Syracuse University

SURFACE

Dissertations - ALL

SURFACE

12-2013

Recombinant Expression, Purification, and Characterization of an HIV-1 Tat-Human Cyclin T1 Chimera

Collin Lesley Fischer

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

Part of the Biochemistry, Biophysics, and Structural Biology Commons

Abstract

Efficient transcription of the human immunodeficiency virus type 1 (HIV-1) requires the interaction of the viral protein Tat with the trans-activation response (TAR) stem-loop of the long-terminal repeat (LTR) portion of nascent viral RNA. The production of viable transcripts is enhanced dramatically by the interaction of HIV-1 Tat with the host protein human Cyclin T1. Interaction with hCycT1 remodels Tat protein contributing a single cysteine residue that is critical to the formation of the second of two zinc fingers (Zn2). Here we suggest that it is the presence of this critical cysteine residue and not the presence of arginine residues from human Cyclin T1 that imparts high affinity and specificity to the interaction with HIV-1 TAR RNA. Crucial structural features of this interaction remain unresolved by NMR or existing crystal structures. Specifically, the structure of the Tat activation domain (AD), and Tat interaction with hCycT1 while bound to HIV-1 TAR RNA remain elusive. Much of the difficulty in obtaining structural data is a result of the notoriously difficult expression of native HIV-1 Tat caused in large part by the high cysteine count, and poor solubility of the Tat protein. This work presents a protocol for the expression and purification of a high affinity recombinant chimeric protein which includes the full 101 amino acid Tat protein fused to an essential minimal portion of CycT1m) necessary for TAR binding in sufficient purity and concentration for structural study by nuclear magnetic resonance (NMR).The elucidation of this critical region has the potential for profound impact in the structural based drug design of HIV-1 therapeutics.

Recombinant Expression, Purification, and Characterization of an HIV-1 Tat-Human Cyclin T1 Chimera

By

Collin L. Fischer

B.A. Oswego State University 2003

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

Syracuse University

December 2013

Acknowledgement

I would like to thank Syracuse University for affording me the opportunity to educate myself, and to pursue my lifelong dream of becoming a scientist. I know of no pursuit which could have been in better agreement with my insatiable curiosity, and my heartfelt desire to leave this world somehow better off for my having been a part of it. I am truly grateful to all who have endowed me with this potential through their support and assistance.

I would like to thank my advisor Dr. Philip Borer who has patiently guided me in my academic efforts, but who has most importantly cultivated my intellectual and professional independence and self-sufficiency. Though there were many difficult moments, I now trust in my own abilities more than I ever imagined I would.

With respect to technical guidance I would like to thank Dr. Mark Braiman for many helpful conversations, digressing through many different topics, and with much of what I gained having been only fully realized later. I trust these conversations will continue to serve me well in the future. I would also like to thank Dr. Michael Cosgrove for his kind assistance on the subject of protein expression. From these conversations I gained not only technical knowledge, but found a kindred difficult protein expression spirit, and the courage to forge ahead. To Dr. Thomas Duncan from Upstate Medical University I am grateful for many hours of kind assistance and instruction in biolayer interferometry and

for the use of his ForteBio Octet instrument. To Dr. Roy Welch I am grateful for his instruction, and his illuminating conversation on the psychology of the graduate student which he delivered at a very timely point in my journey.

To my colleagues around the lab I am grateful to Dr. Mark McPike, Dr. Wei Ouyang, Dr. Lei Chen, Dr. Lingchun Yang, Dr. Gillian Kupakuwana, Dr. Shreyas Athavale, Dr. Chris DeCiantis, Jamie Crill II, Raghuvaran Iyer, and Delores James for their kind assistance, inspiration, and friendship. To Dr. Deborah Kerwood, I am grateful for all these things and more.

On a personal level, I could not have survived without the patience and support of my son Nolan Ford, and my fiancé William Myers. It is my sincere wish that what I have done here will help me to do for you what you both have already done for me.

Finally, to my father James R. Fischer, I am deeply saddened that you are not with me now, but all who knew you know that it was your love of science, your curiosity, and your integrity that have been my inspiration all along.

"We are star stuff..." –Carl Sagan

Table of Contents

ix

Appendices

List of Tables

Table of Figures

xvii

Introduction

The focus of this work is the recombinant expression, purification, and characterization of a difficult to express non-membrane chimeric protein comprised of the HIV-1 Tat regulatory protein fused to an integral portion of its *in vivo* host binding partner the human protein Cyclin T1. The first portion of this work will focus on the rationale behind the importance, design, and potential applications of this particular chimera, as well as the cloning experiments involved in producing the chimera, the conditions found necessary to achieve adequate yield, and the methods required for purification. The latter portion of this work will focus primarily on characterizing binding interaction through the use of binding assays involving the *in vivo* TAR HIV-1 nucleic acid binding partner of the proteins, and the expression and purification protocol modifications necessary to obtain highly concentrated labeled protein for structural experiments. The ultimate goal of this work is to provide data that will aid in the process of drug discovery for the efficacious treatment of HIV-1 by improving our structural understanding of an, as yet, unexploited drug target the HIV-1 Tat regulatory protein.

All of the presently employed HIV therapeutics that target viral proteins target the Env and Pol proteins that are expressed late in the viral infection cycle (1,2). Regulatory proteins Tat and Rev are expressed early in the infection cycle, as is the accessory protein Nef (3). Inhibitors that target proteins early in viral infection cycle have the potential to

be significantly more effective in reducing viral load (2), and may also be less prone to

the development of resistant strains.

Chapter 1 The Tat-TAR Interaction

The HIV-1 regulatory protein Trans-Activator of Transcription (Tat), at its full length, is a 101 amino acid, two zinc finger protein that interacts with the TAR stem-loop located within the 5' Long Terminal Repeat (LTR) region of nascent viral RNA. After synthesis, a nuclear localization signal facilitates transport of Tat into the nucleus (4) where, in concert with other host proteins, Tat facilitates the production of a large number of full length viral RNA transcripts (Figure 1-1). In the absence of Tat the nascent viral RNA transcript is prematurely terminated and the production of viable viral RNA is substantially reduced (5). The potential impact of the interruption of this essential interaction on viral load makes the Tat-TAR interaction an attractive target for drug discovery (6). As of this writing however, there are no FDA approved drugs targeting the inhibition of this important interaction.

1.1. Tat Domains

The Tat protein consists of two exons, and five domain regions (Figure 1-2) (7). The first three domains, from amino acids 1-48, form the Activation Domain (AD) that interacts with the host protein human cyclin T1 comprised of a highly ordered and acidic minimal activation region rich in proline (1-20), a cysteine rich zinc finger region (containing seven highly conserved cysteine residues) (20-40), and a highly conserved core region (40-48). A basic RNA Binding Domain (RBD) region (49-57) (including six arginine and two lysine residues) (Frankel et al. 1988), and a Glutamine rich region (57-72) interact

with HIV TAR RNA (7,8). Exon 2 of the Tat protein (73-101) contains a splicing silencer (ESS) that inhibits splicing of viral mRNA at an upstream 3' splice site (9), and does not play a primary role in transcription (8).

1.2. Tat-P-TEFb Interaction

 In vivo Tat interacts with the human host Positive Transcription Elongation Factor b (P-TEFb), which regulates transcription by RNA polymerase II (RNAP II) (10). The P-TEFb is comprised of a Cyclin dependent kinase CDK9, and the cell cycle regulatory protein Cyclin T1 (11,12). The Cyclin T1 protein is a 726 amino acid catalytic subunit of P-TEFb (Figure 1-3), and contains a purported Tat/TAR-recognition motif (TRM) at amino acids 250-272 (13). However, more recently other researchers have described the TRM as being located between residues 250 and 263 of Cyclin T1 (14). Since it is not yet clear which of these regions is the more accurate the larger region is displayed in Figure 1-3. Cyclin T1 protein binds at the cysteine rich Activation Domain (AD) (1-48) of the Tat regulatory protein (14).

An intramolecular CHCC zinc finger (ZnF1) of Tat is formed by amino acids Cys22, His33, Cys34, and Cys37 (Figure 1-4). A second intermolecular CCCC zinc finger (ZnF2) is formed by Cys261 of Cyclin T1 in concert with three additional cysteine residues Cys25, Cys27, and Cys30 of Tat (Figure 1-5) (15). Both of the Tat zinc fingers are structurally dissimilar to other known zinc fingers and metalloproteins (7,8).

1.3. Recruiting P-TEFb

Acetylation of Tat Lys_{28} by the transcriptional coactivator p300/CREB-binding protein Associated Factor (PCAF) facilitates the recruitment of P-TEFb (16) bringing the CDK9 catalytic subunit in close proximity for hyperphosphorylation of the C terminal domain of RNA Polymerase II (Figure 1-6) (14,15). A large number of intermolecular hydrogen bonds between Tat and the P-TEFb alter Tat folding on contact with the complex (8). This interaction remodels the AD of Tat into an extended conformation with an unusually large area of interface with Cyclin T1 (7,8).

Both Tat and Cyclin T1 bind the TAR RNA stem-loop, purportedly, at a nucleotide bulge formed by a base triple in the stem, and at the apical portion of the stem-loop respectively (15,17). Tat binds the nucleotide bulge thorough an arginine-rich motif (ARM) at residues 49-57 (18). Such arginine-rich binding motifs are also found to regulate RNA binding in several other viral regulatory proteins lending to increased conservation in this region, and a high affinity and specific interaction (19). Binding of the P-TEFb-Tat complex to TAR is enhanced by the interaction between Cyclin T1 and Tat in the bridging of ZnF2 (8) with the contribution of Cyclin T1 C261 (Figure 1-5) (7,15).

1.4. Tat-TAR Structural Study

Much of what we know about the Tat-P-TEFb interaction can be attributed to the mutational and binding assays of Garber et al. 1998 who first suspected the formation of a zinc bridge between Tat and human Cyclin T1. Since then the crystal structure of HIV-

1 Tat complexed with P-TEFb reported by Tahirov et al. 2010 appears to confirm these early conclusions. Though ground breaking work, the Tahirov et al. 2010 structure is reported in the absence of TAR and the conformation of the Tat AD and the Tat-TAR complex remain obscure (8). In the Tahirov structure residues 50-86 of the arginine rich Tat RNA-binding domain are undefined, and the conformation of ZnF2 remains elusive in the absence of residues 252-260 of Cyclin T1 (Figure 1-7) (7,8).

The first NMR structures of the TAR-arginine complex reported by Puglisi et al. 1992, TAR as reported by Aboul-ela et al. 1995, and a 24 residue Tat peptide bound to TAR as reported by Long et al. 1999 provided much of the original insight into this interaction and remain our primary sources of structural detail from NMR. Yet in these as well, crucial structural features remain unclear, specifically that of the Tat Activation Domain and its interaction with Cyclin T1 in the binding of TAR RNA. Elucidation of these critical regions, as might be accomplished by additional NMR or X-ray crystallography work, that includes detailed conformations of both zinc fingers has the potential for profound impact on structure-based drug design.

1.5. TAR -The UCU Bulge

In 1992 Puglisi et al. published the NMR structure of a 31 nucleotide portion of the HIV-1 Transactivation-Response Region (TAR) RNA in complex with the arginine analog argininamide (20). The structure revealed a UCU nucleotide bulge at the Tat protein binding site, and a flexible six nucleotide apical loop. The highly conserved cis-acting

RNA regulatory element TAR binds to the Tat protein at a U23-A27-U38 base triple (Figure 1-8) (20,21). Functional groups of the pyrimidine bulge and the phosphate backbone are repositioned during binding resulting in a conformational change that is an essential feature of specific recognition (22). This dynamic interaction has been described as a "ligand-induced conformational rearrangement" that occurs on a micro- to millisecond timescale (23). The base triple itself undergoes conjugated pi bond stabilization by interaction with the guanidinium groups of arginines within the arginine rich basic RNA binding domain of Tat (20).

 In the unbound state the helical structure of TAR is distorted by the nucleotides of the bulge that stack within the stem. In the bound form the nucleotides of the bulge loop outward, and allow the bases above and below the bulge to stack coaxially (21). Work with conformationally restricted peptide mimics of Tat indicates that looping out of U23 and C24 on Tat binding induces the formation of a binding pocket that places the guanidinium group of arginine in proximity for hydrogen bonding to functional groups of G26 and U23 (22). Supporting this, mutation of critical nucleotides G26 or U23 has been shown to eliminate Tat binding and recognition (20,24).

1.6. TAR - The Apical Loop

The apical region of TAR is a six nucleotide loop that, due to high flexibility, does not lend itself readily to NMR spectroscopy. Dethoff et al. 2008 used mutational analysis, molecular dynamics, and NMR data to conclude that the binding at the bulge and at the

apical loop are largely independent dynamic events with the possibility of some long range interaction between them (25). The CUGGGA nucleotide sequence of the apical loop binds Cyclin T1 of P-TEFb producing additional conformational changes (17). However, much of the finer detail of this interaction remains unclear (20).

We do know from mutational analysis that nucleotides G32, which loops outward, and G34 of the apical loop are essential for Cyclin T1 binding, and that elimination, but not substitution, of A35 diminishes binding substantially (11,26). From NMR data it is apparent that Cyclin T1 interacts directly with U31 of the apical loop (25). Nucleotides A35 and G34 displace one another in alternation between looped in and looped out conformations (25). To date however, considerably more work has focused on characterizing the dynamics of the UCU bulge than on the dynamics of the apical region of TAR (25). Toward the goal of improving rational drug design, much remains to be gained from additional characterization of the apical loop and the recognition of its contribution to the structurally dynamic interaction between TAR RNA, and the Cyclin T1 and Tat proteins.
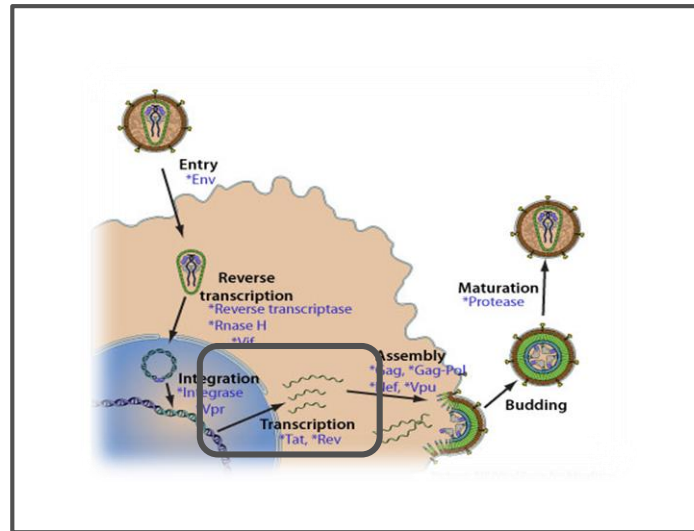
Figure 1-1 The HIV infection cycle.

The HIV infection cycle is comprised of nine principal events: binding, membrane fusion, entry, reverse transcription, integration, transcription, assembly, budding, and maturation. After synthesis, a nuclear localization signal facilitates transport of Tat into the nucleus (4) where, in concert with other host proteins, Tat facilitates the production of a large number of full length viral RNA transcripts.

Accessed online at: http://www.bioafrica.net/proteomics/HIVproteome.html

A
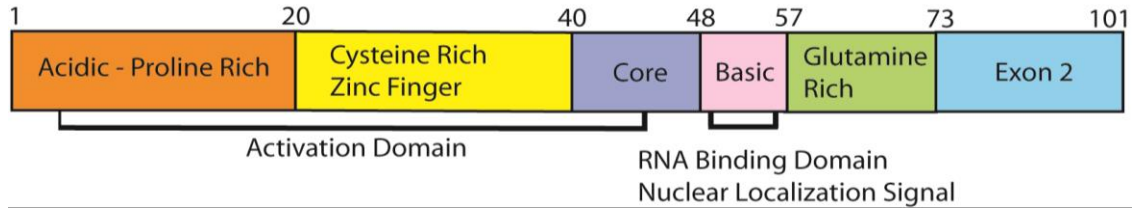
B

MEPVDPNLEP WKHPGSQPRT ACNNCYCKKC CFHCYACFTR KGLGISYGRK

KRRQRRRAPQ DSQTHQASLS KQPASQSRGD PTGPTESKKK VERETETDPF D

Figure 1-2 The Five Domains of the HIV-1 Protein Tat.

The HIV-1 Tat protein consists of two exons, and five domain regions (A) above. The first three domains of Tat, from amino acids 1-48, form the Activation Domain (AD) that interacts with the cellular protein human cyclin T1, increases specificity for TAR (27), and is comprised of a highly ordered and acidic minimal activation region rich in proline (1-20), a cysteine rich zinc finger region (20-40), and a highly conserved core region (40-48). A basic RNA Binding Domain (RBD) region (49-57), and a Glutamine rich region (57-72) interact with HIV TAR RNA. Exon 2 of the Tat protein (73-101) contains a splicing silencer (ESS) that inhibits splicing of viral mRNA at an upstream 3' splice site (9), and does not play a primary role in transcription (8). The amino acid sequence of Tat that corresponds to each of these domains appears in B above.
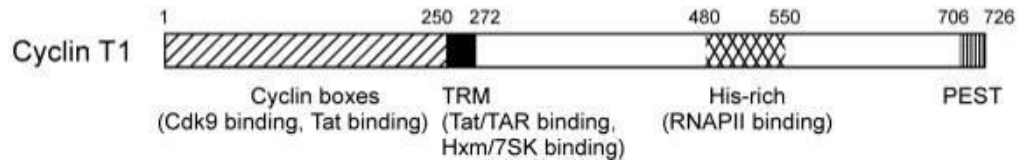
Figure 1-3 Cyclin T1.

The Cyclin T1 protein is a 726 amino acid catalytic subunit of P-TEFb, and contains a

Tat/TAR-recognition motif (TRM) at amino acids 250-272. The Cyclin T1 protein binds

HIV-1 Tat through the cysteine rich Activation Domain (1-48) (14). Reprinted from:

Retrovirology, vol. 5 Page 63, Copyright 2008 (13), with permission from J. Jadlowsky.
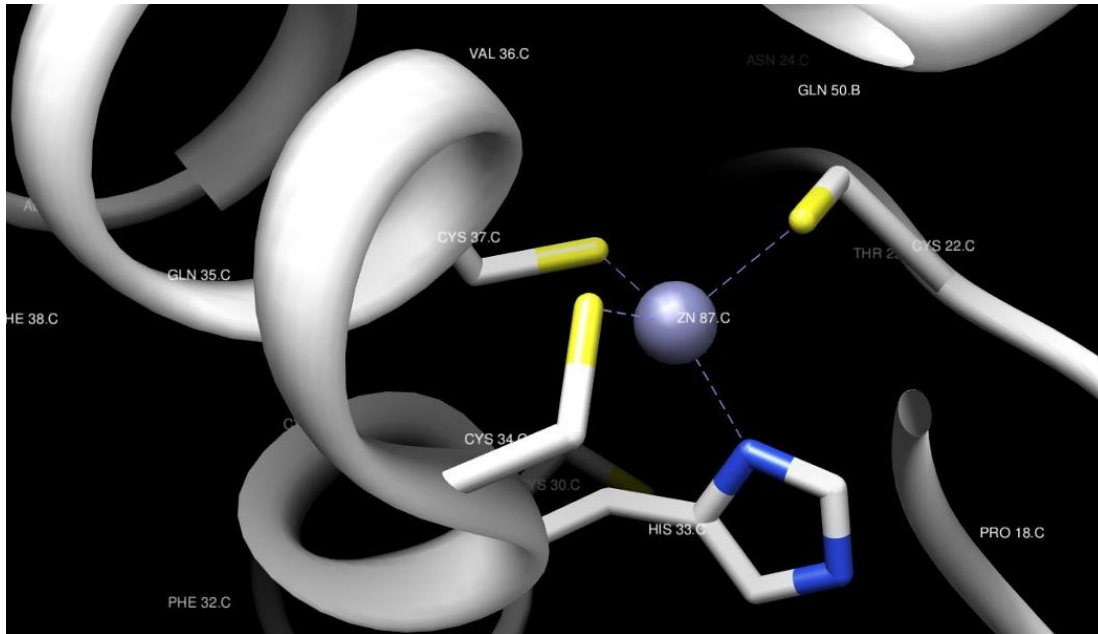
Figure 1-4 An intramolecular CHCC zinc finger (ZnF1)

An intramolecular CHCC zinc finger (ZnF1) is formed by residues Cys22, His33, Cys34, and Cys37 of the HIV-1 regulatory protein Tat. Protein Data Bank File: 3MI9 (7) Crystal structure of HIV-1 Tat complexed with human P-TEFb. Image rendered by C. Fischer from Protein Data Bank File: 3M19 using Insight II molecular modeling software.
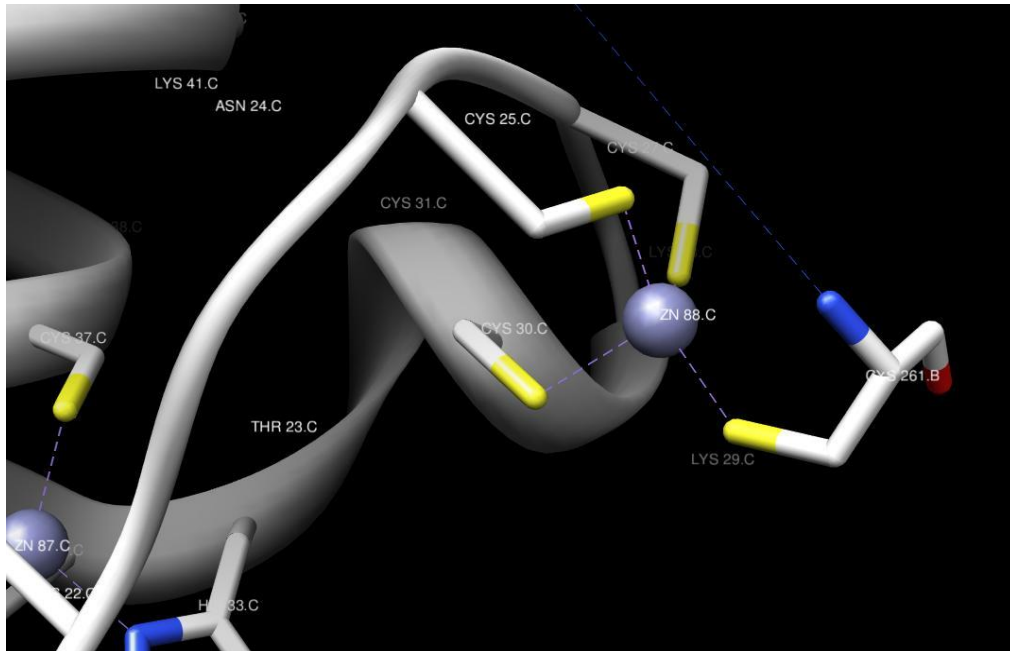
Figure 1-5 A CCCC intermolecular zinc finger (ZnF2)

A CCCC intermolecular zinc finger (ZnF2) is formed by the cysteine 261 of Cyclin T1 in
concert with three additional cysteine residues 25, 27, and 30 from the HIV-1 regulatory
protein Tat. Protein Data Bank File:3MI9 (7) Crystal structure of HIV-1 Tat complexed
with human P-TEFb. Image rendered by C. Fischer from Protein Data Bank File: 3M19
using Insight II molecular modeling software.
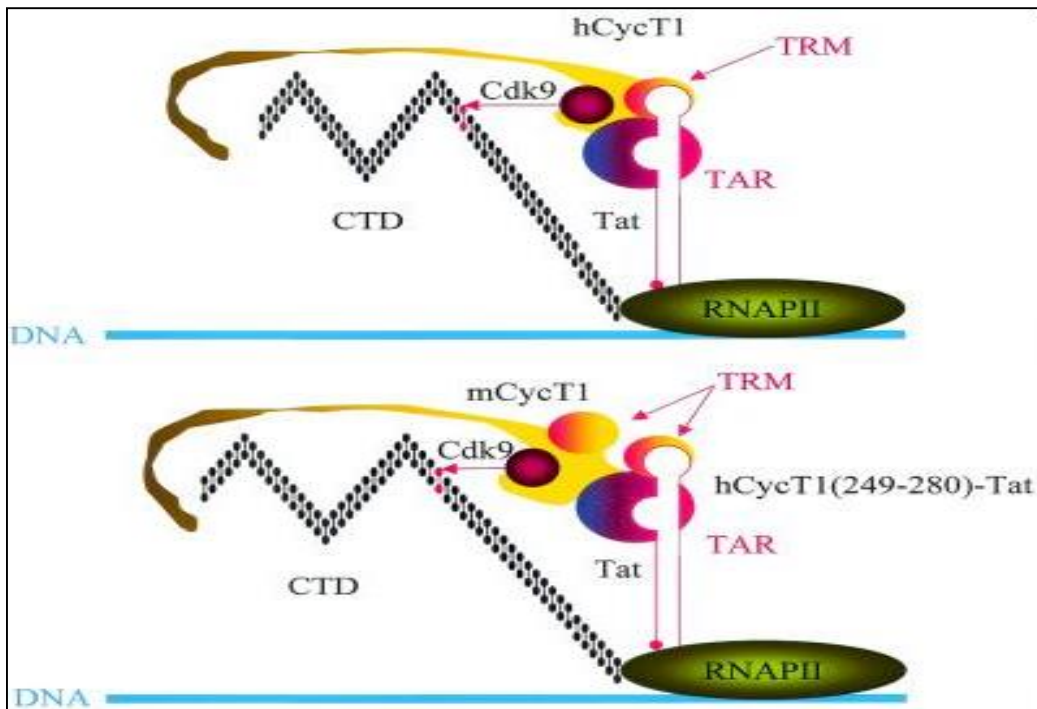
Figure 1-6 HIV-1 Tat recruits P-TEFb by binding Cyclin T1.

