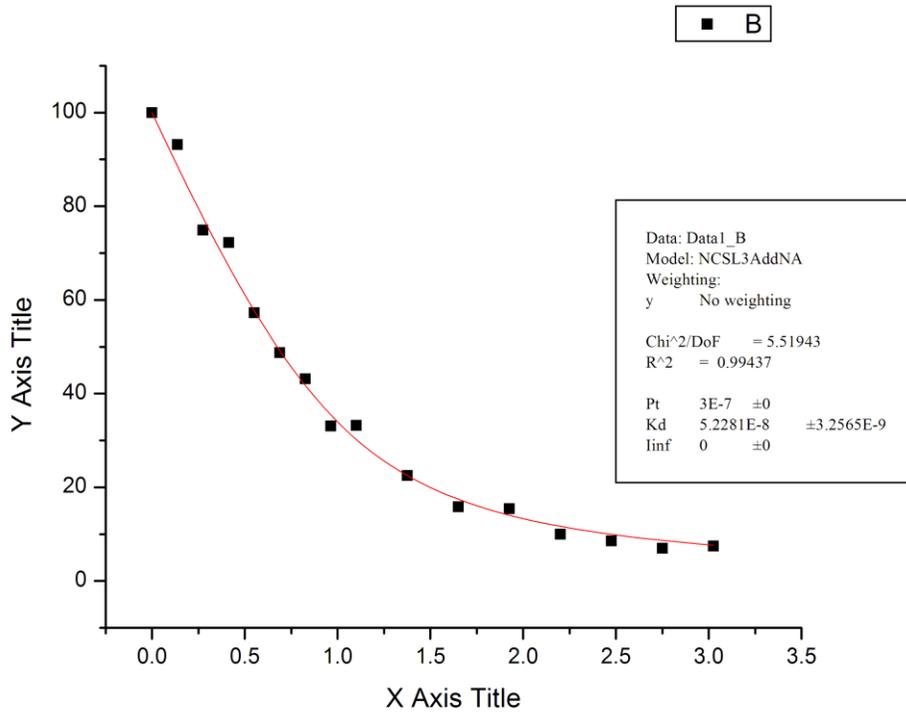
11/22/2010 Q45A



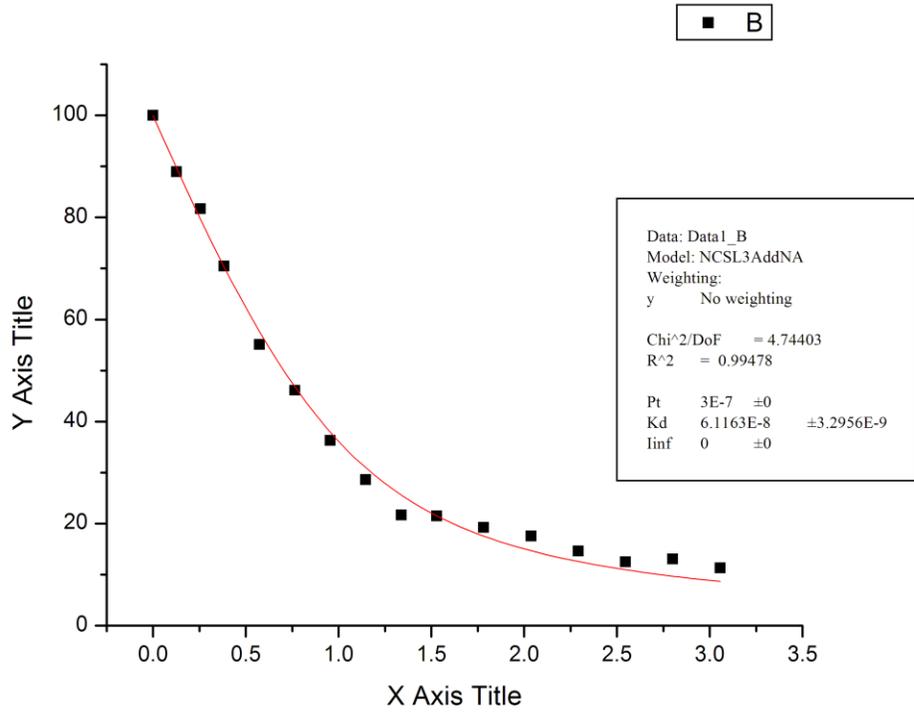
11/23/2010 M46A



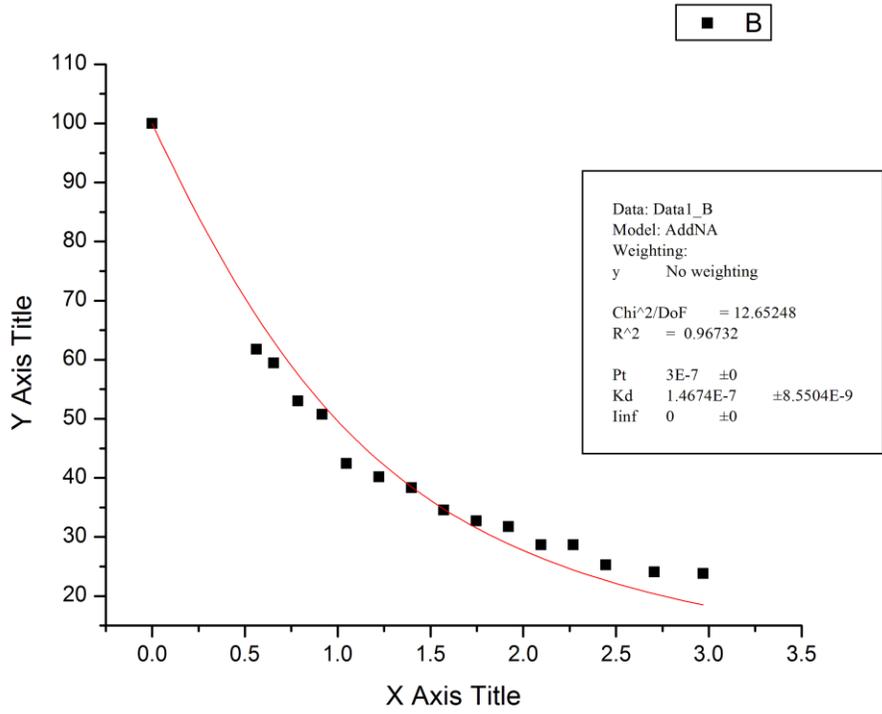
6/8/2011 M46A



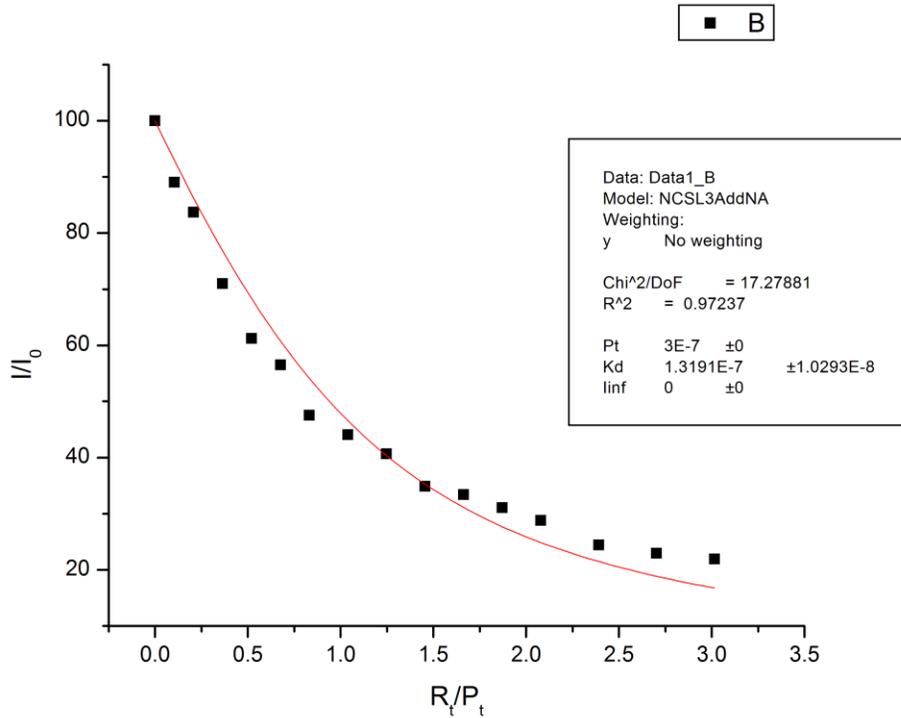
6/9/2011 M46A



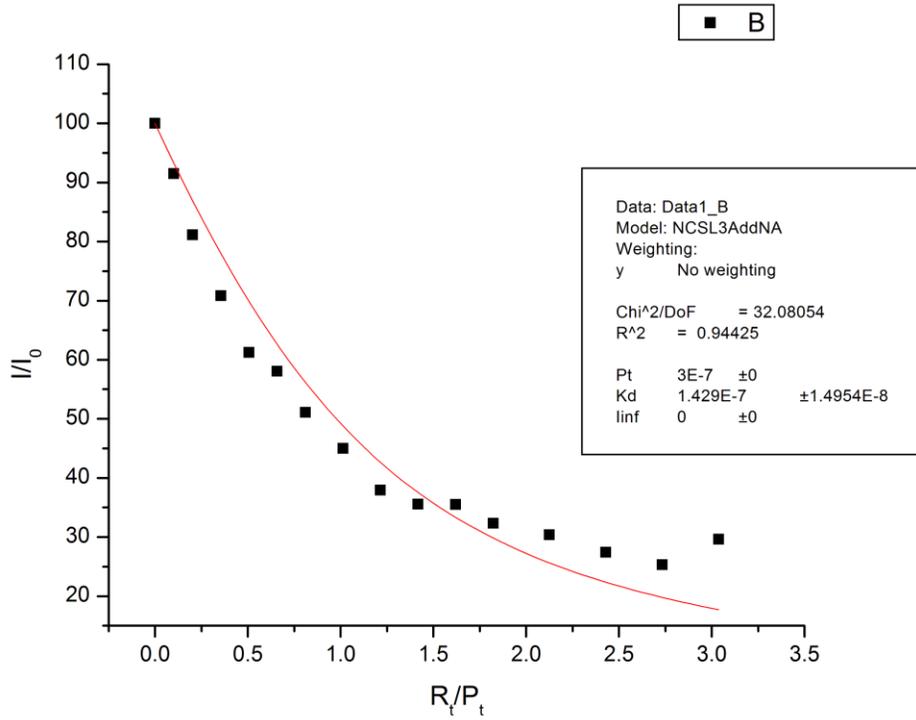
10/12/2010 K14E-E21K



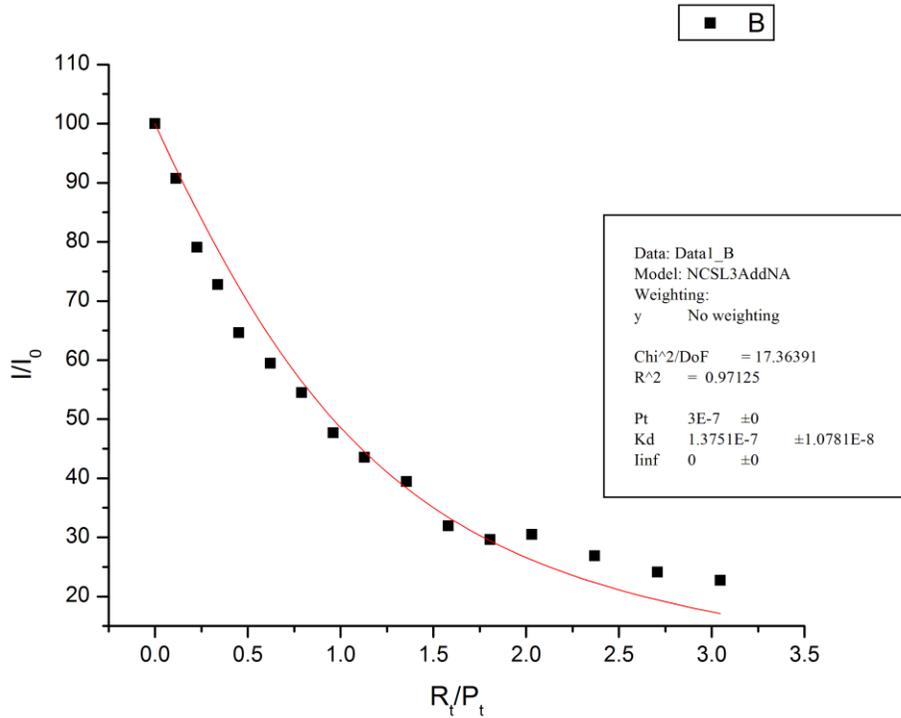
11/1/2010 K14E-E21K



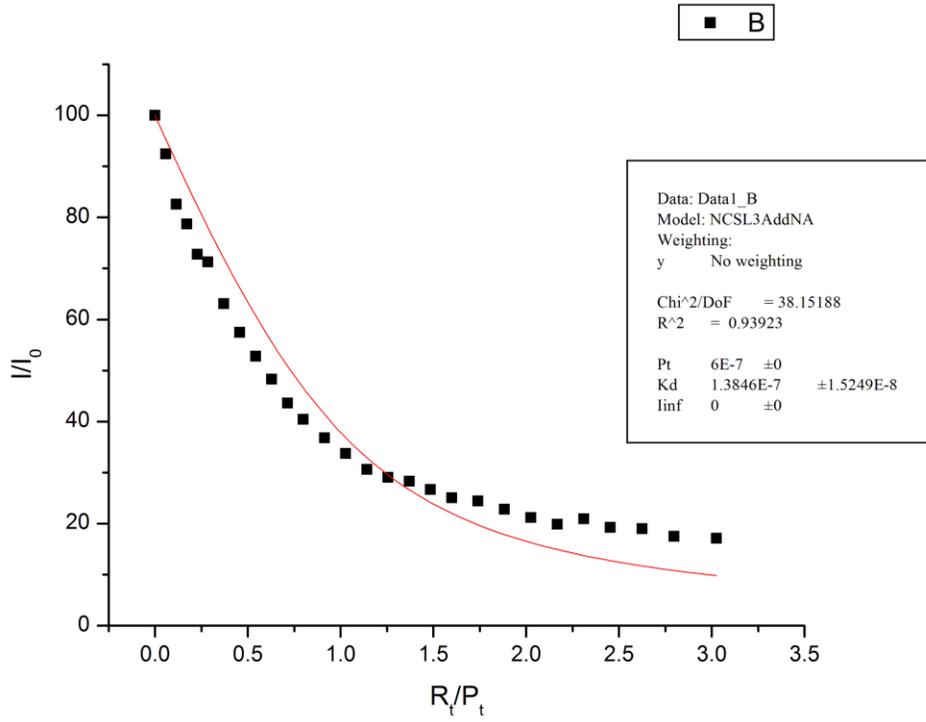
11/4/2010 K14E-E21K



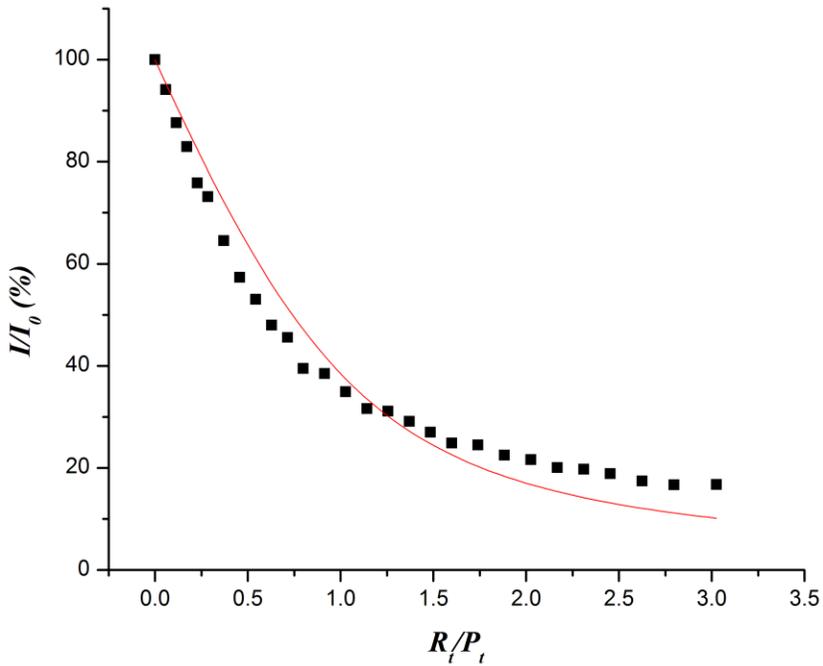
11/22/2010 K14E-E21K

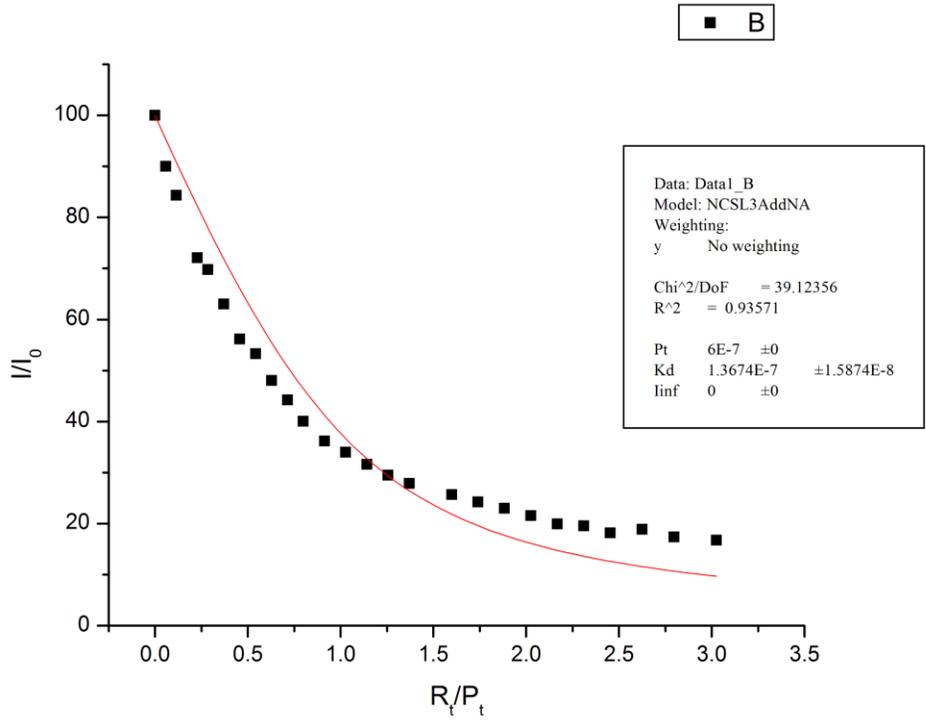


11/29/2010 K14E-E21K-1

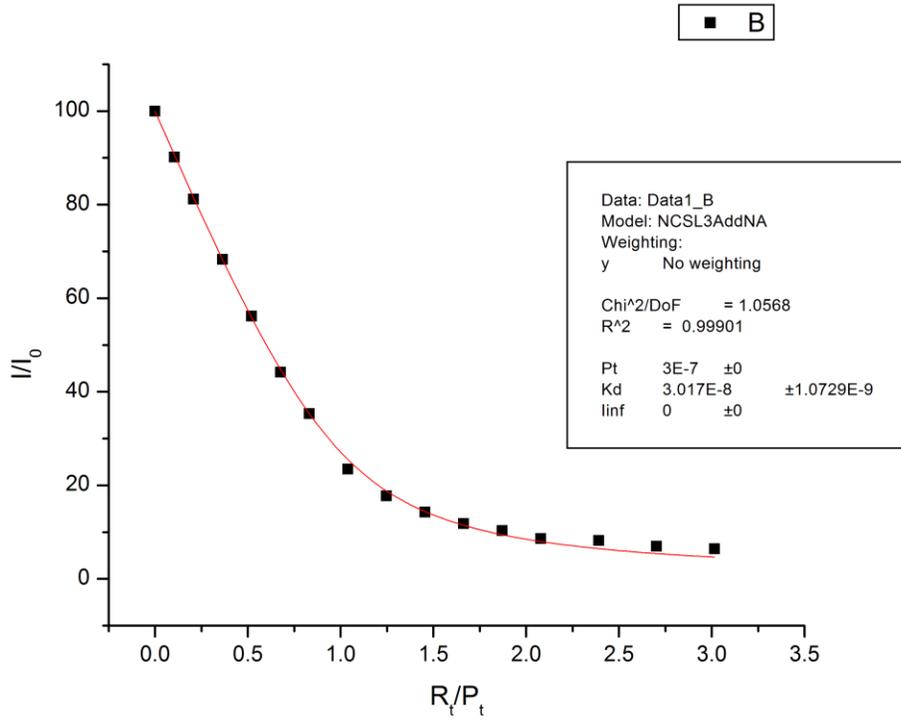


11/29/2010 K14E-E21K-2

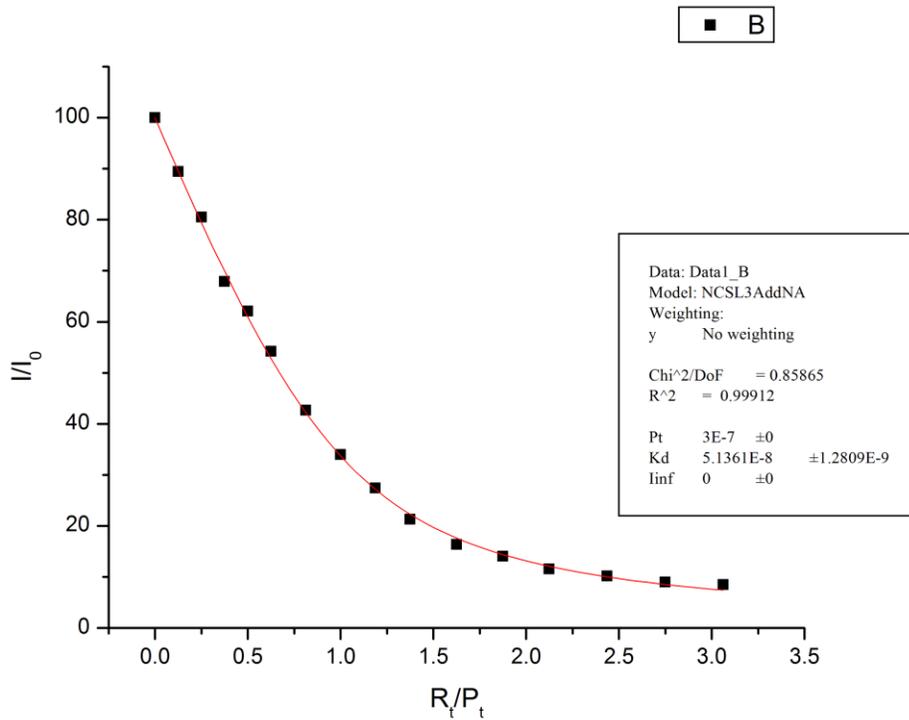




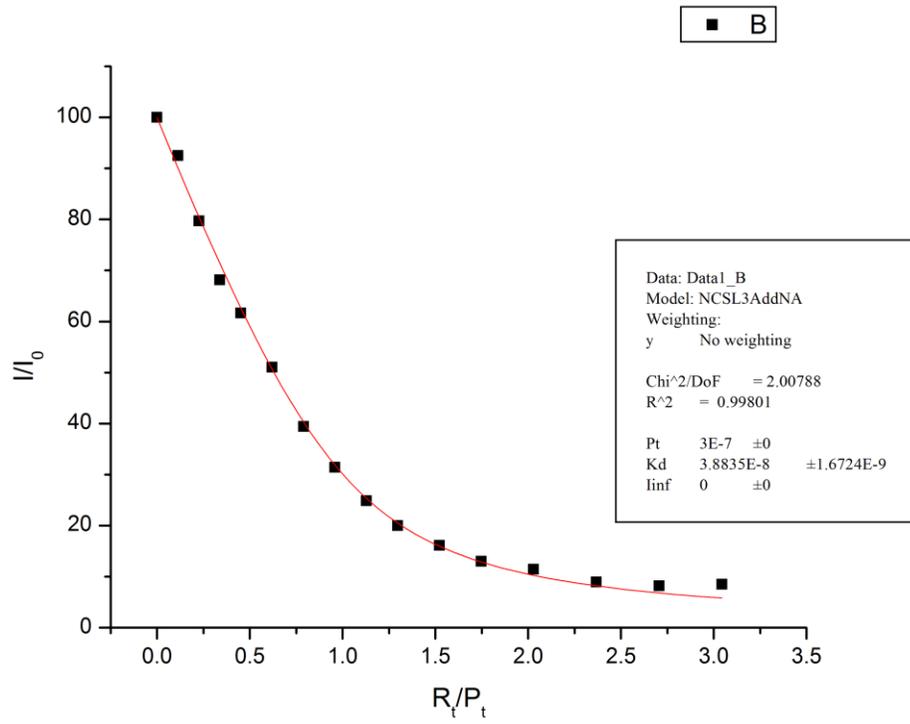
11/1/2010 K33E-E42K



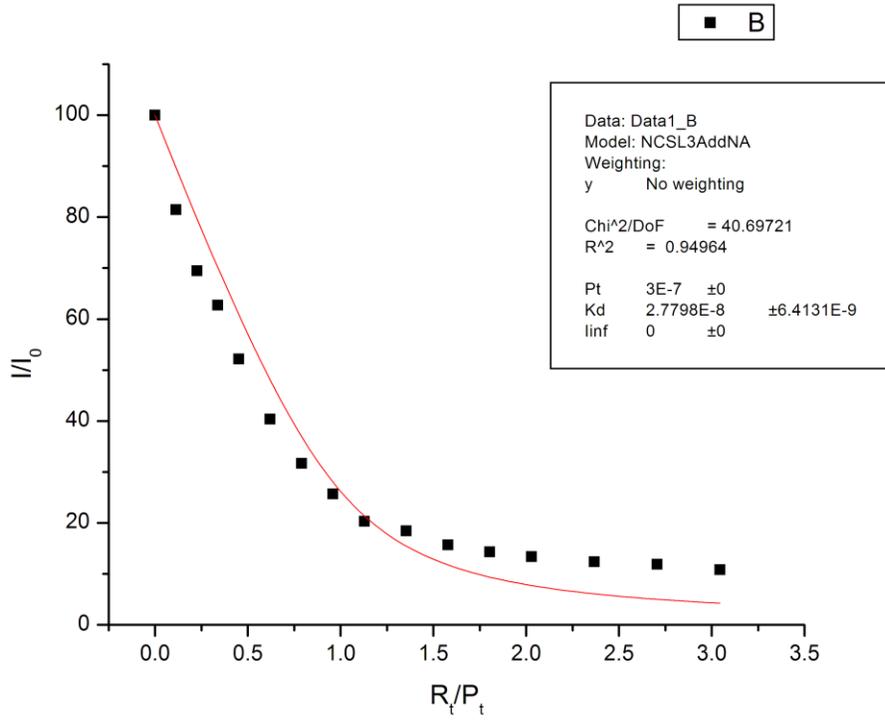
11/18/2010 K33E-E42K



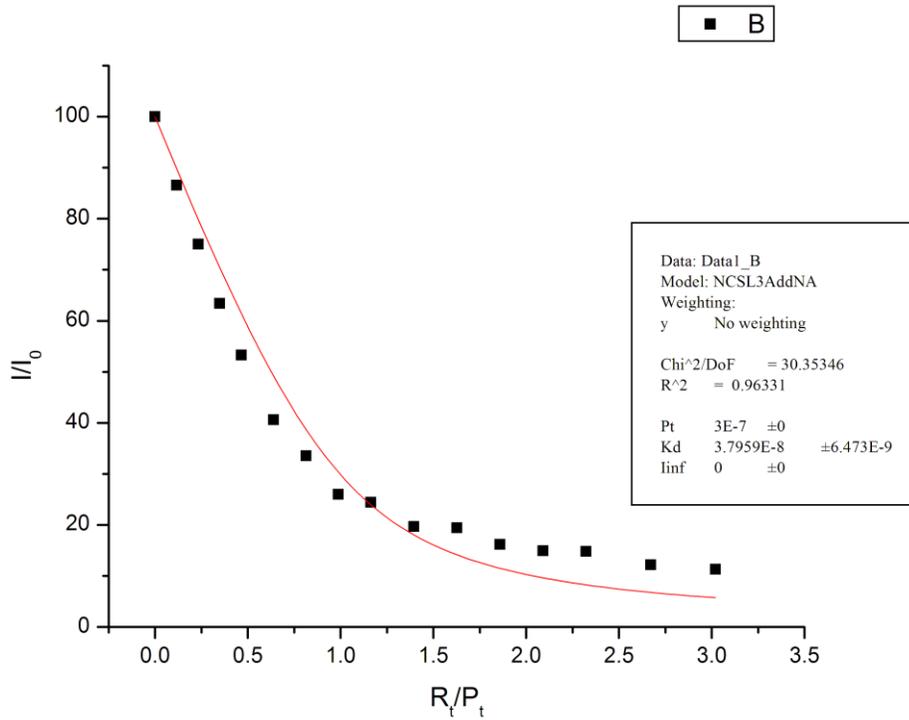
11/22/2010 K33E-E42K



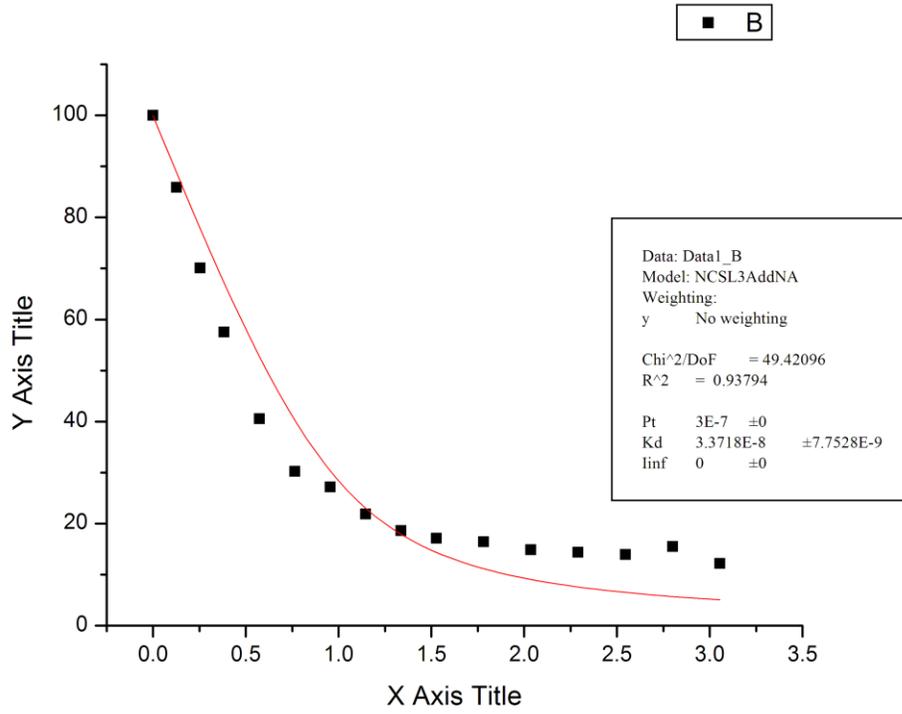
11/7/2010 K38E-E51K



11/23/2010 K38E-E51K