The HIV-1 Tat protein and a portion of Cyclin T1, the Tat/TAR-recognition motif (TRM), purportedly bind the trinucleotide bulge as well as the apical portion, respectively, of the HIV-1 TAR RNA stem-loop. The CDK9 catalytic subunit of P-TEFb is then brought into close proximity to RNA Polymerase II facilitating hyperphosphorylation of the RNA Polymerase II C-terminal domain (CTD). This interaction results in the production of a high number of full length viable transcripts, and thus is an attractive target for therapeutic intervention (14,15). The terms hCycT1, mCycT1, and hCycT1(249-280)-Tat refer to full length 726 amino acid human Cyclin T1, murine Cyclin T1 (724 amino acids) which does not support Tat transactivation, and the minimal chimera of residues 249-280 of human Cyclin T1 fused to 101 amino acid Tat respectively with the latter being the experimental construct as developed by Koh Fujinaga(28). The hCycT1(249-280)-Tat construct is able to activate HIV-1 transcription in murine cells which do not normally transcribe HIV-1 presumably due to the absence of cysteine 261 in murine Cyclin T1 where residue 261 is tyrosine (28). Reprinted from: *Journal of Virology*, vol. 76(24), Pages 12934-9. Copyright 2002 (28) with permission from Koh Fujinaga.
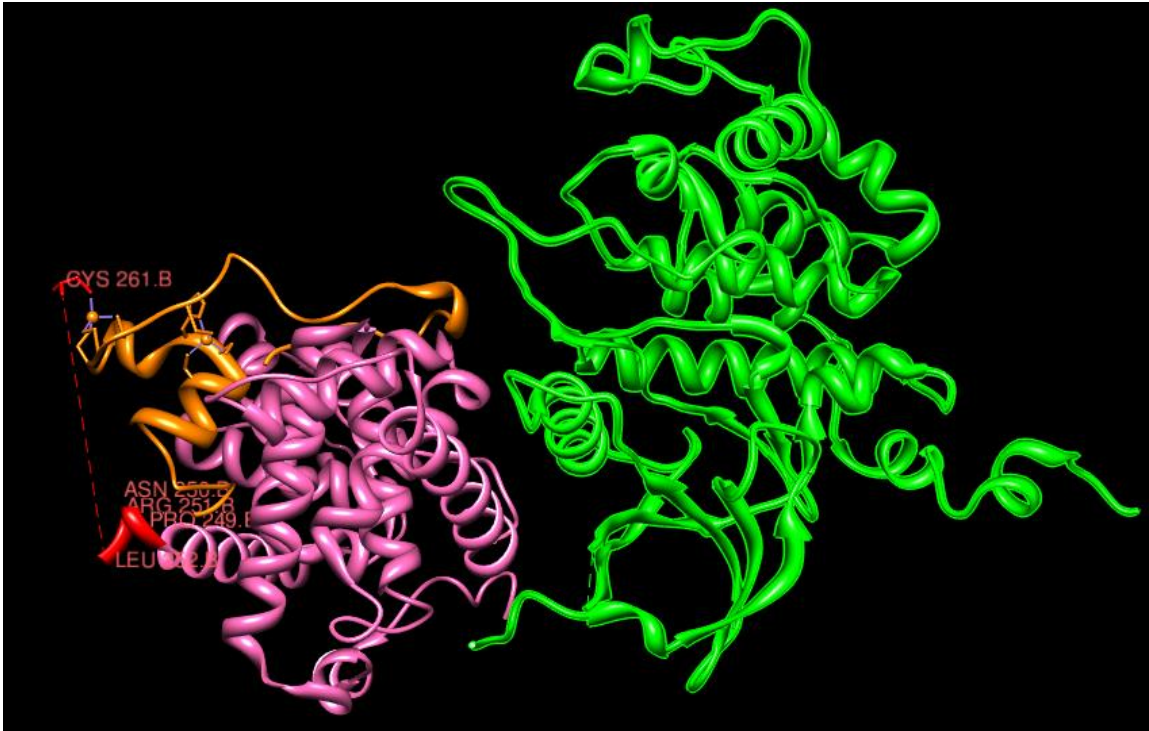
Figure 1-7 Tahirov et al. 2010 Crystal structure of HIV-1 Tat with human P-TEFb.

Tahirov et al. 2010 Crystal structure of HIV-1 Tat complexed with human P-TEFb.

RCSB Protein Data Bank: 3MI9 (7)HIV Tat protein (86 aa) in orange, Cyclin T1 (266 aa)

in pink CDK9 in green (351 aa).

The Tahirov et al. 2010 structure is reported in the absence of TAR. The conformation of

the Tat Activation Domain and the Tat-TAR complex remain obscure (8). Residues 50-

86 of the arginine rich Tat RNA-binding domain are undefined, and the conformation of

ZnF2 remains elusive in the absence of residues 252-260 of Cyclin T1 (in red) (7,8).

G   G
U   G
C   A
C — G
29   36
G — C
A — U
G — C
26   39
U
C
U   23
A — U
22   40
G — C
A — U
C — G
C — G
18   44
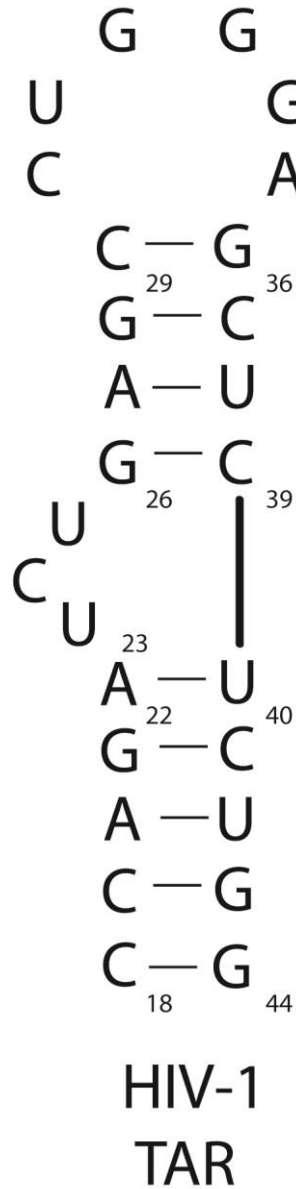
HIV-1
TAR

Figure 1-8 Wild type HIV-1 TAR stem loop.

A U23-C24-U25 trinucleotide bulge is present in the stem of the HIV-1 TAR RNA (at left).The TAR RNA binds HIV-1 Tat protein at a U23-A27-U38 base triple (20,21). Functional groups of the pyrimidine bulge and the phosphate backbone are repositioned during binding, and result in a conformational change that is an essential feature of specific recognition (22).

Chapter 2 Targeting the Tat-P-TEFb-TAR Interaction

The interaction between HIV Tat, human P-TEFb, and HIV TAR RNA is a highly

attractive target for the development of new HIV therapies that may potentially be less

prone to the development of drug resistance because of their potential to inhibit the

essential activity of Tat early in the infection cycle. The Tat-TAR interaction is critical

not only during the exponential phase of virus reproduction, when it substantially

enhances processive elongation of the full length HIV-1 mRNA, but also during the

activation of the integrated provirus that leads to mutation and drug resistant strains (29).

Current therapeutics inhibiting viral entry and formation of the provirus are not effective

at eliminating viral proteins produced early in the infection cycle from provirus already

integrated in the host genome. The neurodegenerative effects of HIV-1 are experienced in

some 50-70% of patients, and begin soon after infection as viral reservoirs are established

in the glial cells of the brain. This neurocognitive impairment persists despite effective

control of viral load (30). The viral protein Tat has been demonstrated to be a potent

neurotoxin and may play a role in this HIV-1 associated neurocognitive disease (HAND)

(30,31). Inhibitors of Tat-TAR interaction have the potential to substantially reduce viral

replication, as well as to improve neurocognitive prognosis.

Attempts to inhibit the Tat-TAR interaction have generally engaged one of three angles

of approach: anti-TAR, anti-Tat, and anti-Tat-P-TEFb. To date, compounds found to

inhibit these interactions have failed to demonstrate sufficient specificity, cell-

penetration, or stability to be effectively employed as therapeutics for the treatment of HIV (32).

2.1. Anti-TAR Agents

The bulk of research on the inhibition of the Tat-P-TEFb-TAR interaction has focused on anti-TAR agents, and specifically on inhibition of Tat binding at the pyrimidine bulge in the TAR RNA stem. Agents found to disrupt the interaction can be categorized into three classes: peptide-based, oligonucleotide-based, and small molecule inhibitors. In the category of peptide-based inhibitors: Tat-derived natural peptides, and Tat-mimetics such as peptoids, and D and β peptides have demonstrated some ability to inhibit the Tat-TAR interaction with the peptide mimetics demonstrating superior resistance to enzymatic degradation (32). Oligonucleotide inhibitors include TAR decoys such as: antisense oligonucleotides, aptamers, and RNA interference (RNAi), and have demonstrated only moderate efficacy in the inhibition of HIV-1 replication (32). Small molecule inhibitors such as: arginine derivatives, quinolones, and others have also displayed some ability to inhibit the Tat-TAR interaction, but as yet no small molecule has demonstrated the sufficiency to warrant further consideration as a drug candidate (32,33).

2.2. Anti-Tat Agents

Small molecules, biopolymers, and antibodies have all demonstrated inhibition of HIV replication by binding directly to Tat but, here also, with insufficient efficacy to produce

viable drugs (32). Recently, a synthetic form of didehydro-Cortistatin A isolated from the marine sponge *Corticium simplex* has shown promise in the early phases of research as an anti-Tat drug, and has been described as a potent suppressor of viral transcription at subnanomolar concentrations (34).

2.3. Anti-Tat-P-TEFb Agents

Targeting the Tat-P-TEFb complex is complicated by the unintended consequence of inhibiting basal cellular transcription that ordinarily requires P-TEFb (32). Selective disruption of the Cyclin T1 subunit of P-TEFb, while highly desirable, is therefore difficult to achieve without unwanted side-effects. Small molecules, antibodies, protein chimeras, intracellular inhibitors, and inhibitors of Tat co-activators of the Tat-P-TEFb interaction have, as yet, all failed to produce feasible drug candidates (32).

2.4. Aptamers as Diagnostic Tools: A Novel Approach to TAR Interaction

Elucidation of the structural features of the TAR-Tat-P-TEFb interaction could offer a great deal toward the facilitation of rational structure based drug design. In the absence of structural detail however, a high-throughput approach could be used to identify drug candidates. Our laboratory utilizes oligonucleotide aptamers in a high-throughput approach toward identification of promising small molecule drug candidates that disrupt interactions between ligands and nucleic acid targets. Figure 2-1 illustrates an aptamer diagnostic tool, here called an "AlloSwitch", in which a chimeric RNA-DNA molecule

comprised of a "probe" strand and a "cover" strand is tethered together by a fixed nucleotide duplex that does not change as the rest of the switch changes its conformation (35-37).

In the absence of the target the equilibrium between the two forms of the switch favors the Hidden (H) form at the left of the figure, where the target's binding site is hidden by base-pairs. The equilibrium shifts toward the right as target is added and binds the probe segment in the Open (O) form.

The RNA probe strand is generally designed with a high degree of sequence and structural similarity to the *in vivo* RNA target (or to an aptamer that binds to the same region of the target), while the DNA cover strand is mostly complementary to the probe strand. When the concentration of target is low, the probe and cover are annealed in the extended "on" form, where a 5' fluorophore is distant from, and therefore not quenched by, a strategically placed downstream quencher. At high concentrations of target, the open-probe form is favored and the fluorescence is efficiently quenched in this "off" state. If a competitive inhibitor is present, the conformation of the switch molecule reverts to the on-state indicating the presence of a potential drug candidate. Other switch formats are being investigated in our lab (35-39).

The sensitivity of an AlloSwitch is related to the three equilibria for high-throughput screening of drug candidates (Figure 2-2). The K1-equilibrium is controlled by the degree of complementarity between the RNA probe strand, and the DNA cover strand. The

greater the complementarity between the probe and cover the more the molecule becomes "locked" in the on form, so less of the switch molecule can be turned off by target binding (K2-equilibrium). The right combination of K1 and K2 usually gives about 90% of the maximum possible decrease in fluorescence on binding the target. This diagnostic tool is then able to detect and rank the affinity of drug candidates via the K3-equilibium.

2.5. Building and Testing the AlloSwitches

Designing and construction of the nucleic acid AlloSwitch requires the careful consideration of thermodynamic, and structural factors, as well as the consideration of the mechanics of oligonucleotide synthesis. Testing the switch requires an experimental ligand that accurately simulates the interaction between the *in vivo* ligand and its nucleic acid target. In the case of the Tat-P-TEFb-TAR interaction, a 33 nucleotide stem-loop portion of TAR complete with pyrimidine bulge can be easily secured from available commercial sources. Recreating the essential features of the Tat-P-TEFb interaction then becomes the challenge that has been the focus of this work. Important consideration must be given to the conformation that unbound Tat adopts (7,8) as well as to the fact that, un-remodeled by interaction with Cyclin T1 and the completion of ZnF2, unbound Tat alone cannot effectively mimic *in vivo* interaction with TAR. Hence, mimicking this interaction in a manner that is structurally accurate requires essential features of both Cyclin T1 and Tat to be present in the experimental target.
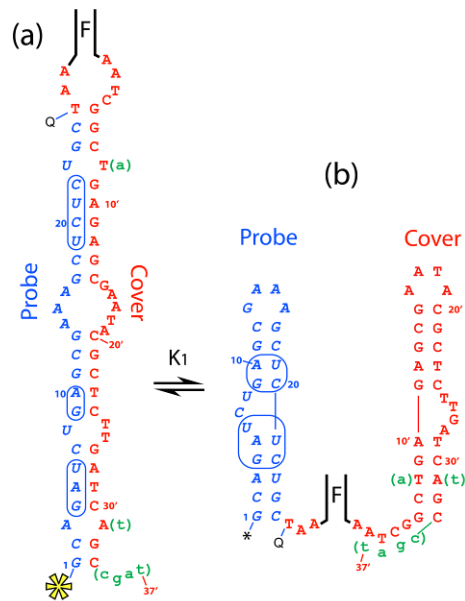
Figure 2-1 An indicator for Tat-TAR Binding

(a) A Tat-TAR indicator,T1, shown in the H conformation where most of the Tat-binding UCU bulge is hidden by base pairs, and (b) in the O-form, where the bulge is open for binding Tat. The critical binding elements for Tat (40) are circled. *RNA* monomers are shown in dark italics, DNA in regular font. The Probe and Cover strands are tethered by a fixed duplex, **F** (see text). Strands are numbered $5' \rightarrow 3'$ and do not include the **F**-sequences. The fluorophore $= *$ and quencher $= Q$ are marked. A second version of the switch has modifications in lower-case letters that include a "balancing" stem extension, ATCG:cgat in the O-form, which increases the stability of the double hairpin O-form.

$$
\begin{array}{c}
\mathbf{H} \overset{K_1}{\rightleftharpoons} \mathbf{O}^* \\
+ \\
\mathbf{L} + \mathbf{X} \overset{K_3}{\rightleftharpoons} \mathbf{L}\cdot\mathbf{X} \\
K_2 \updownarrow \\
\mathbf{L}\cdot\mathbf{O}^*
\end{array}
$$
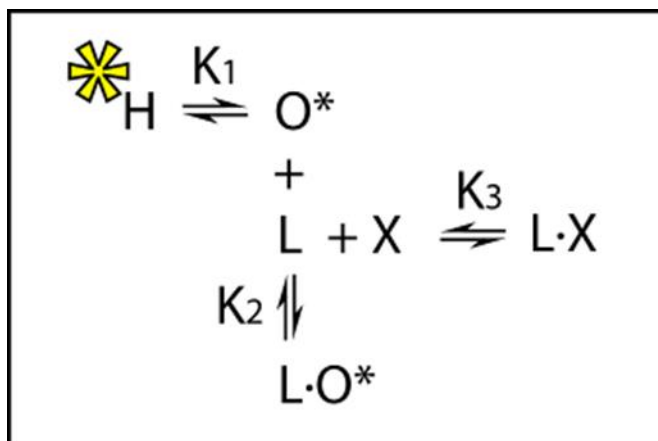
Figure 2-2 High-throughput screen assay format for Tat-TAR blockers.

Indicators have two stable states: **H** and **O**. Interesting competitors, **X**, will block

formation of the **L·O** complex (here, **L** = Tat).

Chapter 3 A Tat-Cyclin T1 Chimera

Embarking on the recombinant expression of a protein frequently presents biochemical challenges that are unique to the specific characteristics of the protein of interest and, the expression strain being used, and which require optimization by what can be a lengthy empirical process. The existing literature on the expression and purification of HIV-1 Tat protein demonstrates that Tat expression in *E. coli* is notoriously difficult to achieve at appreciable yield, solubility, and in native form. This observation is often attributed to the high cysteine content of the Tat activation domain (41). The presence of several atypical codons in the corresponding DNA can compound this problem by stalling expression in *E. coli* (42). Accommodation must be made for the atypical codons either by codon optimization, or by providing the appropriate supplemental tRNA. The presence and maintenance of the two zinc fingers of the Tat protein, and the corresponding secondary structure they induce, present additional considerations for expression and purification which appear to have hindered other work (43).

3.1. Recombinant Tat Expression

Much of the recombinant expression of Tat has been performed in mammalian cells. Where Tat has been expressed in *E. coli,* it is most commonly extracted from inclusion bodies by denaturation and refolding, and is often still hindered by problems with insolubility, precipitation, and aggregation (43,44). Unfortunately, the process of denaturing and refolding is a cumbersome one that, if accomplished successfully, will

then necessitate additional assays in order to confirm the biological activity of the target (41). A high yield of soluble protein from denaturing and refolding protocols is rare (45). Hence, a recombinant expression protocol for adequate yield of the soluble protein in its native form is highly desirable.

In order to facilitate the expression and purification of Tat, researchers frequently opt to express only the first exon of the Tat protein (42,46). The second exon, which is high in basic lysine residues, is not known to play a primary role in viral transcription (8,47,48). Still, difficulty with oxidation, misfolding, insolubility, precipitation, aggregation, or low yield persist even in the expression of truncated Tat in *E. coli* (42,43,49-52). Oxidation of recombinant Tat has been found to produce disulfide bond formation at C27-C30, and is suspected at C31-C34, and C22-C25 (53,54). In some cases, while working with Tat, oxidation was of such great concern that purification and refolding were performed in an anaerobic chamber (43). Even with the use of the anaerobic chamber protein dimers were observed. This can perhaps be explained, however, by the absence of the Cys261 residue contributed by CycT1 rather than by oxidation.

3.2. Human Cyclin T1

The work of Garber et al. 1998 suggested that Tat binding to TAR is mediated by Tat interaction with human Cyclin T1 and specifically by the completion of an *inter*molecular zinc finger by Cys261 of Cyclin T1(15). From an experiment mutating all cysteine and histidine residues of Cyclin T1, with the exception of Cys261, Fujinaga et al. 2002

concluded that Cyclin T1 binds Tat in a Zn^{2+} dependent manner at Cys261, and that no other cysteine or histidine residue of Cyclin T1 between amino acids 1 to 280 was required. This conclusion was largely based on the fact that Tat transactivation was not supported in mutants lacking Cys261(28). However, since the structure of HIV-1 Tat bound to HIV-1 TAR has not yet been solved, many questions remain including whether Cys261 of Cyclin T1 is mediating the interaction of Cyclin T1 with Tat or whether it is in fact directly responsible for binding to TAR. It is also possible that Cys261 could be influential in both interaction between Cyclin T1 and Tat and between the Tat-P-TEFb complex and TAR. Interestingly, it has also been suggested that TAR nucleates and enhances the interaction between Tat and Cyclin T1 (55) perhaps involving Cys261.

Complicating an already rather nebulous image of the interaction, Richter et al. 2002 concluded that residues 252-260 of Cyclin T1 were essential for TAR interaction with one side of the TAR RNA stem-loop and that these same residues enhanced the interaction of Tat residue K50 with the opposite side of the stem-loop (11). Unfortunately the 252-260 residues of Cyclin T1 are missing from the Tahirov crystal structure (Figure 1-7) (7,8). Das et al 2004 however, performed mutagenesis experiments that strongly support a metal binding role for cysteine 261 in the formation of a ternary complex with HIV-1 TAR (27). Also of interest are the basic residues R251 and R254 which have been implicated by some as potential stabilizers of the Tat-P-TEFb-TAR interaction (15). That there is, as yet, little consensus about this interaction is perhaps the only certainty.

Because of the role zinc finger proteins are now known to play in gene transcription, it would seem intuitive that this interaction could be highly influential in binding and recognition between Cyclin T1-Tat and TAR and that the zinc fingers are likely to play an important role in specificity and transactivation. Much of the existing research, however, focuses on the importance of the arginine-rich region of Tat binding to the nucleotide bulge of TAR, and also on arginine residues 251 and 254 of Cyclin T1 as the modulators of TAR interaction rather than on Zn1 and Zn2. Since the original suggestion of the importance of the two zinc fingers in this interaction by Garber et al. 1998 little has been done to assess the effective contribution of the zinc fingers alone. A minimal construct of Cyclin T1 eliminating or mutating R251 and R254 of Cyclin T1 and reducing the contribution of Cyclin T1 (as much as possible) to that solely provided by completion of Zn2 could offer important insight toward this end. Moreover, the mutation or elimination of arginine residues in the arginine-rich region of Tat purported to bind the TAR stem-loop bulge could also yield valuable insight.

Of the 726 residues of Cyclin T1 the first 272, encompassing the entire Cyclin domain, were originally found sufficient to bind Tat (15,28,56). However, interaction between the N-terminal and C-terminal region of Cyclin T1 has an autoinhibitory effect on TAR binding that is removed *in vivo* by Cyclin T1 C-terminal interaction with Tat-SF1 (57). The size and complexity of the dynamic interaction between P-TEFb-Tat-TAR suggests that structural study of the complex would be difficult to accomplish (28), and that a minimal construct could provide much needed structural information. With the intention

of identifying such a minimal construct, Fujinaga et al. 2002 produced a series of

minimal chimeras of N-terminally truncated Cyclin T1 fused to the full length (101

amino acid) Tat.

Working with N-terminally GST tagged chimeras of amino acids 1 through 280 of human

Cyclin T1 fused to the 101 amino acid Tat in NIH 3T3 cells, Fujinaga et al. 2002

measured transactivation of viral RNA for a series of Cyclin T1 N-terminally truncated

variants of the construct. Predictably, the 1-280 amino acid Cyclin T1-Tat chimera was

able to produce the highest level of transactivation at greater than 250 fold increase over

baseline transactivation. However, a minimal construct of amino acids 249-281 of Cyclin

T1 fused to the full 101 amino acid Tat produced an efficient 125 fold increase in

transactivation.

Working with a minimal chimera presents the possibility of acquiring important

structural data from a construct of an appropriate length for 600-800 MHz NMR

structural determination, and so was chosen as the chimera construct for this work.

Subsequent to the work of Fujinaga, circular dichroism experiments performed on a

human Cyclin T1 construct of amino acids 1-272 demonstrated that the 20 amino acids at

the C-terminus of the construct were conformationally flexible or disordered (27).

Similarly, and as mentioned previously (Figure 1-7), residues 252-260 of Cyclin T1 were

absent from the Tahirov crystal structure (7,8), perhaps due to the flexibility and/or

disorder observed by Das et al. 2004. Thus an even shorter construct omitting non-essential and flexible or disordered residues of Cyclin T1 could yield the important advantage of reducing this intractable region, while providing insight into questions about the importance of the completion of Zn2 by cysteine 261, as well as insight into the role of residues R251 and R254 in high affinity TAR binding.

Since the Tat protein is notoriously difficult to express, purportedly due to insolubility and a high number of cysteine residues (43), work on the recombinant production of the chimera began by using the 249-280 hCyclin T1-Tat chimera published by Fujinaga et al. 2002 in an attempt to assess the yield obtainable from this construct.

3.3.pGEX 2TK Plasmid 249-280 hCyclin T1-Tat Chimera

The pGEX 2TK plasmid of the 249-280 human Cyclin T1-Tat chimera was generously provided by Koh Fujinaga from Case Western University. While the details and yield were not published, Fujinaga (personal communication) indicated that the construct had been expressed in *E. coli* BL21 cells (28). Sequencing of the plasmid by Upstate Medial University DNA Core Facility confirmed that the open reading frame of the plasmid coded for the correct portion of human Cyclin T1followed by a 25 amino acid linker (within which a myc tag has been placed for antibody assay), and finally the full length (101 amino acid) Tat protein (Figure 1-2). The molecular weight of the GST-hCycT1-Tat construct is ~ 33.5 kDa, with an extinction coefficient (without disulfide bonds) at 280

nm of ~41250, and pI = 8.1 (Appendix 1). The Tat 101 aa portion of the sequence is identical with the Tat 101 aa protein sequence of accession number AAB59879.1 (101 aa Tat HIV-1 Group M Subtype B isolate ARV2/SF2) by Blast and was confirmed by ClustalX2 alignment.

3.4. Expression of the pGEX 2TK GST hCycT1-Tat Chimera in *E. coli*

The pGEX 2TK (~ 5.0 kb) chimera plasmid utilizes a tac promoter which is a hybrid of the lacUV5 and trp promoters. The tac promoter produces a tightly controlled, high-yield expression of recombinant protein. In general, the tac promoter necessitates the use of BL21 (non-DE3) (non-pLys) expression strain for optimal expression. The BL21(DE3) (pLys) strain designed for use with a T7 promoter would unnecessarily tax the cells with the production of the T7 polymerase, and reduce the yield of the target protein expression (Novagen conversations). While some researchers suggest it is possible to use DE3 strains with plasmids utilizing a tac promoter, Figure 5-1 demonstrates clearly that yield of recombinant protein was considerably lower when using a DE3 strain to express the protease TEV from a pRK793 plasmid with a tac promoter.

Unfortunately, many trials of recombinant expression of the pGEX 2TK plasmid, even when using the appropriate BL21 (non-DE3) (non-pLys) expression strain and following in close accordance with the protocol conditions generously provided by Koh Fujinaga (personal correspondence), produced a lower than anticipated yield (Figure 3-2) that was

insufficient for structural work with NMR. Estimates of crude and soluble protein yield were made using gel electrophoresis since the full length GST protein (already well below approximately 10 mg/L) was unstable when UV absorbing surfactants were removed. Concentration of the target after removal of the GST tag was too low for accurate determination by NanoDrop. A great deal of time and effort was spent varying expression conditions such as: culture temperature, IPTG concentration, media and media additives, among many others, and in the hopes of improving yield, but these efforts were to no avail. A variety of cell lysis techniques including sonication, microfluidizer processing, and freeze thaw cycling also failed to provide any observable improvement in the yield.

Investigation into the potential causes of the low expression yield observed in the BL21 (non-DE3) (non-pLys) expression strain revealed a number of significant factors. Analysis of the DNA sequences of the Cyclin T1 and Tat portions of the chimera alone (not including the adjoining sequence) showed that 15 of the chimera codons are atypical in *E. coli* (discussed in Chapter 7, see Table 7-1), and can therefore stall expression when corresponding tRNA are either unavailable or are in low abundance. However, transformation of the plasmid into the Rosetta 2 *E. coli* expression strain (Novagen Madison, WI), which contains an additional plasmid that codes for the production of several rare tRNA, increased basal protein expression while making only modest improvements in the expression of the target (Figure 3-4).

Solubility also appeared to be a limiting factor early in the expression as much of the expressed protein remained in the insoluble fraction after cell lysis. This was somewhat

surprising in view of the presence of the GST tag. Moreover, cleavage of the GST tag

using thrombin was highly non-specific and resulted in a further reduction in yield of the

cleaved protein after FPLC purification (Figure 3-3). The diminished yield was observed

when using a second benzamidine column to remove thrombin after cleavage, and was

worse still when using a heparin sepharose column for thrombin removal, as in an

unfortunate coincidence heparin bound Tat as well. In all trials with the pGEX 2TK

plasmid the yield of the Tat chimera was far too low for structural work with NMR with

the total yield of the full length chimera (with GST tag attached) at approximately 14

mg/L. Yield of the target after removal of the tag was predictably lower still, and the final

sample was substantially contaminated by co-purified proteins.

249–280 human Cyclin T1----------------------
PN RLKRIWNWRA C̲EAAKKTKAD DRGTDEKTSE

Spacer------**myc**------**Spacer**-------
QTMPEQKLIS EEDLAM**EFLE IDPVD**

HIV-1 Tat (1-101) ---
MEP̲VDP̲NLEP̲ WKHP̲GSQP̲RT AC̲NNC̲YC̲KKC̲ C̲FHC̲YAC̲FTR KGLGISYGRK
    Proline rich domain   **Cysteine rich domain**   **Arginine-rich-**

KRRQ̲RRRAPQ̲ DSQ̲THQ̲ASLS KQ̲PASQ̲SRGD̲ PTGPTESKKK VERETETDPF DLX
(basic) domain   **RGD containing C-terminal domain**

Figure 3-1 Sequence of the pGEX 2TK GST Chimera

The sequence of the pGEX 2TK GST Chimera: shown in green for amino acids 249-280 of human cyclin T1, in cyan for a myc antibody recognition sequence (which, together with 12 surrounding residues shown in black, constitute a spacer between the cyclin T1 domain and Tat) , and in magenta for the 101 amino acid full length HIV-1 Tat protein; key residues are underlined. Blocks of residues are aligned in groups of ten according to the numbering schemes for cyclin T1 and Tat. A residual leucine remains at the C-terminal of the Tat sequence as an artifact of the cloning process and X indicates the stop codon that terminates translation. The three features are recombinantly expressed as a single GST tagged construct The presence of the myc tag permits recognition of the chimera by anti-myc antibodies. The spacer portion of the contruct lends sufficient flexibility for the chimera to form a high-affinity complex with TAR RNA. The GST tag coding region upstream from the cloned insert and positioned at the N-terminus of the recombinant protein. The GE pGEX-2TK GST plasmid contains the GST tag sequence in the generic form prior to cloning. The GST sequence is not shown.
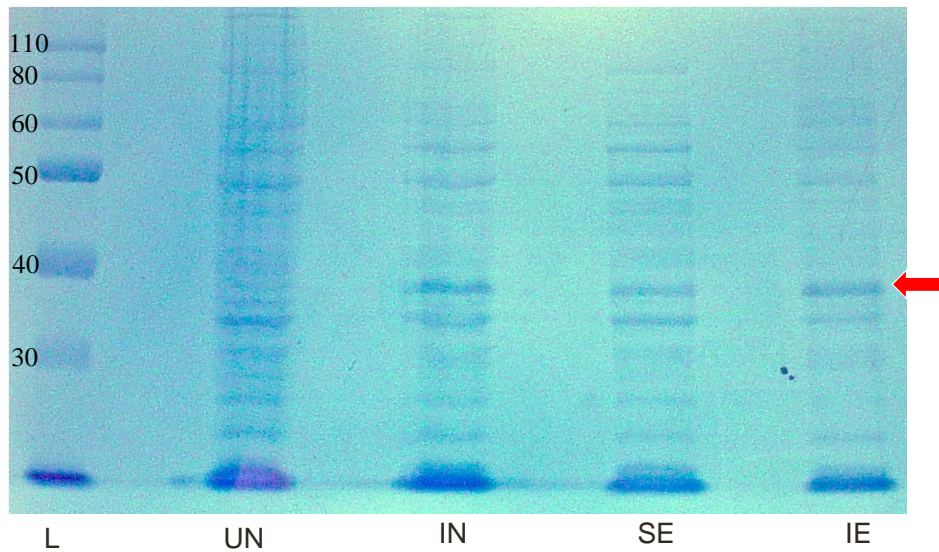
Figure 3-2 Expression of the pGEX 2TK GST hCycT1-Tat chimera

 The expression of the pGEX 2TK GST hCycT1-Tat chimera in BL21 (non-DE3) (non-pLys) in LB media. Red arrow indicates the full ~33.5 kDa GST tagged chimera. From left to right:

(L) ladder   (UN) uninduced

(IN) induced with 1 mM IPTG at 0.6 OD_{600} and harvested after 6 hours

(SE) soluble extract   (IE) insoluble extract
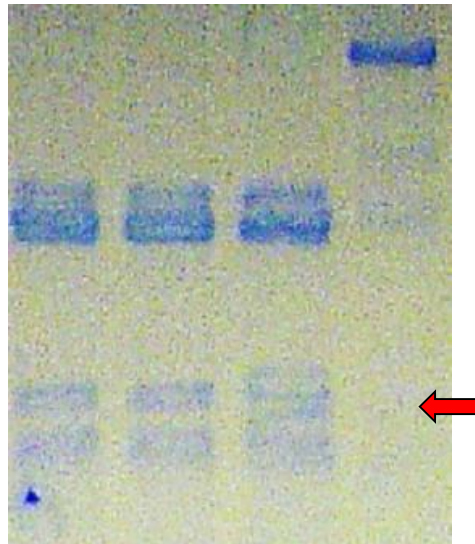
Figure 3-3 Cleavage of full purified GST- hCycT1-Tat chimera

Cleavage of the full purified GST hCycT1-Tat chimera after three hours of incubation
with thrombin at room temperature (RT) from left to right:

6 U Thrombin/mg   12 U Thrombin/mg Tat

18 U Thrombin/mg Tat   Full GST tagged Chimera

Red arrow indicates location of cleaved chimera. The thrombin protease 37,000 kDa was
present at concentrations too low for visible detection on the gel stained with GelCode
Blue coomasie protein stain (this section of the gel omitted).
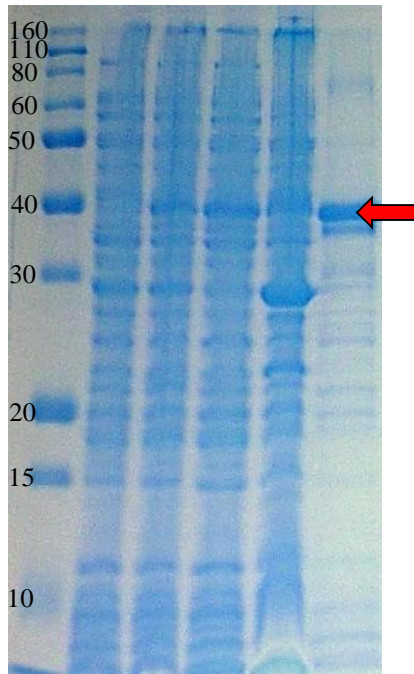
Figure 3-4 Expression of pGEX 2TK GST-hCycT1-Tat in Rosetta 2 cells

Expression of pGEX 2TK GST-hCycT1-Tat full chimera

 in Rosetta 2 cells (Novagen Madison, WI). From left to right:

ladder	uninduced	induced
soluble	insoluble	FPLC purified fraction

Note that the full length chimera and the tag-free chimera (not shown) run at a higher molecular weight than predicted. This artifact was observed consistently throughout all gel electrophoresis experiments, and is likely due to preferential SDS loading and the high number of cysteine residues present in the Tat chimera.

Chapter 4 Re-engineering the hCycT1-Tat Chimera Plasmid

When re-engineering the chimera plasmid the first task was to increase the expression yield and to improve the solubility of the native hCycT1-Tat chimera. To this end a multitude of plasmid attributes were considered for their potential to impact the final yield of the active purified protein. For example, in some cases the expression yield can be improved by an alternate inducible promoter, which provides more tightly regulated expression while mitigating the buildup of toxic recombinant proteins in the cell. Once an adequate expression level is achieved, the final yield can be improved by increasing the solubility of the target protein. One approach to improving solubility is the addition of "solubility enhancing" tags or fusion proteins encoded by the plasmid, and expressed at either the N-terminus or the C-terminus of the recombinant protein.

For many downstream applications the fusion tag may remain on the protein of interest without confounding the assay. With respect to the binding assays and structural work that the purified chimera is intended for, the frequently cumbersome fusion tag must often be removed. As a general rule, each step added to the purification process has the unintended effect of reducing yield. With this in mind, optimizing the cleavage sequence, and employing a highly specific protease enzyme to cleave solubility enhancing tags can contribute substantially to recombinant protein yield by cleaving as much of the tag from the target as possible while minimizing non-specific cleavage.

The second important task in re-engineering the chimera plasmid was to truncate the human Cyclin T1 portion of the chimera. By omitting the flexible region of Cyclin T1 and removing the arginine residues suspected of influencing the affinity of TAR binding the contribution of Cyclin T1 Cys261 may be assessed. Specifically, the role of Cyclin T1 in the binding of such a truncated construct is likely to be solely attributable to Cys261 of Zn2. By including only residues 257-280 of human Cyclin T1, the arginine residues R251 and R254 could be removed while still allowing a few additional residues following Cys261 to afford conformational stability and flexibility.

4.1. Selection of an Appropriate Vector

When selecting a vector the size of the insert, copy number, promoter, selection marker, cloning sites, and other additional attributes of the vector must be thoroughly considered with respect to downstream applications (58). Commercially available plasmids can accommodate inserts approaching 15 kb in size. The chimera DNA insert is comparatively small, at less than 500 base pairs, and can be readily accommodated by most commercially available plasmids. Since the recombinant protein being expressed is considered toxic to *E. coli* in high concentrations, a low copy number plasmid, rather than a high, is a prudent choice (59,60).

4.2. Solubility Enhancing Tags

The initial expression of the glutathione S-transferase (GST) tagged chimera produced a low yield with a high proportion of the target protein in the insoluble fraction. Hence, the primary objective in re-engineering the chimera was to exchange this fusion protein, in the hopes of improving both expression and solubility of the chimera. Oddly, while the GST tag (24 kDa) is often characterized as solubility enhancing, improved solubility as a direct result of the inclusion of the GST tag is rarely observed (61-63). In fact, GST tagged proteins are frequently expressed at even lower levels, and in less soluble form than their untagged counterparts (63). A comparison by Braun and LaBaer 2003 reports the solubility achieved with GST in their own work, and the work of two other research groups, Hammarstrom et al. 2002, and Shih et al. 2002, at 50%, 48%, and 38% solubility, respectively (64,65). Thus, the true advantage of including the GST tag may lie in the ability to purify the protein by affinity chromatography, rather than in any predictable effect on solubility.

Some of the more commonly employed solubility enhancing tags are: maltose-binding protein (MBP) (43 kDa), thioredoxin (Trx) (11 kDa), calmodulin-binding protein (CBP) (4 kDa), cellulose-associated protein (CAP) (17 kDa), NusA (54 kDa), and SUMO (11.5 kDa) (64,66,67). However, little is known about the mechanism by which solubility is achieved with the use of these fusion proteins (68).

Selecting an optimal solubility enhancing tag remains an empirical process that, using traditional methods, can be prohibitively time consuming. Recent advances in high-throughput technology have facilitated parallel cloning, and rapid screening techniques

that expedite this process. As a general rule, larger solubility enhancing tags tend to produce higher expression yields of more soluble proteins, to improve folding, and also to reduced proteolysis of the target (66,69). However, larger tags tend to complicate structure determination necessitating their removal, and reducing yield as an inevitable and undesirable consequence. Folding, function, crystallization, and NMR experiments can all be hindered by the presence of solubility enhancing tags. Thus, keeping the fusion tag at a low molecular weight, and/or removing the tag entirely after purification are part of a repertoire of conventional strategies that can be used to deliberately design target proteins to facilitate downstream applications.

4.3. Gateway® Cloning of hCycT1-Tat into pDEST HisMBP

The fusion protein MBP has demonstrated exceptional efficiency at enhancing both the total expression and the solubility of many target proteins (68). Hammarstrom et al. 2002, Shih et al. 2002, and Braun and LaBaer 2003 reported solubility of target proteins fused to MBP at 70%, 60%, and 90% respectively, far better than these researchers observed with the same targets tagged by GST. In some cases MBP has specifically exhibited the ability to influence proper folding of the target by acting as a molecular chaperone (68). These observations suggested that the MBP fusion protein was an attractive candidate for improving the overall expression yield of the Tat chimera.

One potential drawback to the use of MBP is problematic affinity purification where MBP often does not bind well to amylose resin, or where the target protein interferes with

MBP binding (70-72). The difficulties of MBP affinity purification can be circumvented by the use of a double-affinity fusion system such as that developed by Pryor and Leiting (1997). Addition of a 6 histidine (His$_6$) tag to the C-terminus of MBP permits metal chelating affinity purification, and by employing both tags simultaneously solubility and purification are enhanced beyond what would be achieved by employing either of the tags independently. Building on the work of Pryor and Leiting (1997), the David Waugh Lab designed a similar pDEST-His$_6$MBP plasmid for use in the Gateway Cloning method. In this construct the His$_6$ tag is placed at the N-terminus of MBP. We chose this route for production of the MBP tagged hCycT1-Tat chimera. The pDEST HisMBP plasmid (#11085) was deposited with, and secured from, Addgene for the construction of the hCycT1-Tat pDEST His-MBP. A tac promoter in the pDEST His-MBP plasmid generates a low copy number, and is well-suited to expression of the (possibly) toxic chimera.

4.4. Gateway® Cloning Technology

Gateway® Technology exploits the recombination properties of the bacteriophage lambda by introducing *att* recombination sites flanking the sequence of interest. In the presence of the proprietary Clonase™ enzyme the desired sequence, with flanking att sequences, is recombined into a donor plasmid containing the appropriately complementary att recombination sequences. On recombination the desired sequence is inserted into the donor plasmid in a directional manner replacing the ominous control of cell death (*ccd*) gene and providing stringent selection. Subsequently, and in the presence of the

proprietary enzymes, the insert may be transferred by recombination into a number of commercially available destination plasmids with a variety of attributes that can be tailored to alternate downstream applications. The entire protocol may be accomplished in a few hours with the insert remaining in frame, and appropriately oriented, and without the need for a ligation step (73).

4.5. Cloning Protocol

A nine step procedure was followed for the production of the pDEST His-MBP hCycT1-Tat vector:

1. Introduce AfeI Blunt-end Restriction Sites into the pGEX 2TK hCycT1-Tat plasmid by Site Directed Mutagenesis

2. Cleave pGEX 2TK hCycT1-Tat plasmid

3. Isolate insert from agarose gel

4. PCR amplify insert with N1 and C primer

5. Purify insert with N1 and C attachments from agarose gel

6. PCR amplify N1 C insert with N2 and C primers

7. Purify insert with N1, N2, and C by agarose gel

8. BP reaction to pDONR221 entry vector (Appendix 7)

9. LR reaction to pDEST HisMBP vector (Appendix 7)

4.6. Introduction of AfeI Blunt-end Restriction Sites

Prior to cloning the DNA fragment into the pDEST HisMBP plasmid (Figure 4-1), and beginning with the Cyclin T1 residue 257, the insert was excised from the pGEX-2TK (GE Healthcare Biosciences Pittsburg, PA) plasmid leaving blunt end restriction sites necessary for Gateway® cloning. Two flanking AfeI restriction sites were introduced to the chimeric sequence by site directed mutagenesis (Stratagene La Jolla, CA) (Figure 4-2) (Appendix 6). Primer X online software was used to design the mutagenic primers (Figures 4-3, 4-4). The AfeI restriction site AGC/GCT (serine/alanine) cleaves between the cytosine and guanine residues of the DNA. As an artifact of this procedure the chimeric sequence will have an alanine added to the N-terminus and a residual serine added to the C-terminus of the protein. After site directed mutagenesis the inserts were analyzed by agarose gel electrophoresis (Figure 4-5), excised, purified, and sequenced. The correct placement of the mutations was confirmed by sequencing from the Upstate Medical University DNA Core Facility, and analyzed with Serial Cloner software.

4.7. Addition of the att and TEV sites by PCR

After excising the insert the att recombination sites must be added by polymerase chain reaction (PCR) at each end in order to facilitate recombination into the entry vector. At this point a tobacco etch virus (TEV) protease cleavage site was introduced 3' of the 5' att recombination site (Figure 4-6). The TEV protease is a highly specific protease that provides efficient cleavage of the MBP fusion tag (74). Introduction of the TEV cleavage site will facilitate MBP removal from the chimera at the amino acid sequence

ENLYFQ/G with a single residual glycine amino acid artifact that when added to the alanine artifact produced by the introduction of the AfeI restriction sites will leave 2 residual amino acids at the N-terminus. Following the 101 amino acid Tat sequence, a single additional serine residue will follow a residual leucine that remains at the C-terminus as an artifact of the cloning process (Figure 4-11).

Following the work of Austin et al. 2009 for the construction of a pDEST HisMBP plasmid with a TEV cleavage site, three primers were employed to produce the insert: primer N1, primer C, and primer N2 in a two sequential PCR reactions. In the first reaction the forward primer N1 contains a 5' region coding for the restriction sequence of TEV (Figures 4-6) followed by 20 nucleotides of the coding passenger DNA. The reverse primer, Primer C, has a 5' region encoding the attB2 recombination site followed by 21 nucleotides of the 3' coding region of the passenger DNA (Figure 4-7). Here an excessively long forward primer is obviated by a second PCR reaction in which the first PCR amplicon becomes the primer for the second PCR amplicon N2 (Figure 4-6) (75). Primer C, the reverse primer, is used in both reactions. Primer N2, the forward primer for the second PCR reaction, encodes the attB1 recombination site followed by 19 nucleotides of the coding TEV protease site. The three primers can be employed in a single PCR reaction to obtain a final insert with flanking attB1 and attB2 recombination sites, and a TEV protease site, or the reaction can be done in series amplifying first with Primer N1 and C, purifying the PCR product, and then amplifying a second time with Primer N2 and C.

The reaction was performed in both single step (not shown), and two step reactions with only the later successfully producing the appropriate insert. Concentrations of the primers were consistent with those used by Austin et al. 2009 (Table 1). The PCR thermal cycler settings were as follows: initial melt 5 min at 95°C; 35 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 60 s then hold at 4°C. After agarose gel electrophoresis (Figures 4-7, 4-8) PCR products excised, and then purified with the Pure Link™ PCR clean-up kit (Invitrogen, Carlsbad CA).

4.8. Gateway® Cloning Donor Vector

After PCR the inserts were purified, and introduced by recombination into the donor vector pDONR221 (Invitrogen Carlsbad, CA). The hCycT1-Tat gene was transferred from the attB flanked PCR product into the attP flanking donor vector by recombination in what is called a "Gateway® BP reaction" (referring to the joining of the attB and attP sites) (Figure 4-9) by incubating equimolar amounts of attB-PCR product and donor vector with BP Clonase overnight at 25°C (Invitrogen Carlsbad, CA) in TE buffer at pH 8.0. At the end of the BP reaction Proteinase K solution was added and the reaction was incubated for 10 minutes at 37°C. The pDONR221 vector was then transformed into OmniMAX™ 2-TI^R chemically competent *E. coli*. The negative selection gene ccdB of the pDONR221 plasmid was replaced by recombination of the attB flanked PCR product and only successful recombinants survived on LB plates with 50 ug/ml kanamycin (73).

4.9.Gateway® Cloning Destination Vector

Once the hCycT1-Tat gene was transferred into the pDONR221 donor vector the target

gene was then transferred by LR recombination reaction into the expression or

"destination" vector (more details of the recombination reactions will be described in

connection with Fig. 6-10). The attP recombination sites of the pDONR221 vector

recombined with the attB sites of the PCR insert to form attL sites that recombine into the

attR sites of the destination vector during the LR reaction. The pDONR221 plasmid DNA

(150 ng/ul in TE, pH 8.0) was incubated with 150 ng/ul of pDEST HisMBP in TE, pH

8.0 overnight at 25ºC. At the end of the LR reaction Proteinase K solution was added and

the reaction was incubated for 10 minutes at 37ºC (73). The pDEST His-MBP hCycT1-

Tat vector was then transformed into chemically competent Rosetta Gami B (Novagen

Madison, WI) *E. coli* (Figure 4-8). The pDEST HisMBP hCycT1-Tat plasmid was

sequenced at Upstate Medical DNA Core Facility, and the correct sequence and the

reading frame of the insert were confirmed (Figures 4-10 and 4-11).
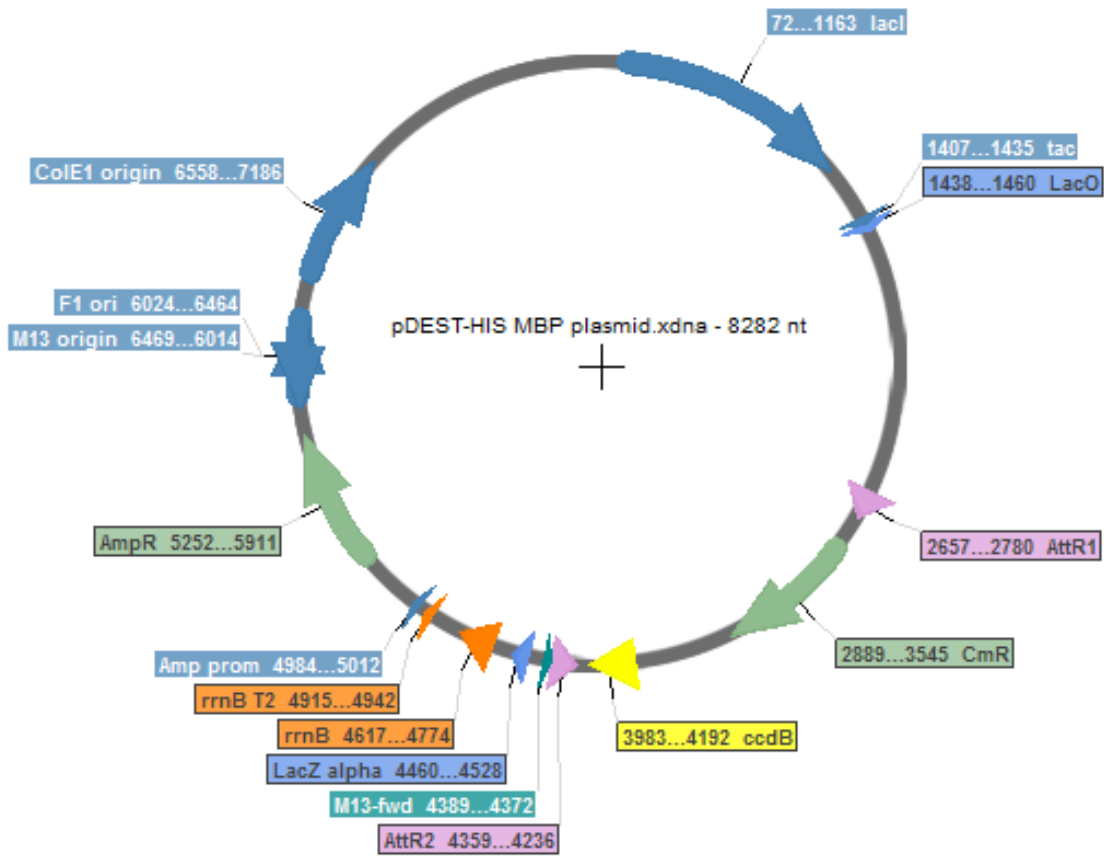
Figure 4-1 pDEST-HisMBP Plasmid.

Addgene#:11085. Deposited by The David Waugh Lab.
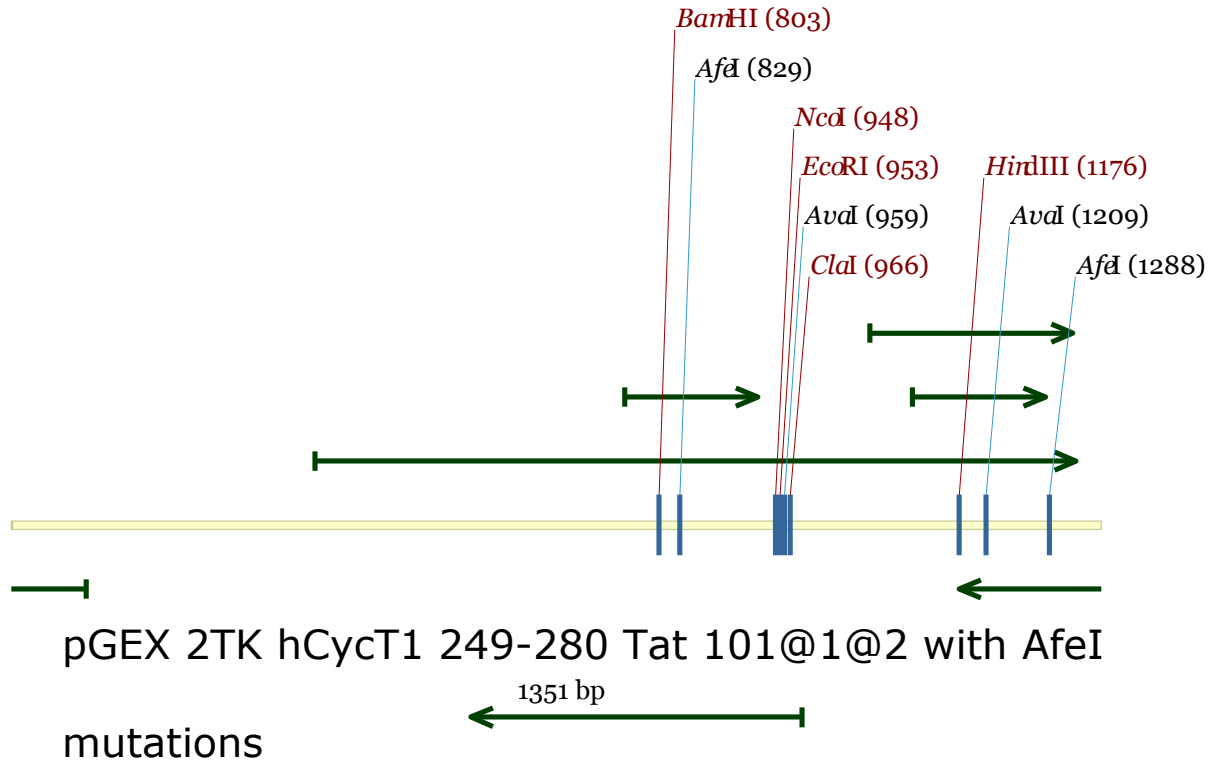
Figure 4-2 Introduction of AfeI restriction sites by site directed mutagenesis.

The location for introduction, by site directed mutagenesis, of two AfeI blunt-end restriction sites to the pGEX 2TK plasmid flanking the hCycT1-Tat chimera DNA sequence. The AfeI sites were placed such that the 5' site cleaved the insert prior to the codon corresponding to residue 257 of the 249-280 portion of Cyclin T1 and the 3' restriction site was placed after the codon corresponding to a single residual leucine residue past the 101 amino acid Tat sequence that remains as an artifact of the cloning process.

Primer pair 1

<center>*****</center>

Forward: 5' CCAACAGGCTCAAACGCAGCGCTAATTGGAGGGCATGCGAG 3'

Reverse: 5' CTCGCATGCCCTCCAATTAGCGCTGCGTTTGAGCCTGTTGG 3'

<center>*****</center>

GC content: 58.54%   Location: 754-794

Melting temp: 77.2°C   Mismatched bases: 5

Length: 41 bp   Mutation: Substitution

5' flanking region: 18 bp   Forward primer MW: 12678.36 Da

3' flanking region: 18 bp   Reverse primer MW: 12535.25 Da

Figure 4-3 Two PCR primers for the introduction of the 5' restriction sites.