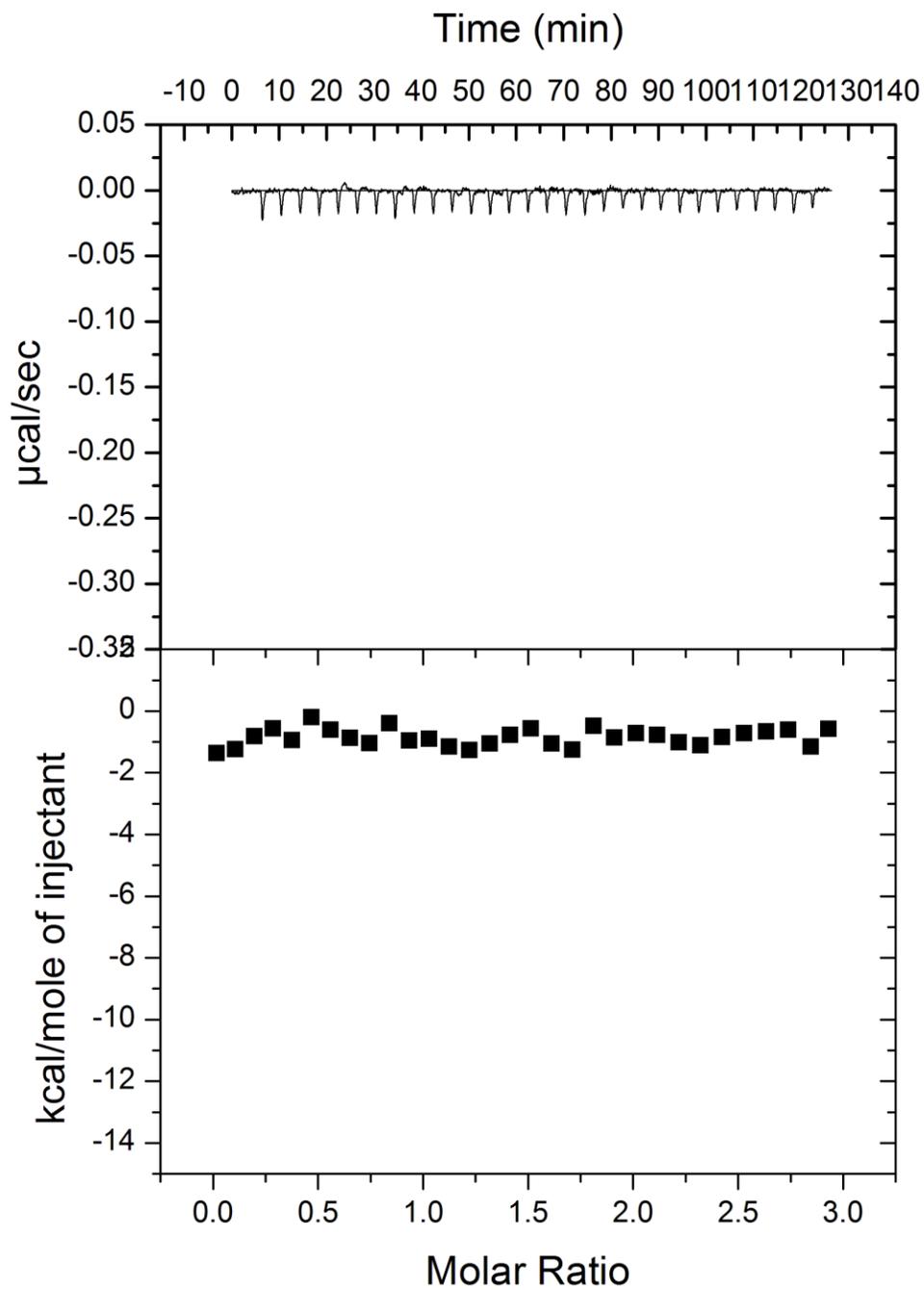


6/9/2011 K38E-E51K

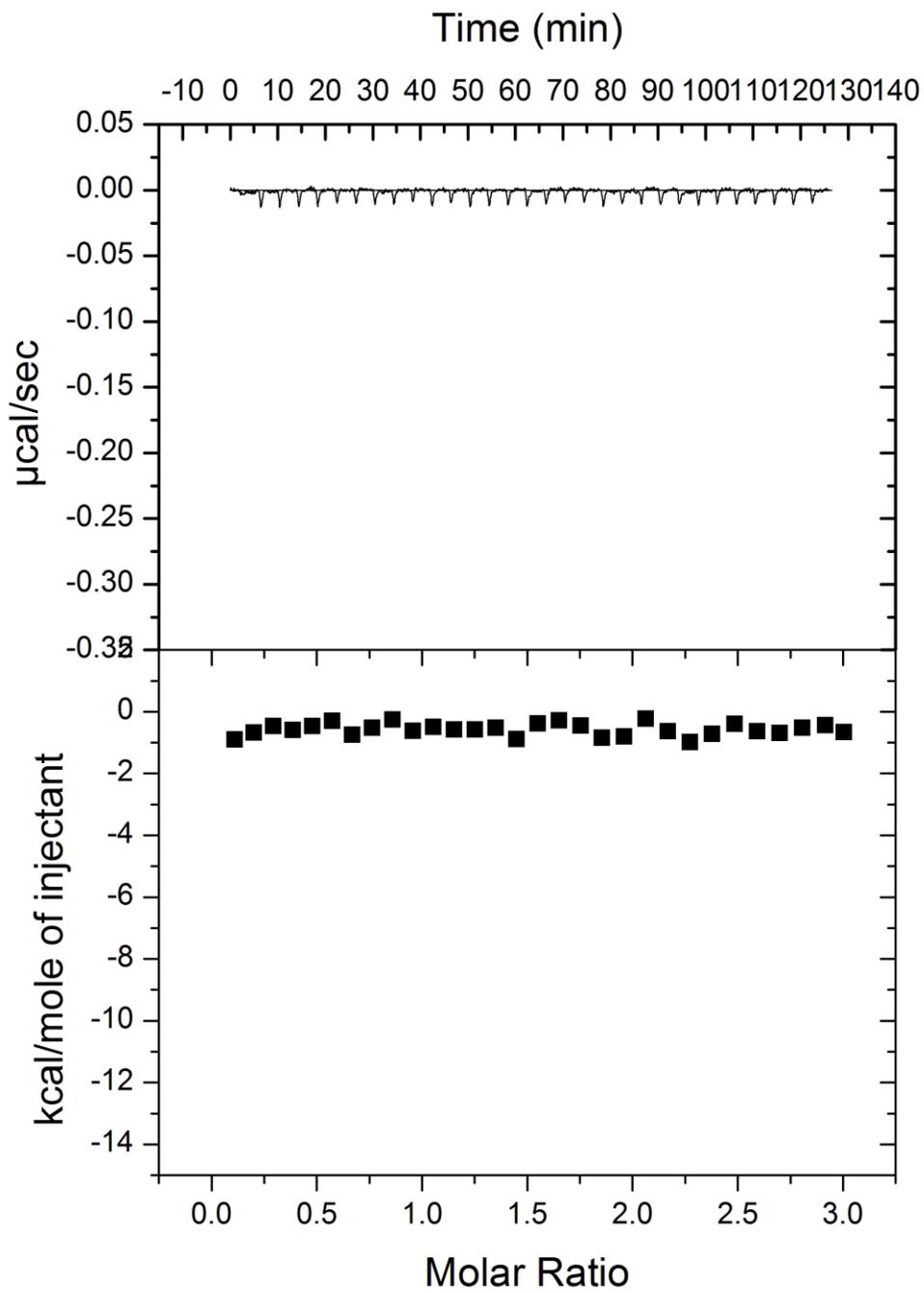


C-3. Individual NCp7-SL3 ITC Curves

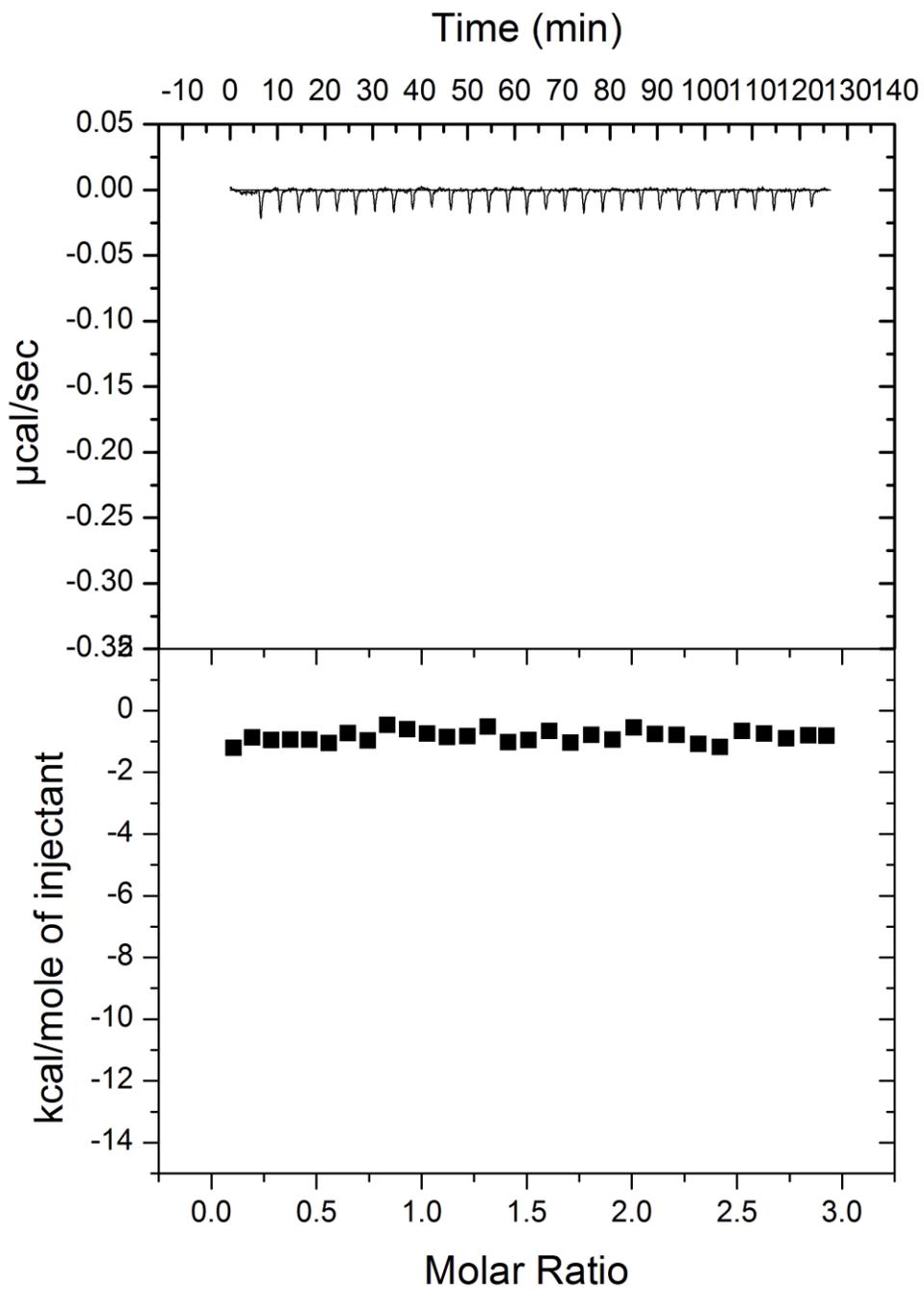
6/9/2011 WT blank titration



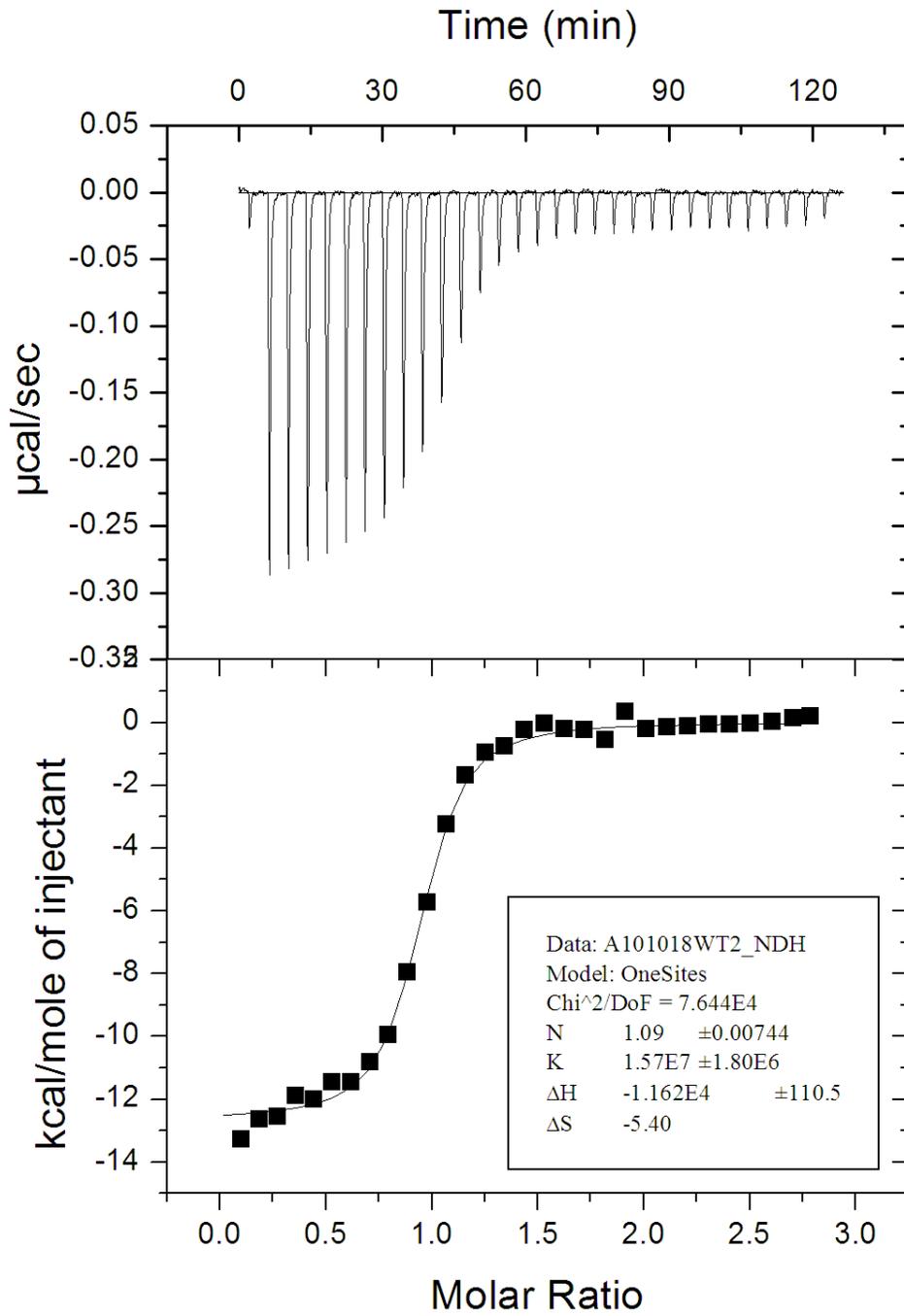
6/9/2011 F16A blank titration



6/12/2011 P31A blank titration

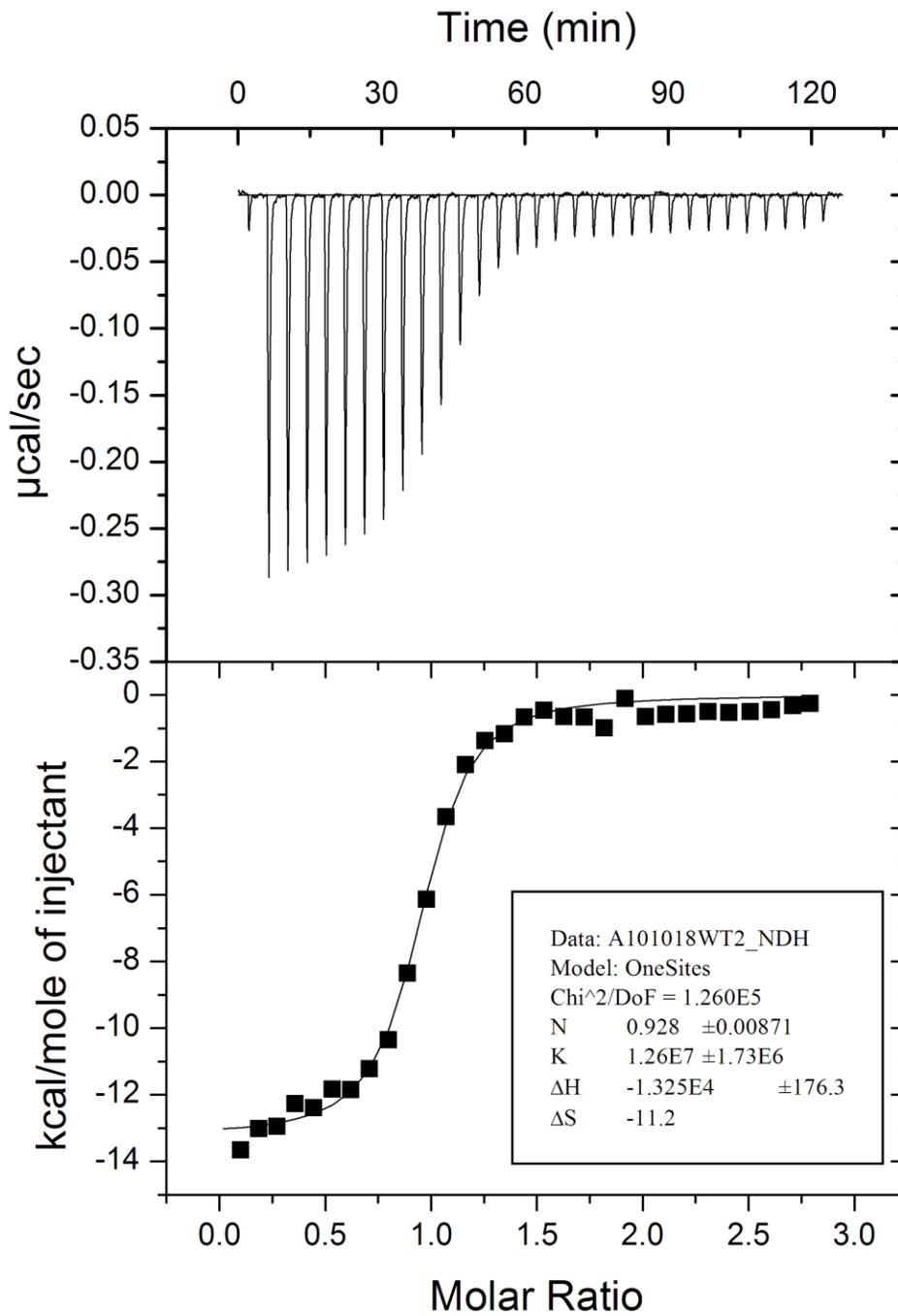


10/18/2010 WT (Use endpoints data as baseline)



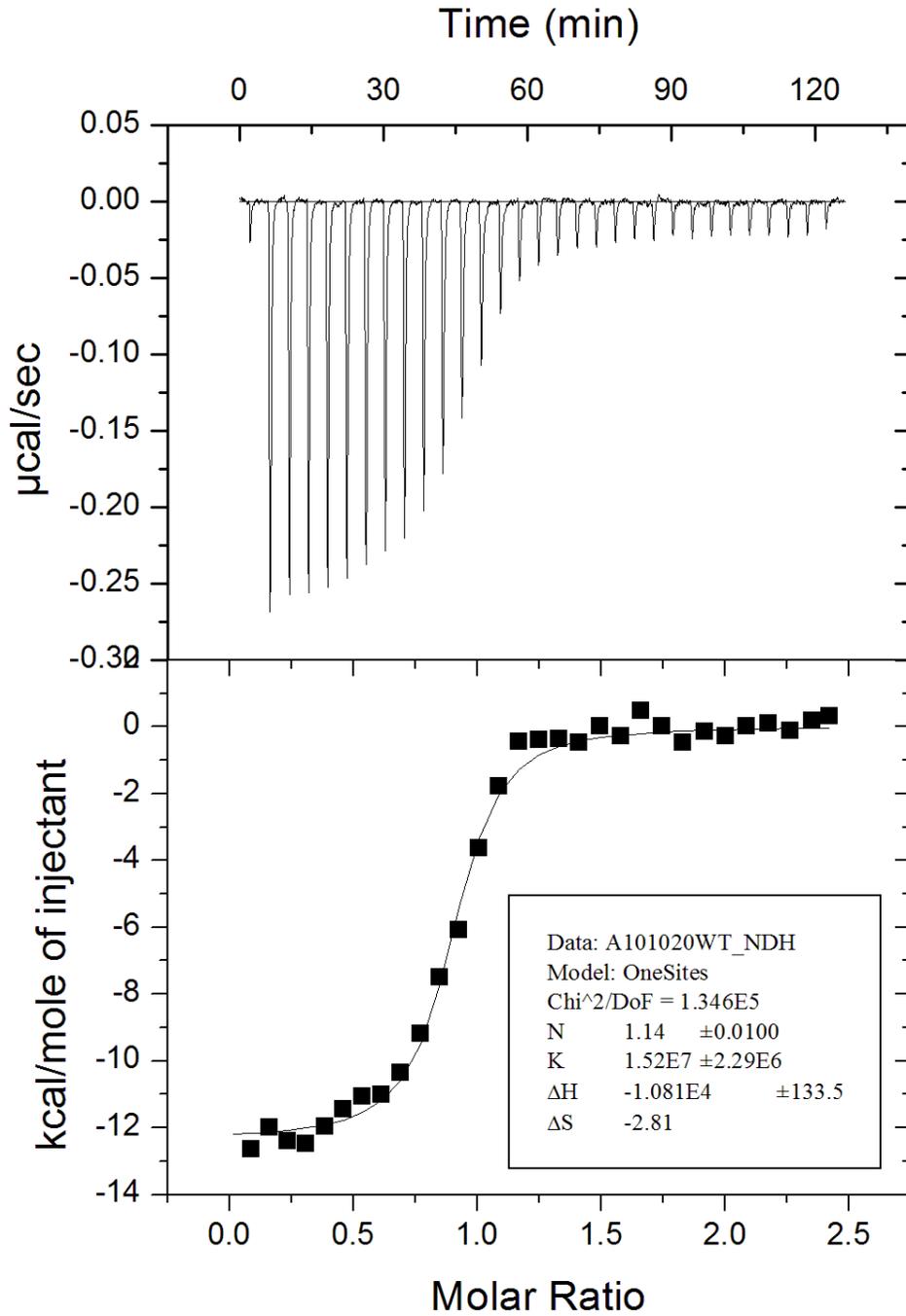
Sample (SL3) = $5.06 \mu\text{M}$, Titrant (NC WT) = $59.69 \mu\text{M}$

10/18/2010 WT (Use blank titration data as baseline)



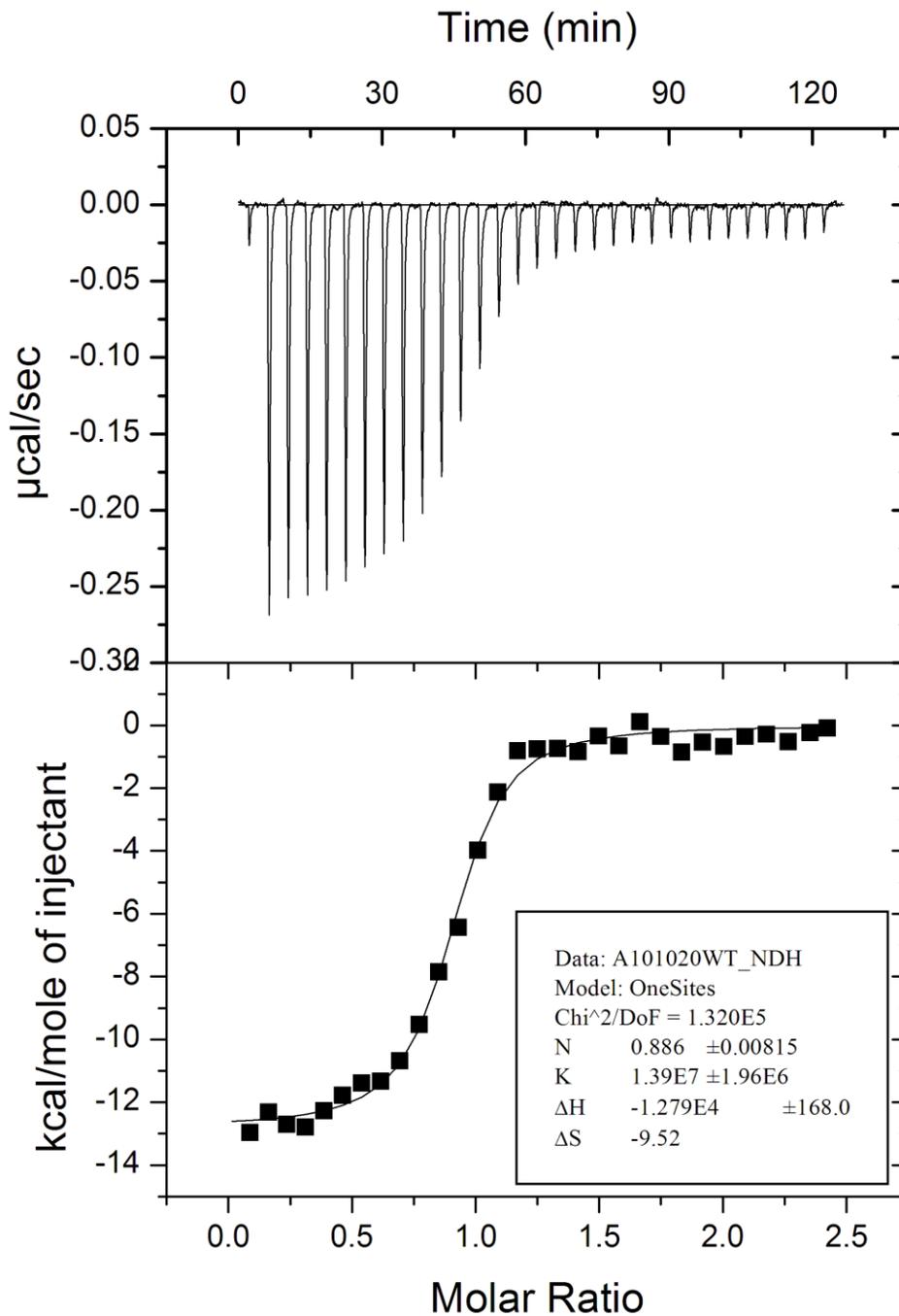
Sample (SL3) = $5.06 \mu\text{M}$, Titrant (NC WT) = $59.69 \mu\text{M}$

10/20/2010 WT (Use endpoints data as baseline)



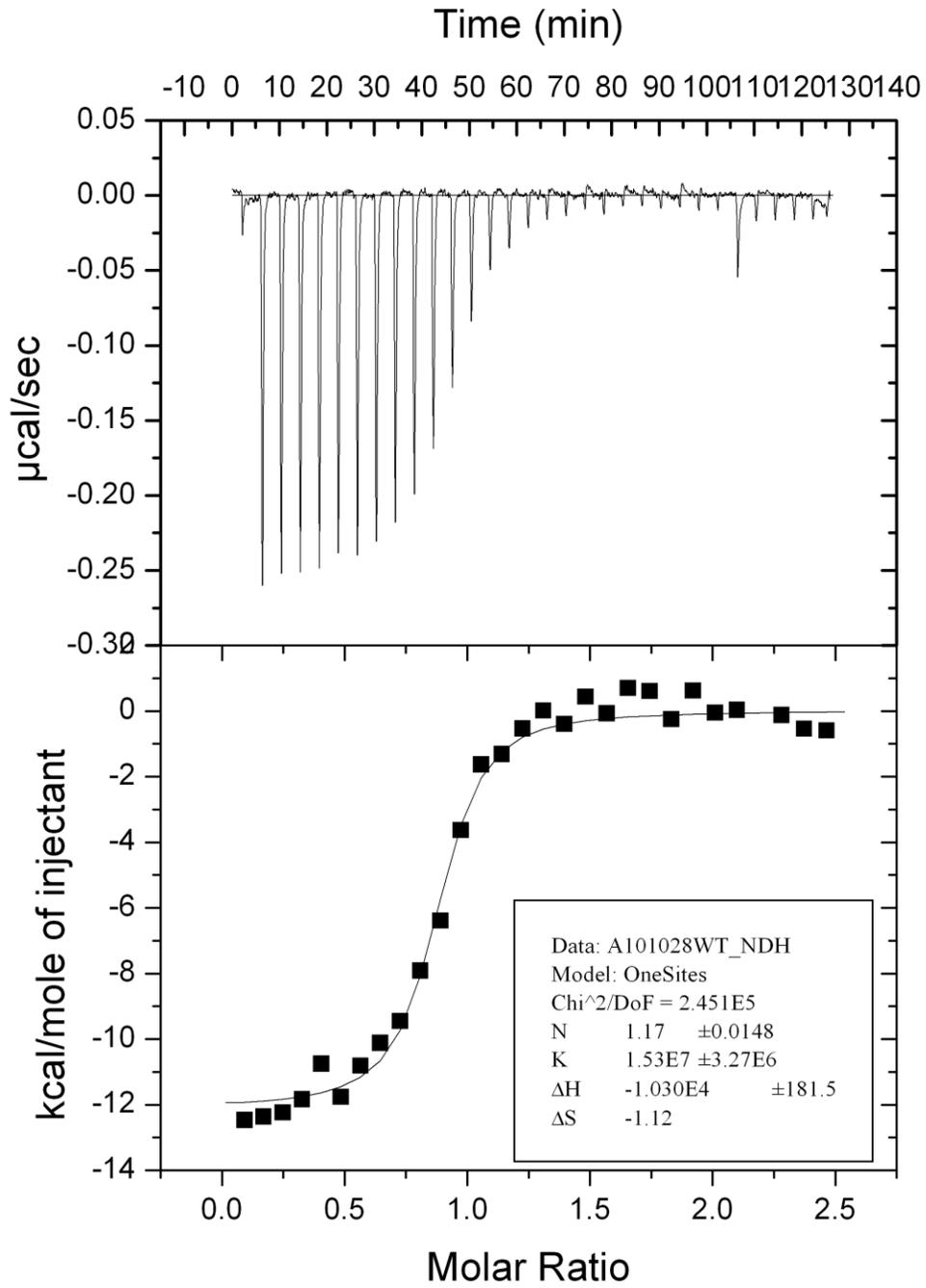
Sample (SL3) = 5.37 μM , Titrant (NC WT) = 54.96 μM

10/20/2010 WT (Use blank titration data as baseline)



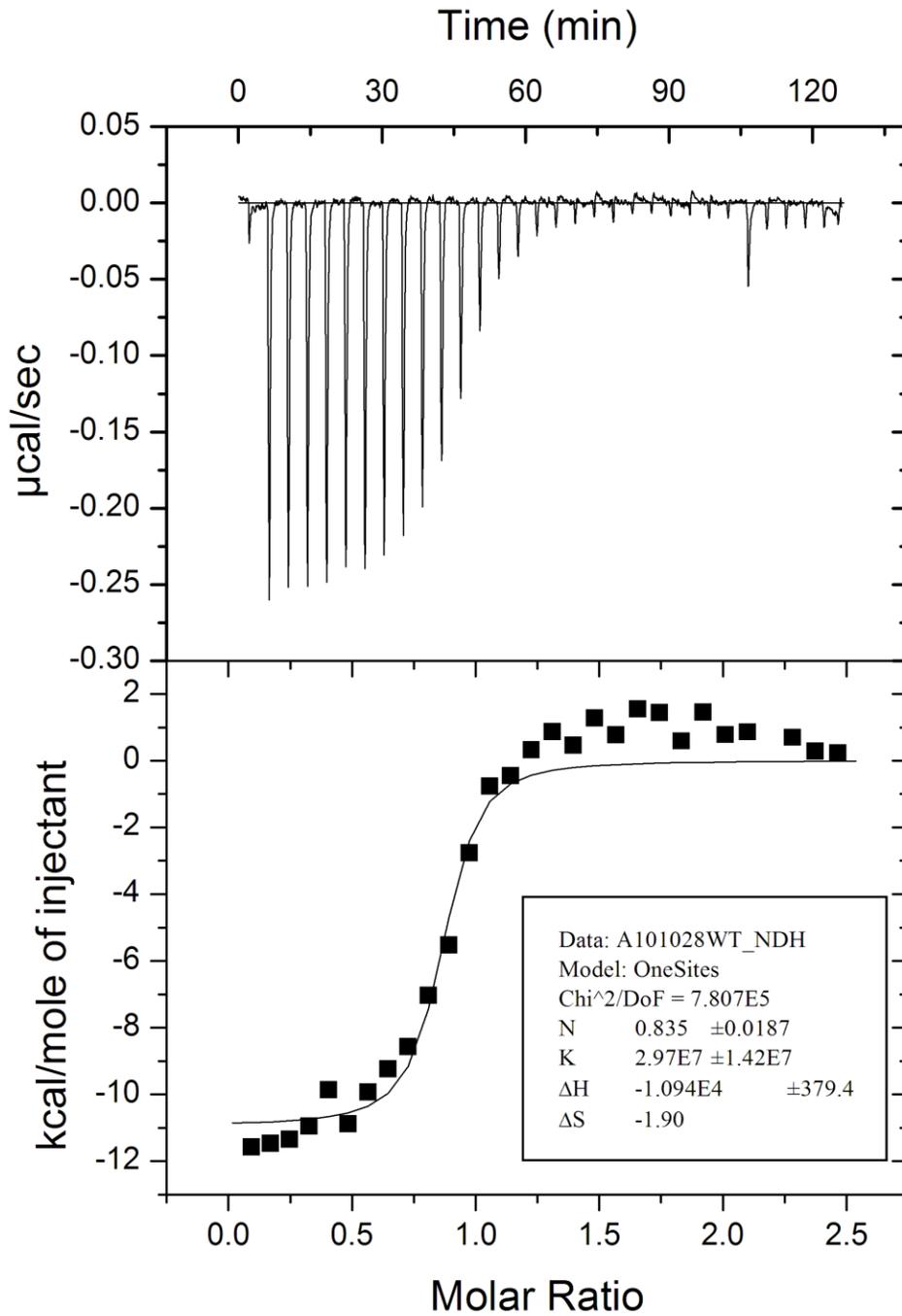
Sample (SL3) = 5.37 μM , Titrant (NC WT) = 54.96 μM

10/28/2010 WT (Use endpoints data as baseline)



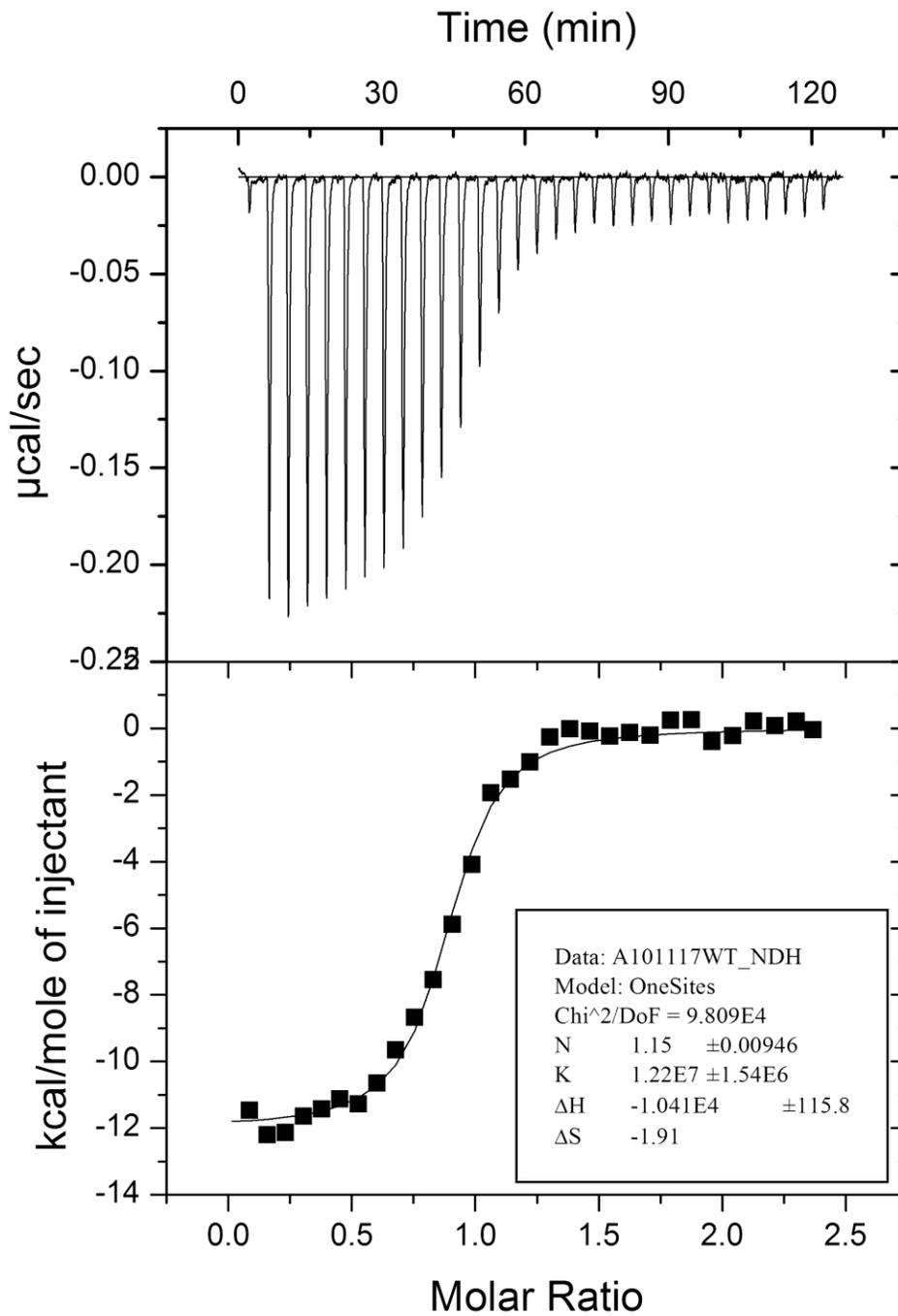
Sample (SL3) = 5.32 μM , Titrant (NC WT) = 57.02 μM

10/28/2010 WT (Use blank titration data as baseline)



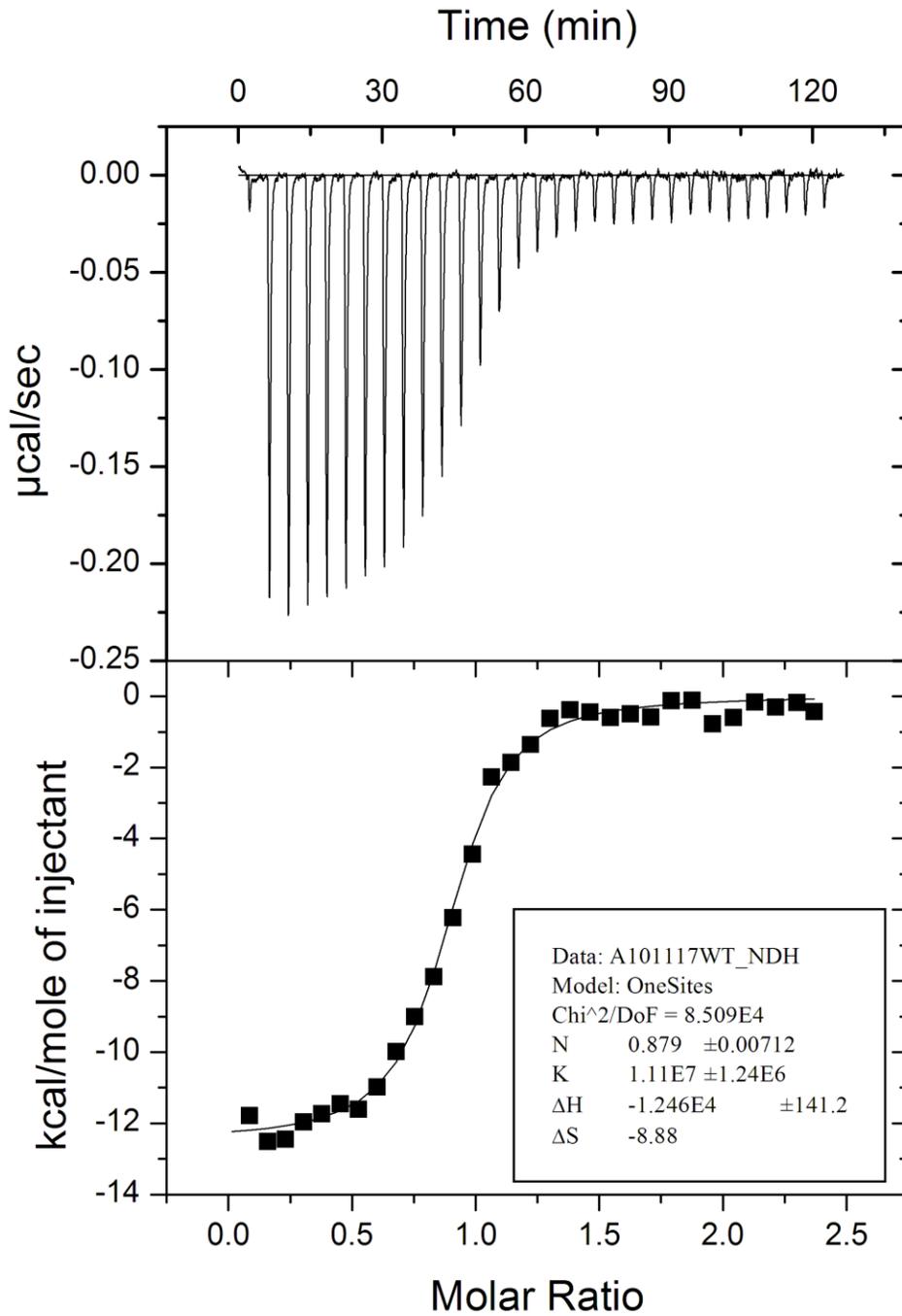
Sample (SL3) = 5.32 μM , Titrant (NC WT) = 57.02 μM

11/17/2010 WT (Use endpoints data as baseline)



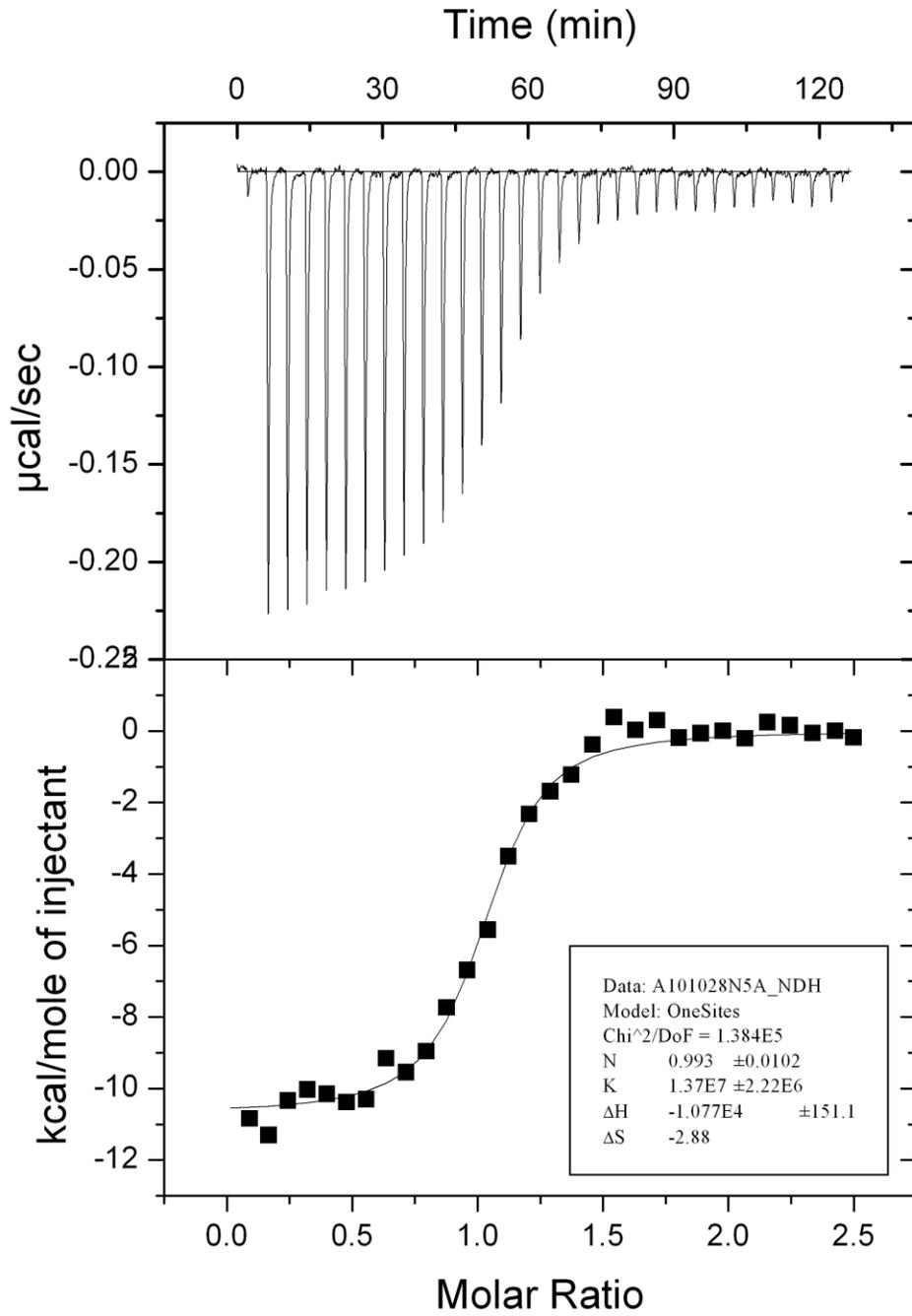
Sample (SL3) = 5.41 μM , Titrant (WT) = 54.10 μM

11/17/2010 WT (Use blank titration data as baseline)

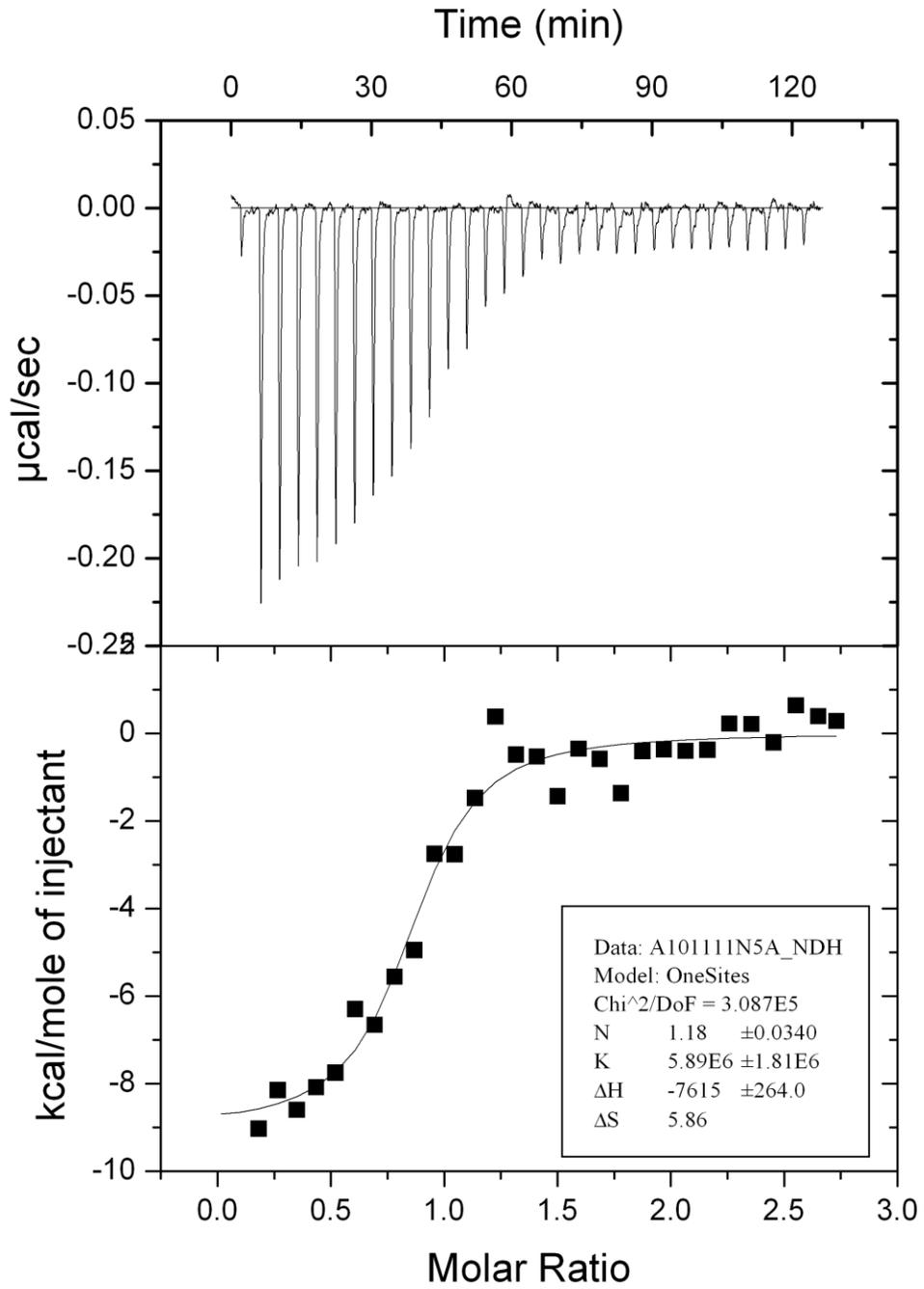


Sample (SL3) = 5.41 μM , Titrant (WT) = 54.10 μM

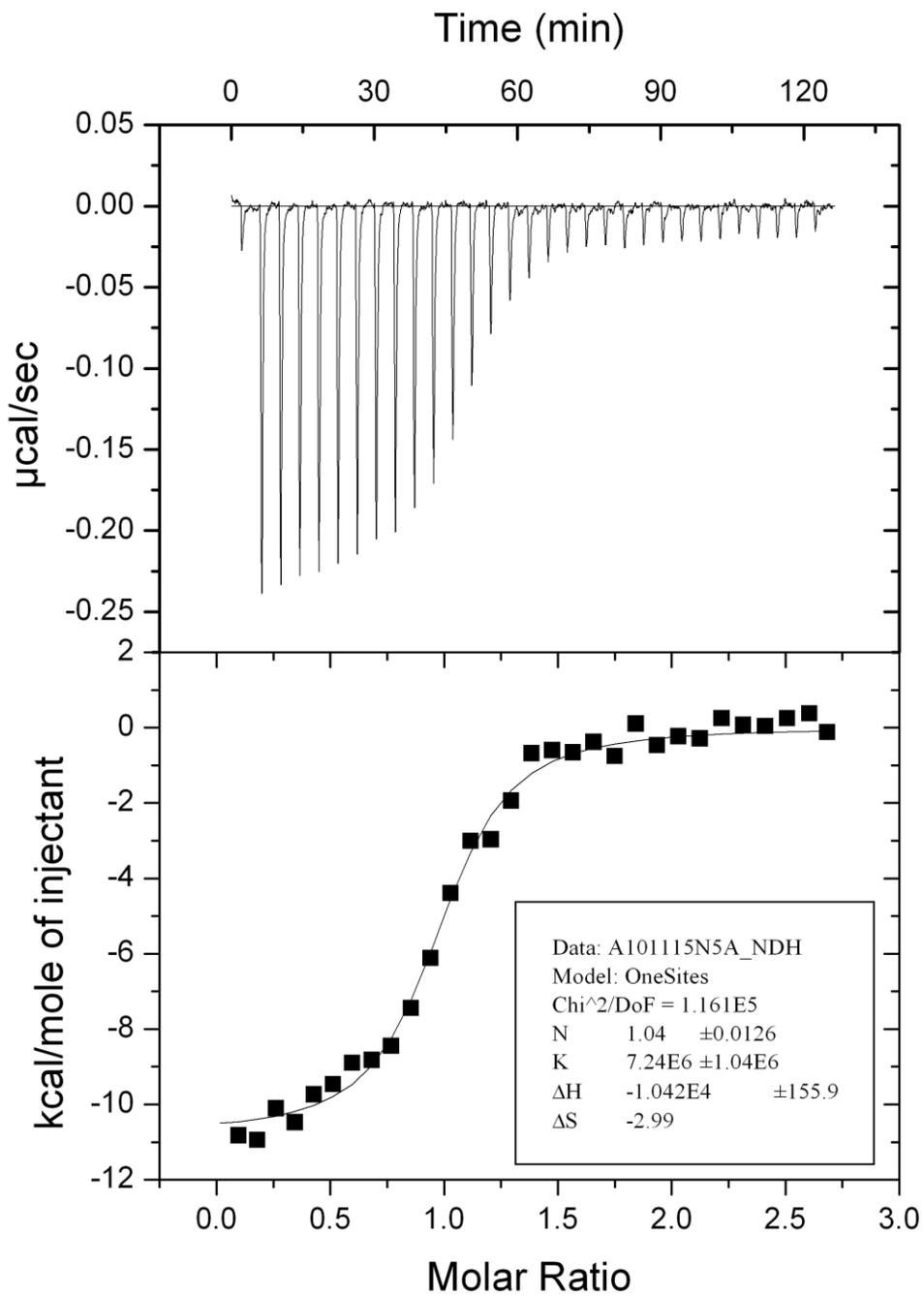
10/28/2010 N5A



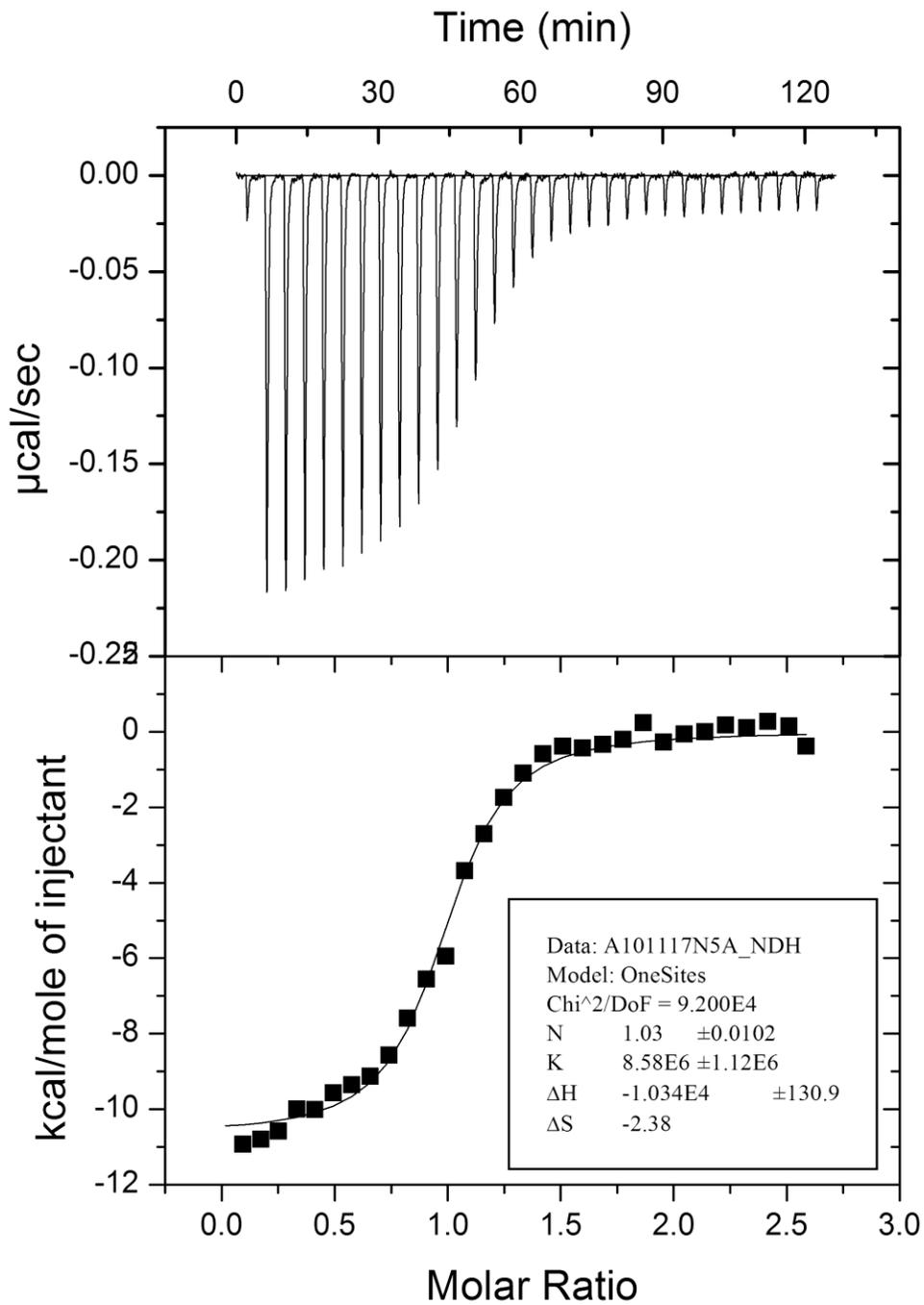
Sample (SL3) = 5.19 μM , Titrant (N5A) = 54.75 μM



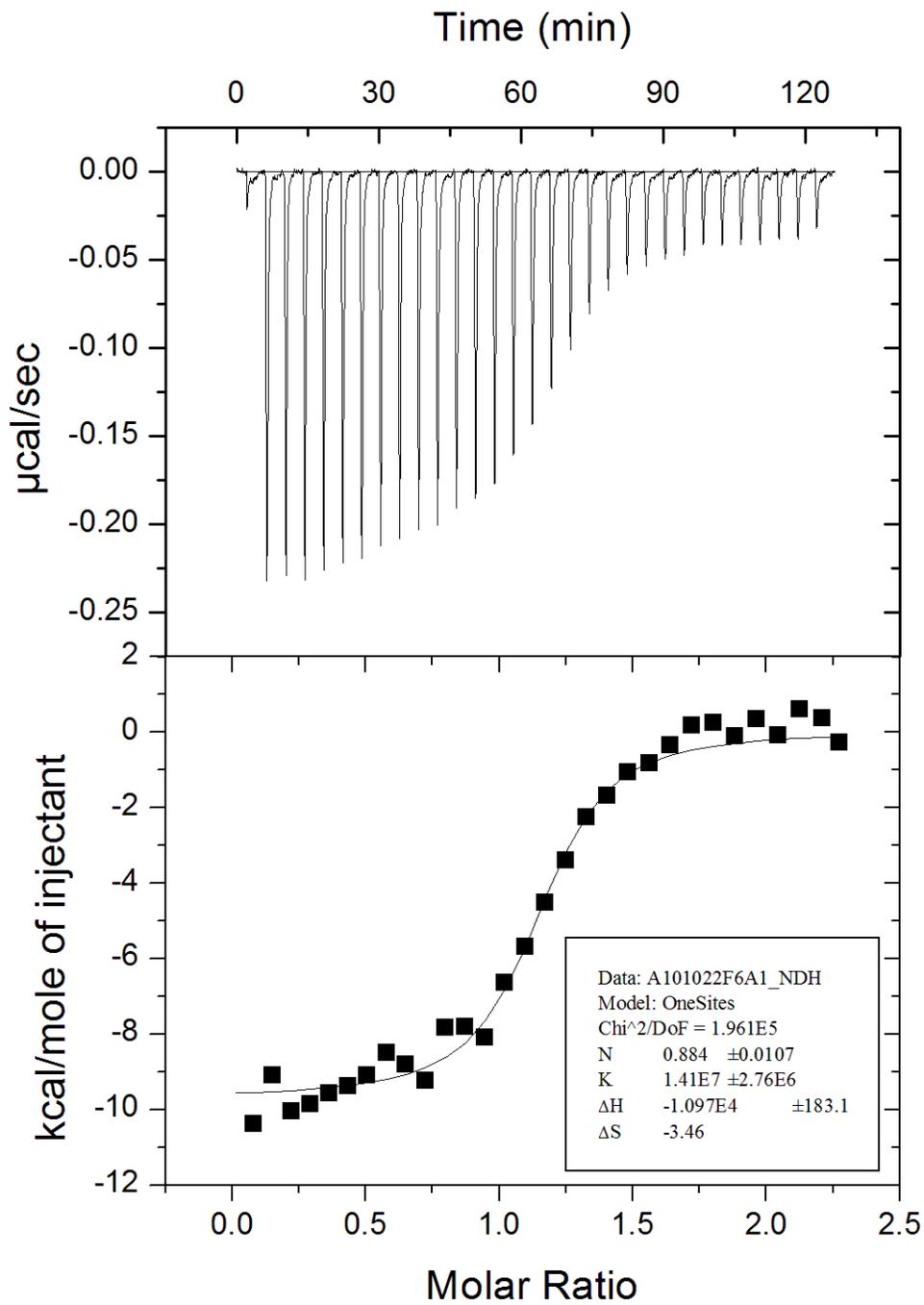
Sample (SL3) = 4.99 µM, Titrant (N5A) = 57.54 µM



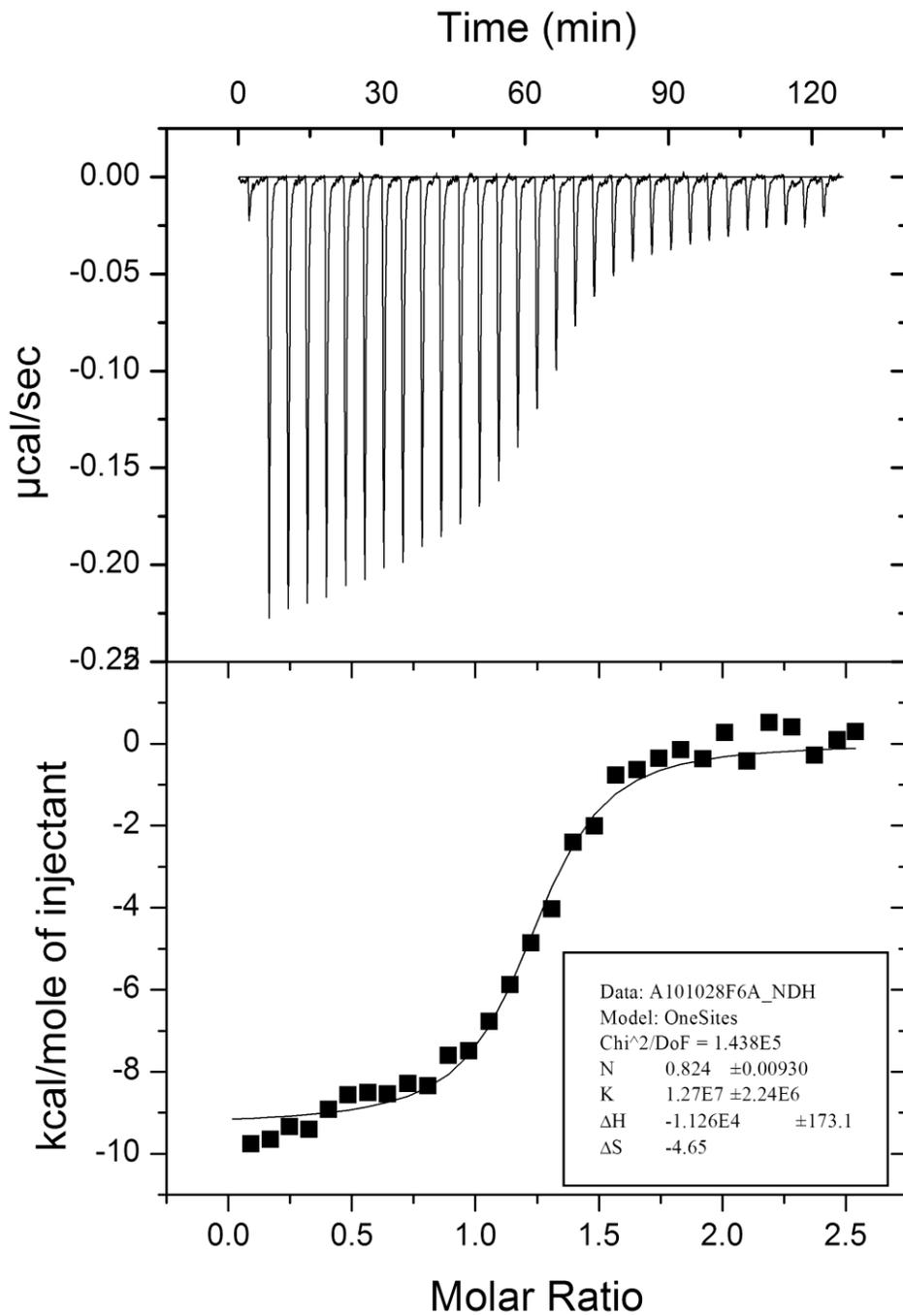
Sample (SL3) = 5.09 µM, Titrant (N5A) = 57.67 µM