Using PrimerX software two primers were designed

to add the same AGC/GCT blunt end restriction site to the 5' end of the insert (Serial

Cloner)

Primer pair 1

******

Forward: 5'

GACAGATCCGTTCGATTTGAGCGCTGTCGAGAGAGCGGCCGCATC 3'

Reverse: 5'

GATGCGGCCGCTCTCTCGACAGCGCTCAAATCGAACGGATCTGTC 3'

******

GC content: 60.00%   Location: 1211-1255

Melting temp: 78.1°C   Mismatched bases: 6

Length: 45 bp   Mutation: Substitution

5' flanking region: 19 bp   Forward primer MW: 13903.14 Da

3' flanking region: 20 bp   Reverse primer MW: 13783.08 Da

Figure 4-4 Two PCR primers for the introduction of the 3' restriction sites.

Using PrimerX software two primers were designed to add the same AGC/GCT blunt end restriction site to the 3' end of the insert (Serial Cloner)
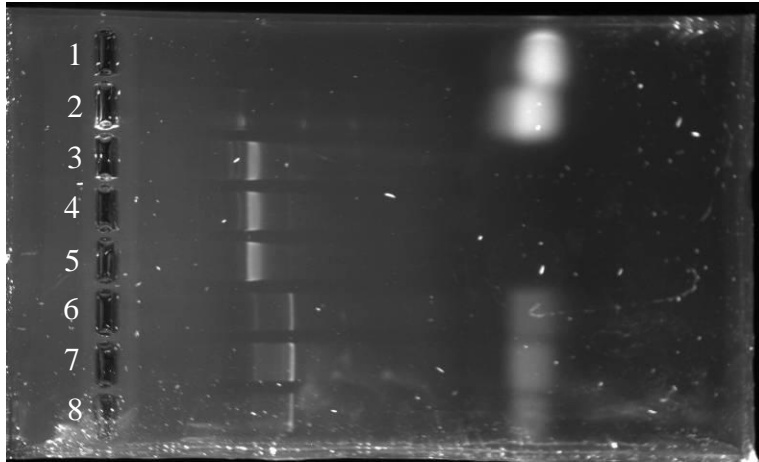
Figure 4-5 Agarose gel electrophoresis of the PCR products for AfeI sites

Agarose gel electrophoresis analysis of the PCR products for the addition of AfeI blunt-end restriction sites by site directed mutagenesis at sites 5' and 3' of the hCycT1-Tat insert. Gel electrophoresis at 0.7% agarose 140 V. From top to bottom:

1 – forward primer for 3' mutation   2 – reverse primer for 3' mutation,

3, 4, 5 – pGEX 2TK plasmid with 5' AfeI sites,

6,7,8 – pGEX 2TK plasmid with both 5' and 3'AfeI sites.
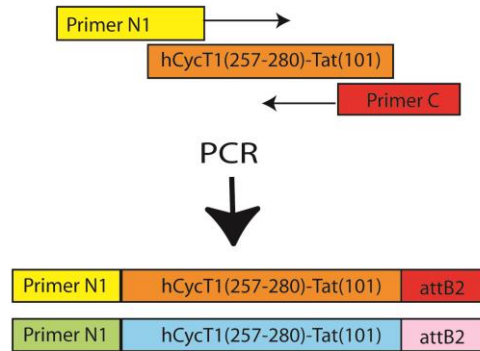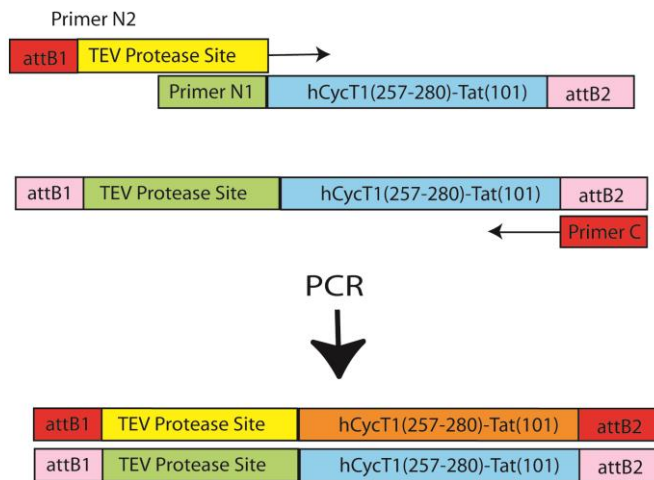
Figure 4-6 Diagram of the sequential PCR reaction for the att flanked insert.

Diagram of the sequential PCR reactions for the production of the HisMBP hCycT1-Tat chimera insert with flanking att, and internal TEV protease sites.

PCR Set-up 2 Step Primer Version

Step 1

	Experiment			Positive Control			Negative Control	
CK	**Reagent**	**Vol (ul)**	**CK**	**Reagent**	**Vol (ul)**	**CK**	**Reagent**	**Vol (ul)**
	DNA Template (3.7 ng/ul)	2.7		DNA Template (5 ng/ul)	2.0		NO DNA replace w/dH2O	2.7
	10 Rxn Buffer	10.0		10 Rxn Buffer	10.0		10 Rxn Buffer	10.0
	dNTP (10 mM)	1.5		dNTP (10 mM)	1.5		dNTP (10 mM)	1.5
	Primer N1 (30 ng/ul)	10.0		Primer Control 1 (100 ng/ul)	3.0		Primer N1 (30 ng/ul)	10.0
	Primer C (100 ng/ul)	3.0		Primer Control 2 (100 ng/ul)	3.0		Primer C (100 ng/ul)	3.0
	Pfu Ultra	1.0		Pfu Ultra	1.0		Pfu Ultra	1.0
	dH$_2$O	71.8		dH$_2$O	79.5		dH$_2$O	71.8
	Total Volume	100.0		Total Volume	100.0		Total Volume	100.0

Step 2

	Experiment			Positive Control			Negative Control	
CK	**Reagent**	**Vol (ul)**	**CK**	**Reagent**	**Vol (ul)**	**CK**	**Reagent**	**Vol (ul)**
	DNA Template (3.9 ng/ul)	2.6		DNA Template (5 ng/ul)	2		NO DNA replace w/dH2O	2.6
	10 Rxn Buffer	10		10 Rxn Buffer	10		10 Rxn Buffer	10
	dNTP (10 mM)	1.5		dNTP (10 mM)	1.5		dNTP (10 mM)	1.5
	Primer N2 (100 ng/ul)	3		Primer Control 1 (100 ng/ul)	3		Primer N2 (100 ng/ul)	3
	Primer C (100 ng/ul)	3		Primer Control 2 (100 ng/ul)	3		Primer C (100 ng/ul)	3
	Pfu Ultra	1		Pfu Ultra	1		Pfu Ultra	1
	dH2O	78.9		dH2O	79.5		dH2O	78.9
	Total Volume	100		Total Volume	100		Total Volume	100

Table 1 Set-up of the PCR reactions for primers N1, N2, and C

Set-up of the sequential PCR reactions for primers N1, N2, and C for the production of

the HisMBP hCycT1-Tat chimera insert.

Primer N1 – 5' - GAG AAC CTG TAC TTC CAG GGT GCT AAT TGG AGG GCA TGC GA – 3'

Primer C - 5' - GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT AGC TCA AAT CGA ACG GAT CTG T – 3'

Primer N2 – 5' - GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC GGA GAA CCT GTA CTT CCA G – 3'

Figure 4-7 Nucleotide Sequences of N1, C, and N2 primers

Sequences of N1, C, and N2 primers for addition of the TEV protease site, and the attB1, and attB2 flanking recombination sequences for Gateway® cloning. For the N1 and N2 primers sequences in blue include the sequence for addition of the TEV protease site. For primer N1 the sequence in black complements the antisense strand of the Tat chimera insert after cleavage by AfeI. For primer C the sequence in green includes the sequence for the addition of the attB2 site, and the portion in black complements the sense strand. The primer N2 sequence in red includes the sequence for addition of the attB1 site, while the sequence in blue is the same as the TEV protease sequence in N1 (with the exception of the final three nucleotides), and complementary to the antisense strand after the first PCR reaction with N1 and C.
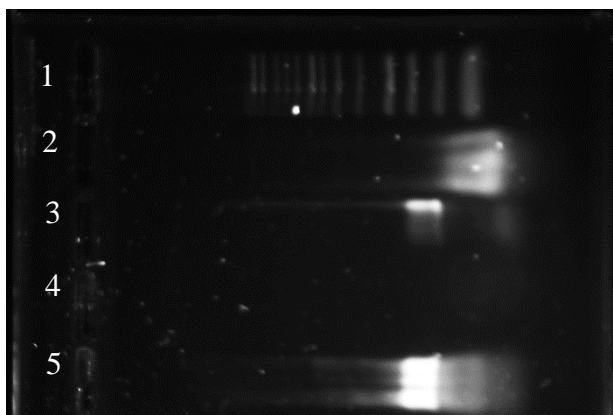
Figure 4-8 Agarose gel electrophoresis of PCR products for the N1 and C primers

An agarose gel electrophoresis analysis of the PCR products for the N1 and C primers adding the TEV protease site and the flanking att B2 recombination site to the hCycT1-Tat chimera DNA insert. Gel electrophoresis 0.7% agarose 0.5 X TBE 140 Volts. From top to bottom:

1 – ladder 1 kb   2 - negative control primers only (no template DNA)

3 – positive control (kit template DNA and primers)   4 – empty

5 – primer N1, primer C, and Tat chimera insert reaction (20 ul).
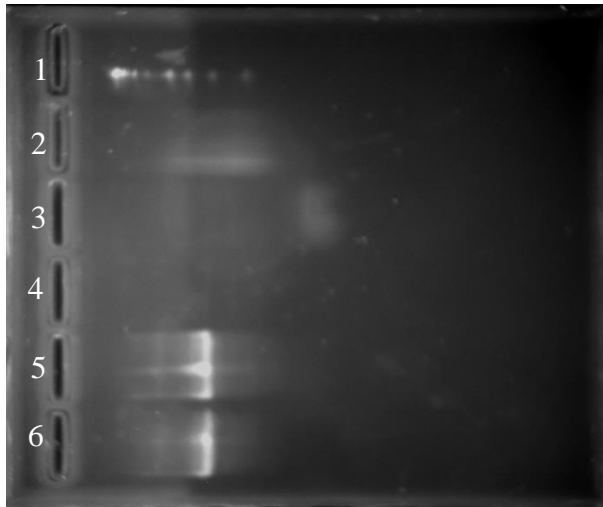
Figure 4-9 Agarose gel electrophoresis of PCR products for the N2 and C primers

An agarose gel electrophoresis analysis of the PCR products for the second round of

PCR with the N2 and C primers adding the attB1 and attB2 sites, respectively to the Tat

chimera DNA insert after the first round of PCR with N1 and C primers. Gel

electrophoresis 0.7% agarose 0.5 X TBE 140 Volts. From top to bottom:

1 – ladder 1 kb   2 – negative control primers only (no template DNA)

3 – positive control (kit template DNA and primers)   4 – empty

5 and 6 – primer N1, primer C, and Tat chimera insert reaction (20 µl)
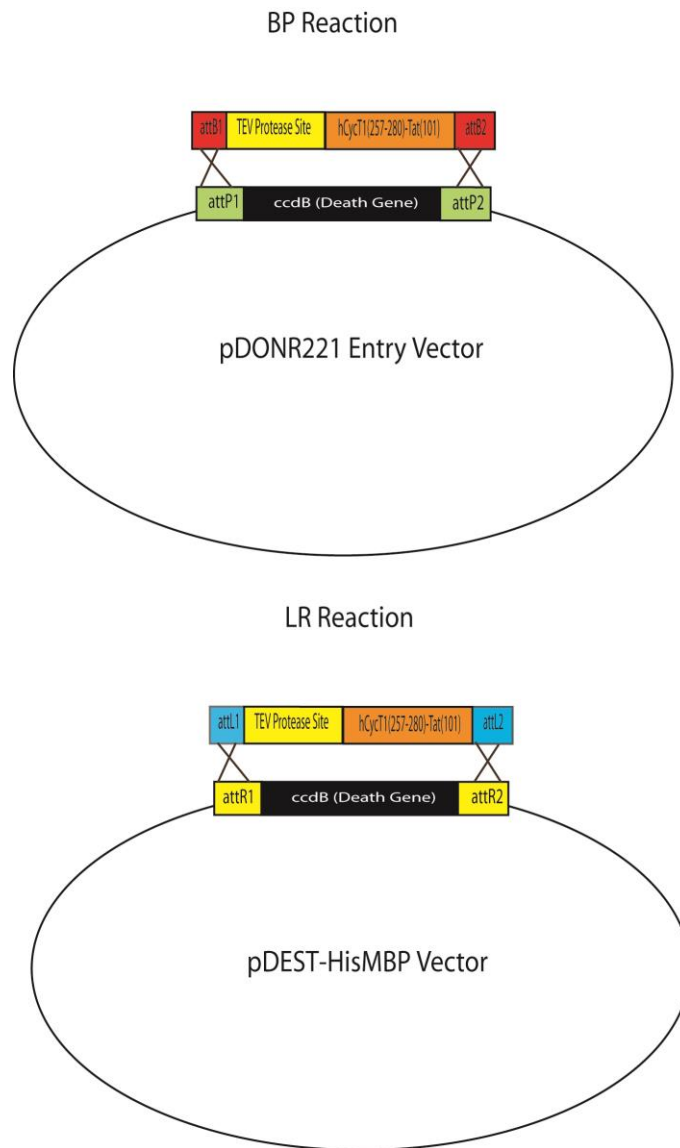
Figure 4-10 Gateway cloning BP and LR reaction

Gateway cloning BP and LR reactions for the production of the pDEST HisMBP (257-280) hCycT1-Tat (101) chimera expression plasmid. In the BP reaction (top) the att flanked insert is recombined into the pDONR221 entry vector. In the LR reaction (bottom) att sites of the donor vector recombine with att sites on the pDEST-HisMBP expression plasmid.

* According to Life Technologies Corporation the letters B, P, L, and R refer to bacterial, phage, left, and right respectively.
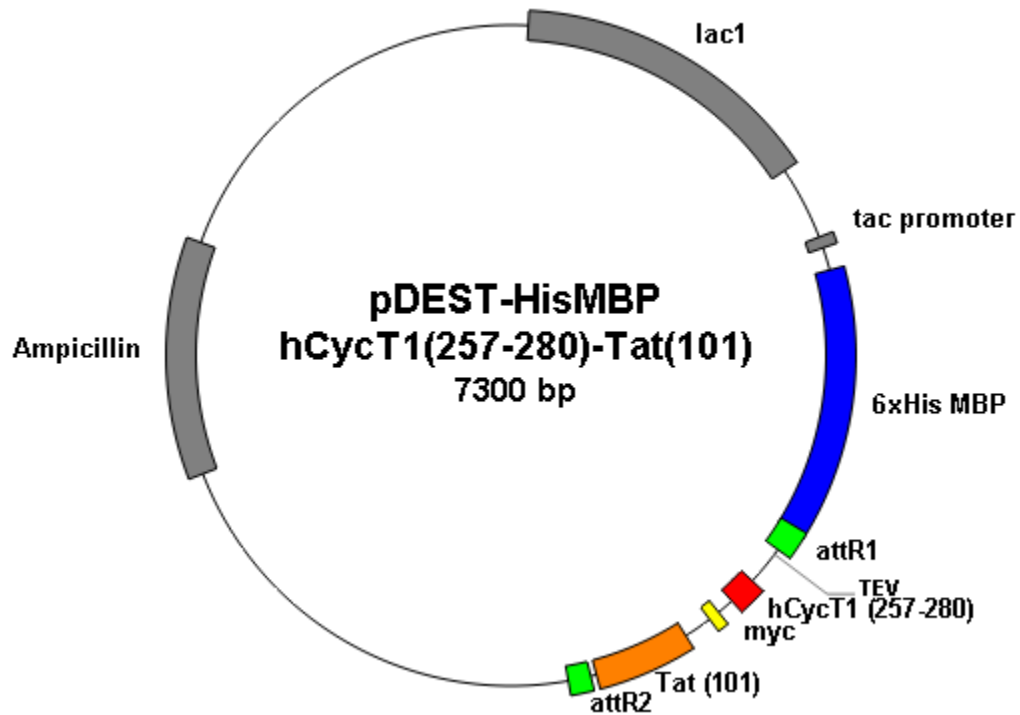
Figure 4-11 Diagram of the pDEST-HisMBP hCycT1-Tat chimera.

Schematic diagram of the pDEST-HisMBP hCycT1-Tat chimera plasmid. The chimera insert is flanked by attR sites compatible with the Gateway® Cloning recombination system. The tightly regulated tac promoter is upstream from the dual HisMBP tag which facilitates purification, and enhances solubility of the Tat chimera. The straight portions between the Tat (orange), myc (yellow), and hCycT1 (red), are linker regions that provide the chimera with the flexibility to adopt secondary structure.

261
|
GA-NWRACEAAKK-TKADDRGTDE-KTSE-QTMPEQKLIS-EEDLAMEFLE-IDPVD-
----------------Cyclin T1-------------------> --------------------Spacer-------------------->

Tat (transactivator of transcription)

```
          10              20  22  25 27  30   3334  37  40
          |               |   |   | |   |    || |   |
MEPVDPNLEP-WKHPGSQPRT-ACNNCYCKKC-CFHCYACFTR-KGLGISYG-
------Acidic/proline rich---------> ---------Cysteine-rich/ZnF--------> ---Core--->

   50            60            70            80            90           100
   |             |             |             |             |            |
RK-KRRQRRR-APQ-DSQTHQASLS-KQPASQSRGD-PTGPTESKKK-VERETETDPF-DL-S
------Basic-----> ------Glutamine-rich-----> -------------------------Exon 2------------------------->
```
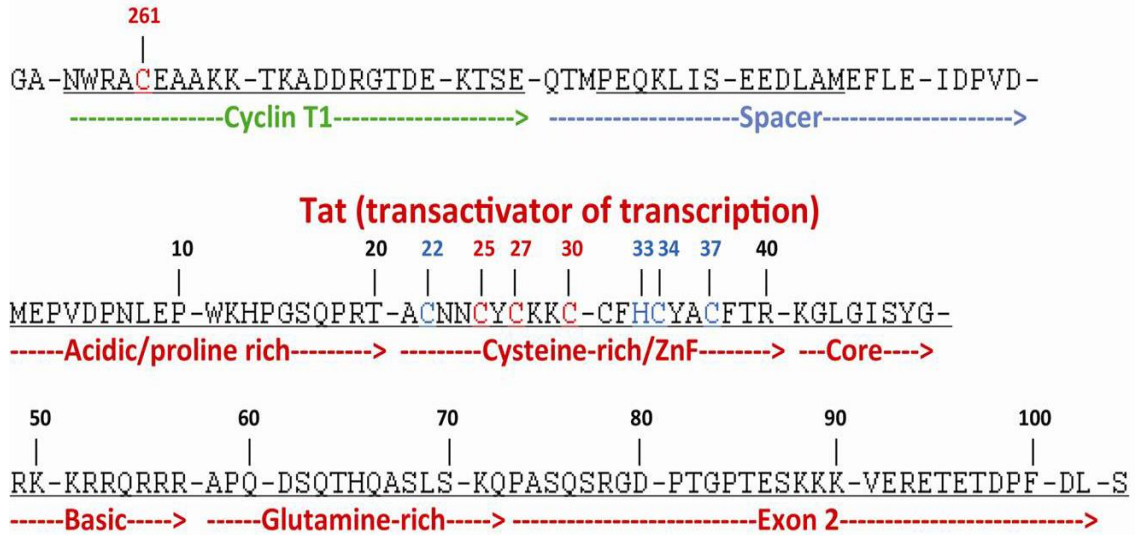
Figure 4-12 pDEST HisMBP 257-280 hCycT1-myc-Tat Protein Sequence

The final translated protein sequence of the HisMBP 257-280 hCycT1-Tat chimera confirmed with DNA sequencing by the Upstate Medical University DNA Core Facility and translated by Serial Cloner software. The sequence begins with glycine and alanine (artifact residues of the cloning process) at the N-terminus of the protein. The 257-280 portion of Cyclin T1 appears underlined in green, followed by a spacer region which contains the myc tag (underlined in blue). The 101 amino acid Tat sequence is followed by residual amino acids leucine and serine which remain (as artifacts of the cloning process) at the C-terminus.

Chapter 5 **A Preliminary Cyclin T1-Tat Chimera**

Once the construction of the pDEST HisMBP hCycT1-Tat chimera was complete, a suitable expression strain was selected. Choosing an appropriate strain for the production of a heterologous protein is essential to achieving high yield, adequate solubility, and proper folding. Due to rapid growth rate, low cost, and the extensive amount of genetic information available *E. coli* strains are regularly chosen for the production of recombinant proteins. In order to achieve the high yield required for structural study of the hCycT1-Tat chimera the chosen strain would first be required to accommodate the tightly controlled tac promoter of the pDEST-HisMBP plasmid.

5.1. The tac Promoter

When expressing recombinant proteins from plasmids with tac promoters it is generally best to select a non-DE3 non-pLysS expression strain (Novagen conversation). Researchers often propose that a DE3 strain can be used with plasmids utilizing tac promoters (42,76). However, here we show that the expression of the pRK793 plasmid (Appendix 8) for production of the TEV protease enzyme (HisMBP-TEV[S219V]) (74) in Rosetta Gami B (Novagen, Madison, WI) (Figure 5-1), a non-DE3 strain of *E. coli,* provides a substantial increase in target expression, as well as a moderate increase in basal expression, over the Rosetta Gami B DE3 strain. This observation can likely be attributed to the fact that the DE3 prophage expresses T7 RNA polymerase which, when utilizing a tac promoter, unnecessarily taxes the cells and should be avoided as tac expression systems employ the

native *E. coli* RNA polymerase. It is also important to note that the pLysS plasmid often provided with DE3 strains produces T7 lysozyme which regulates T7 RNA polymerase reducing basal expression, but is also unnecessarily taxing in, and obviated by, the use of plasmids with tac promoters.

5.2. Codon Usage

The next important consideration in the recombinant expression of the chimera is the presence of codons that are atypical in *E. coli*. Using the UCLA Rare Codon Calculator (http://nihserver.mbi.ucla.edu/RACC/) it was estimated, prior to sequencing, that the cleaved chimera sequence contains at least 15 rare codons that are known to hinder expression in *E. coli* by the stalling and premature termination of translation, incorporation of incorrect amino acids, and by frameshifts that contribute to low expression yield (77) (Table 7-1). Of the rare codons known to be present, 11 are rare arginine codons. There are also two occurrences of a rare leucine codon, and two occurrences of a rare proline codon.

The presence of rare codons in recombinant expression can be addressed by codon optimization, where an equivalent *E. coli* codon is substituted for the atypical codon, or by the use of expression strains that contain additional plasmids coding for the atypical tRNAs such as Rosetta Gami B (Novagen, Madison, WI).

5.3. Disulfide Bonds

In *E. coli* the formation of disulfide bonds is compartmentalized in the periplasm. Subsequently, the reduction of disulfide bonds is accomplished through the thioredoxin, and glutathione/glutaredoxin pathways in the cytoplasm. Mutation of the thioredoxin reductase (*trxB*), and glutathione reductase (*gor*) genes allow disulfide bond formation in heterologous protein to take place in the cytoplasm of *E. coli* by destabilization of the reduction mechanism (78). Substantial increases in active protein yield have been observed in expression strains with both the *trxB* and *gor* mutations with the result being attributed to improved folding where disulfide bond formation in the cytoplasm is facilitated (79). Strains possessing the dual mutation are commercially available.

The high number of cysteine residues in the Tat chimera and the presence of two zinc fingers alone suggest that an environment favorable to disulfide bond formation during expression could potentially augment expression yield. In support of the potential importance of disulfide bond formation, dramatically inhibited Tat activity has been observed in the presence of strong reducing agents leading to speculation about the presence of disulfide bonds in the active protein (44,80,81). Though no such disulfide bonds were reported by Tahirov et al. 2010 in the published structure of Tat complexed with P-TEFb, the conformation of the active site of Tat remains unknown, as does the conformation of Tat when bound to TAR. Moreover, a number of zinc finger proteins have been shown to demonstrate redox sensitivity (82) as well as behavior consistent with that of redox sensory proteins that may alternate conformation between zinc bound states and disulfide bond formation between cysteine residues.

5.4. Outer Membrane and Lon Proteases in *E. coli*

The outer membrane protease T (OmpT) of *E. coli* is a member of the omptin family of integral membrane peptidases implicated in the pathogenicity of several gram-negative bacteria. This highly specific endopeptidase cleaves between two basic amino acids, and demonstrates resistance to extreme denaturing conditions (83). In *E. coli* OmpT has been implicated in protein degradation (84). Expression strains containing an OmpT mutation inactivating this endopeptidase have demonstrated higher yields of heterologous proteins.

The Lon protease of *E. coli* is a highly conserved ATP dependent protease that degrades misfolded, or mutant proteins, and a few specific regulatory proteins (85). In the expression of recombinant proteins it is desirable to choose an expression strain with the Lon protease inactivated by mutation in an effort to prevent possible degradation of the target protein.

5.5. Lactose Permease

In *E. coli* lactose is utilized under the control of the lac operon system where lacY facilitates the transport of lactose into the cell, and lacZ cleaves lactose (86). In *E. coli* protein expression strains containing the lacZY deletion mutation, protein expression levels may be adjusted (called "tuning") throughout all cells in a culture by regulating IPTG concentration at induction. Proteins with solubility issues occasionally exhibit improved solubility with reduced concentrations of IPTG that theoretically allow the protein more time to fold properly. Therefore, in difficult to express proteins it may be

advantageous to employ this mutation. However, when using such strains auto-induction media may not be used as strains with the lacZY mutation do not produce the required allolactose.

5.6.Rosetta Gami B Strain (Novagen, Madison WI)

The commercially available Rosetta Gami B Strain of competent cells (Novagen, Madison, WI) provides a combination of attributes that are well suited to the expression of the recombinant Tat chimera. Rosetta Gami B cells carry an additional plasmid accommodating the expression of six rare codons. This strain, bearing the *trxB* and *gor* mutations, may also improve protein folding and the yield of soluble protein in proteins containing disulfide bonds. Both the OmpT and the Lon proteases have been removed by mutation from the Rosetta Gami B expression strain. Finally, this strain has a "tunable" expression feature, as a result of the lacZY mutation, presenting another potential mechanism for improved folding and solubility.

During the initial expression attempts with the GST tagged chimera virtually no expression was observed in either BL21 DE3 pLysS or BL21 non-DE3 non-pLysS, and very low expression was achieved with the Rosetta 2 (Novagen, Madison, WI) cells that contain a plasmid accommodating expression of atypical codons. After carefully considering the combination of attributes afforded by the strain, the characteristics of the recombinant protein being expressed, and the initial observations during expression of the

GST tagged chimera, the Rosetta Gami B cells were selected for the first attempts at

expression of the HisMBP-hCycT1-Tat chimera.