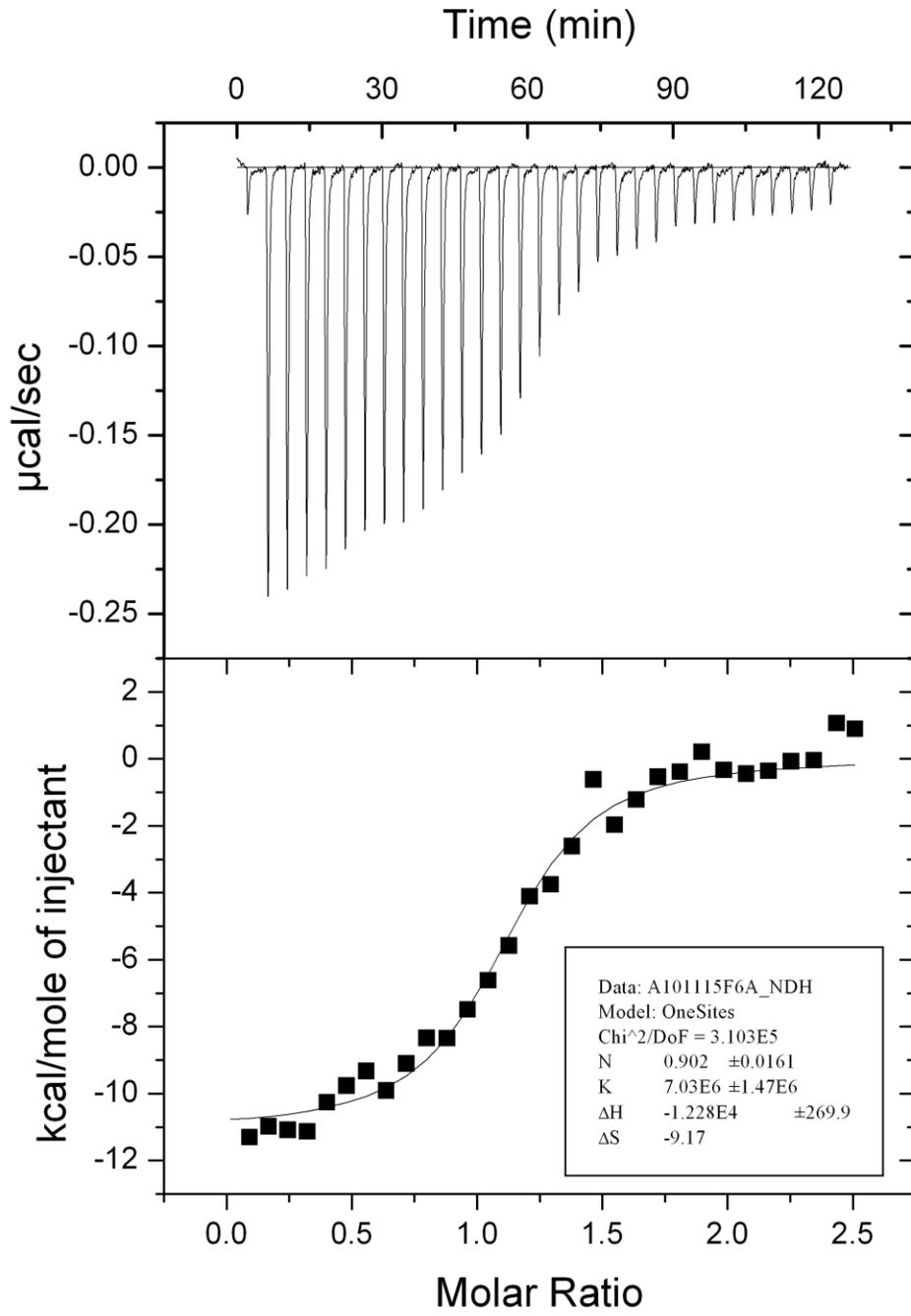
Sample (SL3) = $5.46 \mu\text{M}$, Titrant (WT) = $59.57 \mu\text{M}$



Sample (SL3) = 5.27 μM , Titrant (F6A) = 50.66 μM

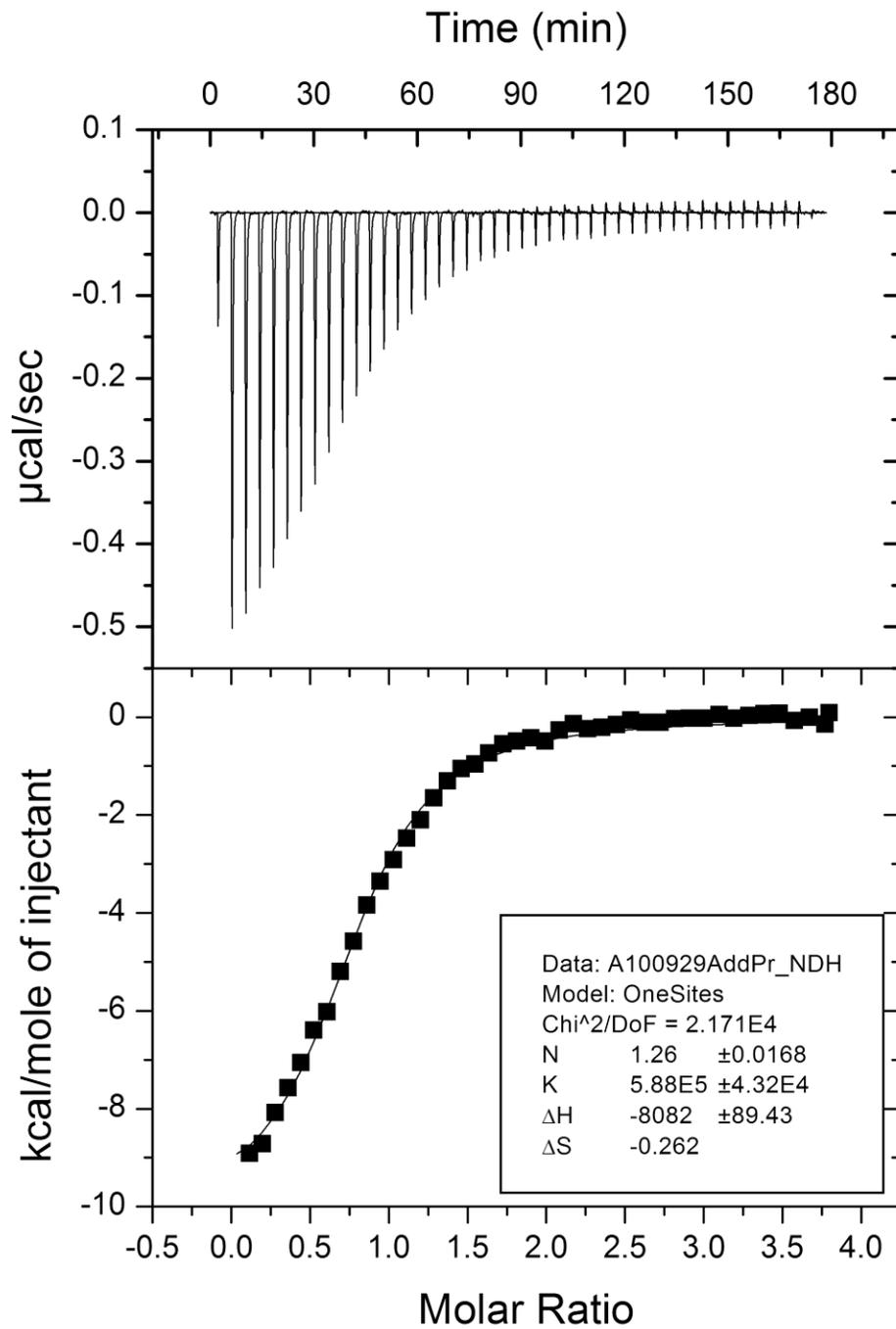


Sample (SL3) = $5.40 \mu\text{M}$, Titrant (F6A) = $57.92 \mu\text{M}$



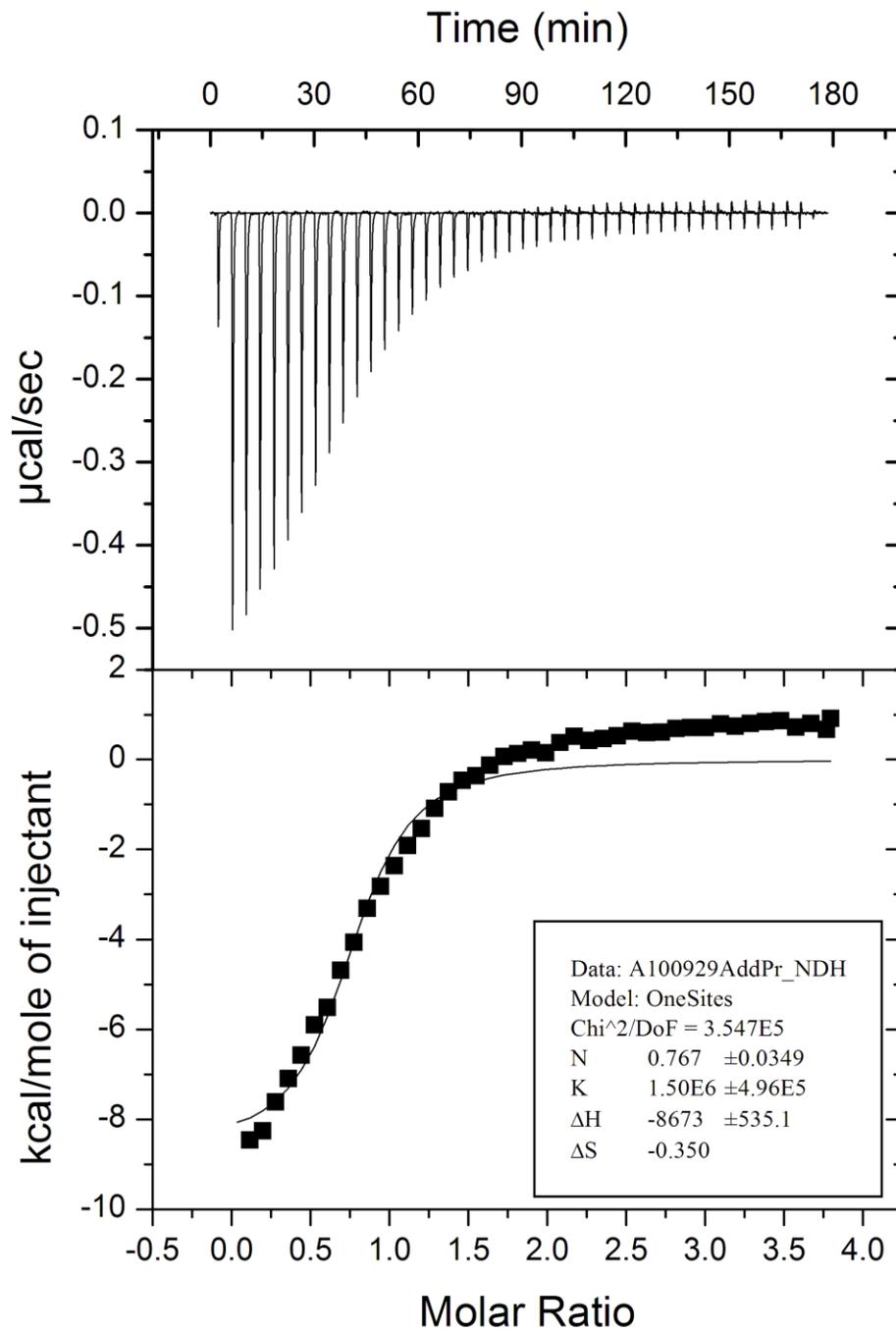
Sample (SL3) = 5.08 μM , Titrant (F6A) = 53.81 μM

9/29/2010 F16A (Use endpoints data as baseline)



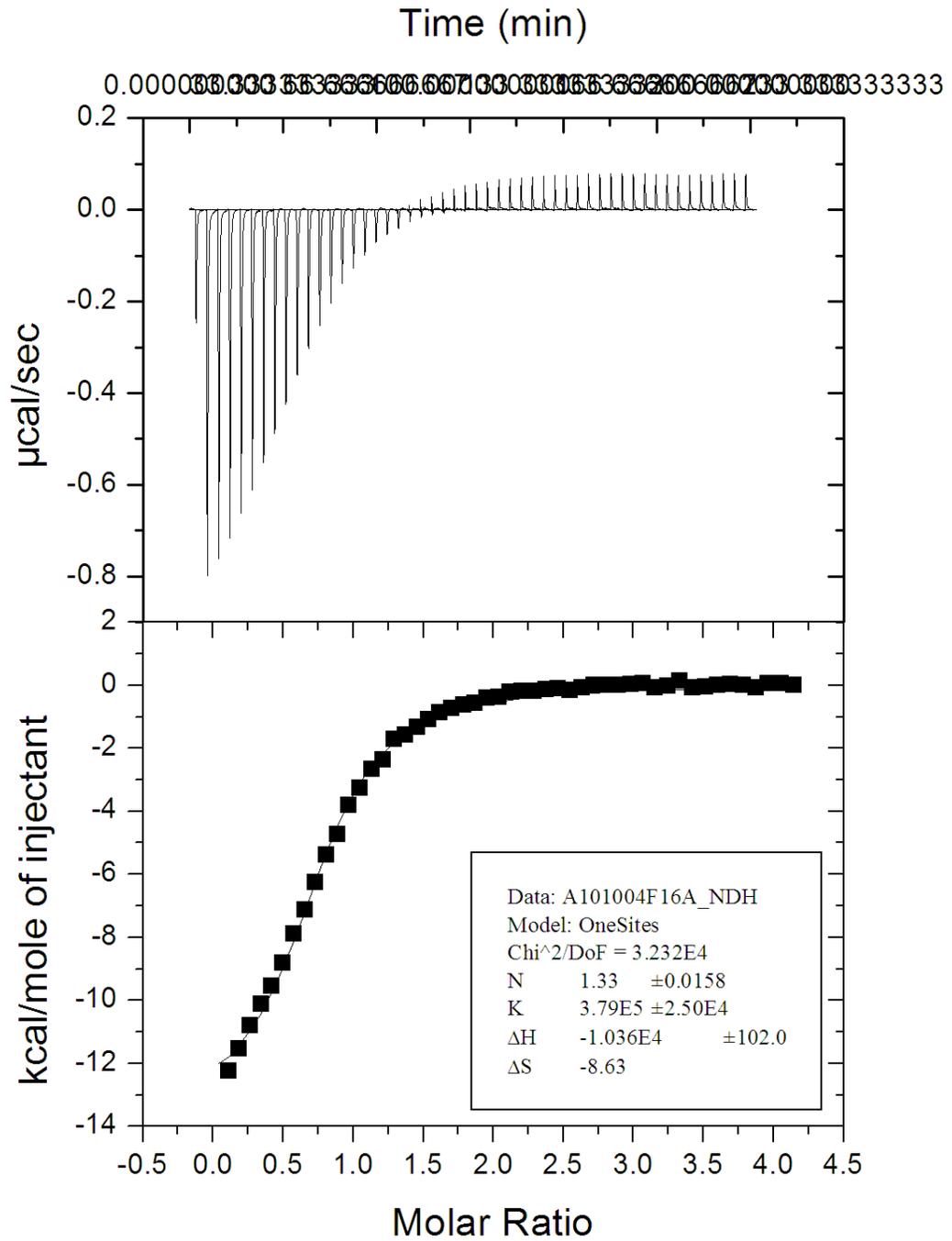
Sample (SL3) = 12.16 μM , Titrant (F16A) = 195.7 μM

9/29/2010 F16A (Use blank titration data as baseline)



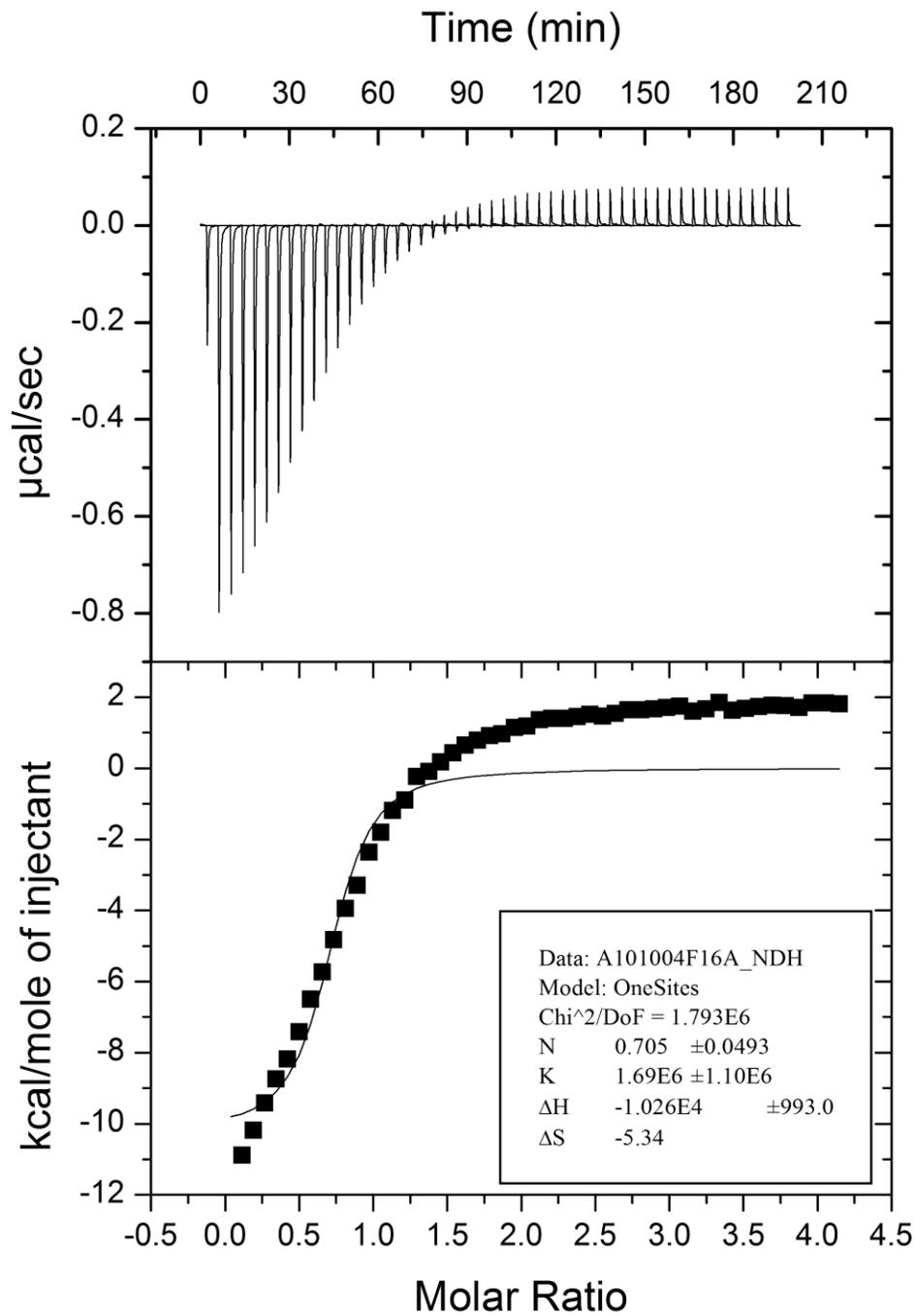
Sample (SL3) = 12.16 μM , Titrant (F16A) = 195.7 μM

10/4/2010 F16A (Use endpoints data as baseline)



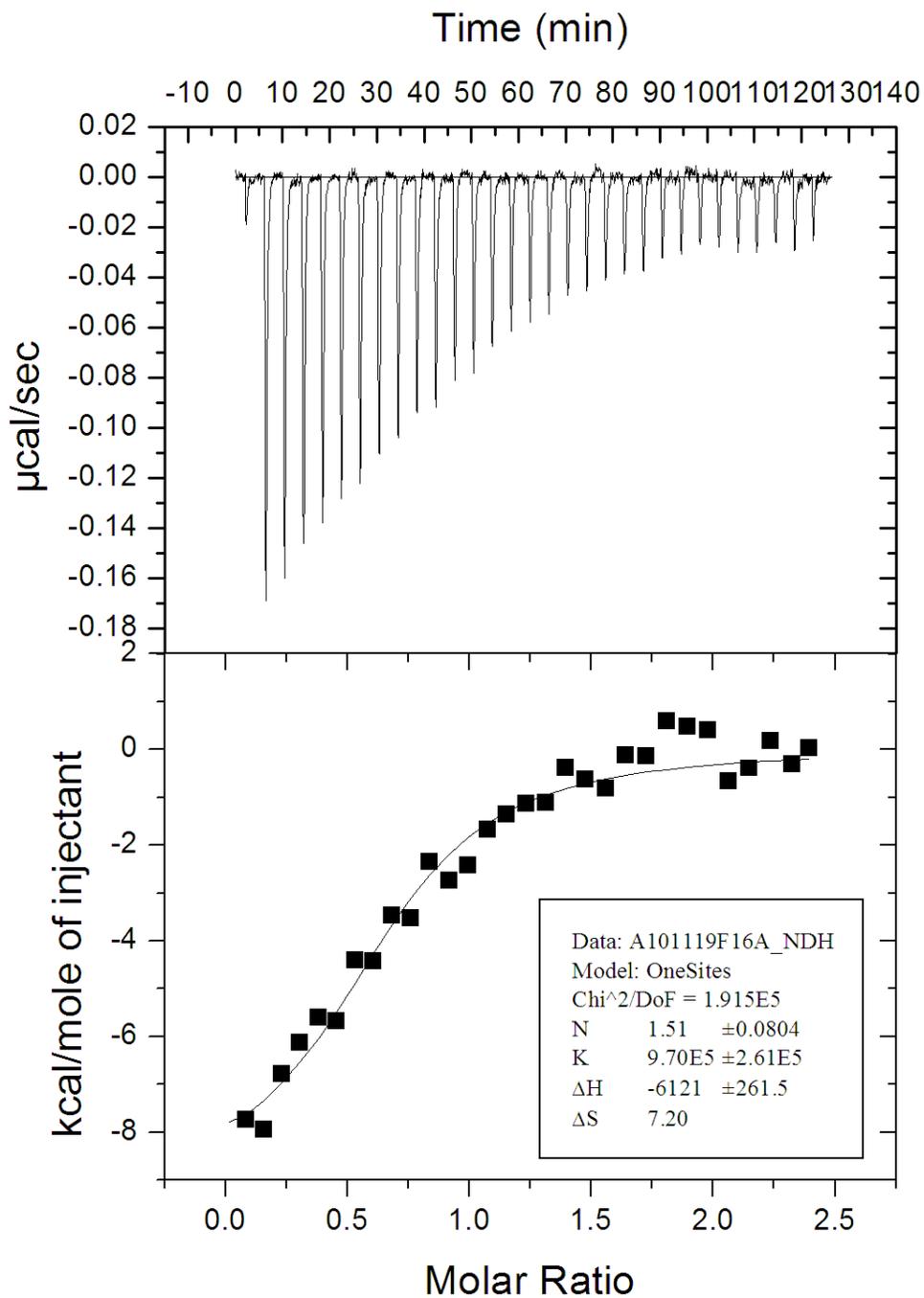
Sample (SL3) = 19.22 μM , Titrant (F16A) = 340.6 μM

10/4/2010 F16A (Use blank titration data as baseline)



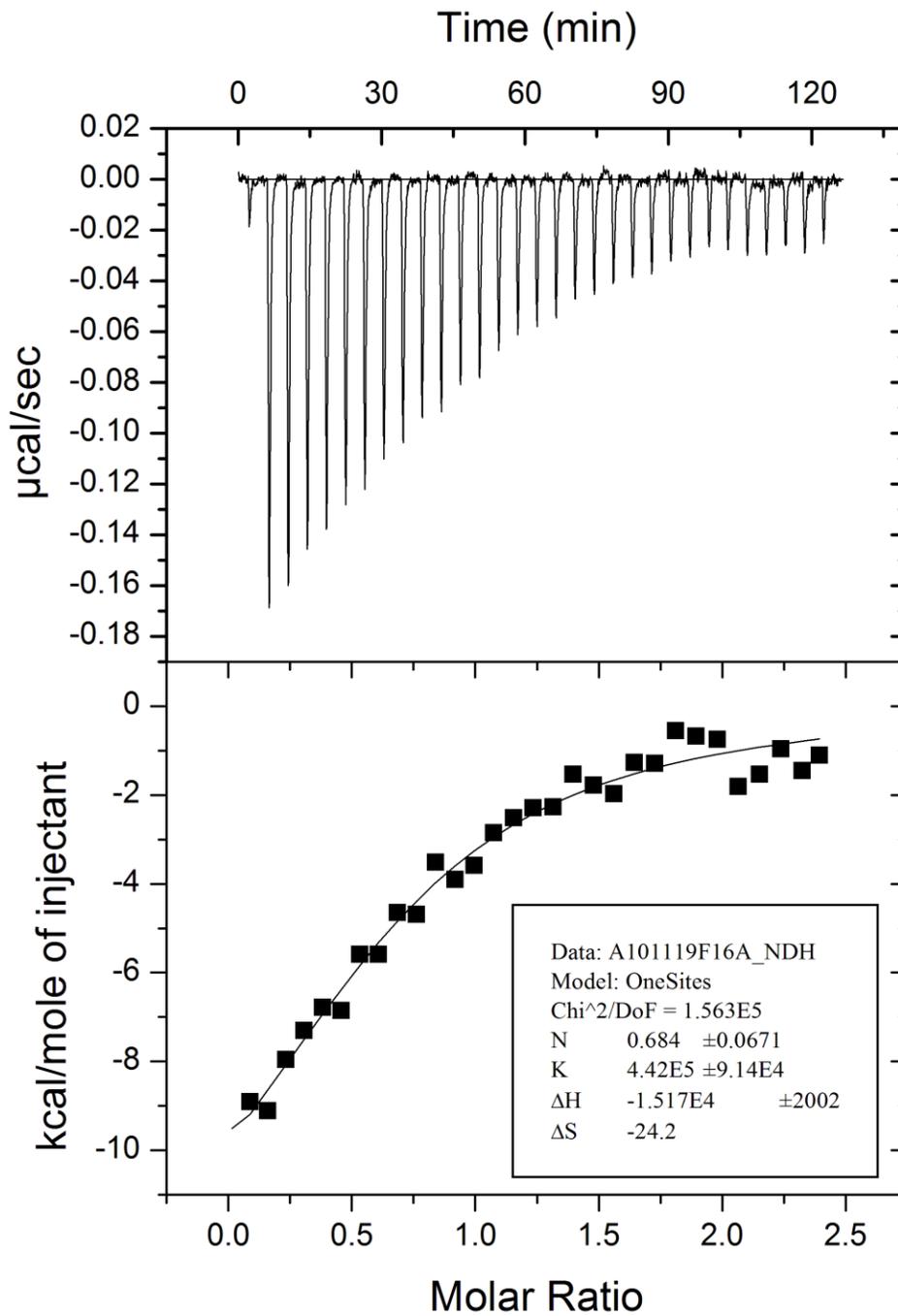
Sample (SL3) = 19.22 μM , Titrant (F16A) = 340.6 μM

11/19/2010 F16A (Use endpoints data as baseline)

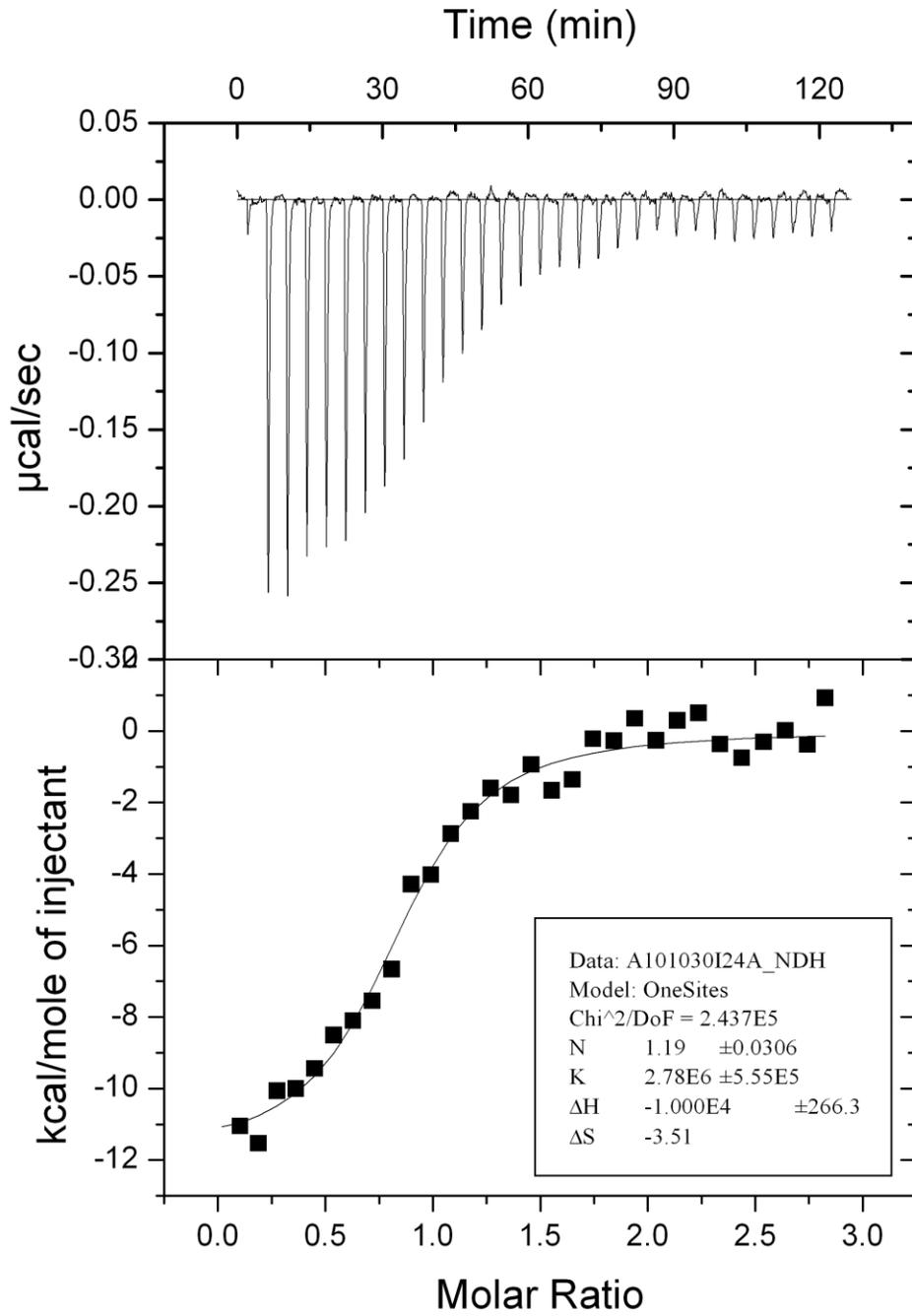


Sample (SL3) = 5.64 μM , Titrant (F16A) = 57.03 μM

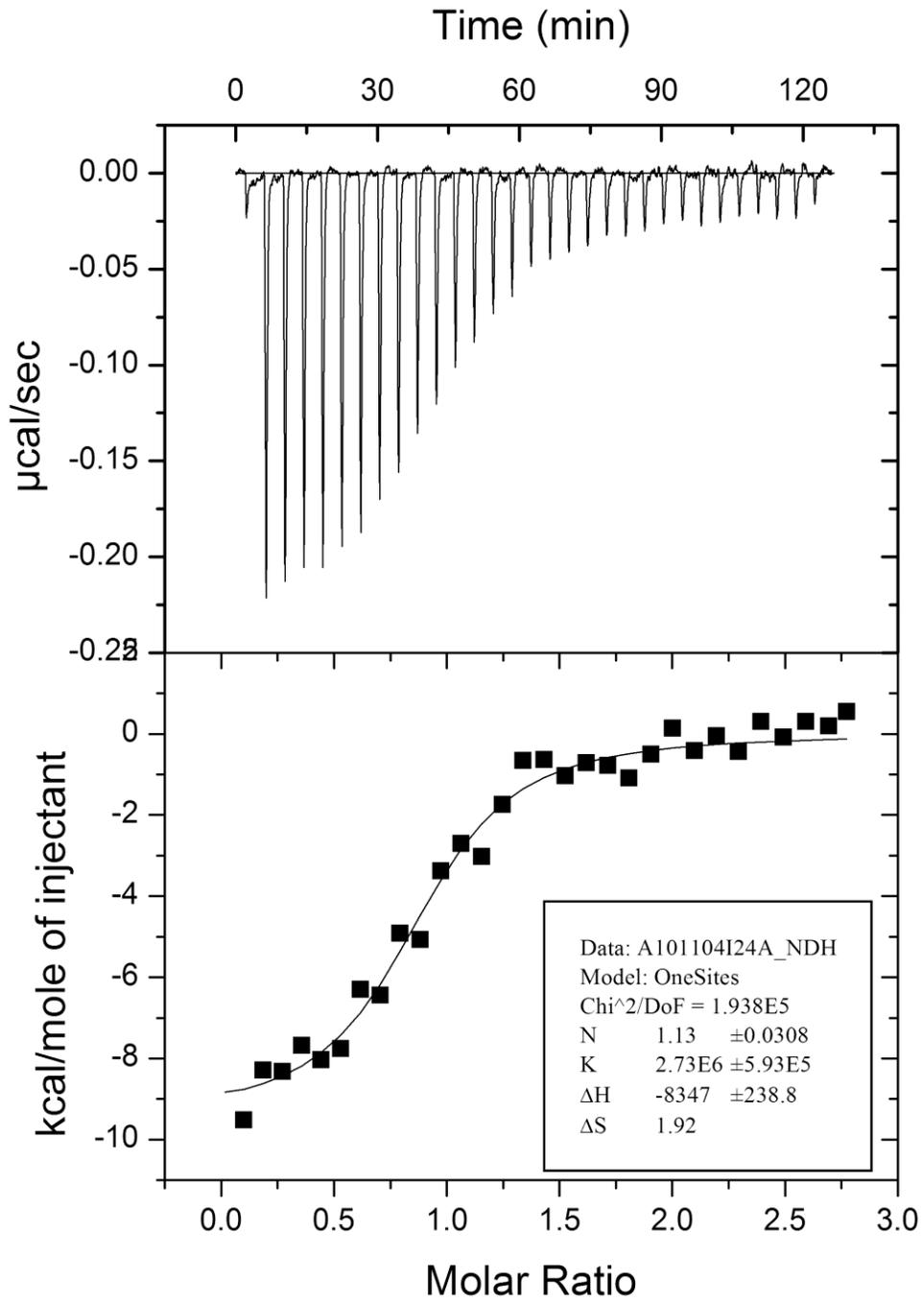
11/19/2010 F16A (Use blank titration data as baseline)