Position	Codon	Amino Acid
5	AGC	Arginine
18	CGA	Arginine
59	CTA	Leucine
61	CCC	Proline
70	AGG	Arginine
91	AGA	Arginine
100	AGG	Arginine
104	AGA	Arginine
106	CGA	Arginine
107	CGA	Arginine
108	AGA	Arginine
120	CTA	Leucine
124	CCC	Proline
129	CGA	Arginine
135	CCC	Proline
144	AGA	Arginine

Table 2 Rare codons and their position in the hCycT1-Tat chimera.

Rare codons in red and green are not accommodated by the Rosetta Gami B strain.

This fact will not prevent expression entirely as these rare tRNA are present in *E. coli*

albeit in lower amounts.
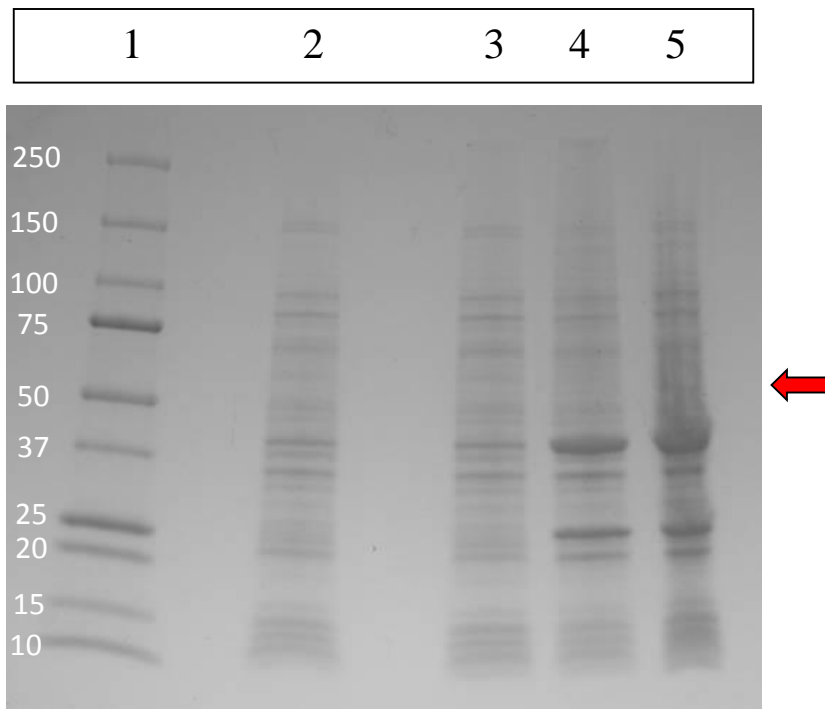
Figure 5-1 Expression of the HisMBP TEV pRK793 plasmid

The expression of the HisMBP TEV pRK793 plasmid (Appendix 8) (for recombinant expression of the Tobacco Etch Virus protease S219V mutant) in Rosetta Gami B DE3 versus non-DE3 strains. From left to right:

1 – ladder   2 – Rosetta Gami B DE3 uninduced

3 – Rosetta Gami B non-DE3 uninduced   4 – Rosetta Gami B DE3 induced

5 - Rosetta Gami B non-DE3 Induced

Chapter 6 Optimizing Protein Expression Yield

Optimizing the yield of recombinant protein expression is an empirical process that requires a dual pronged approach. First the overall crude protein expression yield must be optimized, and second the fraction of the expressed protein available in soluble form must be maximized. Care must be taken to avoid aggregation and precipitation after cleaving solubility enhancing tags, and when the target protein is present in solutions at both low and high concentrations. With an eye toward downstream applications, buffer compatibility with purification and assay requirements and long-term storage conditions must be optimized.

Control over the final soluble expression yield may be exerted at many points throughout the expression, purification, and storage of the protein. The choices of expression strain, growth media, and growth conditions, as well as a long list of potentially helpful techniques and additives that can be employed along the way results in a daunting number of possible combinations of varying efficacy. Predicting which of these techniques will be most effective for a particular protein is rarely possible, and empirical determination can be extremely costly and time-consuming. Hence, wherever possible high-throughput methods of assessing the efficacy of yield and/or solubility enhancing techniques are highly desirable.

While the subject of choosing an expression strain has been discussed previously, it is worth noting here that even within a specific transformed strain target protein expression

levels of some colonies may surpass that of others. Thus, it is often advantageous to screen multiple colonies of the same transformed strain for the purpose of comparing expression yields. In an assay called a Double Colony Selection (DCS) several colonies from a single transformation are assessed for recombinant target expression yield. The colony with the highest yield is then cultured overnight, plated, and in a second round of selection several colonies are again compared by yield of the target protein. In some cases this procedure can substantially increase the yield of recombinant expressions. Periodically re-transforming, and screening the expression strain to be sure that the plasmid is not lost is also an important step in maintaining high yields of protein expressions.

Once the transformed strain has been optimized, the next step is to optimize the growth conditions by providing appropriate nutrients in the form of growth media, and additives where necessary, and by controlling the growth conditions of temperature, aeration, and pH of the growth environment. Many different media formulations are available with a variety of nutrients and additives that may improve both the yield, and the efficiency of the expression protocol.

6.1. Growth Media

The appropriate selection of growth media is necessary for achieving maximal growth rate and cell yield. While *E. coli* are able to synthesize many of the nutrients they require, the production of soluble, stable, and functional recombinant proteins is best achieved

with the addition of trace metals, minerals, and vitamins to the growth medium (87). The

basic components of *E. coli* growth media include water, an amino acid nitrogen source

such as tryptone, a carbon source in the form of a fermentable sugar (such as glucose),

yeast extract (an additional nitrogen source), sodium chloride to regulate the osmotic

environment, and phosphate to provide a source of phosphate for growth and also to

buffer the growth media. Where a specific strain is growing aerobically at a fixed

temperature, both growth rate and yield of cells are dependent on the carbon source (88).

Supplementation of the *E. coli* medium with appropriate nutrients increases the quantity

of *E. coli* cells. Both plasmid and recombinant protein yield are directly proportional to

the quantity of *E. coli* cells.

Another important condition in the optimal growth of *E. coli* is the maintenance of near

neutral pH. While sensitivity to low pH is somewhat lower than sensitivity to high pH,

which can cause cell death, the optimal pH range for *E. coli* growth is 5.5 to 8.5 (87).

Without proper aeration *E. coli* produce acetic acid which lowers the pH of the culture,

and inhibits cell growth. Agitation of the culture during cell growth increases aeration of

the culture, as does the use of a baffle bottom flask. In addition, many media formulations

employ buffers such as potassium phosphate to maintain an optimal pH range during cell

growth.

6.2. LB Media

Lysogenic Broth (LB) also known as Luria-Bertani is a commonly used media formula for the growth of *E. coli* cell cultures. Originally, the formula was supplemented with the addition of 1 mg/ml glucose, but such supplementation is no longer common in contemporary LB formulas. Sezonov et al. 2007 found that steady-state growth of *E. coli* ended when the culture reached an OD_{600} of 0.3, and that the growth of the cell culture was limited by availability of carbon sources which *E. coli* catabolized, not from sugars, but from available amino acids (89). The use of LB media is generally an appropriate choice for recombinant expression of a protein of interest, and tends to produce low basal expression of other *E. coli* cellular proteins.

6.3. Rich Media

Achieving high cell yield and cost effective and efficient recombinant protein production can be enhanced in some expression systems with the use of rich media. Accumulation of acetic acid, a byproduct of glucose metabolism that inhibits cell growth, can be reduced with the use of alternative carbon sources such as glycerol, and also by media with buffering capacity. Many recipes for rich media contain trace metals, trace minerals, and vitamins, as well as proprietary ingredients in commercial formulations which often have animal sources (87). As with most other aspects of recombinant protein expression protocols, the usefulness of any particular media formulation must be determined empirically. One of the potential disadvantages of the use of rich media is an increase in basal expression which can hinder purification.

6.4.Turbo Broth™

One proprietary media formulation, Turbo Broth^{TM} (AthenaES Baltimore, MD), claims to achieve 4 to 5 times higher *E. coli* cell yields than that achieved in LB media alone. This rich media substitutes glycerol for glucose as a carbon source, and is supplemented with trace minerals, vitamins, inorganic compounds, and amino acids. The addition of potassium phosphate maintains the culture pH at 7.2 ± 0.2. Although not quantitatively compared, by visual comparison the induced protein yield of the HisMBP-hCycT1-Tat chimera appears considerably higher with the use of Turbo Broth^{TM} when compared to LB media (Figure 6-1). Predictably however, the basal expression of other *E. coli* cellular proteins also appears to have been considerably increased.

6.5.Auto-induction Media

Auto-induction media contains the carbohydrates lactose, and glucose. Initially *E. coli* use the limited amount of glucose available in the formula as an energy substrate typically until mid or late log phase. When the glucose has been depleted the *E. coli* then use lactose converting it to allolactose with the enzyme β-galactosidase. The lac repressor is released from the DNA by allolactose initiating the expression of the recombinant protein. The advantages of using this formula are that it is not necessary to monitor the culture to determine the induction point, increased cell mass is generally observed, and that optimal initiation of expression tends to increase protein yield (90). However, because of the lacZY mutation the Rosetta Gami B cells will not produce allolactose, and therefore this strain cannot be used for protein expression in auto-induction media.

6.6. Minimal Media

For structural study of proteins using nuclear magnetic resonance (NMR) isotopically labeled proteins must be expressed in *E. coli*. In particular, heteronuclear single quantum coherence (HSQC) requires the protein of interest to be ^{15}N labeled. Isotopic labeling of the protein with ^{15}N is typically achieved by growing *E. coli* in M9 media (Appendix 9) containing $^{15}NH_4Cl$ as the sole nitrogen source. The M9 media is supplemented with glucose, trace metals, vitamins, and other minerals, but in general *E. coli* incorporating ^{15}N tend to grow more slowly in minimal media. Some proteins can be expressed sufficiently with little effort in minimal media, while other less tractable targets can prove problematic. A multitude of minimal media formulas have been developed to improve labeled protein yields in minimal media, and a number of commercial and proprietary formulas are marketed for use in exceptionally difficult expressions. Commercial formulas such as BioExpress® (Cambridge Isotope Laboratories, Inc. Andover, MA), however, can be quite costly (at present in excess of $700 per liter) depending upon the number of isotopes incorporated.

6.7. Growth Phases of *E. coli*

The growth of *E. coli* can be described in four phases: lag phase, log phase, stationary phase, and death phase. When the optical density at 600 nm of a solution of *E. coli* in growth media is plotted as a function of time each of these four phases can be seen to occur in succession. When the culture is first inoculated the rate of growth is typically

during the initial lag phase as the *E. coli* acclimate to fresh media and antibiotics. During the second phase, the logarithmic phase or "log" phase, *E. coli* reproduce exponentially.

It is at the mid-point of this log phase that induction of recombinant protein expression will generally produce optimal yield. During the log phase the *E. coli* are said to be "healthy" as the medium is not yet depleted of nutrients, nor yet full of the toxic byproducts of *E. coli* metabolism. Recombinant protein expression is a process which is heavily taxing to the cell, and which sequesters many cellular resources. Moreover, many recombinant proteins are toxic to *E. coli*. Thus, the induction of expression when the *E. coli* are most fit and reproducing rapidly will generally produce the highest yield. As nutrients are depleted, byproducts contaminate the medium, and crowding occurs. At this point cells enter the stationary phase. In the stationary phase cell density is maintained at a steady state. During the final "death phase" the effects of the nutrient depleted medium and the buildup of the toxic waste byproducts of metabolism combine causing cell death observable as a decrease in culture density.

6.8.Growth Curves for the Tat Chimera in Various Media

Figures 6-2, 6-3, and 6-4, show the growth curves for the pDEST HisMBP hCycT1-Tat chimera in Rosetta Gami B cells (RBG) in LB, Turbo, and BioExpress® media. Figure 6-5 shows the growth curve for the double colony selection mutant (DCS) in BioExpress® media. In each of the growth curves below we see that optimal induction should occur between OD_{600} of 0.6 to 0.8.The mid-point of the mid-log phase for the growth curves of

the Rosetta Gami B (RGB) in LB, Turbo Broth™, and BioExpress®, and for the double

colony selection (DCS) was approximately OD_{600} 0.7.

When Tat chimera expression by RGB transformed with pDEST HisMBP-hCycT1-Tat

was compared in LB, Turbo Broth™, and BioExpress® media, as anticipated, the highest

recombinant protein expression was found in the Turbo Broth™ rich media (Figure 6-7).

Tat chimera expression by DCS in BioExpress® media was compared to expression by

RGB in BioExpress® media. The specifically selected DCS mutant demonstrated higher

expression than RGB in samples taken four hours after induction (Figure 6-7), but lower

levels of expression than RGB in samples taken after 18 hours. This could be due to

protein degradation which is commonly observed in lengthy expression protocols, and

which for some reason the DCS mutant may be more susceptible to. The yield achieved

for both RGB and DCS in the BioExpress® media was comparable to that achieved in

LB (Figure 6-7), a promising result when considering the low levels of recombinant

expression typical of minimal media when compared to nutrient rich media.

6.9. Comparison of Construct, Strain, and Media Changes

There was no consistent observable overexpression of the GST-hCycT1-Tat chimera in

BL21 with either the DE3 pLysS, or non-DE3 non-pLysS strains. This observation is

most likely due to the approximately 14% atypical codons which would not be

accommodated by these strains. When the GST chimera was expressed in Rosetta 2

(Novagen) cells a low level of expression was observed in 1 L of LB induced at OD_{600}

=0.7 with 1.0 mM IPTG and after 6 hours of growth at 28º C (Figure 6-8 A). By visual comparison of the samples in SDS PAGE, expression of the HisMBP chimera construct in Rosetta Gami B in 1 L of Turbo Broth™ was clearly and substantially increased over that of the original GST construct and protocol (Figure 6-8 B)
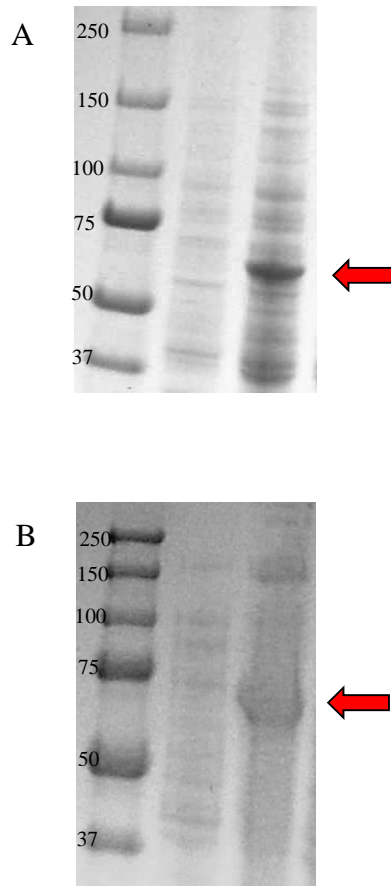
Figure 6-1xpression of the HisMBP Tat Chimera in LB and Turbo Broth

A. Expression of the HisMBP-hCycT1-Tat chimera in LB media

 : 1 – Ladder   2 – Uninduced   3 – Induced.

 B. Expression of the HisMBP-hCycT1-Tat chimera in Turbo Broth^{TM}

: 1 – Ladder   2 – Uninduced   3 – Induced

Both samples from one liter cultures induced at OD_{600} 0.7 and grown at 28°C for 7 hours.
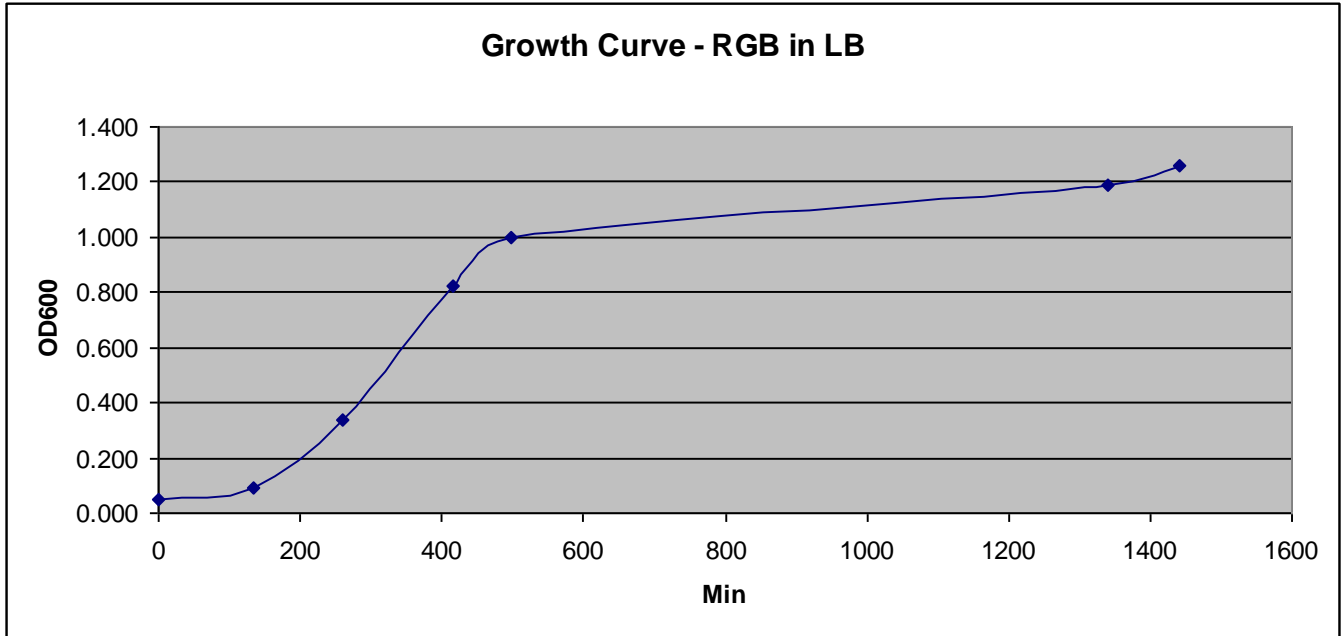
Figure 6-2 Growth Curve of Rosetta Gami B in LB Media

The growth curve of pDEST HisMBP hCycT1-Tat in Rosetta Gami B cells in LB media

at 37°C.
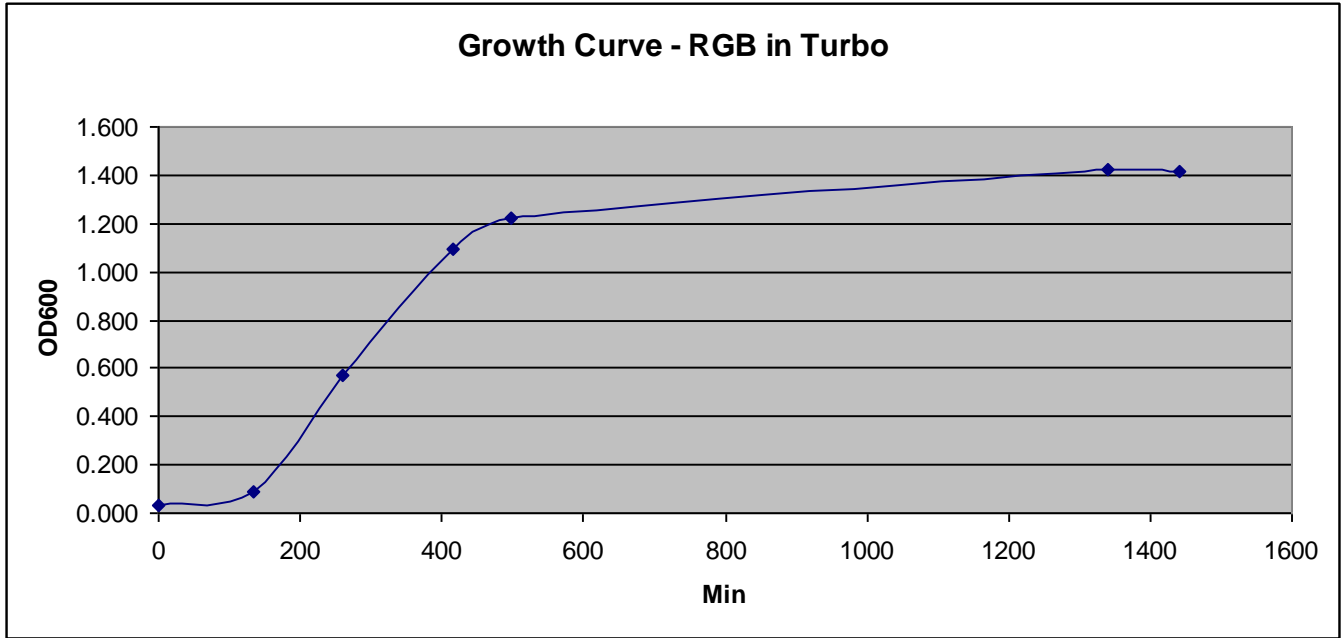
Figure 6-3 Growth Curve of Rosetta Gami B in Turbo Media

The growth curve of pDEST HisMBP hCycT1-Tat in Rosetta Gami B cells in Turbo

media at 37°C.
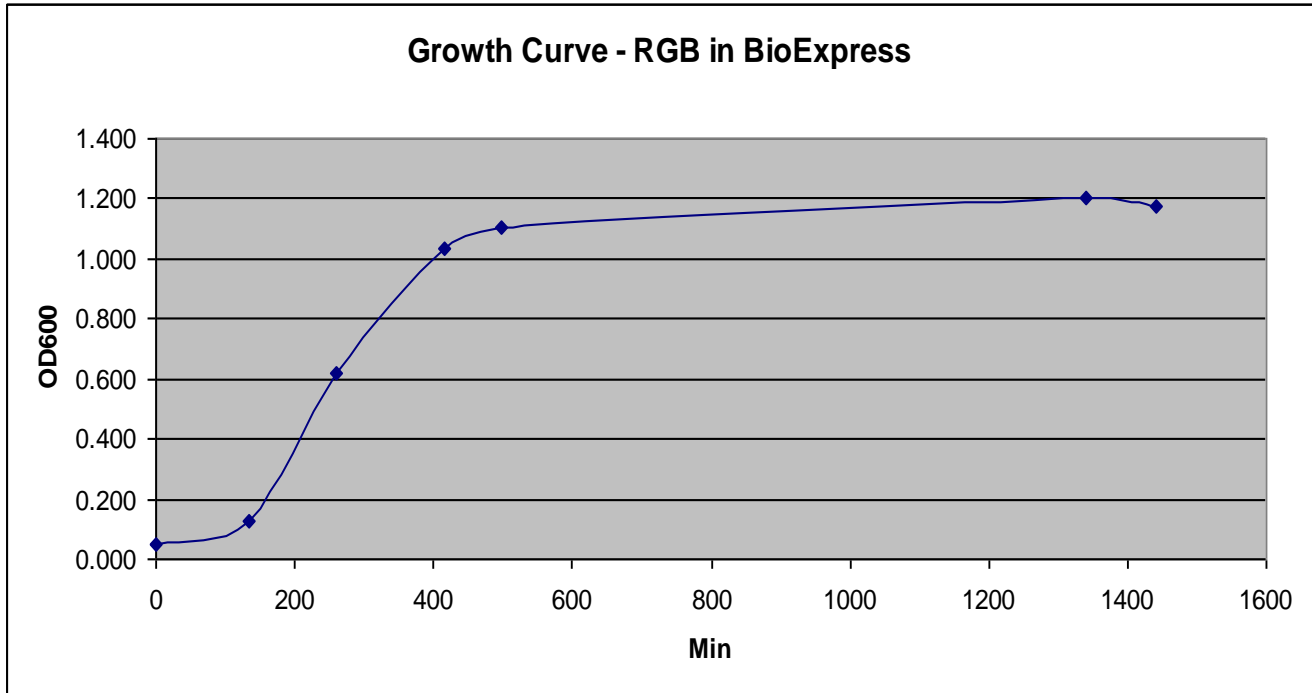
Figure 6-4 Growth Curve of Rosetta Gami B in BioExpress® Media

The Growth curve of pDEST HisMBP hCycT1-Tat in Rosetta Gami B Cells in

BioExpress® Media at 37°C.
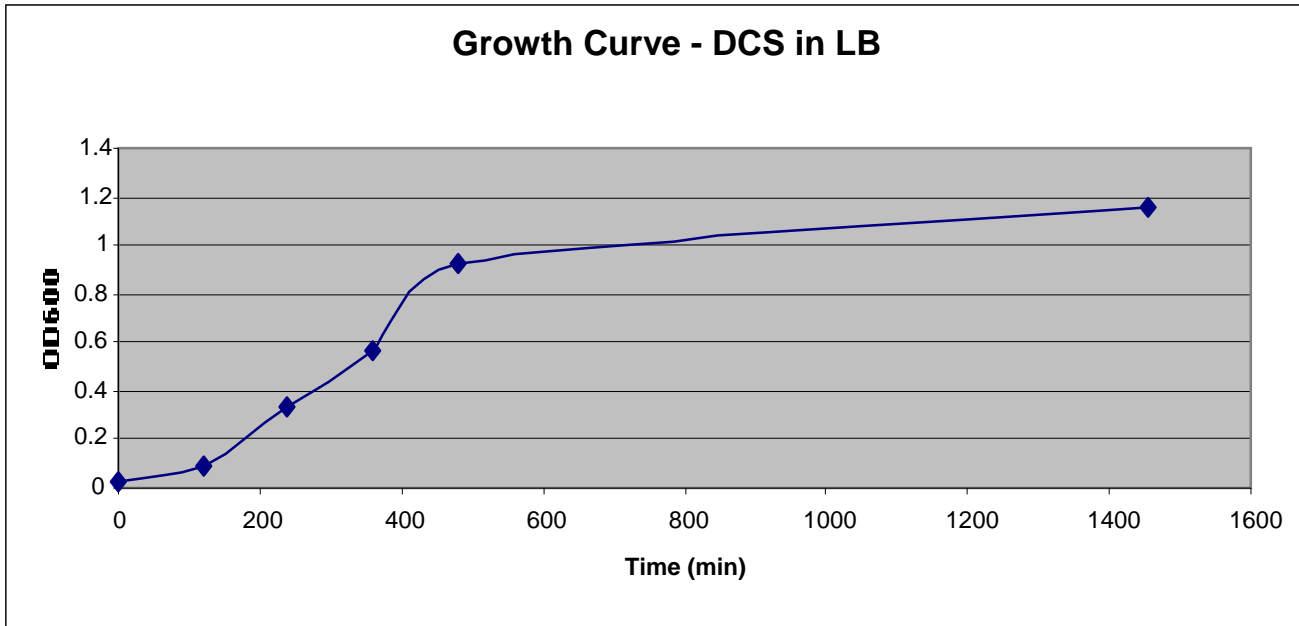
Figure 6-5 Growth Curve of Double Colony Selection in LB Media

The growth curve of pDEST HisMBP hCycT1-Tat in Rosetta Gami B cells Double

Selection Mutant in LB media at 37°C.
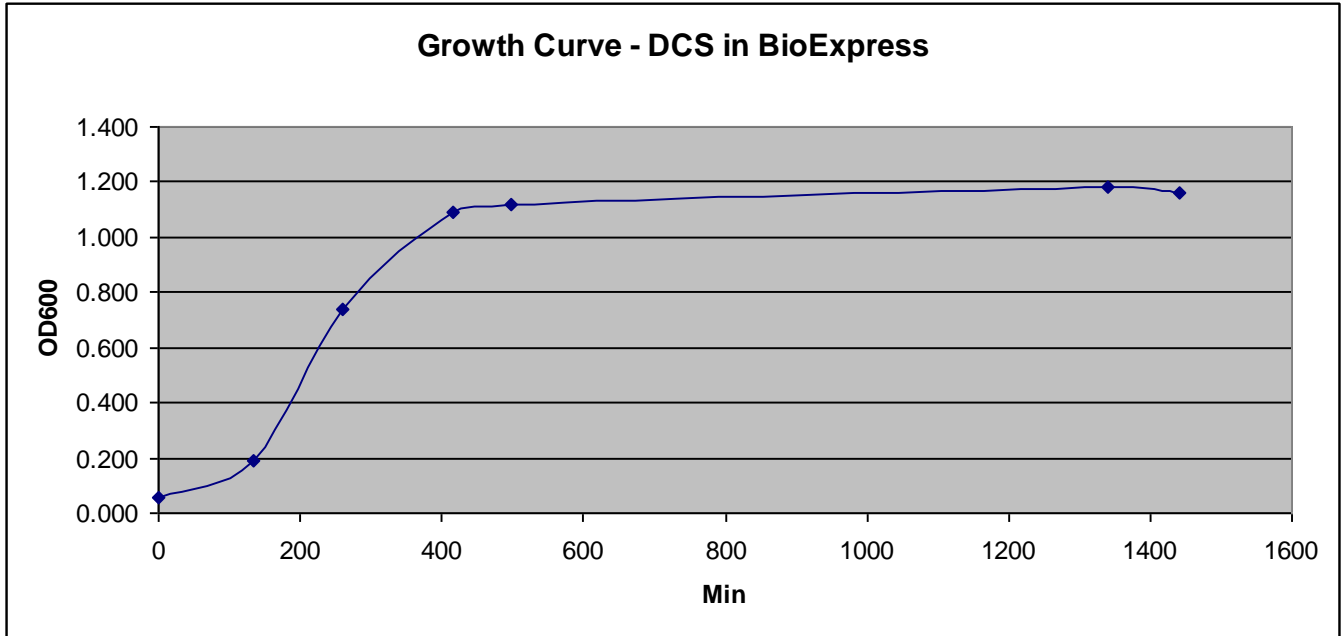
Figure 6-6 Growth Curve of Double Colony Selection in BioExpress® media.

The growth curve of pDEST HisMBP hCycT1-Tat in Rosetta Gami B cells Double

Selection Mutant in BioExpress® media at 37°C.
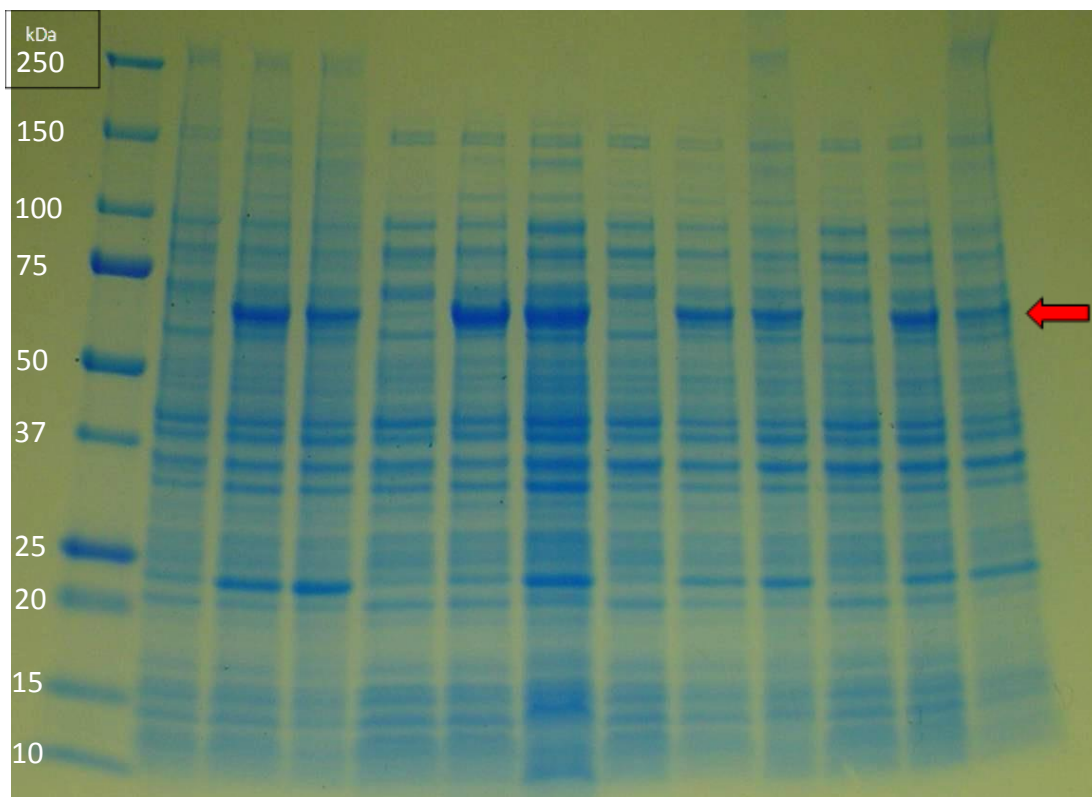
Figure 6-7 Small Scale Media and Strain Comparison in Various Media

Small scale (6 ml) media and strain comparison in various media with induction at OD_{600} =0.7. From left to right:

1 – ladder 2 – uninduced Rosetta Gami B (RGB) in LB

3 – induced RGB in LB after 4 hours 4 – induced RGB in LB after 18 hours

5 – uninduced RGB in Turbo Broth^{TM}

6 – induced RGB in Turbo Broth^{TM} after 4 hours

7 – induced RGB in Turbo Broth^{TM} after 18 hours

8 – uninduced RGB in BioExpress® minimal media

9 – induced RGB in BioExpress® minimal media after 4 hours

10 – induced RGB in BioExpress® minimal media after 18 hours

11 – uninduced double colony selection (DCS) in BioExpress® minimal media

12 – induced DCS in BioExpress® minimal media after 4 hours

13 – induced DCS in BioExpress® minimal media after 18 hours
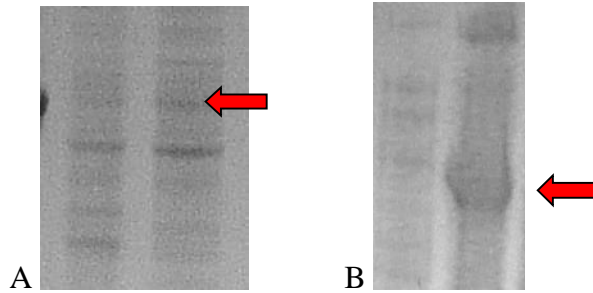
Figure 6-8 Comparison of GST Chimera and HisMBP Chimera Constructs

Comparison of protein expression with the original GST chimera construct and protocol and in the re-engineered HisMBP construct with revised expression strain and protocol.

A. Original GST-hCycT1-Tat chimera in Rosetta 2 cells expressed in LB induced at OD_{600}=0.7 with 1.0 mM IPTG after 6 hours at 28°C (as per original protocol from K. Fujinaga (Appendix 10)).

B. HisMBP-hCycT1-Tat chimera in Rosetta Gami B cells expressed in Turbo Broth™ induced at OD600 =0.7 with 1.0 mM IPTG after 6 hours at 28°C. Samples from 1 ml of 1 L induced culture resuspended in 100 ul 1 X SDS.

Chapter 7 Optimizing Recombinant Protein Solubility

Once a high rate of recombinant protein expression is achieved the next set of challenges entail avoiding proteolytic degradation of the nascent target protein, achieving a sufficient yield of soluble and active protein, and removing the solubility enhancing tag in preparation for downstream applications. While proteolytic degradation can generally be mitigated substantially by the introduction of protease inhibitors to the working solutions, solubility can be considerably less tractable.

The majority of published work addressing techniques shown to improve the solubility of recombinant proteins focuses on the notoriously difficult expression of membrane proteins. Receiving considerably less attention, however, are the 80% of non-membrane proteins which are poor subjects for structural assays primarily due to insolubility (91). Mammalian and other proteins are frequently poorly expressed in bacteria where the absence of post-translational modification and differences in the folding environment introduce formidable challenges to recombinant expression (92). In eukaryotic proteins multiple domains, and the requirement of cofactors, and protein partners tend to hinder recombinant expression efforts (93). In fact, the production of many of the recombinant proteins that would make attractive subjects for research is often abandoned because of the time and expense involved in their pursuit. The introduction of high-throughput protein expression and purification methods is a valuable addition to the field, with the advantage of permitting the screening of a multitude of variable expression, purification,

and storage conditions simultaneously. Unfortunately, this equipment is not yet widely available, and most laboratories still rely on more traditional empirical methods.

7.1. Denaturing and Refolding Recombinant Proteins

In many cases high expression levels of the recombinant protein are observed in the crude extract, but the protein can present in misfolded and inactive forms that aggregate in insoluble inclusion bodies. Rescuing the active form of the protein from insoluble inclusion bodies may or may not be possible, and can involve complicated denaturing protocols utilizing chaotropic agents or acids, followed by difficult refolding procedures into what the researcher hopes will be suitable buffers. Many problematic proteins do not readily refold into active conformations by known *in vitro* folding techniques (94), and high yields of soluble recombinant protein from denaturing and refolding protocols are rarely observed (45). Where refolding is possible, confirming that the form achieved is the proper native and active form of the protein generally requires additional assays that can also be challenging. Moreover, confirming the proper conformation may be complicated by a lack of available structural information for the protein of interest. Thus, wherever possible obviating such complicated denaturing and refolding techniques is highly desirable.

7.2.Increasing Recombinant Protein Solubility

Yield of the purified soluble and active protein may potentially be improved by a plethora of measures, among them: changes in expression conditions such as post-induction temperature, and IPTG concentration, as well as the strategic addition of a variety of potentially solubility enhancing compounds to the growth media. It is also possible to introduce a number of techniques during cell lysis, purification, concentration, and storage procedures that can further enhance solubility, and ultimately improve the final yield of the active protein. Commonly, a multitude of techniques are evaluated empirically, and combined in a single protocol in order to achieve adequate yield of the protein of interest.

7.3.Reducing the Expression Rate

One simple approach to improving the soluble yield of the target protein is to reduce the expression rate in order to allow additional time for proper folding of the native and thermodynamically favored state. Reducing the concentration of the inducing agent (IPTG in this case), and/or decreasing the temperature of the culture post-induction reduces the expression rate of the recombinant protein and frequently results in an increase in the yield of the soluble protein. A range of concentrations of the inducing agent, and temperatures post-induction should be evaluated to determine the optimal conditions for enhanced solubility. Weaker promoters and lower copy number plasmids can also be employed to reduce the expression rate of the recombinant proteins.

7.4. Increasing the Expression of Chaperone Proteins

In vivo protein folding can occur over a timescale ranging from milliseconds to days. During the folding process protein folding intermediates often have exposed hydrophobic surfaces that promote self-association and the formation of aggregates that lead to insoluble inclusion bodies and precipitation (95). Proteins known as chaperone proteins assist in the process of folding nascent proteins and in refolding improperly folded proteins.

In overexpression systems the high concentration of newly formed protein can exacerbate self-association leading to the increased formation of insoluble proteins and inclusion bodies. Moreover, the limited availability of chaperone proteins to assist in the folding and refolding of the newly synthesized proteins will also result in a higher fraction of insoluble protein in the form of inclusion bodies (45,94). Whenever possible, expression of these chaperone proteins should be increased concomitant with overexpression of the target protein in order to accommodate the increased folding requirements of the highly concentrated nascent protein.

In the cytoplasm of *E. coli* Trigger Factor, and the DnaK and GroES complexes are chaperone proteins that assist in the folding, refolding, and the prevention of aggregation in newly synthesized proteins (45). However, which chaperone protein is capable of enhancing the solubility of a particular recombinant protein of interest may differ depending upon characteristics of the target that are usually unknown at the time when optimization begins (94). Once again the empirical determination of optimal expression

conditions by multiple trials with a variety of chaperone proteins, and solubility

enhancing media additives remains a necessary and time consuming process.

7.5. Ethanol

When employed as a media additive, ethanol mimics the heat-shock response in *E. coli*

and has demonstrated efficacy in increasing the expression of heat-shock proteins that

function as molecular chaperones in *E. coli* (92,94-96). Media supplemented with as little

as 1% ethanol demonstrated enhanced heterologous protein expression in *E. coli* (96-98).

Georgiou and Valax 1996 reported that 3 % ethanol added to the growth medium

increased the heat-shock response and the production of chaperone proteins GroES/EL

and DnaK/J and demonstrated a synergistic effect on protein expression. Interestingly,

the enhancing effect of the addition of ethanol on solubility was prominent at the post-

induction temperature of 30ºC, but was markedly reduced at 37 ºC and at 42 ºC (95).

During expression of the HisMBP-hCycT1-Tat chimera the addition of 1% ethanol to the

growth medium approximately 30 minutes prior to induction, and the reduction in culture

temperature from 37ºC to 28ºC for 7 hours of expression produced a substantial increase

in solubility of the recombinant chimera (Figure 7-1). Here the introduction of ethanol

prior to the induction of expression allows some time for the accumulation of chaperone

proteins prior to expression of the target, and appeared to have little detrimental effect on

the growth rate of the culture (data not shown).

However, other work indicates a high yield of soluble protein can also be achieved with the introduction of ethanol prior to inoculation. Barroso et al. 2003 attributes the observed high target solubility in the presence of ethanol to the accumulation of stoichiometric intracellular concentrations of chaperone proteins prior to induction, and the expression of most rather than only some chaperone proteins though the continuous synthesis of heat-shock transcription factor σ^{32} (99). The heat-shock transcription factor σ^{32} controls the induction of some 20 heat-shock genes. The requirement of sequential interaction of nascent proteins with several chaperone proteins is reported by Gragerov et al. 1992, observing extensive aggregation in heat-shock transcription factor σ deficient mutants, and reporting the effects of four heat-shock proteins on the deficient mutant (100).

Where the growth rate of the culture prior to induction is not prohibitively negatively impacted by the addition of ethanol it may be advantageous to include ethanol in the growth media as early as possible, and also at concentrations greater than 1%. Since the addition of 1% (v/v) ethanol at 30 minutes prior to induction produced adequate yield of the soluble protein, the alternatives of introducing ethanol prior to inoculation, and higher concentrations of ethanol were not evaluated. However, the synthesis of heat-shock proteins accelerates rapidly with temperature shift, and is believed to reach steady state within 20 minutes (101). Thus it is possible that introducing ethanol prior to inoculation may offer no further improvement in solubility over that observed with the introduction of ethanol 30 minutes prior to induction.

Increasing the concentration of ethanol could increase the concentration of the soluble

fusion chimera. However, at some point high concentrations of ethanol will produce

diminishing returns for example: where the growth of the culture is taxed by toxic effects

of ethanol, and where translocation of recombinant proteins may be inhibited as observed

by Chaudhuri et al. 2006 at concentrations of ethanol in excess of 2.5% (102). The

efficacy of ethanol in enhancing the solubility of any recombinant protein, its point of

introduction into the expression protocol, and the concentration employed must be all

determined empirically and specifically for each protein target.

7.6. Osmolytes

Osmolytes are small chemically diverse organic metabolites that are made and

accumulate in the intracellular medium of cells in response to osmotic stress (103-105).

Proteins are purportedly stabilized and protected from denaturants in the intracellular

milieu by organic osmolytes that force folding, despite harsh conditions, by interaction

with the protein backbone (103,106). Osmolytic interaction with the protein backbone is

described as a highly unfavorable "solvophobic" interaction that raises the Gibbs free

energy of the denatured state substantially more than it raises the Gibbs free energy of the

native state and in this manner stabilizes the folded form of the protein (103,105). Such

stabilization of folded proteins is highly desirable for structural study. Moreover,

osmolytes may prove useful in influencing the proper folding of proteins *in vivo* as well

as *in vitro*, potentially as therapeutics in the treatment of misfolding diseases such as

cystic fibrosis (106), or as stabilizers of protein therapeutics which are notoriously

hindered by difficulties with stability and long-term storage (107).

Naturally occurring osmolytes can be grouped into three categories: polyols, amino acids,

and the combinations of methylamines, methylsulfonium, and urea (103,108). This

diverse group of compounds can be employed to improve solubility, and stability at many

different points throughout expression, purification, concentration, and storage of

recombinant protein protocols. Here the inclusion of the polyols ethylene glycol during

cell lysis, and glycerol throughout purification and storage of the fusion chimera

appeared to improve the yield of soluble recombinant protein. In the absence of these

compounds low amounts of protein were observed in the soluble extract, and the protein

tended to be lost to precipitation during IMAC purification, and subsequent

concentration.

7.7. Additional Solubility Enhancing Compounds

A number of additional co-solvents are well known to positively influence the solubility

of proteins in solution; among these salts are highly influential and their effects on

solubility were well described by Hofmeister as early as 1888. Detergents, while most

often utilized to improve the solubility of membrane proteins, are not infrequently found

to improve the solubility of non-membrane proteins as well. While detergents often help

to prevent protein aggregation, above relatively low threshold concentrations many

detergents tend to form high molecular weight aggregates that are prohibitively difficult

to remove, can absorb at 280 nm, and are often incompatible with downstream applications. When possible, it is desirable to avoid the use of detergents.

7.8.Salts

In general the solubility of a protein is dependent upon dissolved salts, pH, temperature, and the polarity of the solvent. When working with recombinant proteins the pH is generally kept as close to a biological (pH 7.4) as possible, the buffer is polar, and the temperature is commonly kept low to prevent degradation. The salt concentration of the solution, however, can be modulated to increase the solubility of the protein consistent with the phenomenon known as "salting in" in which the solubility of a protein increases as salt is added to the solution. In salting in the addition of salt ions shield the ionic charges of the protein preventing the aggregation and precipitation of the protein molecules. Increasing the salt concentration past the optimal solubility range however, will produce the opposite "salting out" effect whereby the solubility of the protein decreases as the salt ions and protein molecules compete for molecules of solvent (109) (Figure 7-2).

7.9.The Hofmeister Series

The Hofmeister Series, published in 1888, ranks the effect of ions on protein stability. Kosmotropes are anions of high charge density and have a favorable effect on protein stability. Chaotropes are cations of low charge density and also tend to stabilize protein

structure (Figure 7-3). Though these observations generally hold to be true independent of the protein being studied the underlying basis for this is not well understood, and despite extensive study of this perplexing phenomenon the mechanisms of stabilization remains elusive. So elusive is this mechanism, in fact, that researcher Barry W. Ninham, professor emeritus of the Australian National University who has spent much of his career studying the phenomenon described it as an area of research that is "rediscovered every 10 years" only to be abandoned when it is discovered to be a "Sisyphean task" that he maintains is daunting in its complexity (110).

7.10. Timasheff and Thermodynamics in Protein Stability

Some light was shed on the subject however, nearly one hundred years after the introduction of the Hofmeiser Series when Timasheff and colleagues demonstrated that osmolytes, several other compounds, and ions are preferentially excluded from the immediately surrounding environment of the protein. This exclusion effectively produces a "preferential" hydration zone surrounding and stabilizing the folded state of the protein (105,111-118). The work of Timasheff provided the thermodynamic basis for the ranking of the effect of inorganic salts on solubility by Hofmeiser, and suggested that the solubility enhancing effects of these inorganic salts were dependent upon solute-dependent differences in the preferential interaction of proteins with solvent and co-solvents in solution (119).

The predictable effect of the Hofmeiser Series of ions on the thermodynamic stability of the native state can be used to influence solubility, crystallization, aggregation, and the stability of proteins. Clearly, a better understanding of stabilization mechanisms has tremendous potential for the advancement of protein therapeutics where preventing aggregation and maintaining stability are constant challenges, and in structural study where the formation of crystals for x-ray crystallography is so notoriously difficult. While elucidating the intricacies of the mechanism behind the observed effect of the ions in the Hofmeiser series is beyond the scope of this text, the properties and trends of the co-solvents aforementioned and their effects on solubility have been considered and employed throughout the expression, purification, and storage protocols of the fusion chimera in order to achieve adequate yield for structural work.

The challenges and complexities of elucidating the mechanisms of protein solubility and stability notwithstanding, the author looks forward to progress in this regard. Advances in proteomics, data generated from high-throughput empirical methods of determining solubility and stability, and bioinformatics might be used to compile databases which aid in the determination of these mechanisms.
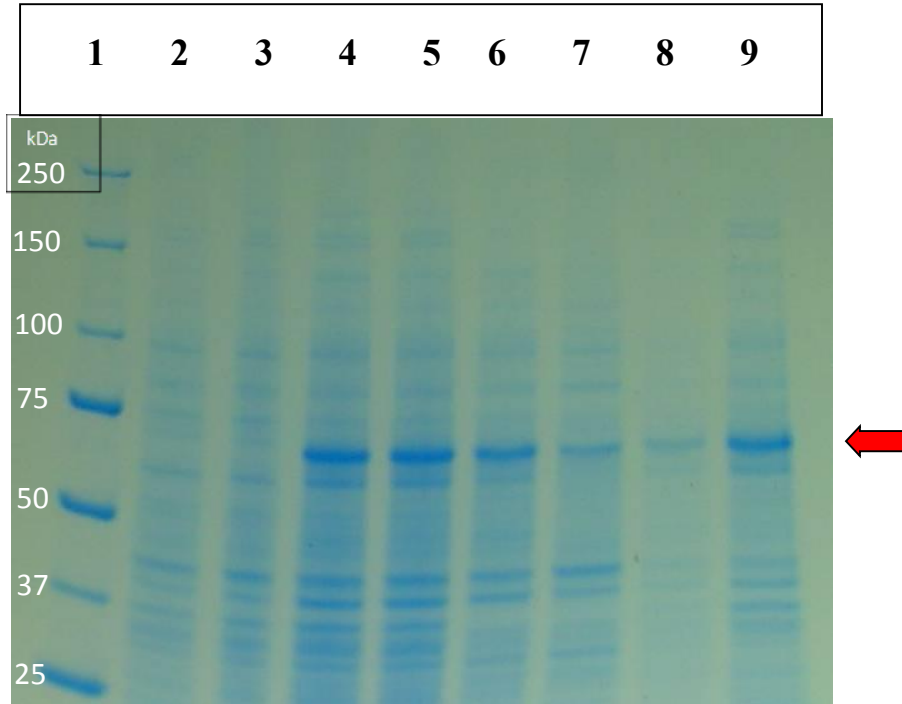
Figure 7-1 Expression of HisMBP Tat Chimera in LB media.

Analysis by SDS PAGE of small scale (25 ml) expression of the pDEST HisMBP

hCycT1-Tat in LB media. Induced with 1 mM IPTG at OD_{600} 0.7 expressed at 28 ℃ for 7

hours. s:

1 – Ladder   2 – Uninduced 1% EtOH   3 – Uninduced no EtOH

4 – Induced 1% EtOH   5 – Induced no EtOH   6 – Soluble 1% EtOH

7 – Soluble no EtOH   8 – Insoluble 1% EtOH   9 – Insoluble no EtOH

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine

buffer. Run condition: 200V, 50 minutes.
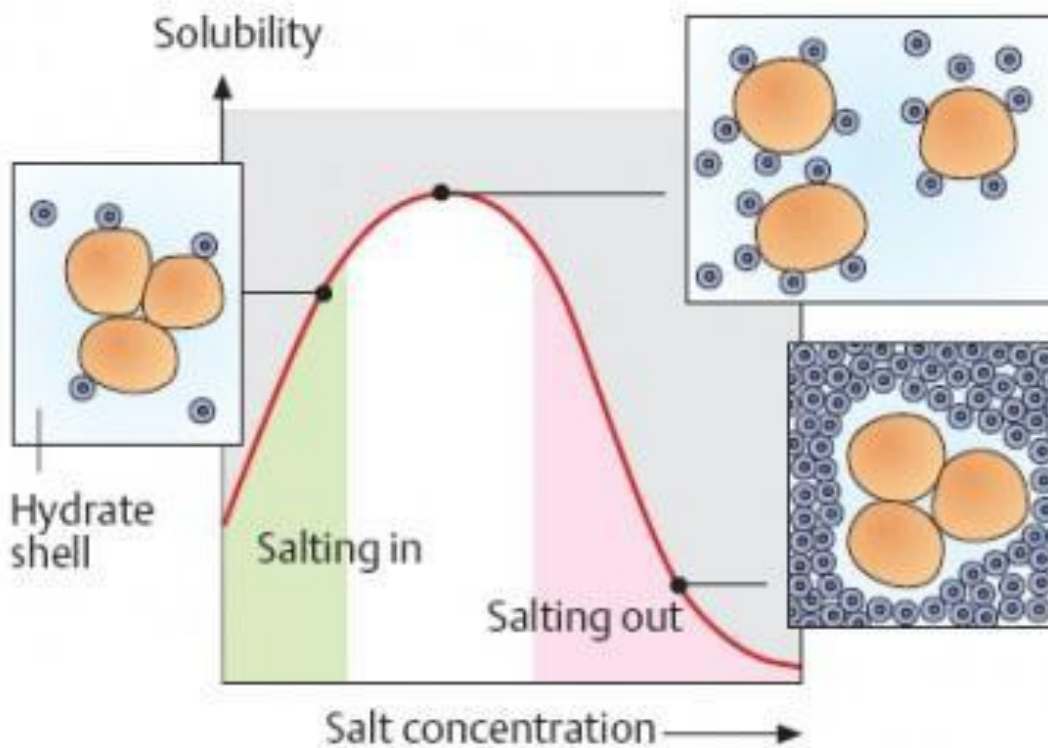
Figure 7-2 The effects of increasing salt concentration on protein solubility.

Salt concentration can be modulated to increase the solubility of protein consistent with the phenomenon known as "salting in" in which the solubility of a protein increases as salt is added to the solution. In salting in the addition of salt ions shield the ionic charges of the protein preventing the aggregation and precipitation of the protein molecules. Increasing the salt concentration past the optimal solubility range however, will produce the opposite "salting out" effect whereby the solubility of the protein decreases as the salt ions and protein molecules compete for molecules of solvent (109). Reprinted from *Color atlas of biochemistry* Koolman, J. and R.H. Rohm Copyright (2005) (120) with permission from Jan Koolman.