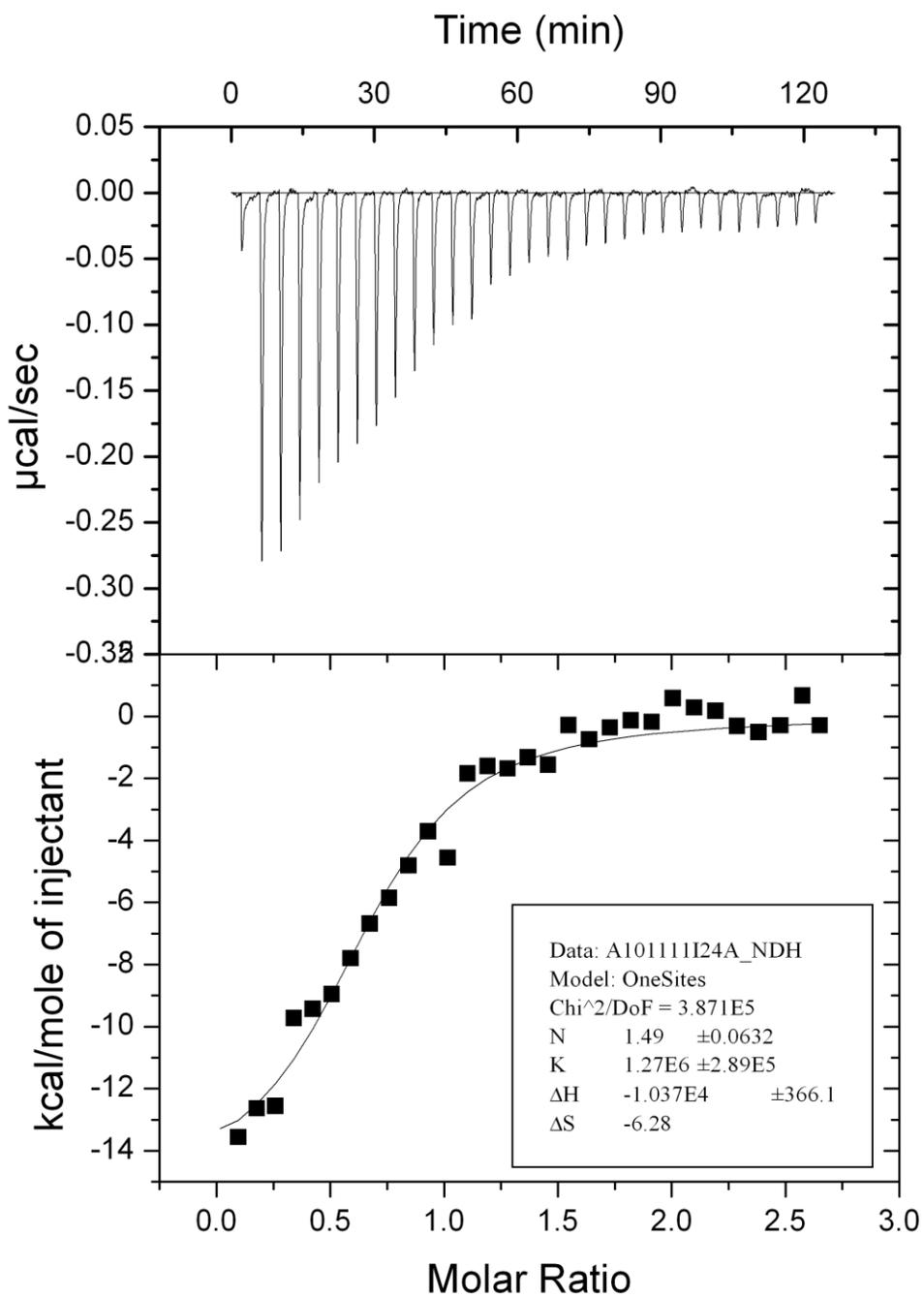
Sample (SL3) = 5.64 μM , Titrant (F16A) = 57.03 μM



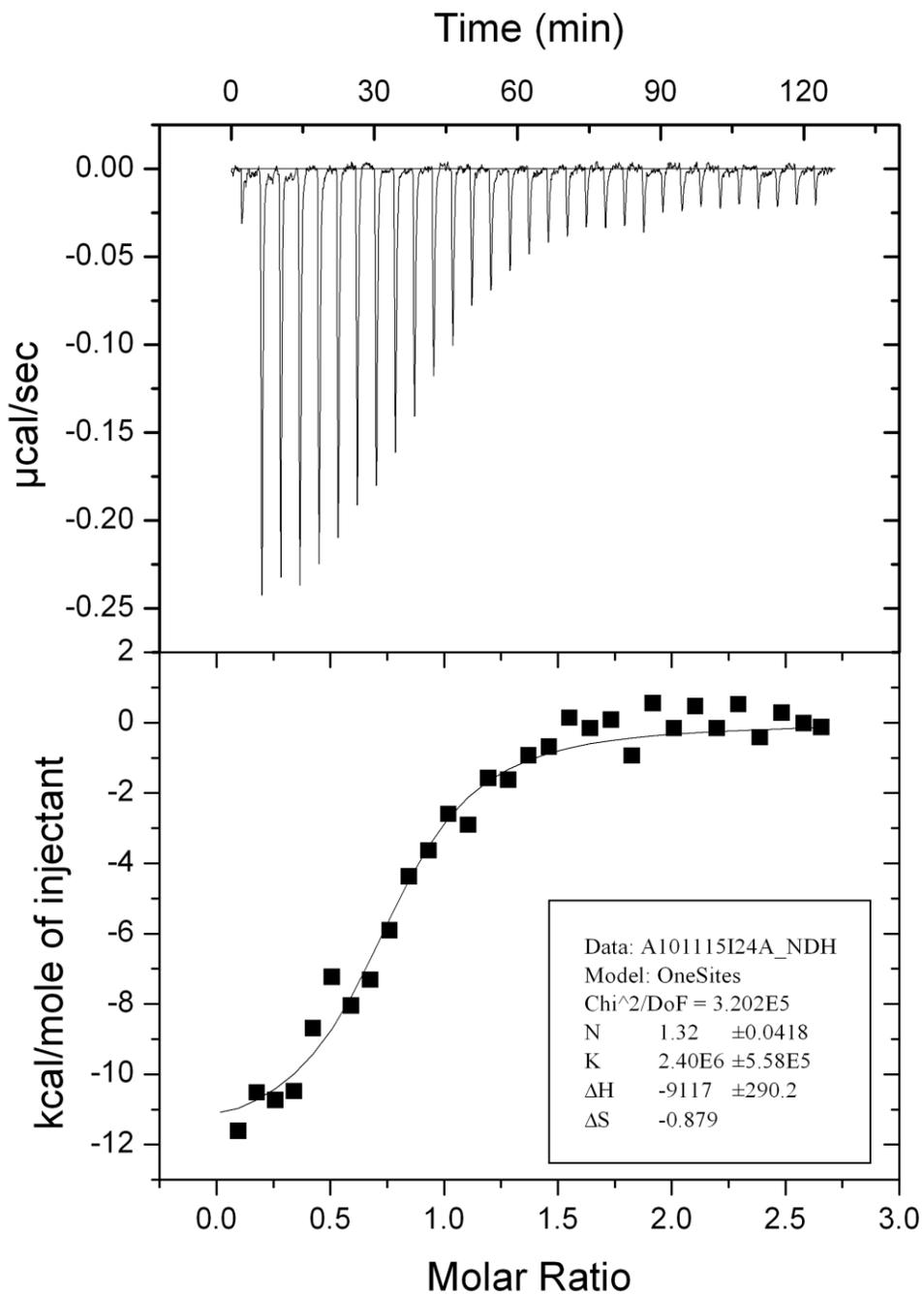
Sample (SL3) = 5.06 µM, Titrant (I24A) = 60.41 µM



Sample (SL3) = 5.46 µM, Titrant (I24A) = 63.91 µM

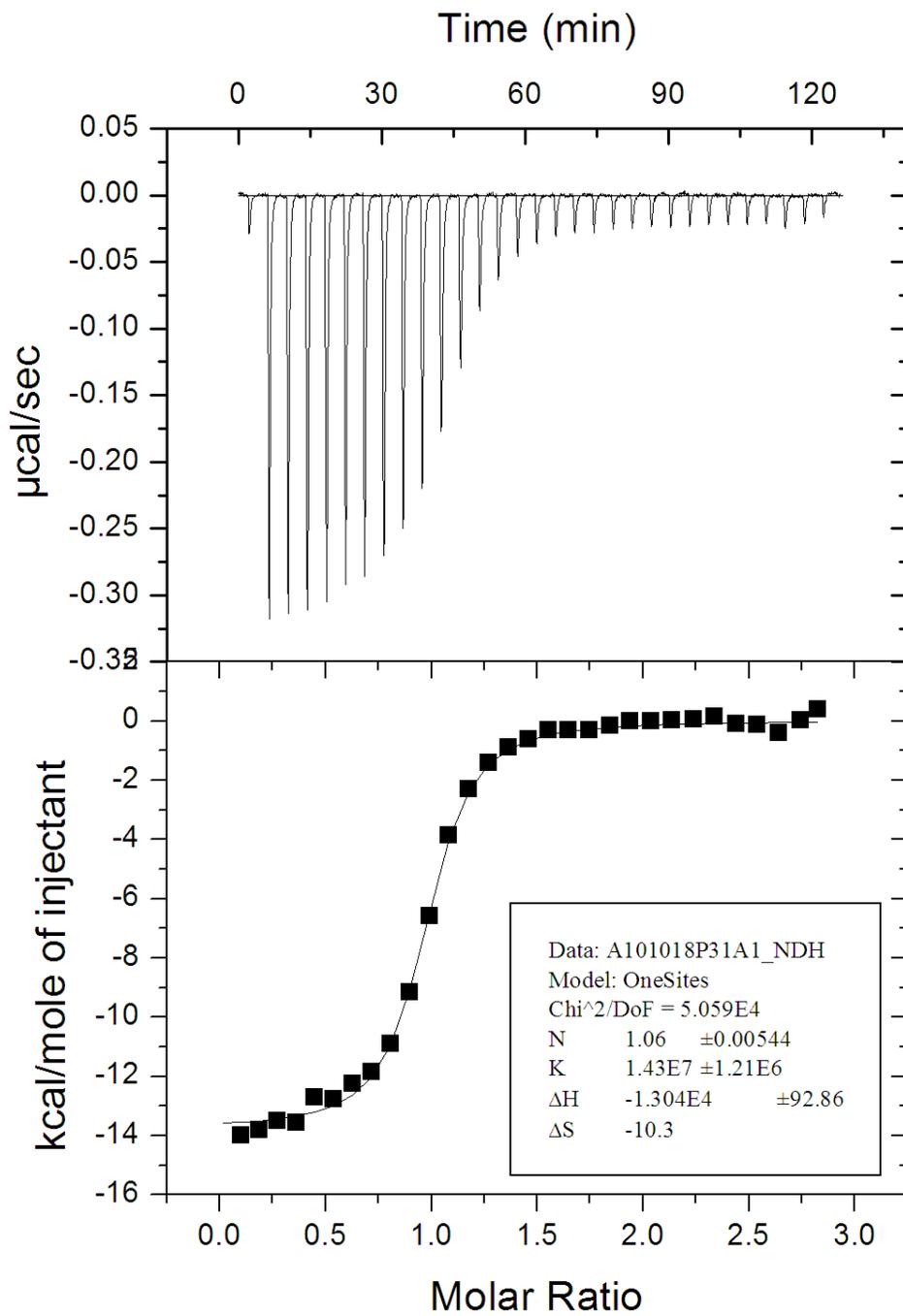


Sample (SL3) = 5.05 µM, Titrant (I24A) = 56.48 µM



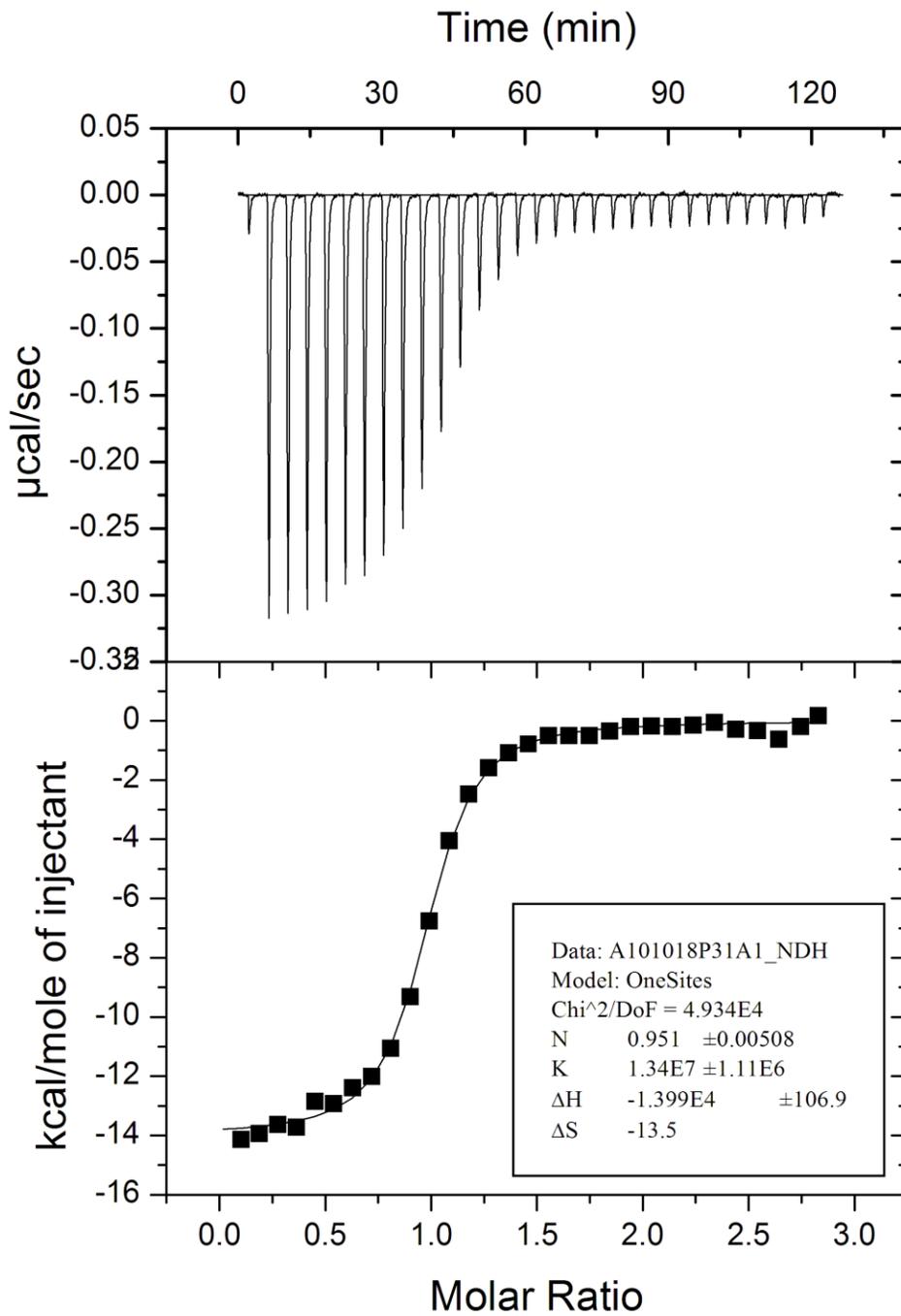
Sample (SL3) = 5.15 μM , Titrant (I24A) = 57.81 μM

10/18/2010 P31A-1 (Use endpoints data as baseline)



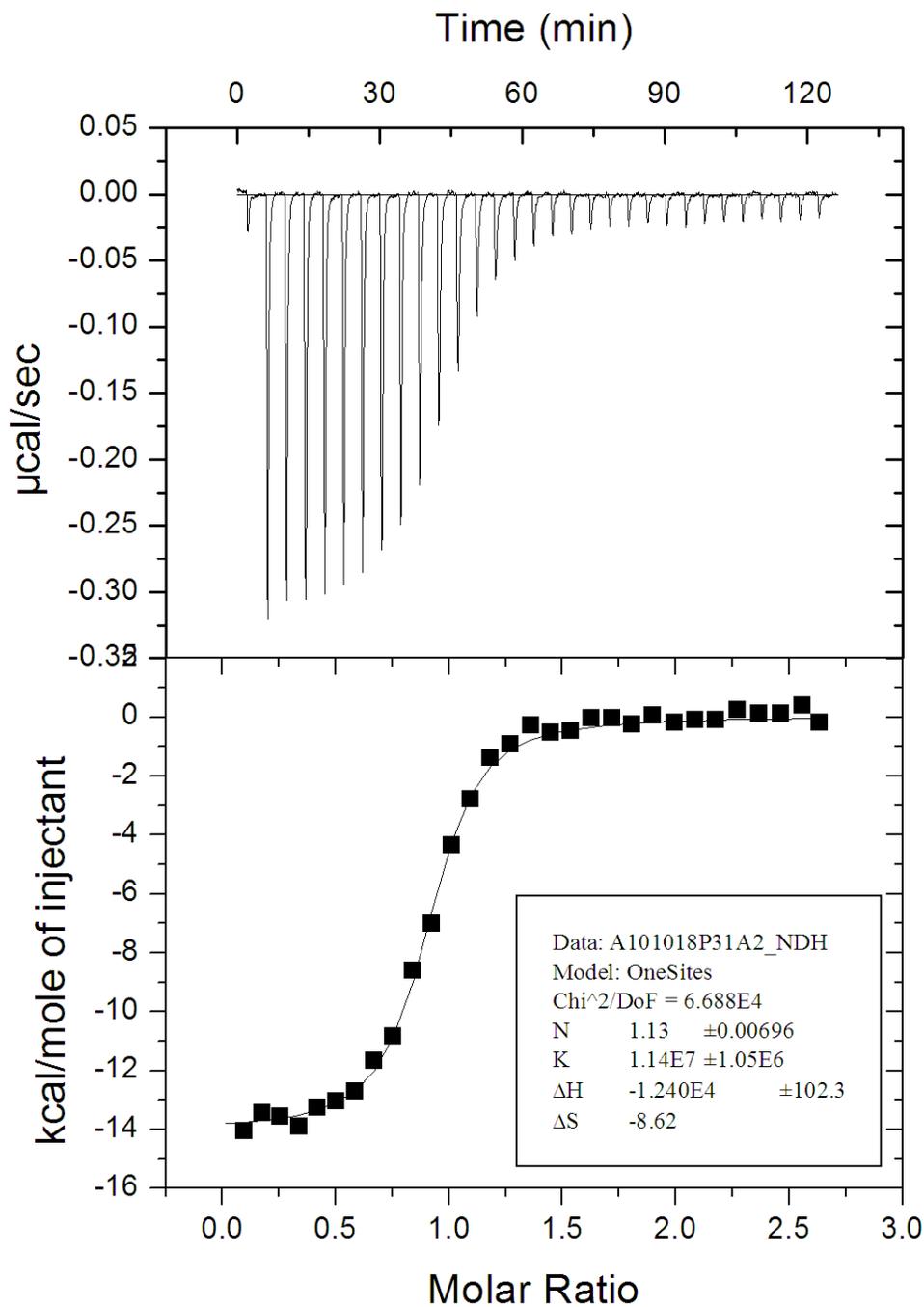
Sample (SL3) = 5.20 μM , Titrant (P31A) = 62.13 μM

10/18/2010 P31A-1 (Use blank titration data as baseline)



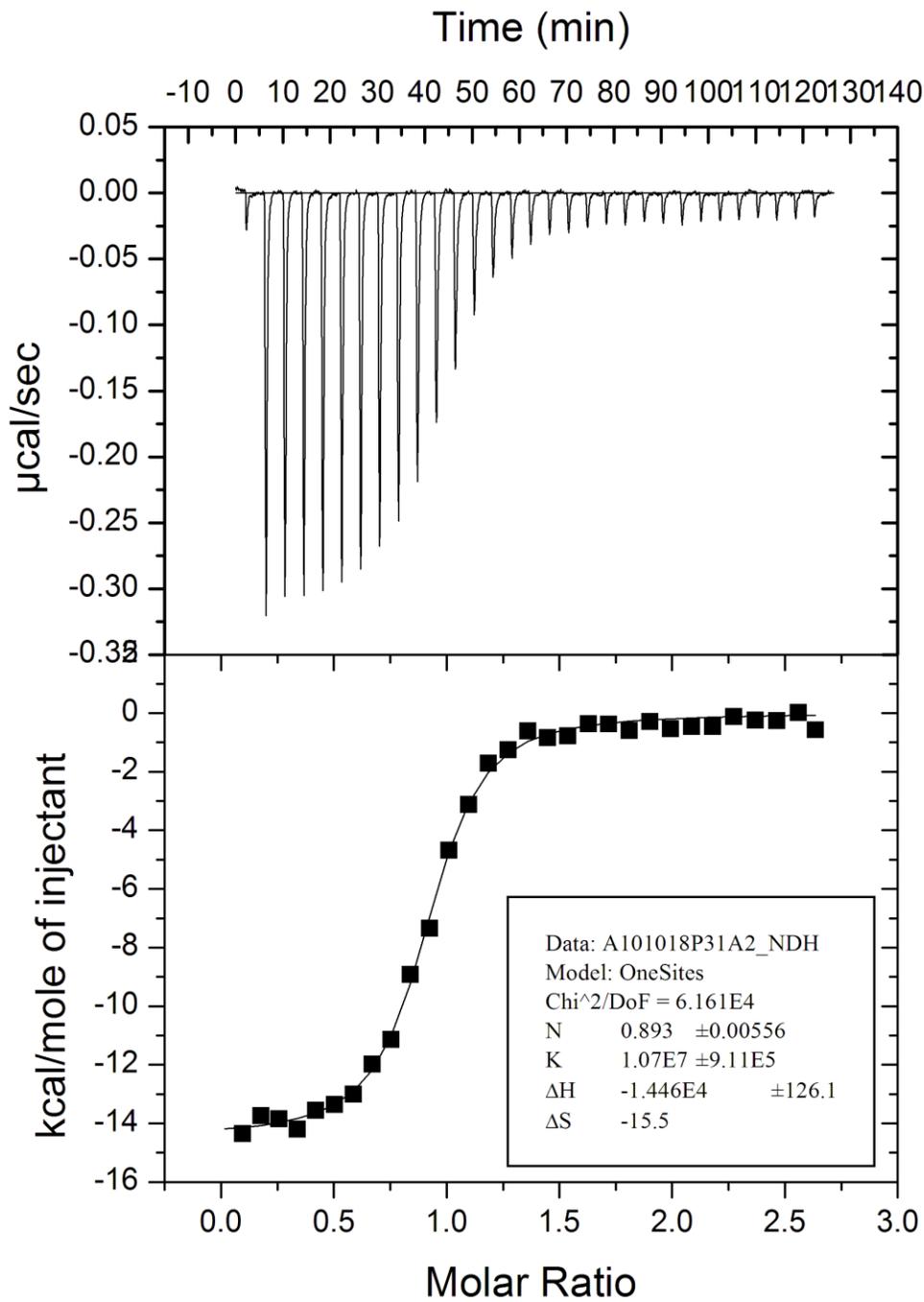
Sample (SL3) = 5.20 μM , Titrant (P31A) = 62.13 μM

10/18/2010 P31A-2 (Use endpoints data as baseline)



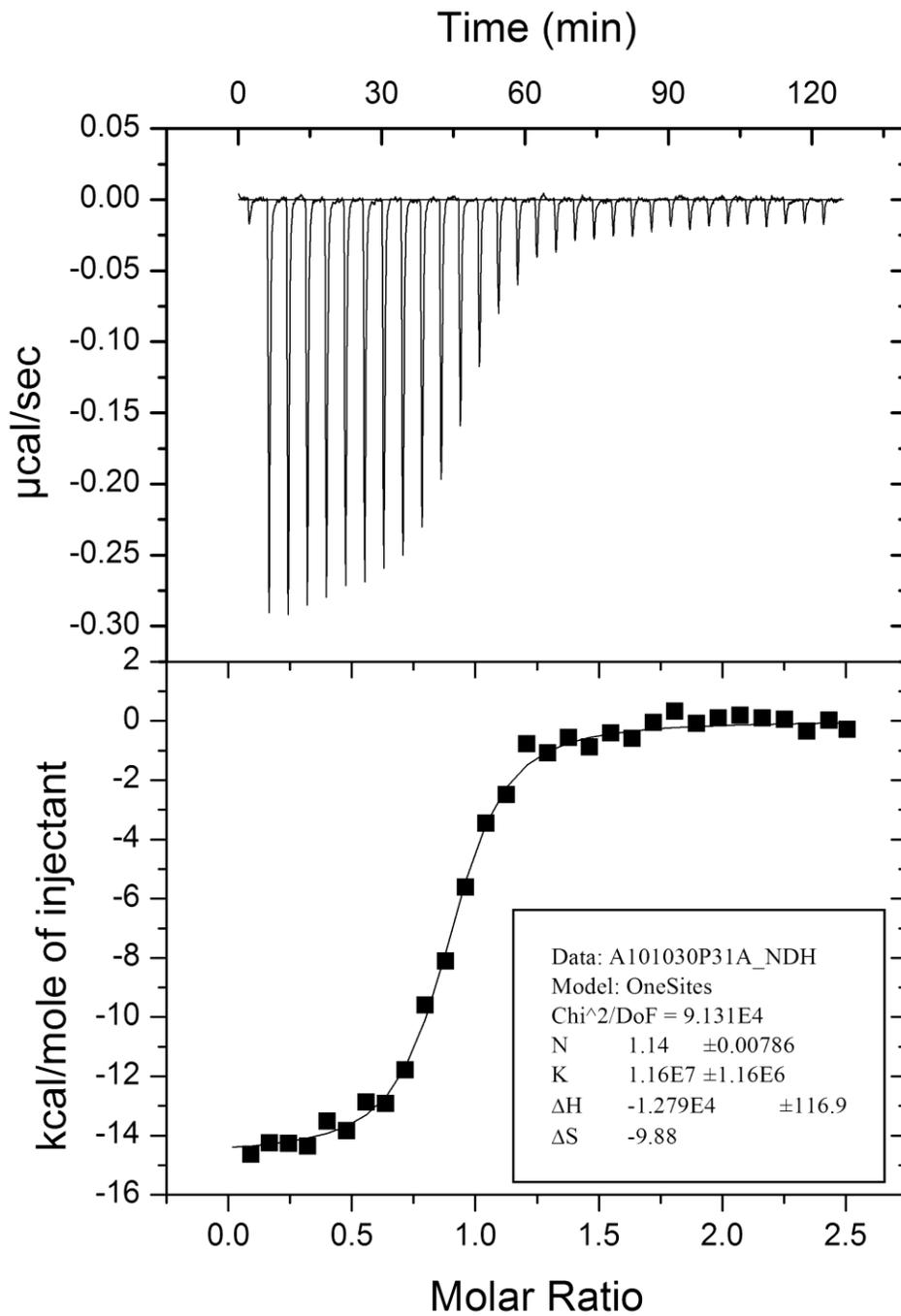
Sample (SL3) = 5.45 μM , Titrant (P31A) = 60.64 μM

10/18/2010 P31A-2 (Use blank titration data as baseline)



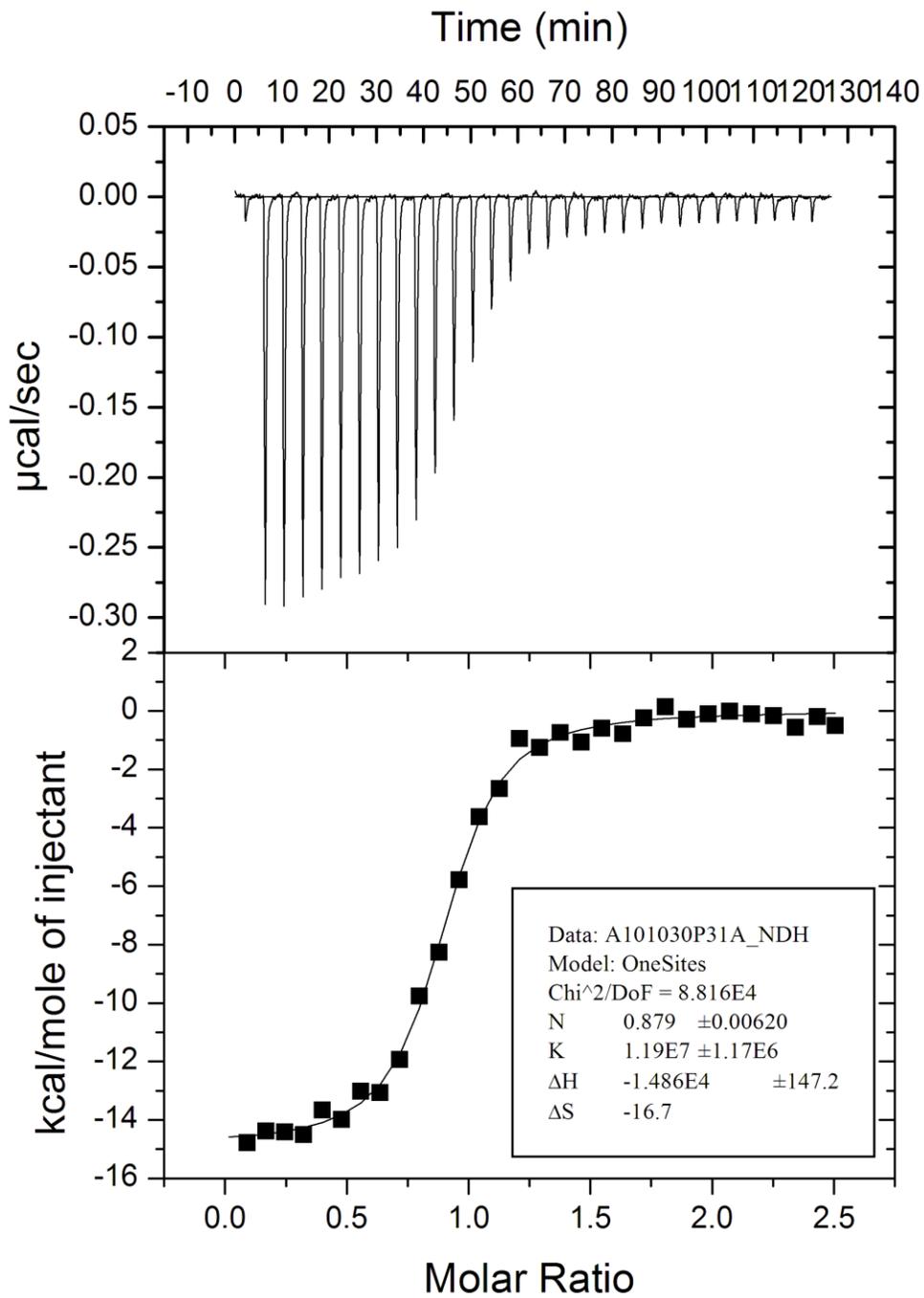
Sample (SL3) = 5.45 μM , Titrant (P31A) = 60.64 μM

10/30/2010 P31A (Use endpoints data as baseline)

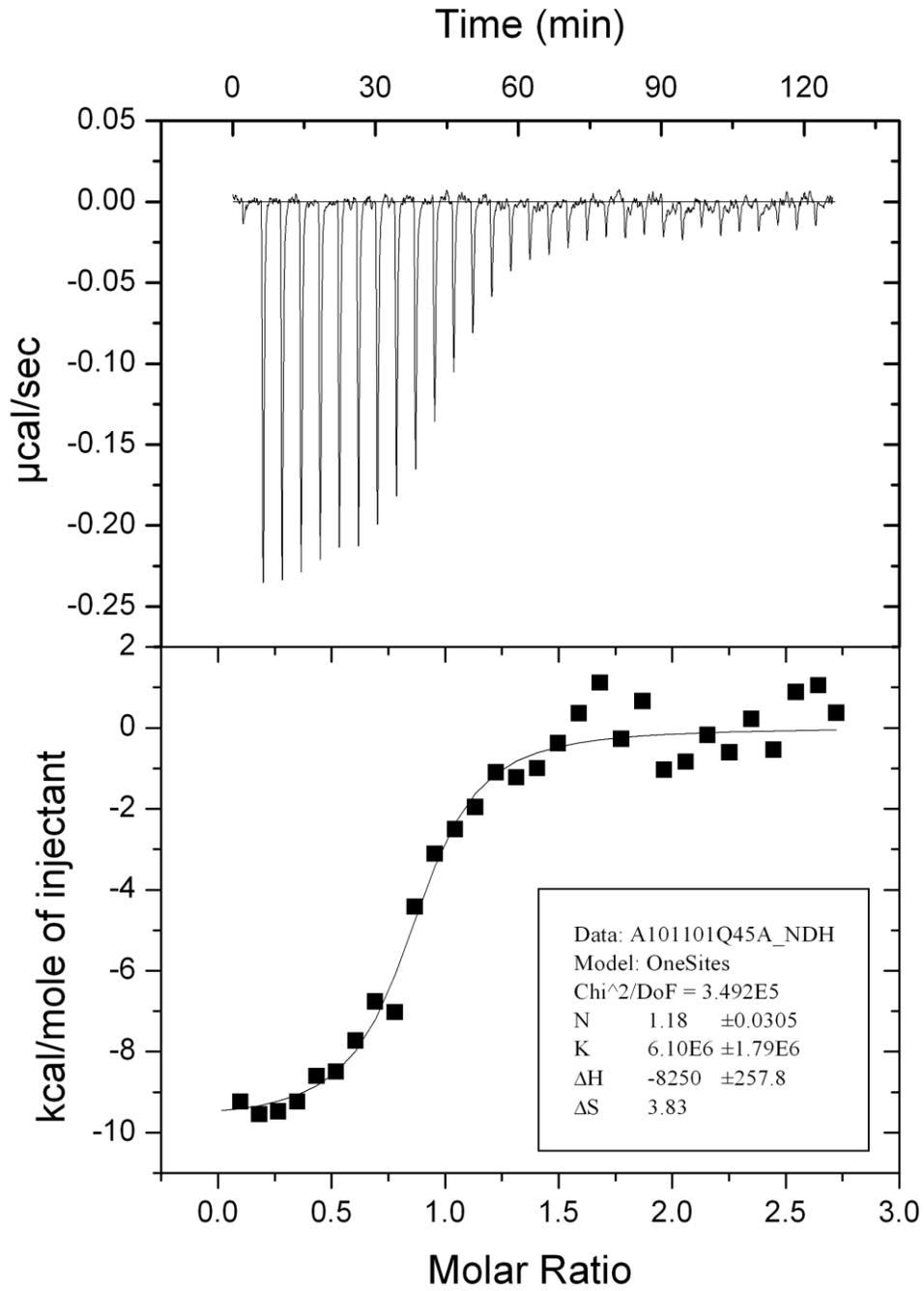


Sample (SL3) = 5.11 μM , Titrant (P31A) = 54.07 μM

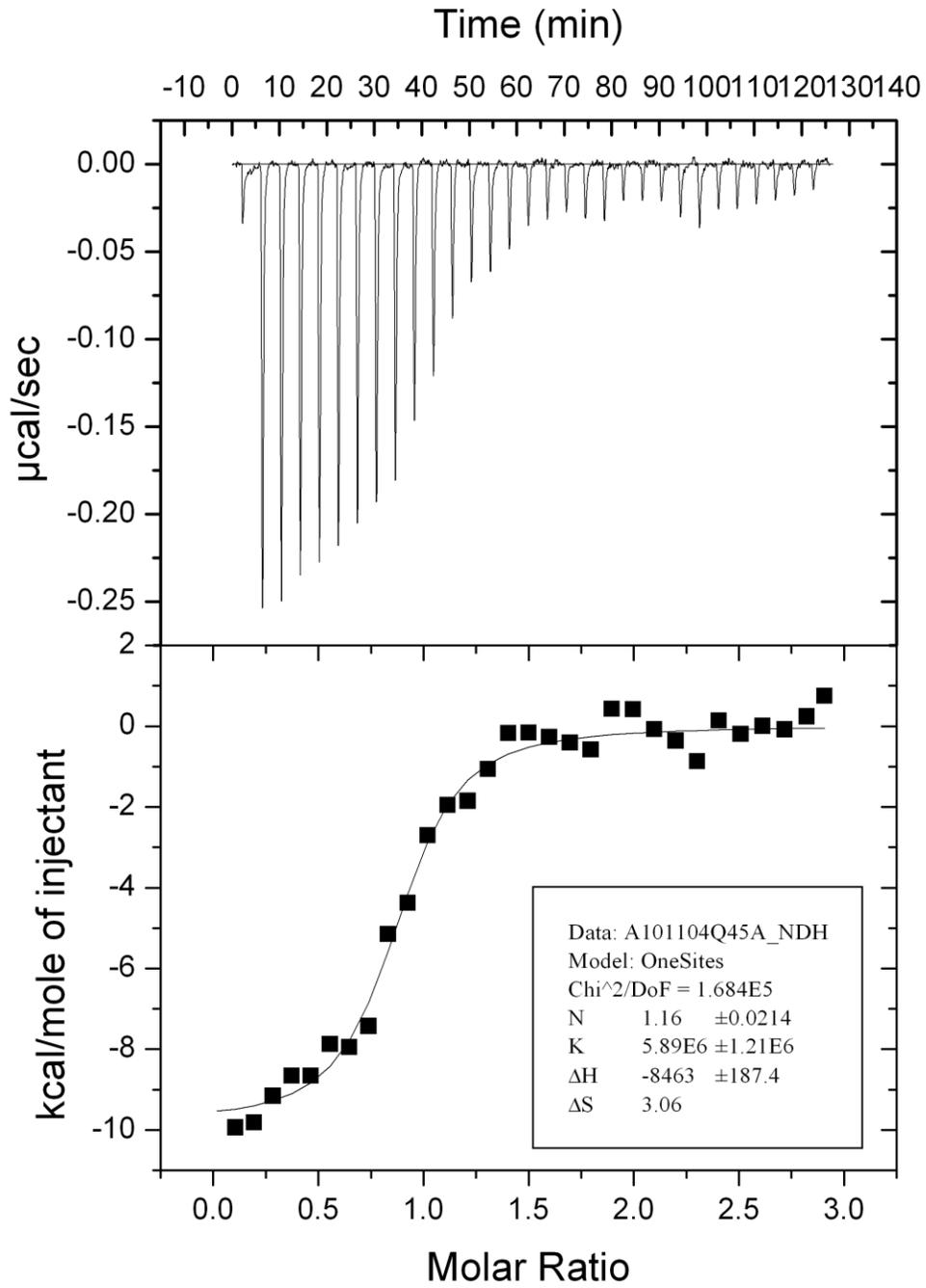
10/30/2010 P31A (Use blank titration data as baseline)



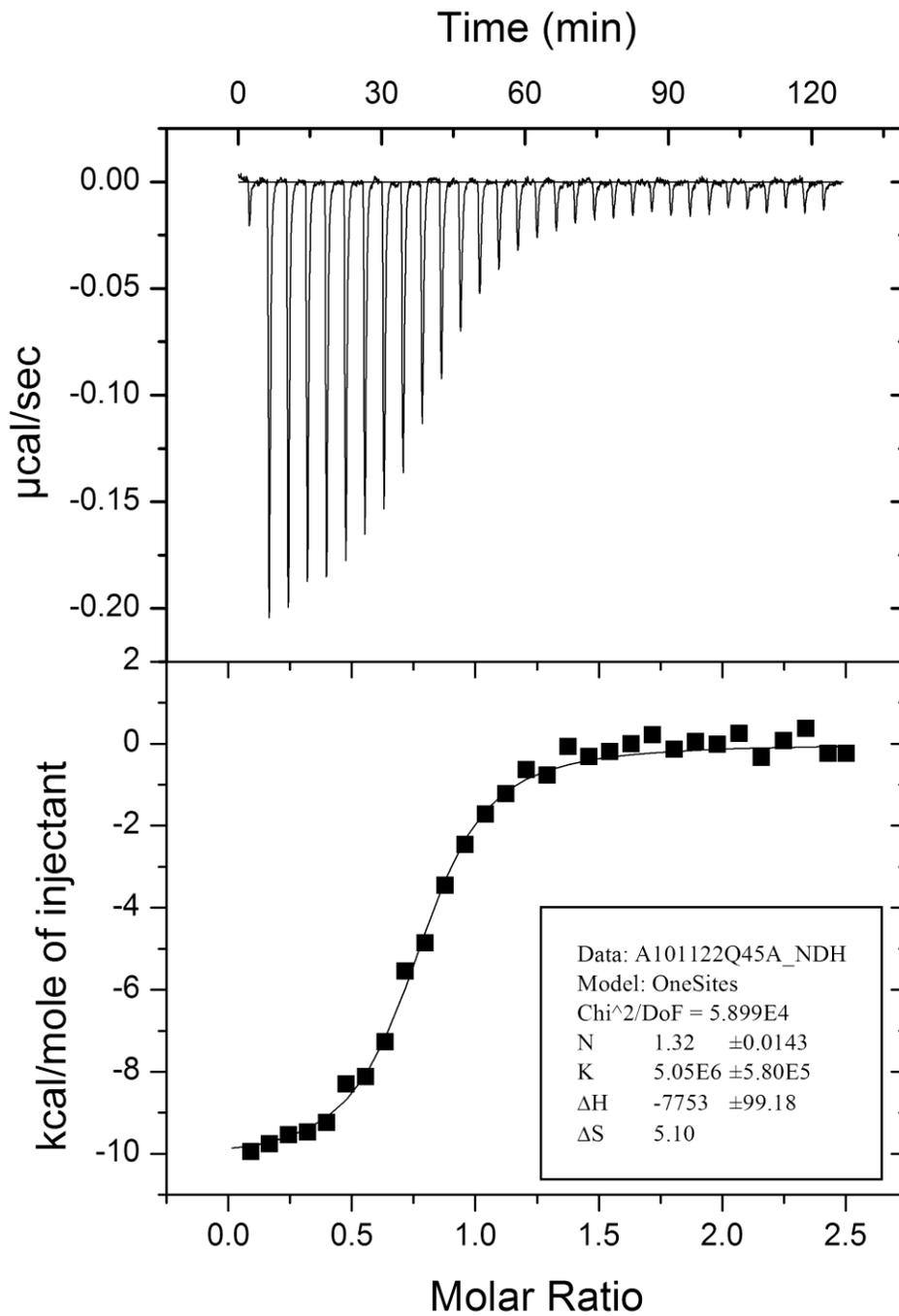
Sample (SL3) = 5.11 μM , Titrant (P31A) = 54.07 μM



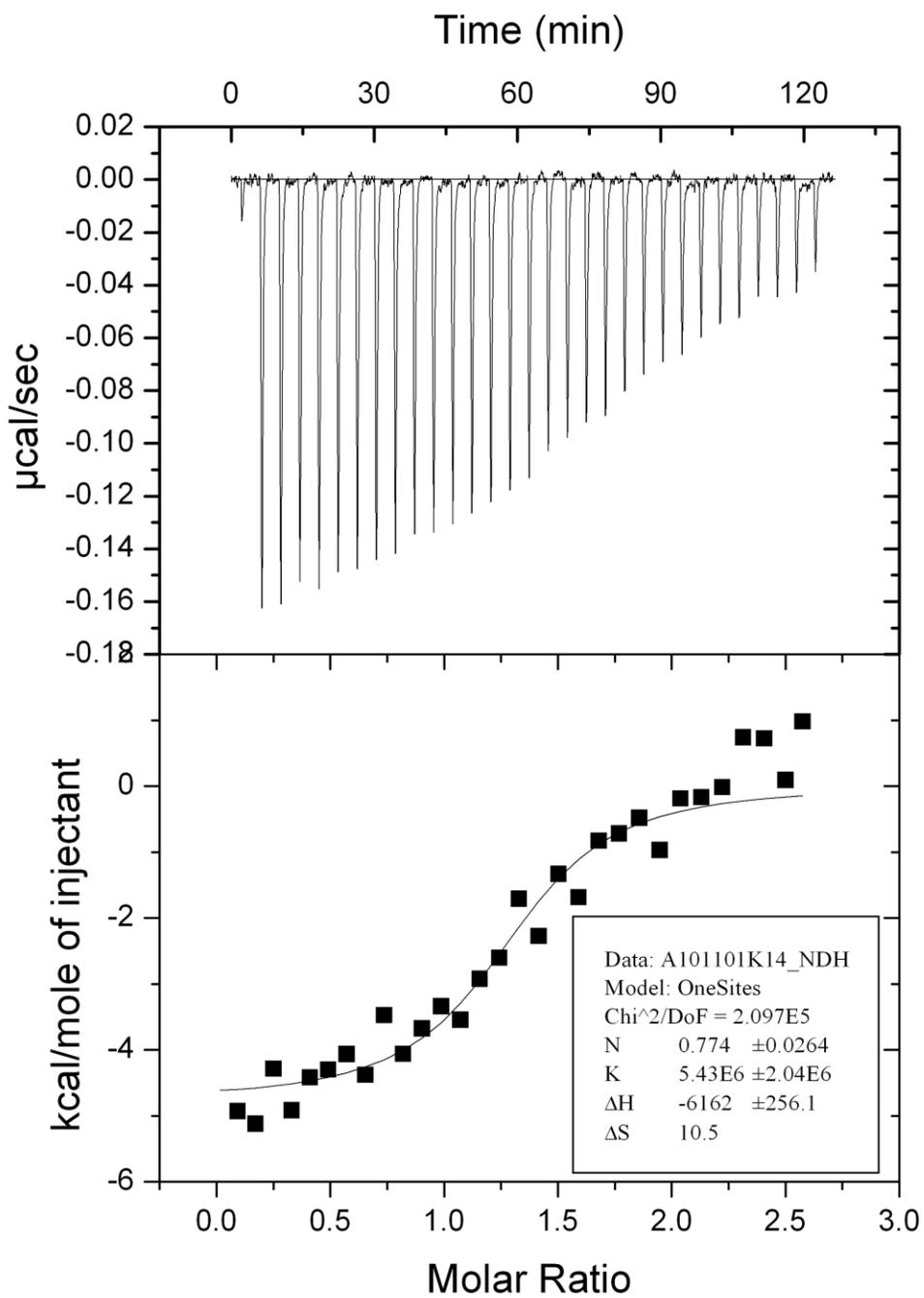
Sample (SL3) = 5.47µM, Titrant (Q45A) = 62.84 µM



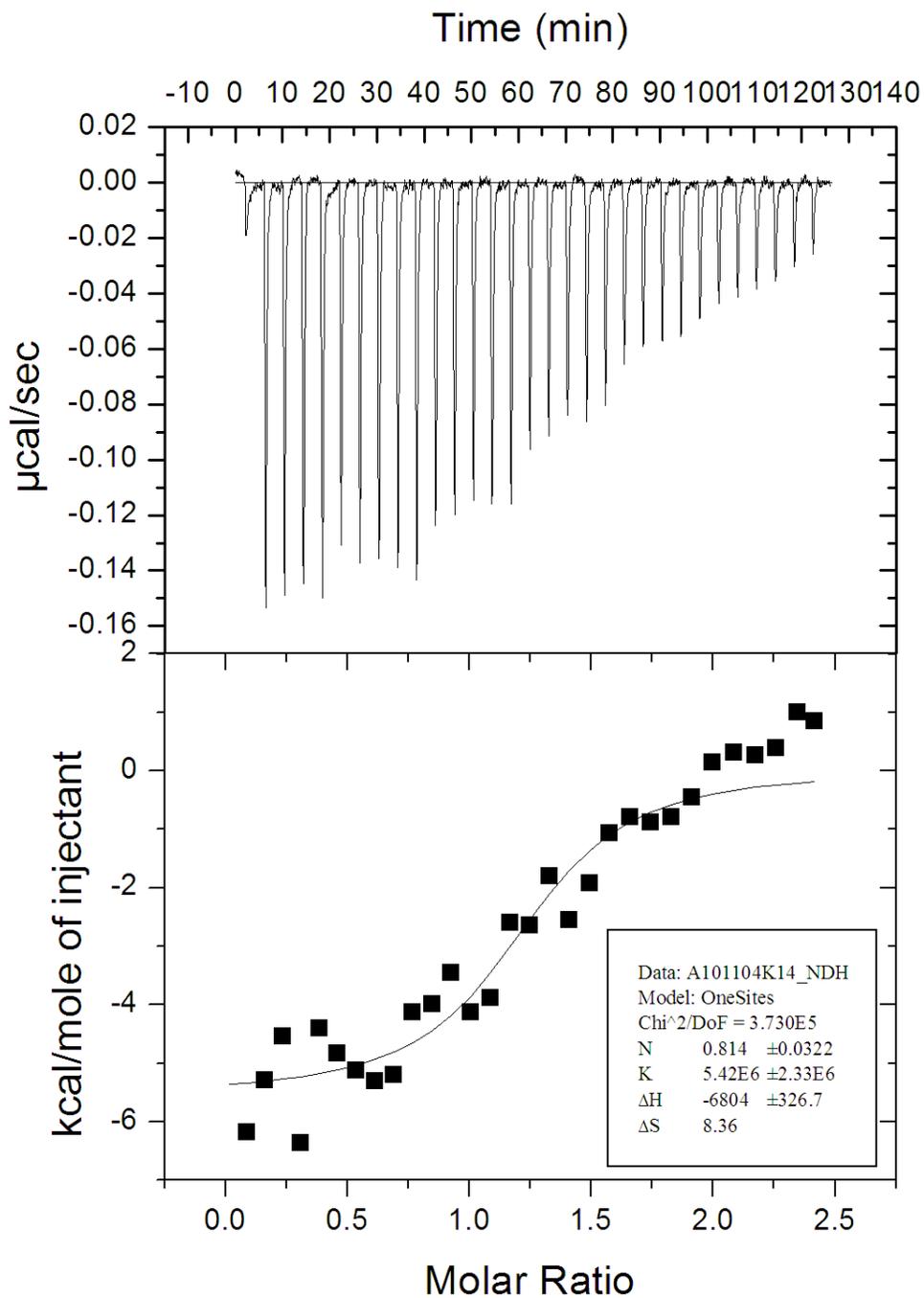
Sample (SL3) = 5.46 μM , Titrant (Q45A) = 67.06 μM



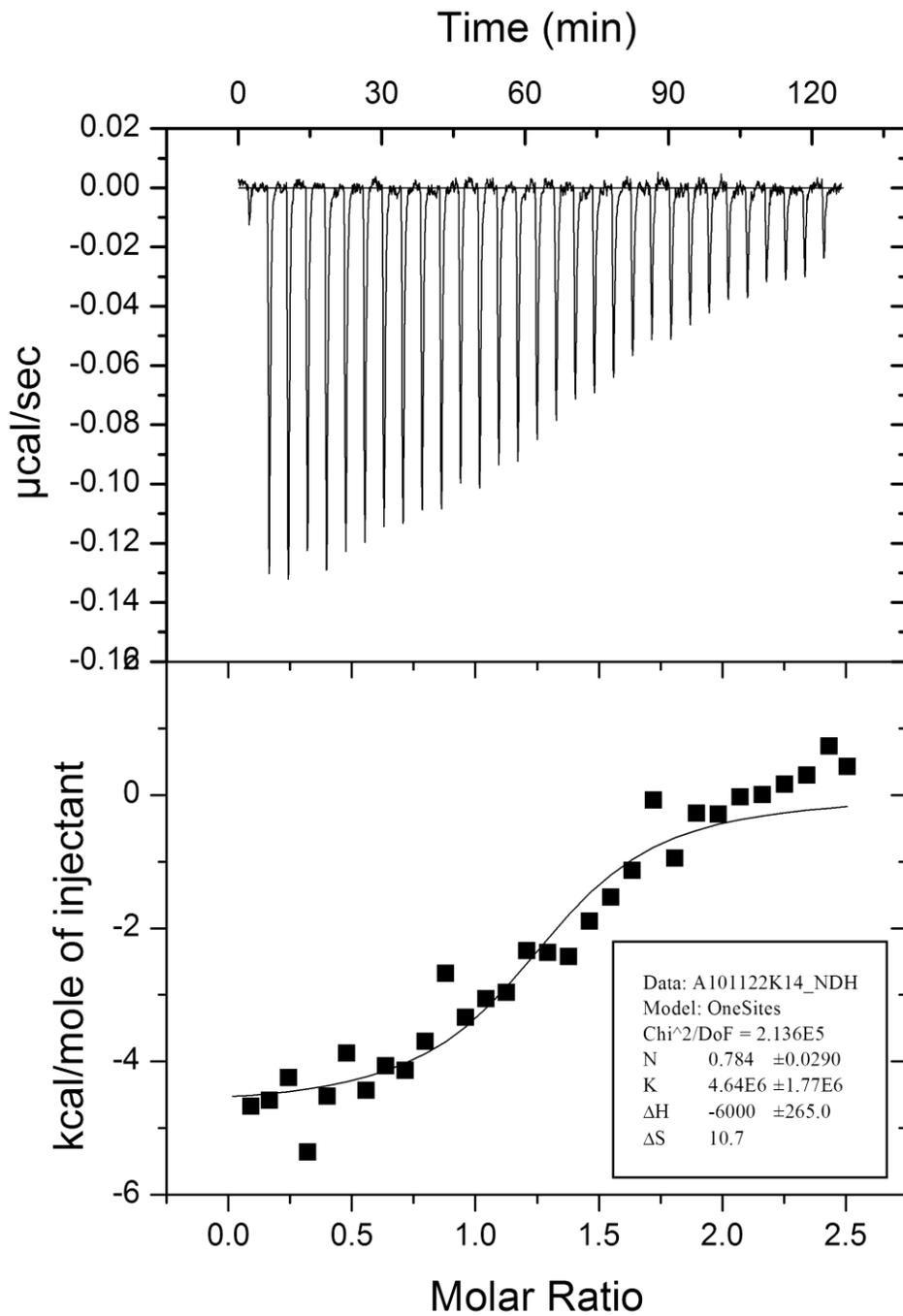
Sample (SL3) = 5.62 μM , Titrant (Q45A) = 59.36 μM



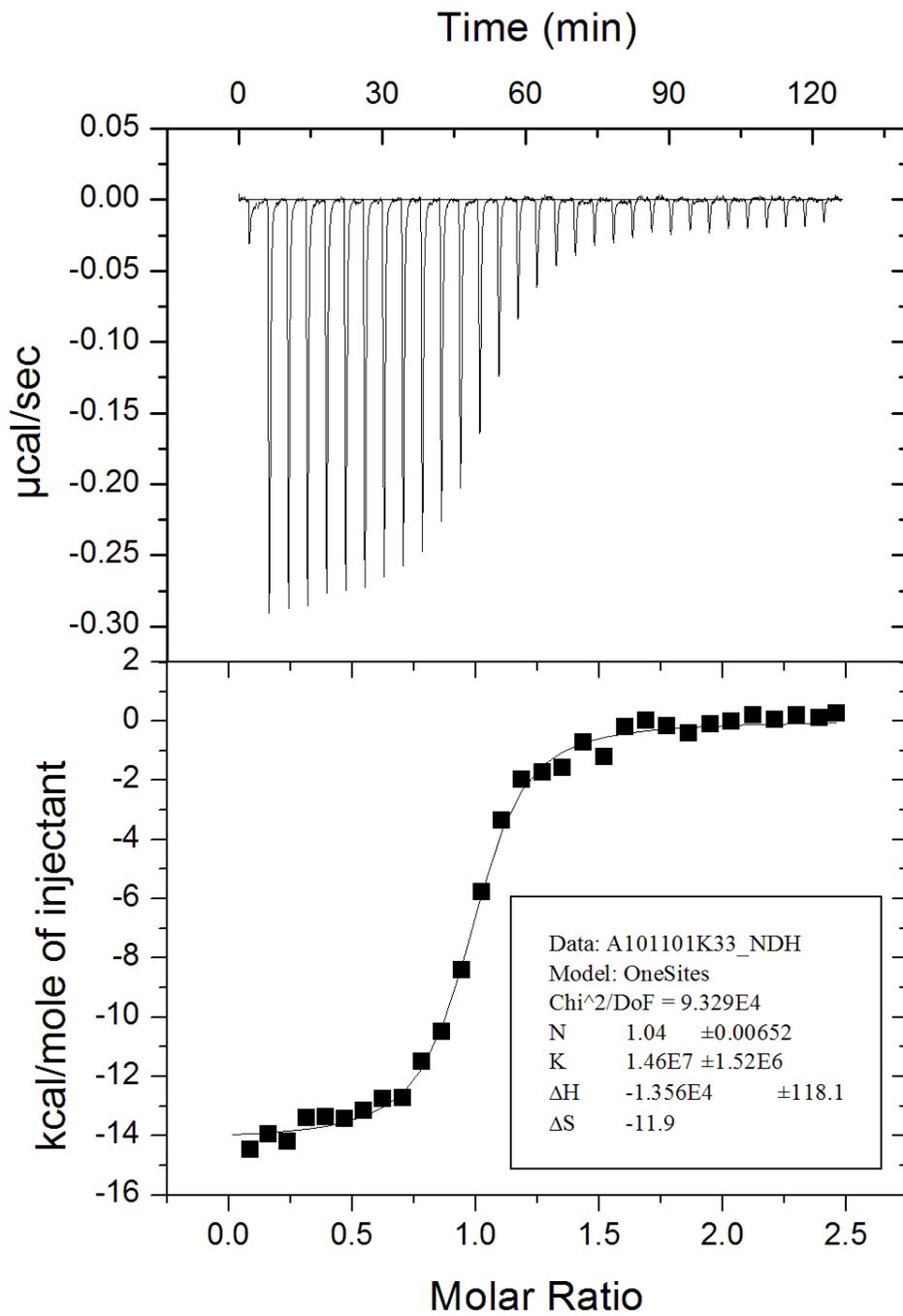
Sample (SL3) = 5.57 µM, Titrant (K14E-E21K) = 60.59 µM



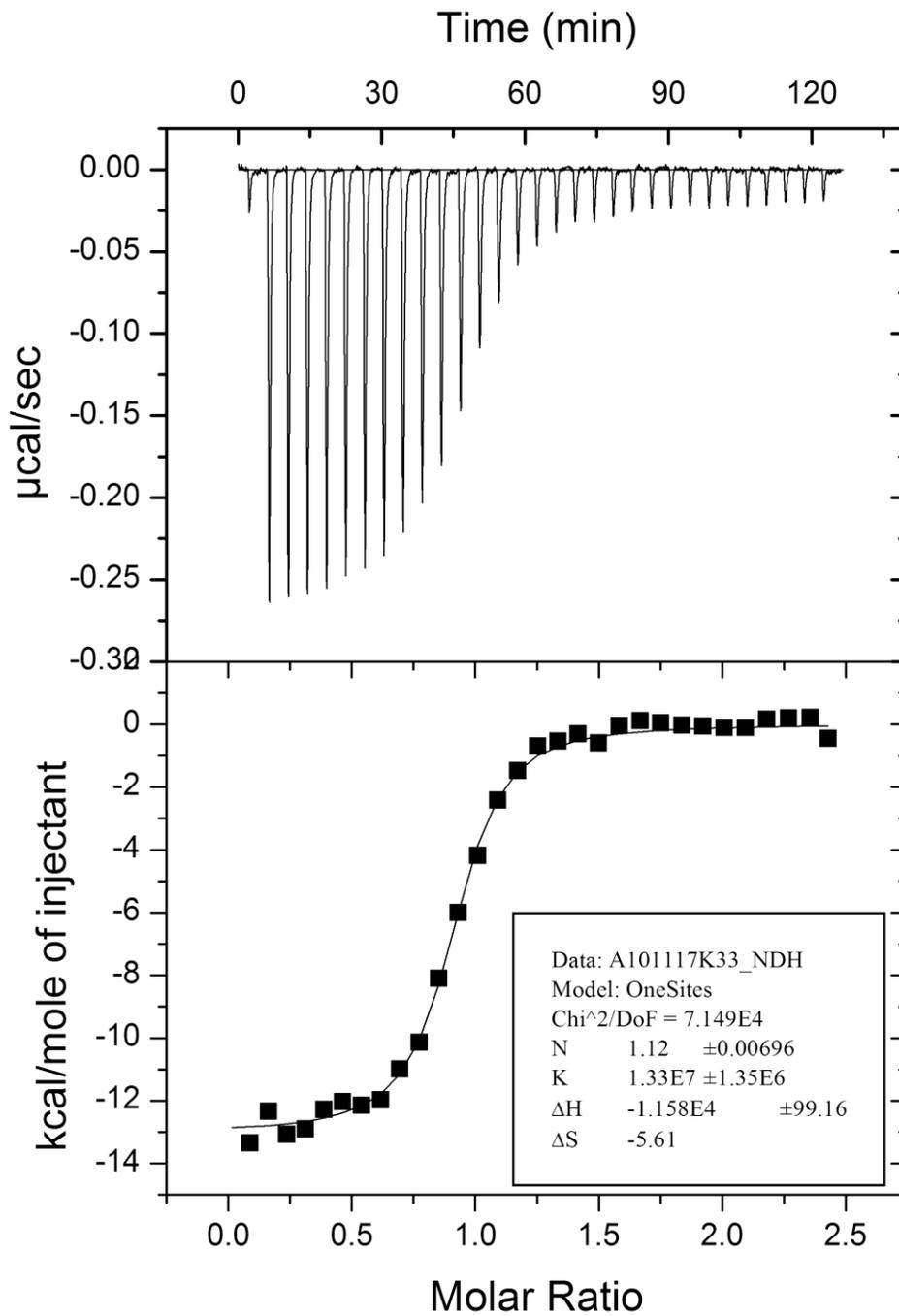
Sample (SL3) = $5.36 \mu\text{M}$, Titrant (K14E-E21K) = $54.8 \mu\text{M}$