.
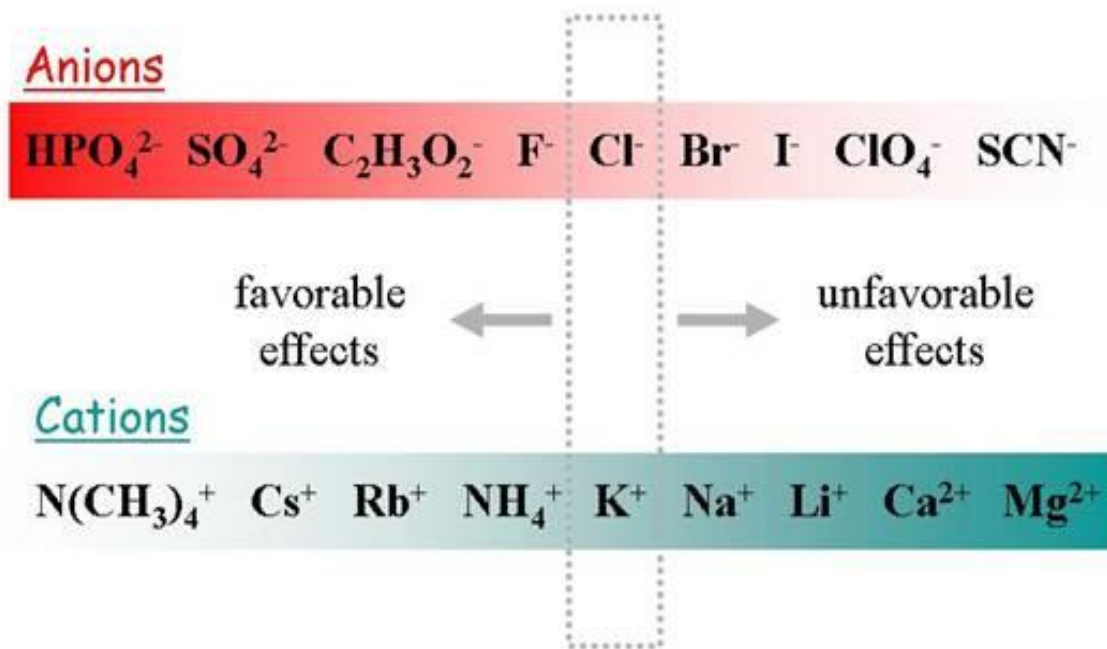
Figure 7-3 The Hofmeister Series of Ions

Reprinted from "Experimental System II: Concentrated Aqueous Solutions & The

Hofmeister Series" with permission from Darryl Eggers. Accessed online at:

http://www.chemistry.sjsu.edu/deggers/new_page_3.htm

Chapter 8 Protein Extraction

Structural determination of a protein of interest requires highly concentrated and purified protein, in native conformation. After expression the target protein must be selectively isolated from other cellular proteins, as well as from the cell membrane, and other cellular debris. Throughout this process, and from lysis to storage, the factors affecting solubility discussed previously must be consistently considered in order to maintain the stability of the protein in solution. With each process added to the protocol in an effort to improve purity, some of the yield is inevitably sacrificed. Optimization of each process in the purification, and the minimization of the number of steps overall, will produce the most efficient and reproducible protocol for high yield recombinant protein production. After many attempts an efficient and reproducible protocol was developed for the expression the HisMBP Tat chimera (Appendix 11).

8.1. Cell Lysis

The cell culture of the fusion chimera in Turbo media was induced with 1 mM IPTG at $OD_{600} = 0.7$, expressed overnight at 28ºC in 1% ethanol, and then harvested by centrifugation. The pellets were washed in ice cold PBS, and then resuspended in 4 ml of ice cold lysis buffer (pH 7.4) per gram of wet cell pellet weight. The lysis buffer consisted of Tris 20 mM, NaCl 0.5 M, $ZnCl_2$ 10 uM, imidazole 20 mM, BME 5 mM, ethylene glycol 20% (v/v), HALT protease inhibitor cocktail EDTA Free (Pierce Rockford, IL) 10 ul/ml, arginine 50 mM, and lysozyme 10 mg/ml.

8.2.Lysis Buffer Design

Many different formulations of the lysis buffer were evaluated. The original protocol received from Koh Fujinaga for the GST tagged chimera specified PBS, but after several trials with PBS, HEPES, and Tris, Tris proved to be the most effective buffer for the HisMBP fusion chimera. Since phosphate ions are known to chelate metal ions; in the interest of maintaining the integrity of the two zinc fingers in the HisMBP fusion chimera, PBS was replaced. Trials with HEPES buffer failed to maintain protein stability during FPLC purification. This is likely due to difficulties related to the use of secondary amines in buffers and the reduction of nickel ions during IMAC purification.

With respect to the use of reducing agents, in initial trials with the HisMBP fusion chimera TCEP was used in the hopes of providing more stable and complete reduction, and less interference with protein quantitation measurements at 280 nm. However, TCEP is inactivated by PBS buffers, and in addition, Bigalke et al 2011 observed that Tat protein precipitates slowly in the presence of TCEP. Replacing TCEP with BME provided improved solubility of the chimera during cell lysis and purification, as well as during concentration, and storage. Improved solubility was apparent from consistent and increasing concentration measurement by Nanodrop, when using BME as opposed to decreasing concentration observed in samples during processing when TCEP was used instead.

After several trials of varying salt concentration, the relatively high 0.5 M concentration

of NaCl proved to be both recommended (GE Life Sciences) and necessary to maintain

the stability of the protein during FPLC purification. Since the yield and purity achieved

with sodium chloride was sufficient for downstream applications, other salts were not

evaluated. However, based on the Hofmeister series and the work of Wei et al. 1998, and

Frankel et al. 1988 the use of potassium chloride in lieu of sodium chloride could provide

additional stability and might make a worthwhile substitution where additional stability is

required (17,43).

A low 10 µM concentration of $ZnCl_2$ was maintained throughout purification and storage

as a precautionary measure to preserve the integrity of the two zinc fingers. Imidazole is

present in the lysis buffer at a relatively low 20 mM concentration for the purpose of

preventing nonspecific binding to the nickel column during IMAC FPLC purification

later in the protocol. The inclusion of ethylene glycol in lysis buffers is generally present

as a cryoprotectant, while here it may have the added benefit of being an osmolyte that

tends to improve both the solubility and the stability of the target. The HALT protease

inhibitor cocktail (Pierce Rockland, IL) provides seven different protease inhibitors, and

effectively inhibited protease related degradation. The EDTA free formula of HALT was

necessary to prevent EDTA chelation of the zinc ions from the fusion chimera.

8.3. Arginine

Arginine is a potent aggregation suppressor that purportedly prevents aggregation by

masking the hydrophobic surface of the protein and prohibiting protein-protein

interaction, specifically such aggregation as can occur due to heating, dilution, and partial

unfolding (121,122). Arginine has been shown to have a solubilizing effect on proteins

within insoluble inclusion bodies (123). Protein aggregation in low concentration

environments, while somewhat less intuitive, is frequently a concern during the early

stages of purification. Arginine is present in the lysis buffer at a concentration of 50 mM.

However, evaluating the effect of arginine on solubility was difficult due to its removal

during FPLC purification to prevent interference with downstream applications.

It should be noted that many other co-solvents were explored, and were not found to be

particularly helpful to any substantial degree. The high number of potentially useful co-

solvents, and the length of time required to complete each expression do not lend

themselves readily to comparison. In the interest of time several factors were often

changed between trials and comparison of the effects of any one component was often

impractical. The appropriate combination of co-solvents determined to achieve adequate

yield may differ markedly from those required to achieve optimal yield, and it is possible

that some of the buffer components were redundant in nature, or even antagonistic rather

than complimentary. Notably however, the addition of osmolytes did appear to have the

most profound effect on enhancing solubility of the fusion chimera when expression

levels were compared in SDS PAGE analysis (Figure 7-1).

8.4. Cell Wall Disruption

A variety of lysis techniques such as sonication, freeze-thaw, liquid homogenization, and

mechanical methods are all commonly used methods of rupturing the cell wall of *E. coli*.

While all but mechanical means were tried in the lysis of the chimera, sonication provided the most consistent and high yield of soluble fusion protein. Sonication breaks down the cell wall by delivering pulses of sonic waves. The sonication process does cause the temperature of the cell solution to become elevated, so several short pulses are typically alternated with equally short intervals at reduced temperature to maintain the low temperature of the solution.

After collecting the cells by centrifugation at 2,500 rpm for 25 minutes at 4°C, the cell pellet was held on ice in 4 ml of lysis buffer per liter for 15 to 20 minutes to begin degradation of the cell wall. The protein extraction reagent B-PER (4 ml/L of culture) (Thermo Fisher Scientific, Inc. Rockford, IL), a non-ionic detergent solution, was added for the final 10 minutes on ice. The cells were then lysed by sonication for several 25 second bursts alternated with 25 seconds on ice until the solution color changed slightly becoming somewhat darker and semitransparent. It is important to avoid overheating the solution thereby degrading the target protein, and also to avoid producing foam during sonication which exposes the protein to air and can lead to aggregation and the formation of inclusion bodies.

8.5. Nucleic Acid Precipitation

Nucleic acids were precipitated with 4% (v/v) polyethyleneimine (PEI) by vortexing lightly followed by a brief hold. In addition to removing nucleic acids, PEI may also improve solubility. The cationic polymer PEI prevents the aggregation, and oxidation, of

proteins, and chelates the metal ions required by proteases to degrade proteins. The solution was then centrifuged at 21,000 rpm for 30 minutes at 4°C to separate the soluble proteins from the cellular debris. The soluble protein was decanted from the insoluble pellet, and the solution was syringe filtered with a 0.2 uM GD/X filter (Whatman Piscataway, NJ) prior to Immobilized Metal ion Affinity Chromatography (IMAC) purification.

Chapter 9 FPLC Purification of the Tat Chimera

The term chromatography, from the Greek for "color-writing", refers to a group of techniques used to separate mixtures. Separation by chromatography was first described by Friedlieb Ferdinand Runge in his 1855 work "Der Bildungstrieb der Stoffe, veranschaulicht in selbstständig gewachsenen Bildern" on early paper chromatography. Runge's work is often considered the precursor to the invention of chromatography (124). The invention of column chromatography is most often credited to M.S. Tswett in 1903 (published in 1905), but unfortunately due to the political climate in Russia, and the criticism of other researchers who were unable to reproduce his results, Tswett's work went largely unrecognized for many years after his report (125,126). In 1952 Martin and Synge received the Nobel Prize jointly for their invention of partition chromatography.

9.1.Column Chromatography

In column chromatography solutions are partitioned into mobile and stationary phases, and particles are separated on the basis of retardation from the mixture based on their relative affinity for each of these phases. A hollow cylindrical column is packed with absorbent material that constitutes the stationary phase, and in biochemical applications is usually a solid or a gel material. The column material is then wetted with the solution that will serve as the mobile phase for the separation. The mixture being separated, the analyte, is applied to the top of the wet column, and travels through the column propelled

by gravity, and is then separated by exploitation of the differences in the particles being separated, and their relative affinities for of the column material, or stationary phase.

9.2.Liquid/Column Chromatography

Liquid/column chromatography refers to all forms of chromatography in which the mobile phase, the analyte, is a liquid, and in which the stationary phase is linked to an inert matrix. Proteins are most often separated and purified with liquid/column chromatography by exploiting particle differences such as size, hydrophobicity, charge, and by the presence or absence of metal binding residues.

9.3.Fast Performance Liquid Chromatography (FPLC)

Prior to the 1970s protein separations by liquid/column chromatography were propelled through columns by gravity in lengthy procedures that frequently produced poor separation. In the late 1970s the addition of pressure during separation, provided by nearly pulse-free pumping systems, led to the development of high performance liquid chromatography. The improved separation was achieved by greatly decreasing particle size, dramatically increasing the surface area of the solid phase. Thus, solutes equilibrate more rapidly between the solid and mobile phases effecting higher resolution of solutes having similar interactions with the solid phase. Higher pressures are required to maintain reasonable flow rates for the mobile phase through the small particles. Fast performance liquid chromatography (FPLC) is similar to high performance liquid chromatography

except that the wetted surfaces of the column, detectors, and tubing are made from glass or fluoropolymers to avoid denaturing proteins on the metal surfaces common in HPLC. This requires lower pressures for FPLC, and particle sizes that are larger than for HPLC to accommodate reasonable flow rates.

During FPLC the solvent velocity is controlled by pumps that control the flow rate of the mobile phase of proteins through one of four types of columns: size-exclusion, ion-exchange, reverse-phase, and affinity. As the name implies, size-exclusion chromatography separates proteins from solutions on the basis of the size of the protein and its speed of migration through porous silica beads of variable size. Ion-exchange column chromatography separates proteins by differences in the net charge of the protein and its affinity for charged column material in high and low salt mobile phases. Reverse-phase chromatography separates proteins based on differences in their hydrophobicity, and therefore by the protein's affinity for the stationary phase. Finally, affinity chromatography exploits the affinity of specific residues of the target protein for a specific ligand or column material that is affixed to the matrix of the stationary phase.

In the experiments that follow the ÄKTA FPLC explorer system and the Unicorn 4.11 system control software (Amersham Pharmacia Biosciences, Uppsala Sweden) were used for all protein purifications of the Tat fusion protein chimera. After many trials an efficient and reproducible protocol was developed for the purification of the HisMBP Tat Chimera (Appendix 12).

9.4. Purification of the GST Chimera by Column Chromatography

Affinity chromatography was used to purify the original GST tagged construct by exploiting the affinity of the fusion tag for the GSTrap FF 5 ml column (GE Healthcare Biosciences Pittsburg, PA). The low expression level and poor solubility of this original construct produced a predictably low yield of the full length protein in the initial purification round. Additional issues arising from removal of the GST tag with thrombin protease compounded the problem and reduced the yield still further. Much of the full length target remained uncleaved after incubation with thrombin at temperatures ranging from 7-15°C and over periods of 2 to 24 hours (data not shown) Poor cleavage was observed in both on column trials and in solution. The low efficiency of thrombin cleavage was likely due, in part, to the fact that the optimal temperature for thrombin cleavage is 22°C (127). However, since high temperatures can adversely affect the native conformation of the target protein, and appeared to be causing precipitation (as evidenced by decreasing concentration measurements of absorbance at 280 nM over time by Nanodrop) higher temperatures were avoided.

Predictably, on column cleavage trials were less effective at removing the GST tag from the full chimera than were the trials in solution (data not shown). In this case the low expression level of the full chimera makes it unlikely that cleavage was hindered by high concentrations of the substrate, as is frequently the problem in on column cleavage protocols. Steric occlusion, concentration issues, and the restricted movement of proteins affixed to the stationary phase are issues known to hinder on column cleavage attempts.

For this reason cleavage of solubility enhancing tags with both the target protein and the protease free in solution is the generally preferred method.

In addition to the problem of incomplete cleavage, was the problem of non-specific cleavage for both the on-column and in-solution trials. Size-exclusion chromatography with the Superdex 75 100/300 (GE Healthcare Biosciences Pittsburg, PA) column failed to adequately separate the tag-free chimera from the GST tag, the uncleaved full chimera, and the non-specifically cleaved fragments despite what should have been sufficient differences in molecular weight for adequate separation. Attempts to separate the proteins by charge with ion-exchange chromatography were also poorly resolved, and several proteins co-purified. When affinity chromatography was used to remove the GST tag, yield of the final target was very low and the solution was also contaminated with several co-purified proteins. Ultimately, the cleaved chimera was inseparable from the full length GST fusion construct, and from a number of non-specifically cleaved fragments, even after the addition of a several polishing steps.

9.5. Affinity Chromatography with the HisMBP Tagged Chimera

Re-engineering the chimera not only improved the overall expression yield and solubility dramatically, but the HisMBP tagged chimera proved considerably more tractable throughout purification (Figure 9-1). The full chimera was purified by affinity chromatography utilizing the affinity of the His_6 tag for the nickel residues of the HisTrap

HP column (GE Healthcare Biosciences Pittsburg, PA). An adequate yield and purity was readily achieved and reproducible.

9.6. Buffer Selection

Selecting appropriate buffers for all aspects of the purification proved challenging. Phosphate buffered saline (PBS) was the initial buffer used for the GST tagged chimera in the inherited protocol (Appendix 10). However, concerns over the chelating of metal ions by PBS with specific regard to the maintenance of the zinc molecules within the two zinc fingers of the chimera, and also concerns over the incompatibility of PBS with the reducing agent tris(2-carboxyethyl)phosphine) (TCEP) led to a change in the buffering component.

Trials with the buffer HEPES at 20 mM were unsuccessful as the full chimera prematurely eluted from the HisTrap HP column. This premature elution could be related to the fact that as a Qiagen technical article reports "Buffers with secondary or tertiary amines may reduce nickel ions" (128). While Qiagen recommends that such buffers as Tris, HEPES, and MOPS can be used with nickel columns at concentrations below 100 mM, trials with a 20 mM concentration of HEPES were unsuccessful. Ultimately, a Tris buffer of 20 mM and pH 7.4 produced the adequate and reproducible yield required.

9.7. Reducing Agents

The reducing agent dithiothreitol (DTT) is known to remove metal ions, as are most chelating agents including: ethylenediaminetetraacetic acid (EDTA), and sodium citrate. As much as possible, all agents known to remove metal ions were avoided in order to preserve the integrity of the two zinc fingers of the Tat chimera. Many trials were conducted with the reducing agent TCEP because it is often suggested that it does not interfere with concentration measurements made at 280 nM, and because, while more expensive, it is known to be a strong and reasonably stable reducing agent. Interestingly however, Bigalke et al. 2011 observed the slow precipitation of Tat in the presence of TCEP. When TCEP 0.3 mM was replaced by 5 mM BME, the stability of the chimera was noticeably improved in the form of higher and more stable final concentrations of both the full and the cleaved chimera (52).

The importance of the electrochemical environment in the successful expression and purification should not be underestimated as, in general, with the purification of Tat most of the concern has been over keeping the chimera reduced (43). Observations suggest that there may be a point of diminishing returns at which high concentrations of reducing agent may denature the chimera perhaps by disturbing the zinc fingers, or by altering some other as yet unknown feature. The use of reducing and/or chelating agents such as EDTA and DTT may be the cause of at least some of the dimerization and aggregation reported in other work on Tat. Moreover, *in vivo* the actual conformation of proteins in general may be somewhat more dynamic than currently contemplated (129).

9.8.Salt

The FPLC purification of the Tat chimera was successful when using a 0.5 M sodium chloride concentration in both the binding and elution buffers, and since this was consistent with GE Healthcare Biosciences recommendations for the HisTrap HP column higher concentrations were not explored. It is worth noting, however, that Frankel, Bredt et al. 1988 found an even higher 0.7 M potassium chloride concentration was needed to maintain the solubility of Tat alone (86 amino acid construct), and so such higher concentrations during purification might be worth investigating (43). It was necessary to lower the salt concentration considerably later on while cleaving the tag from the chimera with TEV protease, and during binding assays. Thus, the improved solubility provided by increasing the salt concentration may prove unsustainable for downstream applications. It may be useful, however, to employ higher salt concentrations when attempting to use spin columns to concentrate the Tat chimera which has proved to be somewhat problematic.

9.9.Glycerol

The osmolyte glycerol was present in all FPLC buffers during purification in order to enhance the solubility of the chimera. A concentration of 5% glycerol was sufficient to maintain solubility of the chimera throughout the FPLC purification, including after removal of the solubility enhancing tag and did not appear to hinder TEV cleavage. The concentration of glycerol was increased to as much as 20% when storing the purified protein long term, as at this point glycerol also serves as a cryoprotectant. When

conducting binding assays the glycerol was removed from the solution by dialysis or by buffer exchange using Vivaspin® 20 20 ml 5,000 MWCO centrifugal concentrator columns (GE Healthcare Bio-Sciences AB Uppsala, Sweden) to prevent any potential interference by glycerol in the binding interaction.

9.10. Imidazole

A low concentration of 20 mM imidazole was present in both the lysis and the binding buffer in order to maintain consistency between the two buffers while loading the crude protein extract onto the column, and during the column wash. This low concentration of imidazole serves to prevent non-specific binding of proteins to the column. A wash buffer of 60 mM was employed in several trials to remove any low affinity binding proteins, but no further improvement in purification was observed with this increased concentration. Subsequent washes were performed with the 20 mM imidazole binding buffer. An elution gradient of up to 500 mM imidazole was used to elute the full His_6 tagged chimera from the HisTrap HP column. The full chimera typically eluted from the column when 66% of the 500 mM imidazole elution buffer, and 33% of the 20 mM imidazole binding buffer was reached. Alternatively, the full chimera can be successfully eluted with 100% 500 mM imidazole with acceptable purity.

9.11. Sodium Azide

The anti-microbial agent sodium azide was added to the binding and elution buffers at 0.02% to prevent degradation of the protein by contaminating microorganisms during the several days from cell lysis through purification, dialysis, and downstream assays when the protein could not be frozen.

9.12. Protocol

Following the protocol outlined in Figure 9-2, the syringe filtered cell lysate was applied to the HisTrap HP column using a 50 ml Superloop at a flow rate of 0.2-0.4 ml/min. or slower. The column was then washed with binding buffer for five column volumes (50 ml) at a flow rate of 0.4-0.6 ml/min. or until a stable baseline was achieved. After allowing ample time to wash at the stable baseline the full length chimera was eluted by setting a gradient exchanging the 20 mM binding buffer on pump A for the 500 mM elution buffer on pump B over 30 minutes at a flow rate of 2 ml/min (Figure 9-3).

9.13. Results

With this protocol 40 mg of the 60 kDa full chimera were purified from 17 grams of wet cell pellet harvested from one liter of Turbo media which was induced at OD600 = 0.7 with 1.0 mM IPTG and expressed overnight at 30°C. Figure 9-4 shows the uninduced, induced, soluble, and insoluble fractions of the expression prior to FPLC purification, as well as the purified fractions, and the loading and elution flow through waste solutions.

Lanes 10 through 14 are purified fractions A1 though A5 of the full chimera and correspond to the fractions indicated in the chromatogram in Figure 9-3. Lanes 11 and 12 of Figure 9-4 show the highly concentrated fractions of the 60.8 kDa full chimera. Additional bands between 100 and 150 kDa may be dimers of the full chimera and seem to be more prevalent in concentrated fractions.

Lanes 6, 7, and 8 (Figure 9-4) are samples of the loading flow through materials that did not bind to the column. The highly concentrated ~60 kDa band here appears to indicate that the column was overwhelmed by the highly concentrated recombinant chimera. Lane 9 is a sample of materials eluted from the column prior to the peak (X1) and indicates that a fair amount of protein was eluted prior to the observance of the peak. This is likely to be an artifact of the AKTA system and indicates that wherever possible fraction collection should begin slightly in advance of the appearance of the peak. The highly concentrated fractions A1 though A5 were pooled and then dialyzed into a low imidazole buffer for cleavage of the HisMBP tag, and its subsequent removal by reapplication to the HisTrap HP column.
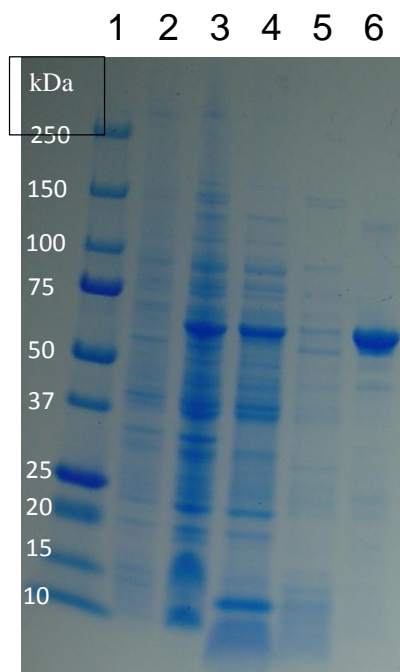
Figure 9-16 The HisMBP Tat chimera expressed in the Rosetta Gami B

The HisMBP Tat chimera expressed in the Rosetta Gami B strain in LB media and FPLC

purified. Induction at 0.7 OD with 0.8 mM IPTG. Incubation at 30°C post induction for 7

hours before harvest. From left to right:

 1 – ladder   2 – uninduced   3 – induced

 4 – soluble   5 – insoluble   6 – full chimera

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine

buffer. Run condition: 200V, 50 minutes.
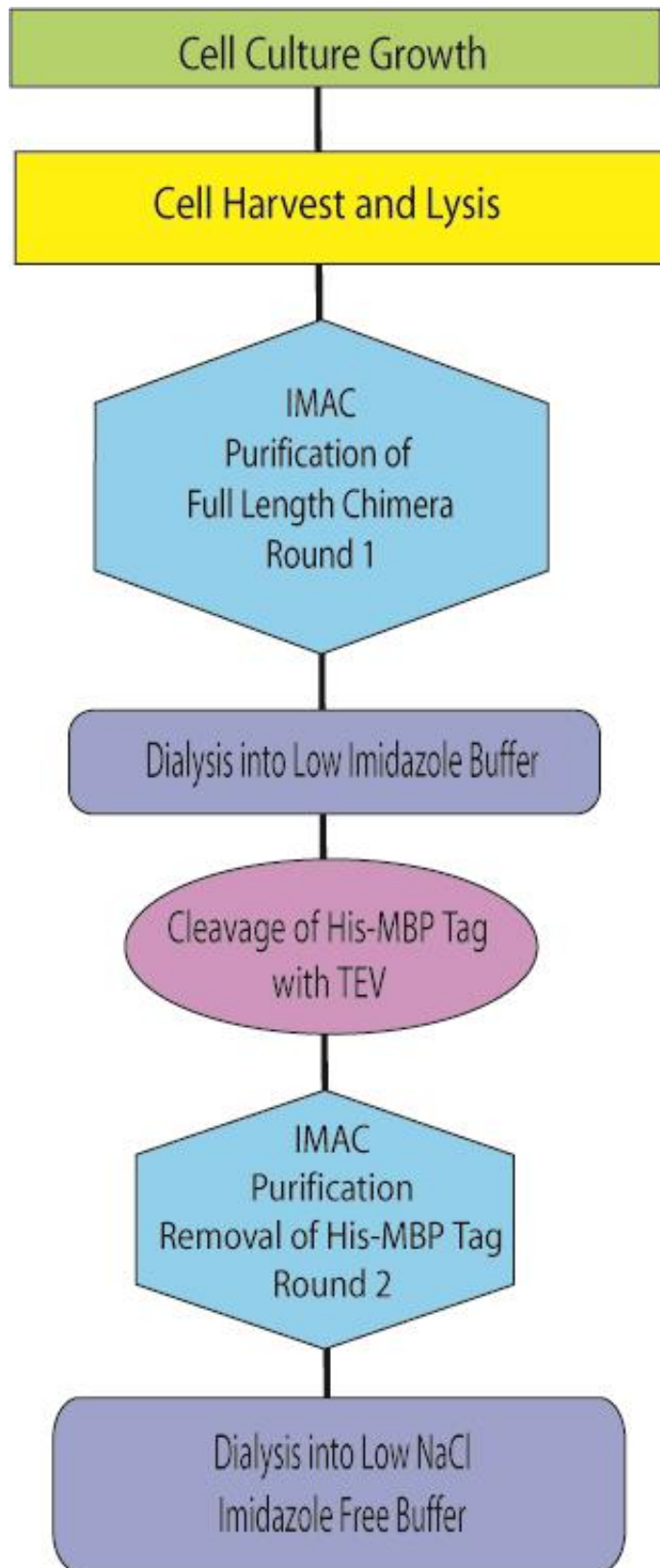
Figure 9-2 Recombinant protein purification flow chart.