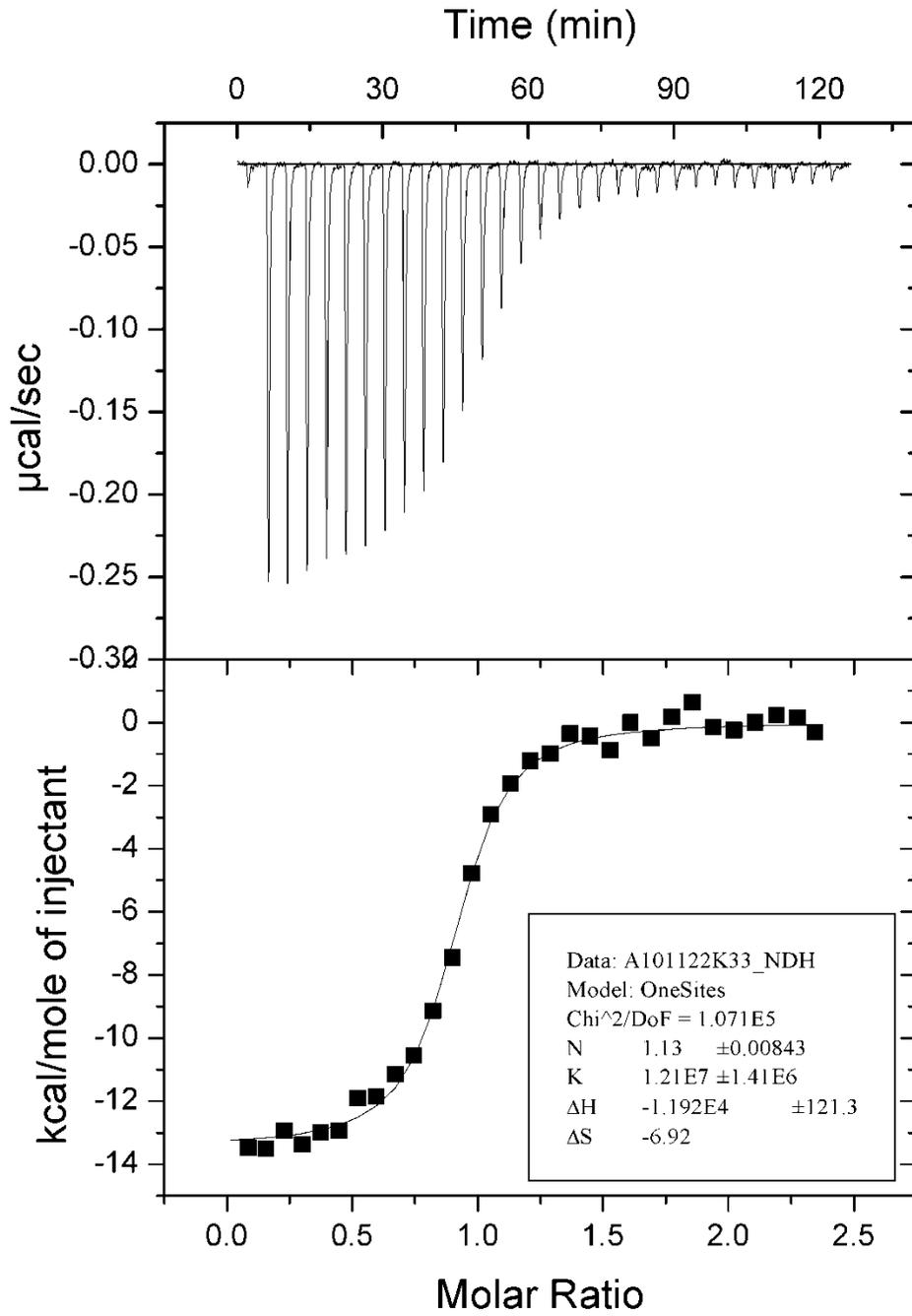
Sample (SL3) = 5.51 µM, Titrant (K14E-E21K) = 58.31 µM



Sample (SL3) = 5.45 μM , Titrant (K33E-E42K) = 56.68 μM



Sample (SL3) = 5.68 µM, Titrant (K33E-E42K) = 58.19 µM



Sample (SL3) = 5.61 μM , Titrant (K33E-E42K) = 55.56 μM

Appendix D. Recipes and Protocols

Used in the Development of Methods

for Aptamer Discovery

D-1. Recipes for buffers and solutions

5X PBS

Per liter:

NaCl	40g
KCl	1g
Na ₂ HPO ₄	7.2g
KH ₂ PO ₄	1.2g

Adjust pH to 7.4 with HCl. Filter and store at RT.

1X PBST

PBS	1X
Tween-20	0.05%
MgCl ₂	5mM

Bind and Wash buffer (for MyOne beads)

Tris-HCl	5mM
EDTA	0.5mM
NaCl	1.0M
pH 7.5	

1X TE

Tris-HCl	10mM
EDTA	1 mM
pH 8.0	

NXS

Tris-HCl	10mM, pH 7.5
EDTA	1mM, pH 8.0
NaCl	100mM
Triton X-100	1% (v/v)
SDS	1% (w/v)

NX

Tris-HCl	10mM, pH 7.5
EDTA	1mM, pH 8.0
NaCl	100mM

Triton X-100 1% (v/v)

0.1 M NaOH

Per liter:

NaOH 4g

Filtered sterilize and store air tight at room temperature.

20x SSPE

Per liter:

NaCl 175.3g

NaH₂PO₄·H₂O 27.6g

EDTA 7.4g

Dissolve in 800ml of H₂O, adjust pH to 7.4 with NaOH, adjust volume to 1 liter.

Filtered sterilize. Store at RT.

6X SSPE w. Triton X-100

Adjust 20X SSPE stock solution to 6X with dH₂O.

Add Triton X-100 to 0.01%

10X Sodium Borate Buffer (0.5M)

Per liter:

Boric Acid 30.92g

Dissolved in 800ml of ddH₂O, adjust pH to 8.5 using NaOH

Filtered sterilize and store at room temperature.

D-2. Protocol for bridge amplification and detection

Materials:

- SuperStreptavidin substrates slide (ArrayIt)
- Micro Spotting Solution Plus (ArrayIt)
- SuperStreptavidin blocking buffer (ArrayIt)
- 25ul Frame-seal incubation chambers (Bio-Rad)
- 1X PBST
- Templates and primers (biotinylated)
- dNTP, TurboPfu, 10X Pfu buffer
- Sybr555 DNA stain (Invitrogen)

Procedure:

1. Dissolve biotinylated primer and template oligos in Micro Spotting Solution Plus at desired final concentration. Mix the primer and template oligos in designed ratio.
2. Attach a 25- μ l frame-seal incubation chamber (Bio-Rad) onto a SuperStreptavidin Substrates slide (ArrayIt, Sunnyvale, CA).
3. Spot DNA samples inside the incubation chamber by pipetting. Let the sample hybridize with the substrates for 30 minutes at room temperature.

4. Block with 25 μ l SuperStreptavidin blocking buffer for 1 hour at room temperature.
5. Wash twice with 1x PBST for 5 minutes each time. Spin dry the slide.
6. Prepare PCR mix as below:

Turbo Pfu, 2.5 U/ μ L	2 μ L
10mM dNTP	2.5 μ L
10X Pfu Buffer	2.5 μ L
Nuclease free H ₂ O	18 μ L
<hr/>	
Final Volume:	25 μ L

7. Fill the chamber with the PCR mix. Seal the chamber with a plastic cover slip.
8. Run PCR thermal cycles as below:
 - 1) 92°C 1 minute
 - 2) 92°C 30 seconds
 - 3) 55°C 30 seconds
 - 4) 72°C 30 seconds
 - 5) Go back to 2) for a total of 40 cycles
 - 6) 72°C 1 minute
 - 7) Hold at 4°C
9. After amplification, remove the cover slip rinse the slide twice with 1X PBST for 5 minutes each time.
10. Stain the slide with 20 μ l of Sybr555 DNA stain (Invitrogen) in 27ml of 1X PBST at room temperature for 5 minutes.

11. Wash the slide twice with 1X PBST for 5 minutes each time.
12. Spin dry the slide (optional).
13. Scan the stained slides in a GenePix 4400A microarray scanner (Axon™ Instruments, Union City, CA, USA). Record images using GENEPIX PRO 5.0 software (Axon).

D-3. Protocol for emulsion PCR and detection

Materials:

- Dynal MyOne streptavidin C1 beads
- Bind and Wash buffer
- Dynal MPC-S magnet
- 1X TE, NXS, NX, 0.1 M NaOH, 1X PBS, 1X PBST, 6x SSPE w/ 0.01% Triton X-100
- Templates and primers, capping and bridging oligos
- dNTP, 10X Taq buffer without MgCl₂, 50mM MgCl₂ (Invitrogen)
- hot-start Taq (5U/μl, Denville)
- Tegosoft DEC (Degussa)
- light mineral oil (Sigma #5904-500ml)
- ABIL WE09 (Degussa)
- Bis(Sulfosuccinimidyl) suberate (BS3)
- Teflon coated glass slide (Erie Scientific) (pre-treated with aminosilane)
- 2X Self-seal reagent (Bio-Rad)
- SuperStreptavidin blocking buffer (ArrayIt)

Procedure:

1. Bead-loading

- 1) Take 100 μ l of Dynal MyOne paramagnetic streptavidin beads (10mg/ml) from stock vial, transfer into a 1.5ml eppendorf tube.
- 2) Add 100ul Bind and Wash buffer into the beads. Mix and remove all liquid on a magnetic particle concentrator (Dynal MPC-S magnet).
- 3) Wash twice in 200 μ l of Bind and Wash buffer, and re-suspend in 198 μ l of Bind and Wash buffer.
- 4) Add 2ul 1mM 5'-dual-biotin labeled forward primer. Mix well by pipetting and incubate the mixture at room temperature for 20 minutes on a Dynabeads rotary mixer, or periodically mix by pipetting or flicking.
- 5) Remove all liquid, and wash twice in 200 μ l of Bind and Wash buffer, then once in 200ul of 1X TE.
- 6) Remove all liquid and re-suspend beads in 200 μ l of 1X TE. These beads should now be around 5×10^9 beads/ml.

2. Emulsion PCR

- 1) Prepare the aqueous and oil phase as below:

Aqueous phase ^a:

10X Taq buffer without MgCl ₂ (Invitrogen)	12 μ l
50 mM MgCl ₂	45 μ l
10 mM dNTP mix	17 μ l
2 mM reverse primer	1 μ l

Pre-loaded MyOne beads in 1X TE	7.5 μ l
Hot-start Taq (5U/ μ l, Denville)	7 μ l
Template DNA at appropriate concentration	1 μ l
Nuclease free H ₂ O	22.5 μ l
<hr/>	
Total volume:	120 μ l

Oil phase ^b:

Tegosoft DEC (Degussa)	500 μ l
light mineral oil (Sigma #5904-500ml)	138 μ l
ABIL WE09 (Degussa)	48 μ l
<hr/>	
Total volume:	686 μ l

^a For aqueous phase, add all components except Taq polymerase, mix well by vortexing. Then add Taq and mix well again.

^b For oil phase, add all components to a 15 ml polypropylene falcon tube and vortex well to mix.

2) Add aqueous phase to oil phase, vortex well for 5 min. Distribute mix into PCR tubes at approximately 50 μ l per tube.

3) Run thermal cycle as below:

- | | |
|----------|------------|
| (1) 94°C | 10 minutes |
| (2) 94°C | 15 seconds |

- (3) 57°C 30 seconds
- (4) 70°C 75 seconds
- (5) Go back to (2) for a total of 120 cycles
- (6) 70°C 2 minutes
- (7) 4°C hold

3. Breaking emulsion

- 1) Add 100 µl of isopropanol to each tube or well containing the amplified PCR mixture. Mix well by pipetting.
- 2) Pool the isopropanol/PCR mixes into a 15 ml polypropylene falcon tube. Vortex for 1 minute.
- 3) Spin down the beads for 30 seconds at 4000 rpm in a table-top microcentrifuge. Remove supernatant by decanting
- 4) Add 2 ml of isopropanol and re-suspend the pellet by pipetting. Make sure no visible clumps in the suspension. Add 1 ml of isopropanol. Vortex for 30 seconds.
- 5) Spin for 30 seconds at 4000 rpm in a table-top microcentrifuge.
- 6) Remove supernatant as much as possible. Re-suspend the pellet in 3 ml of NXS and mix well by vortexing. Make sure that the pellet is VERY well re-suspended.
- 7) Spin down the beads for 30 seconds at 3000 rpm in table-top centrifuge
- 8) Remove supernatant using a magnetic particle separator

- 9) Wash the pellet once in 1 ml of NXS and then twice in 1 ml of 1X TE.
- 10) Re-suspend in 500 μ l of 0.1 M NaOH. Incubate the suspension at room temperature on a rotary mixer for 5 minutes.
- 11) Wash once with 500 μ l of 0.1 M NaOH, then twice with 500 μ l of 1X TE.
- 12) Re-suspend and combine beads in 30 μ l of 1X TE.

4. Bead Capping

- 1) Prepare the hybridization mixture as below:

6x SSPE with 0.01% Triton X-100	180 μ l
1mM Bridging oligo 1	1.8 μ l
1mM Bridging oligo 2	1.8 μ l

- 2) Re-suspend beads in the hybridization mixture solution, mix well and incubate for 10 minutes at 56°C.
- 3) Mix again, and then incubate the mixture on a rotary mixer for 10 more minutes at room temperature.
- 4) Wash the beads twice with 200 μ l of TE. Prepare ligation mix as below:

10x T4 Ligase Buffer	10 μ l
1mM Capping oligo 1	2 μ l
1mM Capping oligo 2	2 μ l
T4 DNA ligase (3U/ μ l)	2 μ l
Nuclease free dH ₂ O	84 μ l
<hr/>	
Total volume:	100 μ l

- 5) Remove all liquid from beads and re-suspend them in the ligation mix solution. Incubate the mixture on a rotary mixer for 1 hour at room temperature.
- 6) Remove liquid and wash beads once with 200 μ l of 1X TE, once with 200 μ l of NX, then twice with 200 μ l of 1X TE.
- 7) Transfer beads to a new tube, wash once with 200 μ l of 1X TE, then re-suspend the beads in 30 μ l of 1X TE.

5. Bead attachment

- 1) Wash beads three times in 50 μ l of 1X PBS.
- 2) Prepare fresh BS3 solution (2.86 mg of Bis(Sulfosuccinimidyl) suberate in 1 ml of 1X PBS).
- 3) Remove all liquid from beads, quickly re-suspend the beads in 35 μ l of fresh BS3 solution
- 4) Apply the beads onto the glass surface of the center circle on a Teflon coated glass slide (Erie Scientific) that has been pre-treated with aminosilane. Cover the circle with a cover slip.
- 5) Incubate the slide with beads at room temperature for 1 hour without disturbing.
- 6) Wash the slide twice with 1X TE for 5 minutes each time to remove any unattached beads.

6. Staining the beads with TBA probe (on slide)

- 1) Add 35 μ l of 6x SSPE w/ 0.01% Triton X-100 to each circle, cover with cover slip.
- 2) Incubate the slide at 56°C for 10 minutes.
- 3) Wash immediately with 1x TE twice for 2 minutes on labquake. Add 35 μ l of the following staining solution:

6x SSPE w/ 0.01% Triton X-100	149 μ l
1mM 5'-6-FAM-TBA comp oligo	1 μ l
<hr/>	
Total volume:	150 μ l

- 4) Cover with a cover slip. Incubate the slide on a thermal cycler using a heat cycle setting as below:

56°C	2 minutes
40°C	10 minutes
25°C	10 minutes

- 5) Wash twice with 1x TE for 5 minutes each time on a labquake.
- 6) Spin dry the slide. Add 25 μ l of 1x self-seal reagent in TE, cover with a cover slip.

7. Staining the beads with TBA probe (unattached, in tube)

- 1) Wash beads wash twice with 200 μ l of 6X SSPE.

2) Prepare the staining solution as below:

6X SSPE	149 μ l
1mM 5'-6-FAM-TBA comp oligo	1 μ l
<hr/>	
Total volume:	150 μ l

3) Re-suspend beads in 150 μ l of the staining solution. Incubate in a thermal cycler using the same heating cycles as described above.

4) Wash the stained beads twice with 200 μ l of 1XTE.

5) Re-suspend the beads in 25 μ l of 1X self-seal reagent in 1X TE, apply the suspension onto a glass slide, and cover with a cover slip.

8. Staining the beads with thrombin (on slide)

1) Block the slide with SuperStreptavidin blocking buffer for 1 hour at room temperature.

2) Prepare staining solution as below:

1X PBST	150 μ l
Thrombin (Cy3 or Cy5 labeled)	500 nM (final concentration)
<hr/>	
Total volume:	150 μ l

3) Apply 25 μ l of the staining solution onto each circle and cover with a cover slip. Incubate the slide overnight at room temperature.

4) Wash the stained slide twice with 1X PBST and twice with 1X PBS for 10 minutes each time.

5) Add 25 μ l of 1x self-seal reagent in TE, cover with a cover slip.

9. Staining the beads with thrombin (unattached, in tube)

- 1) Wash the beads twice with 200 μ l of 1X PBST.
- 2) Prepare the staining solution as below:

1X PBST	150 μ l
Thrombin (Cy3 or Cy5 labeled)	500 nM (final concentration)
<hr/>	
Total volume:	150 μ l

- 3) Re-suspend beads in 150 μ l of the staining solution. Incubate overnight at room temperature.
- 4) Wash the stained beads twice with 200 μ l of 1X PBST and twice with 200 μ l of 1X PBS for 5 minutes each time.
- 5) Re-suspend the beads in 25 μ l of 1x self-seal reagent in 1X PBS, apply the suspension onto a glass slide, cover with a cover slip.

10. Visualization using Nikon TE2000 microscope

- 1) Turn on Nikon TE2000 microscope in the following order:
 - a. X-Cite mercury bulb
 - b. Nikon Scope
 - c. Spot Camera
- 2) Light path of different ports are:

Port 1: eye;

Port 5: camera;

Port 3: both eye and camera.

- 3) Emulsion beads are only visible under a 100X objective. Use oil specific for Nikon microscopes.
- 4) Lower the platform before load the sample slide. Slowly raise the platform using the adjusting screw. When the objective hits the oil drop, a sudden flash of light can be seen. Start micro-tuning at that point.
- 5) Take one bright field image and one fluorescence image of the same microscopic field.
- 6) Use ImageJ to view and measure the photos.

D-4. Protocol for aminosilane treatment of glass slides

Materials:

- Teflon coated glass slide (Erie Scientific)
- Glass rack for microslides
- 0.5% Triton X-100 in dH₂O
- 100% ethanol, dry acetone
- aminopropyl triethoxysilane (Thermo Pierce)

Procedure:

1. Load glass slides onto a glass rack with handle.
2. Immerse glass slides in solution of 0.5% Triton X-100 in dH₂O. Wash by manually agitating the rack for 1 minute.
3. Pour all liquid off and wash thoroughly under running water to remove all detergent. Rinse with dH₂O.
4. Pour off all liquid and dry by immersion in 100% ethanol. Remove slides from ethanol and allow all ethanol to evaporate (aided by use of a Dust-off compressed air canister if possible).

5. Prepare 2% aminopropyl triethoxysilane (Pierce) solution in dry acetone in a glass beaker. Pipette with glass pipettes (the solution will react with polystyrene).
6. Immerse rack of completely dry (and ethanol-free) glass slides in silane solution for 30 seconds. Use the handle to manually agitate the rack.
7. Remove and immediately immerse in fresh dry acetone rinse for 1 minute. Use the handle to manually agitate the rack.
8. Remove rack and blow-dry partially with compressed air (to remove the bulk of the acetone).
9. Completely air-dry and store under vacuum at RT.
10. When cleaning beakers containing acetone and aminopropyl triethoxysilane, be sure to rinse with fresh acetone to remove all traces of aminopropyl triethoxysilane. If present, it will form a white precipitate upon evaporation of acetone.

D-5. Protocol for labeling protein with Cy3 or Cy5

Materials:

- Protein
- 0.05M Borate buffer (pH 8.5)
- DyLight 549 or 649 microscale antibody labeling kit (Pierce)
- Illustra microspin column (GE)

Procedure:

1. Dialyze the protein sample against 0.05M Borate buffer (pH 8.5) overnight.
2. Take 100 μ l of dialyzed protein (around 1mg/ml), add to one vial of DyLight 549 or 649 dye (DyLight 549 equals Cy3, DyLight 649 equals Cy5), pipet to mix.
3. Incubate the mixture at room temperature for 1 hour.
4. Take 100 μ l of purification resin (from the labeling kit), add to an Illustra microspin column (GE).
5. Spin the resin at 1000g for 30 seconds.
6. Change collection tube, apply the protein-dye mixture onto the center of the resin column.
7. Spin the column at 1000g for 30 seconds.

8. Collect the flow through, measure protein concentration and label efficiency by UV-Vis absorption.

References

1. Friedland, G.H. and R.S. Klein, *Transmission of the human immunodeficiency virus*. N Engl J Med, 1987. **317**(18): p. 1125-35.
2. *Pneumocystis pneumonia--Los Angeles*. MMWR Morb Mortal Wkly Rep, 1981. **30**(21): p. 250-2.
3. UNAIDS., *Global report: UNAIDS report on the global AIDS epidemic 2010*, Geneva: UNAIDS. p. 1-359.
4. Trono, D., et al., *HIV persistence and the prospect of long-term drug-free remissions for HIV-infected individuals*. SCIENCE, 2010. **329**(5988): p. 174-80.
5. Coffin, J.M., S.H. Hughes, and H. Varmus, *Retroviruses* 1997, Plainview, N.Y.: Cold Spring Harbor Laboratory Press. xv, 843 p.
6. Kanki, P.J., J.R. Hopper, and M. Essex, *The origins of HIV-1 and HTLV-4/HIV-2*. Ann N Y Acad Sci, 1987. **511**: p. 370-5.
7. Marlink, R., *Lessons from the second AIDS virus, HIV-2*. Aids, 1996. **10**(7): p. 689-99.
8. Ho, D.D., *Time to hit HIV, early and hard*. N Engl J Med, 1995. **333**(7): p. 450-1.
9. Hammer, S.M., et al., *Antiretroviral treatment of adult HIV infection: 2008 recommendations of the International AIDS Society-USA panel*. JAMA, 2008. **300**(5): p. 555-70.
10. Panel on Antiretroviral Guidelines for Adults and Adolescents, *Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents*, December 1, 2009, Department of Health and Human Services. p. 1-161.
11. Duffy, S., L.A. Shackelton, and E.C. Holmes, *Rates of evolutionary change in viruses: patterns and determinants*. Nat Rev Genet, 2008. **9**(4): p. 267-76.
12. Johnson, V.A., et al., *Update of the drug resistance mutations in HIV-1: December 2009*. Top HIV Med, 2009. **17**(5): p. 138-45.
13. Marcelin, A.G., et al., *Resistance to novel drug classes*. Curr Opin HIV AIDS, 2009. **4**(6): p. 531-7.
14. Das, A.T. and B. Berkhout, *HIV-1 evolution: frustrating therapies, but disclosing molecular mechanisms*. Philos Trans R Soc Lond B Biol Sci, 2010. **365**(1548): p. 1965-73.
15. Menendez-Arias, L., *Molecular basis of human immunodeficiency virus drug resistance: an update*. Antiviral Res, 2010. **85**(1): p. 210-31.
16. Drag, M. and G.S. Salvesen, *Emerging principles in protease-based drug discovery*. Nat Rev Drug Discov, 2010. **9**(9): p. 690-701.
17. Walker, L.M. and D.R. Burton, *Rational antibody-based HIV-1 vaccine design: current approaches and future directions*. Curr Opin Immunol, 2010. **22**(3): p. 358-66.
18. Este, J.A. and T. Cihlar, *Current status and challenges of antiretroviral research and therapy*. Antiviral Res, 2010. **85**(1): p. 25-33.

19. Wensing, A.M., N.M. van Maarseveen, and M. Nijhuis, *Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance*. Antiviral Res, 2010. **85**(1): p. 59-74.
20. Naider, F. and J. Anglister, *Peptides in the treatment of AIDS*. Curr Opin Struct Biol, 2009. **19**(4): p. 473-82.
21. D'Souza, V. and M.F. Summers, *How retroviruses select their genomes*. Nat Rev Microbiol, 2005. **3**(8): p. 643-55.
22. Capon, D.J. and R.H. Ward, *The CD4-gp120 interaction and AIDS pathogenesis*. Annu Rev Immunol, 1991. **9**: p. 649-78.
23. Myszka, D.G., et al., *Energetics of the HIV gp120-CD4 binding reaction*. Proc Natl Acad Sci U S A, 2000. **97**(16): p. 9026-31.
24. Liu, J., et al., *Molecular architecture of native HIV-1 gp120 trimers*. Nature, 2008. **455**(7209): p. 109-13.
25. Wu, L., et al., *CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5*. Nature, 1996. **384**(6605): p. 179-83.
26. Mondor, I., et al., *Interactions among HIV gp120, CD4, and CXCR4: dependence on CD4 expression level, gp120 viral origin, conservation of the gp120 COOH- and NH2-termini and V1/V2 and V3 loops, and sensitivity to neutralizing antibodies*. Virology, 1998. **248**(2): p. 394-405.
27. Berger, E.A., P.M. Murphy, and J.M. Farber, *Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease*. Annu Rev Immunol, 1999. **17**: p. 657-700.
28. Melikyan, G.B., et al., *Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion*. J Cell Biol, 2000. **151**(2): p. 413-23.
29. Arhel, N., *Revisiting HIV-1 uncoating*. Retrovirology, 2010. **7**: p. 96.
30. Nisole, S. and A. Saib, *Early steps of retrovirus replicative cycle*. Retrovirology, 2004. **1**: p. 9.
31. Delelis, O., et al., *Integrase and integration: biochemical activities of HIV-1 integrase*. Retrovirology, 2008. **5**: p. 114.
32. Coiras, M., et al., *Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs*. Nat Rev Microbiol, 2009. **7**(11): p. 798-812.
33. Geeraert, L., G. Kraus, and R.J. Pomerantz, *Hide-and-peek: the challenge of viral persistence in HIV-1 infection*. Annu Rev Med, 2008. **59**: p. 487-501.
34. Colin, L. and C. Van Lint, *Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies*. Retrovirology, 2009. **6**: p. 111.
35. Frankel, A.D. and J.A. Young, *HIV-1: fifteen proteins and an RNA*. Annu Rev Biochem, 1998. **67**: p. 1-25.
36. Freed, E.O., *HIV-1 gag proteins: diverse functions in the virus life cycle*. Virology, 1998. **251**(1): p. 1-15.
37. Gottlinger, H.G., *The HIV-1 assembly machine*. Aids, 2001. **15 Suppl 5**: p. S13-20.
38. Linial, M.L. and A.D. Miller, *Retroviral RNA packaging: sequence requirements and implications*. Curr Top Microbiol Immunol, 1990. **157**: p. 125-52.

39. Gelderblom, H.R., *Assembly and morphology of HIV: potential effect of structure on viral function*. *Aids*, 1991. **5**(6): p. 617-37.
40. Malim, M.H. and M. Emerman, *HIV-1 accessory proteins--ensuring viral survival in a hostile environment*. *Cell Host Microbe*, 2008. **3**(6): p. 388-98.
41. Darlix, J.L., et al., *First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses*. *J Mol Biol*, 1995. **254**(4): p. 523-37.
42. Lapadat-Tapolsky, M., et al., *Possible roles of HIV-1 nucleocapsid protein in the specificity of proviral DNA synthesis and in its variability*. *J Mol Biol*, 1997. **268**(2): p. 250-60.
43. Dawson, L. and X.F. Yu, *The role of nucleocapsid of HIV-1 in virus assembly*. *Virology*, 1998. **251**(1): p. 141-57.
44. Druillennec, S., et al., *Evidence of interactions between the nucleocapsid protein NCp7 and the reverse transcriptase of HIV-1*. *J Biol Chem*, 1999. **274**(16): p. 11283-8.
45. Darlix, J.L., et al., *Nucleocapsid protein of human immunodeficiency virus as a model protein with chaperoning functions and as a target for antiviral drugs*. *Adv Pharmacol*, 2000. **48**: p. 345-72.
46. Krishnamoorthy, G., et al., *DNA condensation by the nucleocapsid protein of HIV-1: a mechanism ensuring DNA protection*. *Nucleic Acids Res*, 2003. **31**(18): p. 5425-32.
47. Azoulay, J., et al., *Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations*. *J Mol Biol*, 2003. **326**(3): p. 691-700.
48. Buckman, J.S., W.J. Bosche, and R.J. Gorelick, *Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA*. *J Virol*, 2003. **77**(2): p. 1469-80.
49. Alce, T.M. and W. Popik, *APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein*. *J Biol Chem*, 2004. **279**(33): p. 34083-6.
50. Egele, C., et al., *HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions*. *J Mol Biol*, 2004. **342**(2): p. 453-66.
51. Bampi, C., et al., *The chaperoning and assistance roles of the HIV-1 nucleocapsid protein in proviral DNA synthesis and maintenance*. *Int J Biochem Cell Biol*, 2004. **36**(9): p. 1668-86.
52. Cruceanu, M., et al., *Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins*. *Nucleic Acids Res*, 2006. **34**(2): p. 593-605.
53. Bourbigot, S., et al., *How the HIV-1 nucleocapsid protein binds and destabilises the (-)primer binding site during reverse transcription*. *J Mol Biol*, 2008. **383**(5): p. 1112-28.
54. Kafaie, J., et al., *Role of capsid sequence and immature nucleocapsid proteins p9 and p15 in Human Immunodeficiency Virus type 1 genomic RNA dimerization*. *Virology*, 2009. **385**(1): p. 233-44.

55. Dussupt, V., et al., *The nucleocapsid region of HIV-1 Gag cooperates with the PTAP and LYPXnL late domains to recruit the cellular machinery necessary for viral budding*. PLoS Pathog, 2009. **5**(3): p. e1000339.
56. Dussupt, V., et al., *Basic Residues in the Nucleocapsid Domain of Gag Are Critical for Late Events of HIV-1 Budding*. J Virol, 2011. **85**(5): p. 2304-15.
57. De Guzman, R.N., et al., *Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element*. Science, 1998. **279**(5349): p. 384-8.
58. Wills, J.W. and R.C. Craven, *Form, function, and use of retroviral gag proteins*. Aids, 1991. **5**(6): p. 639-54.
59. Oertle, S. and P.F. Spahr, *Role of the gag polyprotein precursor in packaging and maturation of Rous sarcoma virus genomic RNA*. J Virol, 1990. **64**(12): p. 5757-63.
60. Damgaard, C.K., H. Dyhr-Mikkelsen, and J. Kjems, *Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions*. Nucleic Acids Res, 1998. **26**(16): p. 3667-76.
61. Gallo, R.C., & Jay, G., eds., *The Human Retroviruses* 1991: Academic Press, New York.
62. Clever, J., C. Sasseti, and T.G. Parslow, *RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type I*. J Virol, 1995. **69**(4): p. 2101-9.
63. Clever, J.L., D.A. Eckstein, and T.G. Parslow, *Genetic dissociation of the encapsidation and reverse transcription functions in the 5' R region of human immunodeficiency virus type I*. J Virol, 1999. **73**(1): p. 101-9.
64. Clever, J.L. and T.G. Parslow, *Mutant human immunodeficiency virus type I genomes with defects in RNA dimerization or encapsidation*. J Virol, 1997. **71**(5): p. 3407-14.
65. McBride, M.S. and A.T. Panganiban, *The human immunodeficiency virus type I encapsidation site is a multipartite RNA element composed of functional hairpin structures*. J Virol, 1996. **70**(5): p. 2963-73.
66. McBride, M.S., M.D. Schwartz, and A.T. Panganiban, *Efficient encapsidation of human immunodeficiency virus type I vectors and further characterization of cis elements required for encapsidation*. J Virol, 1997. **71**(6): p. 4544-54.
67. Henderson, L.E., et al., *Gag proteins of the highly replicative MN strain of human immunodeficiency virus type I: posttranslational modifications, proteolytic processings, and complete amino acid sequences*. J Virol, 1992. **66**(4): p. 1856-65.
68. Williams, M.C., et al., *Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching*. Proc Natl Acad Sci U S A, 2001. **98**(11): p. 6121-6.
69. Muriaux, D., et al., *NCp7 activates HIV-1 RNA dimerization by converting a transient loop-loop complex into a stable dimer*. J Biol Chem, 1996. **271**(52): p. 33686-92.
70. Li, X., et al., *Human immunodeficiency virus Type I nucleocapsid protein (NCp7) directs specific initiation of minus-strand DNA synthesis primed by human*

- tRNA(Lys3) in vitro: studies of viral RNA molecules mutated in regions that flank the primer binding site.* J Virol, 1996. **70**(8): p. 4996-5004.
71. Guo, J., et al., *Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer.* J Virol, 2000. **74**(19): p. 8980-8.
 72. Zhang, Y., et al., *Analysis of the assembly function of the human immunodeficiency virus type 1 gag protein nucleocapsid domain.* J Virol, 1998. **72**(3): p. 1782-9.
 73. Zabransky, A., E. Hunter, and M. Sakalian, *Identification of a minimal HIV-1 gag domain sufficient for self-association.* Virology, 2002. **294**(1): p. 141-50.
 74. Lener, D., et al., *Involvement of HIV-1 nucleocapsid protein in the recruitment of reverse transcriptase into nucleoprotein complexes formed in vitro.* J Biol Chem, 1998. **273**(50): p. 33781-6.
 75. Henriet, S., et al., *Vif is a RNA chaperone that could temporally regulate RNA dimerization and the early steps of HIV-1 reverse transcription.* Nucleic Acids Res, 2007. **35**(15): p. 5141-53.
 76. Didierlaurent, L., et al., *The conserved N-terminal basic residues and zinc-finger motifs of HIV-1 nucleocapsid restrict the viral cDNA synthesis during virus formation and maturation.* Nucleic Acids Res, 2008. **36**(14): p. 4745-53.
 77. Houzet, L., et al., *Nucleocapsid mutations turn HIV-1 into a DNA-containing virus.* Nucleic Acids Res, 2008. **36**(7): p. 2311-9.
 78. de Rocquigny, H., et al., *The zinc fingers of HIV nucleocapsid protein NCp7 direct interactions with the viral regulatory protein Vpr.* J Biol Chem, 1997. **272**(49): p. 30753-9.
 79. de Rocquigny, H., et al., *Interactions of the C-terminus of viral protein R with nucleic acids are modulated by its N-terminus.* Eur J Biochem, 2000. **267**(12): p. 3654-60.
 80. Carteau, S., R.J. Gorelick, and F.D. Bushman, *Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein.* J Virol, 1999. **73**(8): p. 6670-9.
 81. Poljak, L., et al., *Analysis of NCp7-dependent activation of HIV-1 cDNA integration and its conservation among retroviral nucleocapsid proteins.* J Mol Biol, 2003. **329**(3): p. 411-21.
 82. Liu, B., et al., *Interaction of the human immunodeficiency virus type 1 nucleocapsid with actin.* J Virol, 1999. **73**(4): p. 2901-8.
 83. Chatel-Chaix, L., et al., *Identification of Staufen in the human immunodeficiency virus type 1 Gag ribonucleoprotein complex and a role in generating infectious viral particles.* Mol Cell Biol, 2004. **24**(7): p. 2637-48.
 84. Ott, D.E., *Potential roles of cellular proteins in HIV-1.* Rev Med Virol, 2002. **12**(6): p. 359-74.
 85. Cen, S., et al., *The interaction between HIV-1 Gag and APOBEC3G.* J Biol Chem, 2004. **279**(32): p. 33177-84.
 86. Berthoux, L., C. Pechoux, and J.L. Darlix, *Multiple effects of an anti-human immunodeficiency virus nucleocapsid inhibitor on virus morphology and replication.* J Virol, 1999. **73**(12): p. 10000-9.

87. Lin, Y. and P.N. Borer, *Variation in 1700 NCp7 sequences (Unpublished Data)*. 2002.
88. South, T.L. and M.F. Summers, *Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC*. *Protein Sci*, 1993. **2**(1): p. 3-19.
89. Summers, M.F., et al., *Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1*. *Protein Sci*, 1992. **1**(5): p. 563-74.
90. Delahunty, M.D., et al., *Nucleic acid interactive properties of a peptide corresponding to the N-terminal zinc finger domain of HIV-1 nucleocapsid protein*. *Biochemistry*, 1992. **31**(28): p. 6461-9.
91. Gorelick, R.J., et al., *Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence*. *Proc Natl Acad Sci U S A*, 1988. **85**(22): p. 8420-4.
92. Berg, J.M. and Y. Shi, *The galvanization of biology: a growing appreciation for the roles of zinc*. *SCIENCE*, 1996. **271**(5252): p. 1081-5.
93. Ramboarina, S., et al., *Target specificity of human immunodeficiency virus type 1 NCp7 requires an intact conformation of its CCHC N-terminal zinc finger*. *J Virol*, 2004. **78**(12): p. 6682-7.
94. Mark-Danieli, M., et al., *Single point mutations in the zinc finger motifs of the human immunodeficiency virus type 1 nucleocapsid alter RNA binding specificities of the gag protein and enhance packaging and infectivity*. *J Virol*, 2005. **79**(12): p. 7756-67.
95. Grigorov, B., et al., *Intracellular HIV-1 Gag localization is impaired by mutations in the nucleocapsid zinc fingers*. *Retrovirology*, 2007. **4**: p. 54.
96. Aldovini, A. and R.A. Young, *Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus*. *J Virol*, 1990. **64**(5): p. 1920-6.
97. Dupraz, P., et al., *Point mutations in the proximal Cys-His box of Rous sarcoma virus nucleocapsid protein*. *J Virol*, 1990. **64**(10): p. 4978-87.
98. Lee, N., R.J. Gorelick, and K. Musier-Forsyth, *Zinc finger-dependent HIV-1 nucleocapsid protein-TAR RNA interactions*. *Nucleic Acids Res*, 2003. **31**(16): p. 4847-55.
99. Heath, M.J., et al., *Differing roles of the N- and C-terminal zinc fingers in human immunodeficiency virus nucleocapsid protein-enhanced nucleic acid annealing*. *J Biol Chem*, 2003. **278**(33): p. 30755-63.
100. Summers, M.F., *Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance*. *J Cell Biochem*, 1991. **45**(1): p. 41-8.
101. Demene, H., et al., *¹H NMR structure and biological studies of the His23-->Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity*. *Biochemistry*, 1994. **33**(39): p. 11707-16.

102. Dannull, J., et al., *Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues.* Embo J, 1994. **13**(7): p. 1525-33.
103. Tanchou, V., et al., *Role of the N-terminal zinc finger of human immunodeficiency virus type 1 nucleocapsid protein in virus structure and replication.* J Virol, 1998. **72**(5): p. 4442-7.
104. Schmalzbauer, E., et al., *Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro.* J Virol, 1996. **70**(2): p. 771-7.
105. Muller, G., et al., *Amino acid requirements of the nucleocapsid protein of HIV-1 for increasing catalytic activity of a Ki-ras ribozyme in vitro.* J Mol Biol, 1994. **242**(4): p. 422-9.
106. Cimarrelli, A., et al., *Basic residues in human immunodeficiency virus type 1 nucleocapsid promote virion assembly via interaction with RNA.* J Virol, 2000. **74**(7): p. 3046-57.
107. Lee, E.G. and M.L. Linial, *Basic residues of the retroviral nucleocapsid play different roles in gag-gag and Gag-Psi RNA interactions.* J Virol, 2004. **78**(16): p. 8486-95.
108. Ottmann, M., C. Gabus, and J.L. Darlix, *The central globular domain of the nucleocapsid protein of human immunodeficiency virus type 1 is critical for virion structure and infectivity.* J Virol, 1995. **69**(3): p. 1778-84.
109. Morellet, N., et al., *Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR.* J Mol Biol, 1994. **235**(1): p. 287-301.
110. Huang, Y., et al., *Effect of mutations in the nucleocapsid protein (NCp7) upon Pr160(gag-pol) and tRNA(Lys) incorporation into human immunodeficiency virus type 1.* J Virol, 1997. **71**(6): p. 4378-84.
111. Shubsda, M.F., et al., *Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein.* Biochemistry, 2002. **41**(16): p. 5276-82.
112. Mehellou, Y. and E. De Clercq, *Twenty-six years of anti-HIV drug discovery: where do we stand and where do we go?* J Med Chem, 2010. **53**(2): p. 521-38.
113. De Clercq, E., *New developments in anti-HIV chemotherapy.* Biochim Biophys Acta, 2002. **1587**(2-3): p. 258-75.
114. Rice, W.G., et al., *Inhibition of multiple phases of human immunodeficiency virus type 1 replication by a dithiane compound that attacks the conserved zinc fingers of retroviral nucleocapsid proteins.* Antimicrob Agents Chemother, 1997. **41**(2): p. 419-26.
115. Rice, W.G., et al., *Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein.* Nat Med, 1997. **3**(3): p. 341-5.
116. Vandeveld, M., et al., *ADA, a potential anti-HIV drug.* AIDS Res Hum Retroviruses, 1996. **12**(7): p. 567-8.
117. Huang, M., et al., *Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins.* J Med Chem, 1998. **41**(9): p. 1371-81.
118. Stephen, A.G., et al., *The nucleocapsid protein as a target for novel anti-HIV drugs.* Curr. Drug Disc., 2003(August): p. 33-36.

119. Goldschmidt, V., et al., *The nucleocapsid protein of HIV-1 as a promising therapeutic target for antiviral drugs*. HIV Therapy, 2010. **4**(2): p. 179-198.
120. Ji, X., G.J. Klarmann, and B.D. Preston, *Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity in vitro*. Biochemistry, 1996. **35**(1): p. 132-43.
121. Tanchou, V., et al., *Formation of stable and functional HIV-1 nucleoprotein complexes in vitro*. J Mol Biol, 1995. **252**(5): p. 563-71.
122. Vuilleumier, C., et al., *Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study*. Biochemistry, 1999. **38**(51): p. 16816-25.
123. Shubsda, M.F., et al., *Binding of human immunodeficiency virus type 1 nucleocapsid protein to psi-RNA-SL3*. Biophys Chem, 2000. **87**(2-3): p. 149-65.
124. Athavale, S.S., et al., *Effects of the nature and concentration of salt on the interaction of the HIV-1 nucleocapsid protein with SL3 RNA*. Biochemistry, 2010. **49**(17): p. 3525-33.
125. Kuiken, C., et al., eds. *HIV Sequence Compendium 2010*. 2010, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 10-03684.
126. Berkhout, B., K. Arts, and T.E. Abbink, *Ribosomal scanning on the 5'-untranslated region of the human immunodeficiency virus RNA genome*. Nucleic Acids Res, 2011.
127. Russell, R.S., C. Liang, and M.A. Wainberg, *Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably?* Retrovirology, 2004. **1**: p. 23.
128. Paillart, J.C., et al., *First snapshots of the HIV-1 RNA structure in infected cells and in virions*. J Biol Chem, 2004. **279**(46): p. 48397-403.
129. Watts, J.M., et al., *Architecture and secondary structure of an entire HIV-1 RNA genome*. Nature, 2009. **460**(7256): p. 711-6.
130. Berkhout, B., *Structure and function of the human immunodeficiency virus leader RNA*. Prog Nucleic Acid Res Mol Biol, 1996. **54**: p. 1-34.
131. Rana, T.M. and K.T. Jeang, *Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA*. Arch Biochem Biophys, 1999. **365**(2): p. 175-85.
132. Bannwarth, S. and A. Gatignol, *HIV-1 TAR RNA: the target of molecular interactions between the virus and its host*. Curr HIV Res, 2005. **3**(1): p. 61-71.
133. Ashe, M.P., L.H. Pearson, and N.J. Proudfoot, *The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site*. Embo J, 1997. **16**(18): p. 5752-63.
134. Das, A.T., B. Klaver, and B. Berkhout, *A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection*. J Virol, 1999. **73**(1): p. 81-91.
135. Kleiman, L., *tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1*. IUBMB Life, 2002. **53**(2): p. 107-14.
136. Hargittai, M.R., et al., *Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site*. J Mol Biol, 2004. **337**(4): p. 951-68.

137. Skripkin, E., et al., *Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro*. Proc Natl Acad Sci U S A, 1994. **91**(11): p. 4945-9.
138. Berkhout, B. and J.L. van Wamel, *Role of the DIS hairpin in replication of human immunodeficiency virus type 1*. J Virol, 1996. **70**(10): p. 6723-32.
139. Harrison, G.P. and A.M. Lever, *The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure*. J Virol, 1992. **66**(7): p. 4144-53.
140. Baudin, F., et al., *Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains*. J Mol Biol, 1993. **229**(2): p. 382-97.
141. Hayashi, T., et al., *RNA packaging signal of human immunodeficiency virus type 1*. Virology, 1992. **188**(2): p. 590-9.
142. Berglund, J.A., B. Charpentier, and M. Rosbash, *A high affinity binding site for the HIV-1 nucleocapsid protein*. Nucleic Acids Res, 1997. **25**(5): p. 1042-9.
143. Hayashi, T., Y. Ueno, and T. Okamoto, *Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1*. FEBS Lett, 1993. **327**(2): p. 213-8.
144. Dietz, J., et al., *Inhibition of HIV-1 by a peptide ligand of the genomic RNA packaging signal Psi*. ChemMedChem, 2008. **3**(5): p. 749-55.
145. Russell, R.S., et al., *Sequences downstream of the 5' splice donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA*. J Virol, 2003. **77**(1): p. 84-96.
146. Fisher, R.J., et al., *Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides*. Nucleic acids research, 2006. **34**(2): p. 472-84.
147. Gorelick, R.J., et al., *The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent*. JOURNAL OF VIROLOGY, 1993. **67**(7): p. 4027-36.
148. Dorfman, T., et al., *Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein*. J Virol, 1993. **67**(10): p. 6159-69.
149. Berthoux, L., et al., *Mutations in the N-terminal domain of human immunodeficiency virus type 1 nucleocapsid protein affect virion core structure and proviral DNA synthesis*. J Virol, 1997. **71**(9): p. 6973-81.
150. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J Comput Chem, 2004. **25**(13): p. 1605-12.
151. Carter, P., *Site-directed mutagenesis*. Biochem J, 1986. **237**(1): p. 1-7.
152. Stratagene, *QuikChange II Site-Directed Mutagenesis Kit Instruction Manual*2005. p. 1-22.
153. Marinus, M.G., *DNA methylation in Escherichia coli*. Annu Rev Genet, 1987. **21**: p. 113-31.
154. Razin, A. and A.D. Riggs, *DNA methylation and gene function*. SCIENCE, 1980. **210**(4470): p. 604-10.
155. Szyf, M., et al., *Studies on the biological role of DNA methylation: V. The pattern of E.coli DNA methylation*. Nucleic Acids Res, 1982. **10**(22): p. 7247-59.

156. Geier, G.E. and P. Modrich, *Recognition sequence of the dam methylase of Escherichia coli K12 and mode of cleavage of Dpn I endonuclease*. J Biol Chem, 1979. **254**(4): p. 1408-13.
157. Cunningham, B.C. and J.A. Wells, *High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis*. SCIENCE, 1989. **244**(4908): p. 1081-5.
158. You, J.C. and C.S. McHenry, *HIV nucleocapsid protein. Expression in Escherichia coli, purification, and characterization*. J Biol Chem, 1993. **268**(22): p. 16519-27.
159. Lee, B.M., et al., *Dynamical behavior of the HIV-1 nucleocapsid protein*. J Mol Biol, 1998. **279**(3): p. 633-49.
160. Getz, E.B., et al., *A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry*. Anal Biochem, 1999. **273**(1): p. 73-80.
161. Roepstorff, P., *MALDI-TOF mass spectrometry in protein chemistry*. EXS, 2000. **88**: p. 81-97.
162. Gross, J.H., *Mass Spectrometry - A Textbook*. 2nd ed2011, Heidelberg: Springer. 1-770.
163. Roepstorff, P., *Mass spectrometry in the analysis of peptides and proteins, past and present*. Methods Mol Biol, 1996. **61**: p. 1-7.
164. Philo, J.S., *Is any measurement method optimal for all aggregate sizes and types?* AAPS J, 2006. **8**(3): p. E564-71.
165. Paoletti, A.C., et al., *Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1*. Biochemistry, 2002. **41**(51): p. 15423-8.
166. Sambrook, J. and D.W. Russell, *Molecular cloning : a laboratory manual*. 3rd ed2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
167. Studier, F.W., et al., *Use of T7 RNA polymerase to direct expression of cloned genes*. Methods Enzymol, 1990. **185**: p. 60-89.
168. QIAGEN, *QIAprep Miniprep Handbook*2006: QIAGEN. 1-54.
169. Chenna, R., et al., *Multiple sequence alignment with the Clustal series of programs*. Nucleic Acids Res, 2003. **31**(13): p. 3497-500.
170. Stratagene, *BL21(DE3) Competent Cells, BL21(DE3)pLysS Competent Cells, and BL21 Competent Cells Instruction Manual*2001. p. 1-14.
171. Putnam, C. *Protein Calculator v3.3*. Available from: <http://www.scripps.edu/cgi-bin/cdputnam/protcalc3>.
172. Gill, S.C. and P.H. von Hippel, *Calculation of protein extinction coefficients from amino acid sequence data*. Anal Biochem, 1989. **182**(2): p. 319-26.
173. Schuck, P., *Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling*. Biophys J, 2000. **78**(3): p. 1606-19.
174. Sharp, P.M., et al., *Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity*. Nucleic Acids Res, 1988. **16**(17): p. 8207-11.
175. Haas, J., E.C. Park, and B. Seed, *Codon usage limitation in the expression of HIV-1 envelope glycoprotein*. Curr Biol, 1996. **6**(3): p. 315-24.

176. Novagen pET Vector Table. Available from: http://www.emdchemicals.com/life-science-research/vector-table-novagen-pet-vector-table/c_HdSb.s1O77QAAAEhPqsLdcab.
177. Burgess, R.R., *Protein precipitation techniques*. Methods Enzymol, 2009. **463**: p. 331-42.
178. Eftink, M.R., *Intrinsic Fluorescence of Proteins*, in *Topics in Fluorescence Spectroscopy*, J.R. Lakowicz, Editor 2002, Springer US. p. 1-15.
179. Burstein, E.A., N.S. Vedenkina, and M.N. Ivkova, *Fluorescence and the location of tryptophan residues in protein molecules*. Photochem Photobiol, 1973. **18**(4): p. 263-79.
180. Vivian, J.T. and P.R. Callis, *Mechanisms of tryptophan fluorescence shifts in proteins*. Biophys J, 2001. **80**(5): p. 2093-109.
181. Feig, A.L., *Applications of isothermal titration calorimetry in RNA biochemistry and biophysics*. Biopolymers, 2007. **87**(5-6): p. 293-301.
182. Freyer, M.W. and E.A. Lewis, *Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions*. Methods Cell Biol, 2008. **84**: p. 79-113.
183. MicroCal, L., *VP-ITC MicroCalorimeter User's Manual*, ed. L. MicroCal2002, Northampton, MA: MicroCal, LLC.
184. Wiseman, T., et al., *Rapid measurement of binding constants and heats of binding using a new titration calorimeter*. Anal Biochem, 1989. **179**(1): p. 131-7.
185. Pierce, M.M., C.S. Raman, and B.T. Nall, *Isothermal titration calorimetry of protein-protein interactions*. Methods, 1999. **19**(2): p. 213-21.
186. Salim, N.N. and A.L. Feig, *Isothermal titration calorimetry of RNA*. Methods, 2009. **47**(3): p. 198-205.
187. Prats, A.C., et al., *Viral RNA annealing activities of the nucleocapsid protein of Moloney murine leukemia virus are zinc independent*. Nucleic Acids Res, 1991. **19**(13): p. 3533-41.
188. Surovoy, A., et al., *Conformational and nucleic acid binding studies on the synthetic nucleocapsid protein of HIV-1*. J Mol Biol, 1993. **229**(1): p. 94-104.
189. Mori, M., et al., *Molecular dynamics and DFT study on HIV-1 nucleocapsid protein-7 in complex with viral genome*. J Chem Inf Model, 2010. **50**(4): p. 638-50.
190. Dong, F. and H.X. Zhou, *Electrostatic contributions to T4 lysozyme stability: solvent-exposed charges versus semi-buried salt bridges*. Biophysical journal, 2002. **83**(3): p. 1341-7.
191. MicroCal, L., *ITC Data Analysis in Origin Tutorial Guide Version 7.0*, ed. L. MicroCal2004, Northampton, MA: MicroCal, LLC.
192. DeCiantis, C.L., *A FRET based bistable oligonucleotide switch, AlloSwitch, designed for specific recognition of HIV-1 NCp7 and use in High Throughput Screening*, in *Chemistry Department2008*, Syracuse University: Syracuse. p. 190.
193. Crick, F.H., et al., *General nature of the genetic code for proteins*. Nature, 1961. **192**: p. 1227-32.
194. Crick, F., *Central dogma of molecular biology*. Nature, 1970. **227**(5258): p. 561-3.

195. Joyce, G.F., *Nucleic acid enzymes: playing with a fuller deck*. Proc Natl Acad Sci U S A, 1998. **95**(11): p. 5845-7.
196. Fiammengo, R. and A. Jaschke, *Nucleic acid enzymes*. Curr Opin Biotechnol, 2005. **16**(6): p. 614-21.
197. Ellington, A.D., *RNA selection. Aptamers achieve the desired recognition*. Curr Biol, 1994. **4**(5): p. 427-9.
198. Syed, M.A. and S. Pervaiz, *Advances in aptamers*. Oligonucleotides, 2010. **20**(5): p. 215-24.
199. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands*. Nature, 1990. **346**(6287): p. 818-22.
200. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. SCIENCE, 1990. **249**(4968): p. 505-10.
201. Bock, L.C., et al., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin*. Nature, 1992. **355**(6360): p. 564-6.
202. Potyrailo, R.A., et al., *Adapting selected nucleic acid ligands (aptamers) to biosensors*. Anal Chem, 1998. **70**(16): p. 3419-25.
203. Stojanovic, M.N., P. de Prada, and D.W. Landry, *Aptamer-based folding fluorescent sensor for cocaine*. J Am Chem Soc, 2001. **123**(21): p. 4928-31.
204. Pandana, H., K.H. Aschenbach, and R.D. Gomez, *Systematic aptamer-gold nanoparticle colorimetry for protein detection: Thrombin*. Ieee Sensors Journal, 2008. **8**(5-6): p. 661-666.
205. Hicke, B.J., et al., *Tumor targeting by an aptamer*. J Nucl Med, 2006. **47**(4): p. 668-78.
206. Gnanam, A.J., et al., *Development of aptamers specific for potential diagnostic targets in Burkholderia pseudomallei*. Trans R Soc Trop Med Hyg, 2008. **102 Suppl 1**: p. S55-7.
207. Ng, E.W. and A.P. Adamis, *Anti-VEGF aptamer (pegaptanib) therapy for ocular vascular diseases*. Ann N Y Acad Sci, 2006. **1082**: p. 151-71.
208. Chen, F., et al., *Aptamer from whole-bacterium SELEX as new therapeutic reagent against virulent Mycobacterium tuberculosis*. Biochem Biophys Res Commun, 2007. **357**(3): p. 743-8.
209. Thomas, M., et al., *Selective targeting and inhibition of yeast RNA polymerase II by RNA aptamers*. J Biol Chem, 1997. **272**(44): p. 27980-6.
210. Tuerk, C., S. MacDougall, and L. Gold, *RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase*. Proc Natl Acad Sci U S A, 1992. **89**(15): p. 6988-92.
211. Gold, L., et al., *Diversity of oligonucleotide functions*. Annu Rev Biochem, 1995. **64**: p. 763-97.
212. Bruno, J.G. and J.L. Kiel, *In vitro selection of DNA aptamers to anthrax spores with electrochemiluminescence detection*. Biosens Bioelectron, 1999. **14**(5): p. 457-64.
213. Charlton, J., J. Sennello, and D. Smith, *In vivo imaging of inflammation using an aptamer inhibitor of human neutrophil elastase*. Chem Biol, 1997. **4**(11): p. 809-16.

214. Fitter, S. and R. James, *Deconvolution of a complex target using DNA aptamers*. J Biol Chem, 2005. **280**(40): p. 34193-201.
215. Hesselberth, J., et al., *In vitro selection of nucleic acids for diagnostic applications*. J Biotechnol, 2000. **74**(1): p. 15-25.
216. Homann, M. and H.U. Goring, *Combinatorial selection of high affinity RNA ligands to live African trypanosomes*. Nucleic Acids Res, 1999. **27**(9): p. 2006-14.
217. Jayasena, S.D., *Aptamers: an emerging class of molecules that rival antibodies in diagnostics*. Clin Chem, 1999. **45**(9): p. 1628-50.
218. Morris, K.N., et al., *High affinity ligands from in vitro selection: complex targets*. Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2902-7.
219. Proske, D., et al., *Aptamers--basic research, drug development, and clinical applications*. Appl Microbiol Biotechnol, 2005. **69**(4): p. 367-74.
220. Gopinath, S.C.B., *Aptamers*. Encyclopedia of Analytical Chemistry 2006: John Wiley & Sons, Ltd.
221. Nimjee, S.M., C.P. Rusconi, and B.A. Sullenger, *Aptamers: an emerging class of therapeutics*. Annu Rev Med, 2005. **56**: p. 555-83.
222. Macaya, R.F., et al., *Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution*. Proc Natl Acad Sci U S A, 1993. **90**(8): p. 3745-9.
223. Feigon, J., T. Dieckmann, and F.W. Smith, *Aptamer structures from A to zeta*. Chem Biol, 1996. **3**(8): p. 611-7.
224. Dieckmann, T., et al., *Solution structure of an ATP-binding RNA aptamer reveals a novel fold*. Rna, 1996. **2**(7): p. 628-40.
225. Jiang, F., et al., *Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex*. Nature, 1996. **382**(6587): p. 183-6.
226. Watson, J.D., et al., *Molecular Biology of the Gene*. 6th Ed ed 2007: Benjamin Cummings. 880.
227. Pan, T., A. Loria, and K. Zhong, *Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA*. Proc Natl Acad Sci U S A, 1995. **92**(26): p. 12510-4.
228. Kramer, F.R., et al., *Evolution in vitro: sequence and phenotype of a mutant RNA resistant to ethidium bromide*. J Mol Biol, 1974. **89**(4): p. 719-36.
229. *Nucleic Acid-based Technologies*. Available from: <http://www.molgen.mpg.de/~nabt/background.html>.
230. Kupakuwana, G.V., et al., *Acyclic Selection of Aptamers for Human alpha-Thrombin Using Over-represented Libraries and Deep Sequencing*. PLoS One, 2011. **6**(5): p. 1-11.
231. McPike, M.P., *Personal communication*, 2011.
232. Chen, L., *Application of High Throughput Sequencing in Selection of RNA Aptamers*, in *Program of Structural Biology, Biochemistry and Biophysics 2011*, Syracuse University. College of Arts and Sciences.: Syracuse, NY.
233. Mardis, E.R., *Next-generation DNA sequencing methods*. Annu Rev Genomics Hum Genet, 2008. **9**: p. 387-402.
234. Metzker, M.L., *Sequencing technologies - the next generation*. Nat Rev Genet, 2010. **11**(1): p. 31-46.

235. Diehl, F., et al., *BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions*. Nat Methods, 2006. **3**(7): p. 551-9.
236. Li, M., et al., *BEAMing up for detection and quantification of rare sequence variants*. Nat Methods, 2006. **3**(2): p. 95-7.
237. Macaya, R.F., et al., *Structural and functional characterization of potent antithrombotic oligonucleotides possessing both quadruplex and duplex motifs*. Biochemistry, 1995. **34**(13): p. 4478-92.
238. Connor, A.C. and L.B. McGown, *Aptamer stationary phase for protein capture in affinity capillary chromatography*. J Chromatogr A, 2006. **1111**(2): p. 115-9.
239. Joyce, M.V. and L.B. McGown, *Detection of G-quartet structure in a DNA aptamer stationary phase using a fluorescent dye*. Appl Spectrosc, 2004. **58**(7): p. 831-5.
240. Cai, H., T.M.-H. Lee, and I.M. Hsing, *Label-free protein recognition using an aptamer-based impedance measurement assay*. Sensors and Actuators B: Chemical, 2006. **114**(1): p. 433-437.
241. Church, G.M. and C. Benoit. *Polony Sequencing Protocols*. 2008; Available from: http://openwetware.org/wiki/Church_Lab:PoloProt.
242. Moore, M.J. and C.C. Query, *Joining of RNAs by splinted ligation*. Methods Enzymol, 2000. **317**: p. 109-23.
243. Porreca, G.J., J. Shendure, and G.M. Church, *Polony DNA sequencing*. Curr Protoc Mol Biol, 2006. **Chapter 7**: p. Unit 7 8.
244. Mitra, R.D. and G.M. Church, *In situ localized amplification and contact replication of many individual DNA molecules*. Nucleic Acids Res, 1999. **27**(24): p. e34.
245. Pemov, A., et al., *DNA analysis with multiplex microarray-enhanced PCR*. Nucleic Acids Res, 2005. **33**(2): p. e11.
246. Shapero, M.H., et al., *SNP genotyping by multiplexed solid-phase amplification and fluorescent minisequencing*. Genome Res, 2001. **11**(11): p. 1926-34.
247. Shendure, J., et al., *Accurate multiplex polony sequencing of an evolved bacterial genome*. SCIENCE, 2005. **309**(5741): p. 1728-32.
248. Chen, L., *Effect of Fluorophore- or Biotin-labeling on the Affinity of Thrombin for Thrombin-binding Aptamer (Unpublished Data)*. 2011, Syracuse University: Syracuse, NY.
249. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction*. Nucleic Acids Res, 2003. **31**(13): p. 3406-15.

Biographical Data

NAME OF AUTHOR: Wei Ouyang

PLACE OF BIRTH: Pingxiang, Jiangxi Province, P. R. China

DATE OF BIRTH: November 5th, 1977

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Syracuse University, Syracuse, New York

Nanjing University, Nanjing, Jiangsu Province, P. R. China

DEGREES AWARDED:

Master of Science in Plant Sciences, 2001, Nanjing University

Bachelor of Science in Biological Pharmacy, 1998, Nanjing University

AWARDS AND HONORS:

Syracuse University	CASE Renaissance Scholarship	2008-2011
---------------------	------------------------------	-----------

Syracuse University	Graduate Assistantship	2003-2008
---------------------	------------------------	-----------

Nanjing University	Graduate Assistantship	1998-2001
--------------------	------------------------	-----------

Nanjing University	People's Scholarship	1994-1998
--------------------	----------------------	-----------

PROFESSIONAL EXPERIENCE:

Teaching Assistant	Dept. of Chemistry, Syracuse University	2005-2006
--------------------	---	-----------

Research Associate	Shanghai Mendel Gene Research Co., Ltd.	2002-2003
--------------------	---	-----------

Research Associate	Shanghai HealthDigit Co., Ltd., China	2001-2002
--------------------	---------------------------------------	-----------