Figure 9-3 FPLC Chromatograph of the elution of the full chimera.

The sharp peak starting in fraction A2 eluted approximately one column volume after reaching the full 500 mM imidazole concentration. The SDS Page analysis of fractions X1 through A5 can be seen in Figure 9-4. The UV remains elevated after the peak due to 280 nm absorbance by the 500 mM imidazole elution buffer.
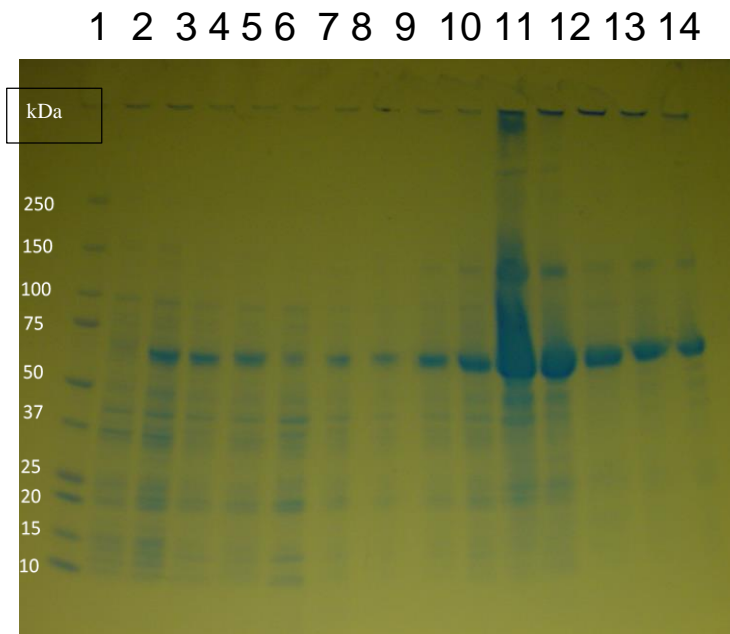
Figure 9-4 SDS PAGE of the Tat Chimera Expression and Purification

The SDS PAGE analysis of the overnight expression and purification of the full chimera in Turbo media. The expression and FPLC purification fractions of the full chimera (60.8 kDa), s from left to right:

1 – ladder

2 – uninduced

3 – induced

4 - soluble

5 – insoluble,

6 – loading flow thorough

7 – loading flow through

8 – loading flow through

9 – elution flow through X1

10 – full chimera A1

11 – full chimera A2

12 – full chimera A3

13 – full chimera A4

14 – full chimera A5

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine buffer. Run condition: 200V, 50 minutes.

Chapter 10 Removal of the Fusion Tag

While fusion tags are often necessary to facilitate expression of recombinant proteins at high yield, they also may confound downstream applications, particularly applications that explore binding interactions, biological activity, and structural detail. Removal of the solubility enhancing tag can introduce an entirely new set of challenges, not the least of which is maintaining the solubility of the recombinant protein after this important solubility enhancing component has been removed.

Choosing a protease enzyme that is highly specific and cleaves only at a rare series of amino acids is a helpful means of preventing non-specific cleavage of the target. Some proteases such as thrombin are notorious for cleaving non-specifically (74,130). Moreover, non-specific cleavage can be exacerbated by long incubation periods, high temperatures, excessive concentrations of proteolytic enzyme, and a variety of other factors specific to the chosen enzyme.

In some cases non-specific cleavage by thrombin can be mitigated by the prompt removal of the enzyme with affinity purification using benzamidine columns, and perhaps more so by removal with heparin sepharose columns (130). However, in the case of the GST-hCycT1-Tat chimera, non-specific cleavage was still observed even with prompt removal of thrombin by benzamidine, and in an unfortunate coincidence the Tat chimera also bound to heparin rendering this manner of protease removal useless (131,132).

When developing a new recombinant protein expression and purification protocol one should carefully consider the available protease enzymes and choose an enzyme that is specific, efficient, cost effective, and compatible with downstream applications whenever possible. The efficiency of the cleavage can be an extremely important, albeit somewhat less obvious, consideration in achieving a high yield of recombinant protein after purification.

10.1. Cleavage of HisMBP Tag with TEV

Having inherited the pGEX 2TK chimera plasmid with a thrombin protease site for the purpose of cleavage of the original GST tag, and after many disappointing attempts to isolate the tag-free chimera, the highly specific protease TEV was chosen to be the protease enzyme for cleavage of the HisMBP tag from the re-engineered chimera. The protease sequence ENLYFQ/G was introduced into the chimera sequence between the MBP tag and the hCycT1 portion of the chimera in order to facilitate tag removal (Figure 4-6). This peptide sequence is readily cleaved by the highly active and highly specific cysteine protease TEV from the tobacco etch virus.

Unfortunately, wild type TEV is auto-inactivated rapidly *in vivo* and catalytic activity is diminished markedly and in direct proportion to the concentration of the enzyme (74,133). Self-cleavage of the enzyme and the multiple fragments that result complicate purification, as well as the removal of the enzyme from the cleavage reaction. Further, the activity of the enzyme continues to diminish during storage (74). Autoproteolysis

takes place between Met218 and Ser219 of the TEV enzyme and may be an intramolecular event (74). Kapust et al. 2001 mutated the TEV enzyme with the substitution S219V, and found that self-cleavage was eliminated and the mutated enzyme was some 100 fold more stable (74).

10.2. Results: Cleavage of the MBP Tag with GST Tagged TEV

The initial attempts to remove the MBP tag from the fusion chimera employed a GST-His$_6$ tagged TEV protease (~ 50 kDa), which was generously donated by Michael Cosgrove, Upstate Medical University, NY. However, incubation of 1 mg of the enzyme per 20 mg of the fusion chimera produced precipitation with nutation during incubation, and appeared to produce incomplete cleavage after 24, and 48 hours at room temperature even without nutation (Figure 10-1). Predictably, at 14°C the reaction was even less efficient leaving considerably more of the full fusion chimera uncleaved after 24, 48, and even 72 hours (Figure 10-2).

In order to optimize the yield of CycT1m-Tat, a TEV protease that was more efficient, especially at lower temperatures, was required. Kapust et al. 1999 compared the solubilizing effects of three fusion tags: maltose-binding protein (MBP), glutathione S-transferase (GST), and thioredoxin (TRX), and found MBP to be, by far, the more effective solubilizing fusion partner (68). Further, Kapust et al. 1999 compared the solubility of GST tagged TEV to that of MBP tagged TEV, and found that less than 20% of the GST-TEV was soluble as opposed to greater than 60% of the MBP-TEV (68).

- 122 -

Having observed such poor solubility in the recombinant expression of TEV, Kapust et al. 2001 constructed a plasmid for the recombinant expression of MBP-His$_6$-TEV(S219V)-Arg in an effort to simultaneously improve both the stability of the enzyme and the solubility (74). Fortunately, the pRK793 plasmid was deposited with Addgene (Cambridge, MA) from which it was purchased for a nominal fee. In this interesting construct the MBP tag is self-cleaved from the recombinant enzyme after expression, and the protease can then be purified by affinity for either its remaining N-terminus His$_6$ or by the C-terminus Arg tag (74).

10.3. Expression of His$_6$-TEV(S219V)-Arg Protease in *E. coli*

Following the work of Tropea, Cherry et al. 2009 the pRK793 plasmid was expressed consistent with the published protocol with an exception (Appendix 13). The expression strain used by Tropea et al. 2009 is the BL21 (DE3) CodonPlus-RIL(134). Though it is not uncommon to recommend the use of a DE3 expression strain with a plasmid containing a tac promoter (135), the DE3 strain is designed for the T7 promoter system and expresses T7 RNA polymerase. When working with the tac promoter the native *E. coli* RNA polymerase is used, and so production of T7 RNA polymerase places an entirely unnecessary stress on the cell (Novagen personal conversation). Since so much conflicting information on this exists the two cell lines were compared in the recombinant expression of the TEV enzyme from the pRK793 plasmid with a tac promoter (Figure 10-3). Here the expression is compared in the Rosetta Gami B cell line for both the DE3

Rosetta Gami B cells and the non-DE3 Rosetta Gami B cells. Comparison of lanes four and five clearly indicates that the non-DE3 cell line produces a substantially higher concentration of the TEV protein in the crude induced cell extract, as well as a slight increase in basal expression levels.

When performing a large scale expression of TEV the Rosetta 2 expression strain (Novagen, Madison, WI) was used. This strain of *E. coli* is similar to the Rosetta Gami B strain but does not provide the folding accommodation for disulfide bond formation in the cytoplasm, which was not required here. Three liters of LB media containing 100 ug/ml ampicillin, 30 ug/ml chloramphenicol, and 0.2% glucose were inoculated with 25 ml of an overnight culture of Rosetta 2 (non-DE3) cells containing the pRK793 plasmid. The culture was grown at 37°C until OD_{600} ~0.5 when the temperature was reduced to 30°C and the culture was induced with 1 mM IPTG. The TEV protein was expressed for 6 hours post induction. On harvesting the 3 liter culture yielded 23.4 g of wet cell pellet weight which was slightly above the range of 30-40 g per 6 L reported by Tropea, Cherry et al. 2009 (134).

The MBP-His$_6$-TEV(S219V)-Arg$_5$ was purified by affinity chromatography with a HisTrap HP nickel column. The Tropea et al. 2009 protocol contains an additional gel filtration polishing step at this point. Since the appropriate column was not available the gel filtration step was omitted. From the 23.4 g of wet cell pellet weight the 41 mg of purified MBP-His$_6$-TEV(S219V)-Arg$_5$ recovered were of sufficient concentration and purity for downstream applications (Figure 10-4). The FPLC fractions were pooled, and

concentrated with Vivaspin® 20 20 ml MWCO 5,000 columns (GE Healthcare Bio-

Sciences AB Uppsala, Sweden).

10.4.Determination of Optimal Cleavage Conditions

In order to determine the optimal concentration of the His_6-TEV protease for cleavage of

the HisMBP Tat chimera six different concentrations of the protease were incubated

overnight at either room temperature or 4°C (Figure 10-5). Ratios of milligrams to

milligrams 100:1, 50:1, 25:1, 12.5:1, 6:1, and 1:1 are compared with full cleavage in all

room temperature samples, and nearly complete cleavage in only the 1:1 sample at 4°C.

In Figure 10-5 the lowest concentration ratio of the His_6-TEV(S219V)-Arg_5 to target to

achieve efficient cleavage of the HisMBP chimera appeared to be between 50:1 and 25:1

in the room temperature trials. Using the midpoint concentration of 37.5:1, cleavage

reactions were compared at room temperature, and at 15°C at intervals of one hour for

between 1 to 5 hours (Figure 10-6). Here the goal was to achieve the greatest amount of

cleavage in the shortest amount of time, with the least amount of enzyme, and at the

lowest incubation temperature possible. All cleavage reactions in this comparison

achieved efficient cleavage, and so the ratio of 37.5:1 mg/mg, the temperature of 15°C,

and the incubation time of at least one hour were set as the cleavage conditions for the

remaining experiments.

GST-His$_6$ TEV ~50 kDa ➡️

➡️ Full length fusion chimera 60.8 kDa

➡️ MBP 42 kDa

➡️ Cleaved chimera 17.5 kDa
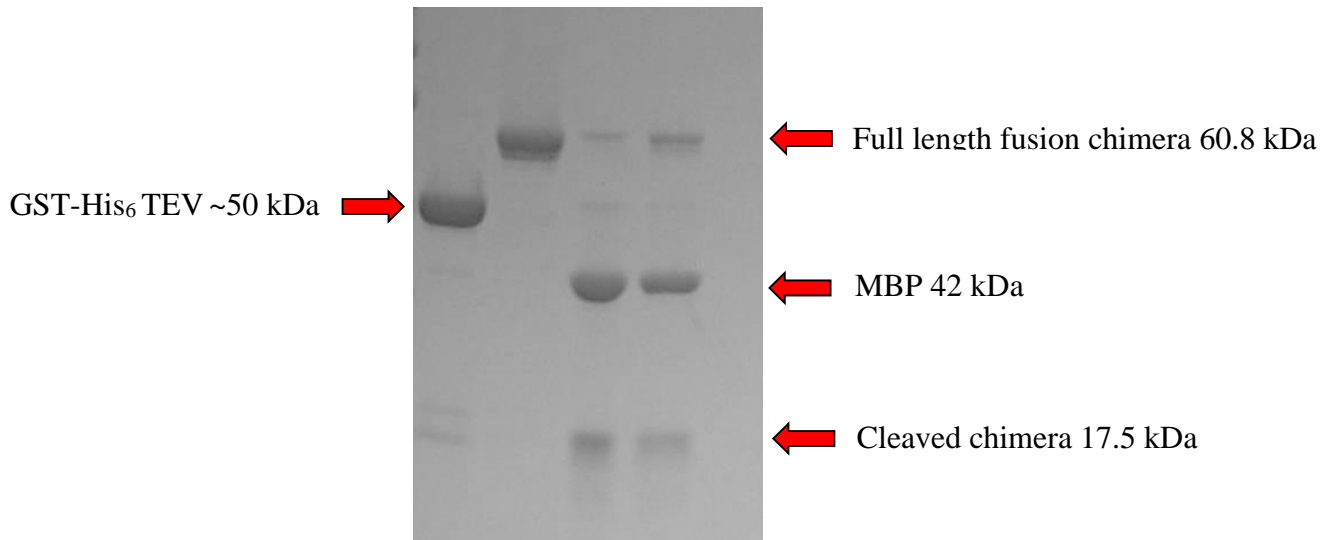
Figure 10-1 Cleavage of the fusion chimera with GST-His$_6$-TEV protease

The cleavage of the full fusion chimera with GST-His$_6$-TEV protease at a ratio of 1:20 milligrams of protease to substrate at room temperature (RT). Cleavage reactions from left to right:

1 – GST-His$_6$-TEV control   2 – Full fusion chimera control

3 – Cleavage reaction 48 hours   4 – Cleavage reaction 24 hours

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine buffer. Run condition: 200V, 50 minutes.

Full length fusion chimera 60.8 kDa

GST-His$_6$ TEV ~ 50 kDa

MBP 40 kDa

Cleaved chimera 17.5 kDa
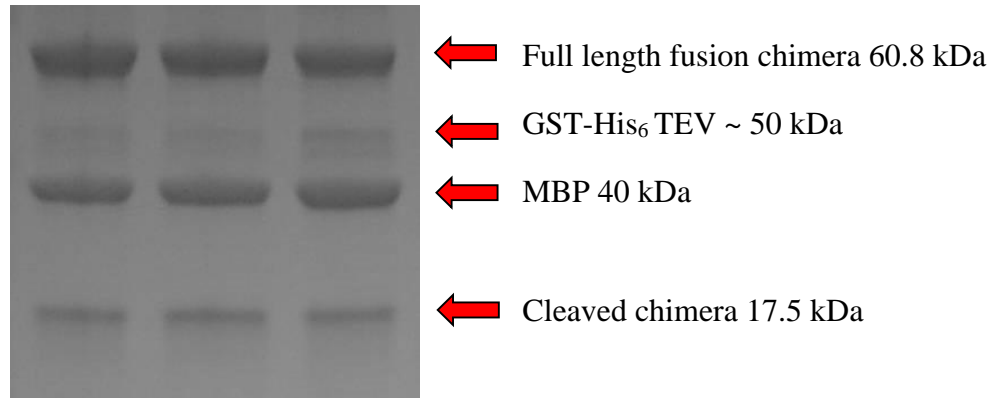
Figure 10-2 Cleavage of the full fusion chimera with GST-His$_6$-TEV

The cleavage of the full fusion chimera with GST-His$_6$-TEV protease at a ratio of 1:20

milligrams of protease to substrate with incubation at 14°C. Cleavage reactions from left

to right:

1 – cleavage reaction 28 hours   2 – cleavage reaction 48 hours

3 – cleavage reaction 72 hours

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine

buffer. Run condition: 200V, 50 minutes.
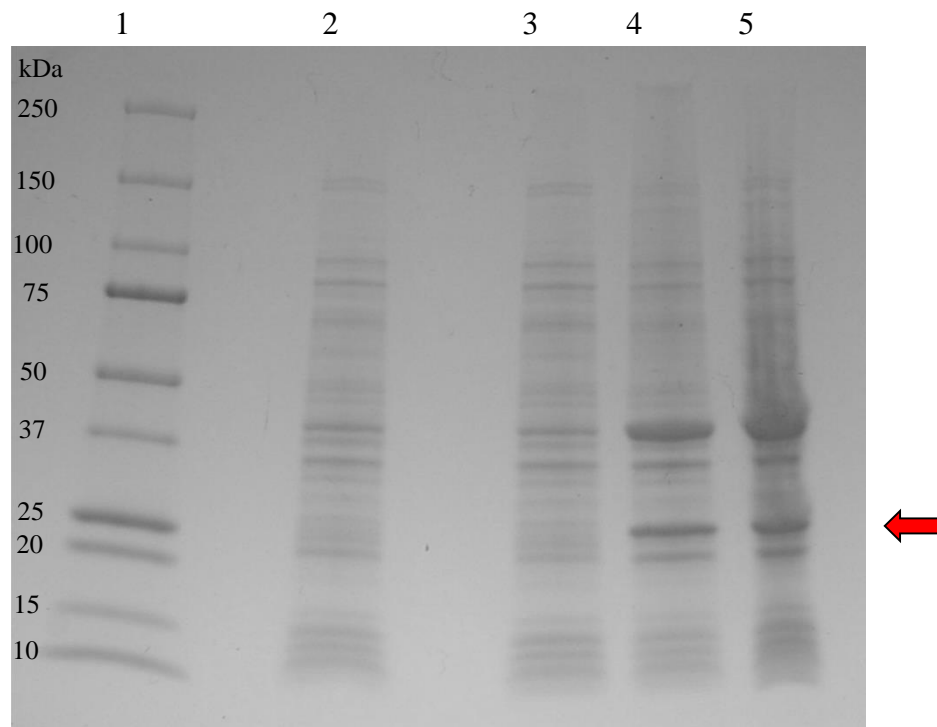
Figure 10-3 Comparison of expression conditions for the pRK793 plasmid

A comparison of the expression conditions for the pRK793 plasmid for the recombinant

expression of His$_6$-TEV(S219V)-Arg$_5$ in *E. coli* DE3 vs. non-DE3 cell lines. Expression

samples from left to right:

1 – ladder   2 – uninduced Rosetta Gami B DE3

3 – uninduced Rosetta Gami B non-DE3   4 – induced Rosetta Gami B DE3

5 – induced Rosetta Gami B non-DE3

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine

buffer. Run condition: 200V, 50 minutes.
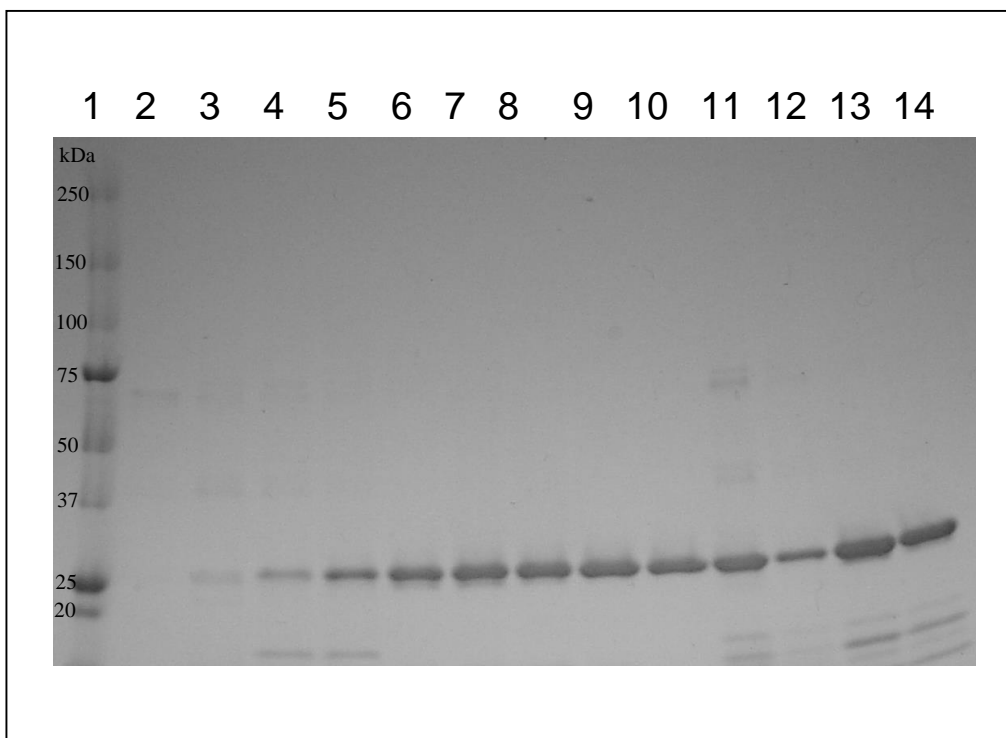
Figure 10-4 PAGE analysis of FPLC fractions for His_6-TEV(S219V)-Arg_5.

The FPLC purification fractions of His_6-TEV(S219V)-Arg_5, lanes from left to right:

1 – ladder   2 - empty   3 – A1   4 – A2

5 – A3   6 – A4   7 – A5   8 – A6

9 – A7   10 – A8   11 – A9   12 - concentrated TEV 1

13 – concentrated TEV 2   14 – concentrated TEV 3

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine

buffer. Run condition: 200V, 50 minutes.
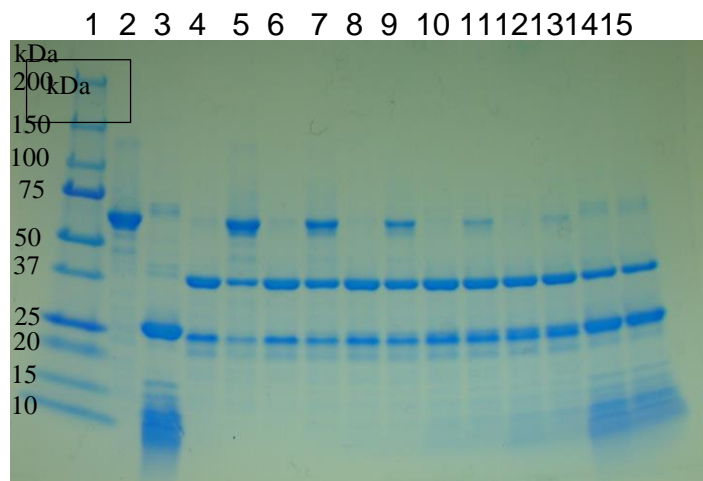
Figure 10-5 Conditions comparison for His_6-TEV(S219V)-Arg_5

Comparison of cleavage with increasing concentrations of the His_6-TEV(S219V)-Arg_5 protease at room temperature (RT) and 4°C incubated overnight with the HisMBP Tat chimera from left to right:

1 – ladder   2 – full MBP chimera   3 – His_6-TEV(S219V)-Arg_5 4 – 100:1 RT

5 – 100:1 4°C   6 – 50:1 RT   7 – 50:1 4°C   8 – 25:1 RT

9 – 25:1 4°C   10 – 12.5:1 RT   11 – 12.5:1 4°C   12 – 6:1 RT

13 – 6:1 4°C   14 – 1:1 RT   15 – 1:1 4°C

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine buffer. Run condition: 200V, 50 minutes.
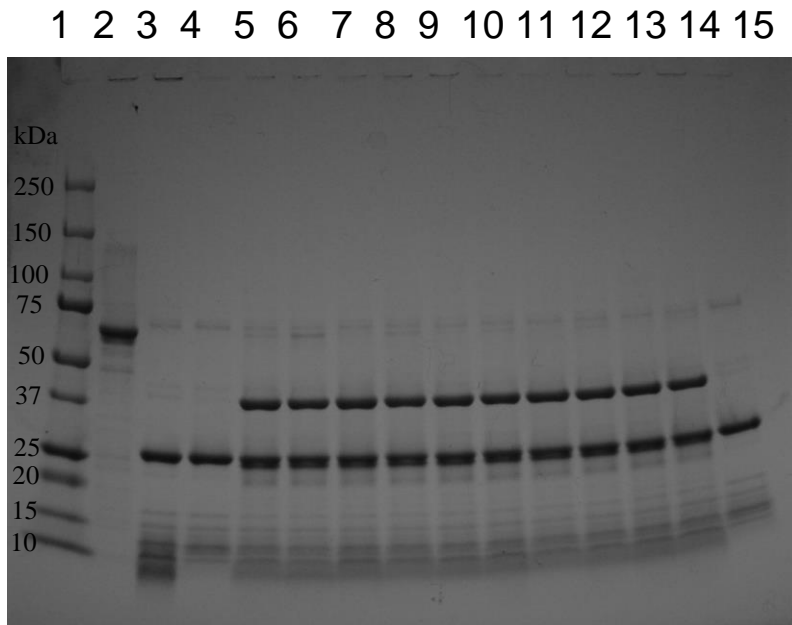
Figure 10-6 Comparison of cleavage ratio of 37.5:1 for His$_6$-TEV(S219V)-Arg$_5$

Cleavage reactions at a concentration ratio of 37.5:1 mg/mg for both room temperature

(RT) and 15°C at intervals of 1 hour from 1 to 5 hours from left to right:

1 – ladder   2 – full MBP tagged chimera   3 – His$_6$-TEV(S219V)-Arg$_5$

4 –FPLC purified His$_6$-TEV(S219V)-Arg$_5$ (control)

5 – cleavage reaction RT (1 hour)   6 – cleavage reaction 15°C (1 hour)

7 - cleavage reaction RT (2 hours)   8 – cleavage reaction 15°C (2 hours)

9 – cleavage reaction RT (3 hours)   10 – cleavage reaction 15°C (3 hours)

11 – cleavage reaction RT (4 hours)   12 – cleavage reaction 15°C (4 hours)

13 – cleavage reaction RT (5 hours)   14 – cleavage reaction 15°C (5 hours)

15 –FPLC purified His$_6$-TEV(S219V)-Arg$_5$ (control) with additional purification by

10,000 MWCO spin column.

Chapter 11 FPLC Purification of CycT1m-Tat

Once the HisMBP tag has been sufficiently cleaved from the full chimera the tag and the His$_6$-TEV protease can be removed from the solution by affinity chromatography with the HisTrap HP column. In contrast to the first round of purification, in this step the His$_6$ tagged proteins that bind to the column will be the unwanted cleaved tag and protease; the desired target (cleaved chimera, CycT1m-Tat) will be captured in the flow through. Since a reasonable degree of purity is achieved in the first purification step of the full chimera, and the concentration of His$_6$-TEV is low relative to the chimera, the bulk of material bound to the column will be the cleaved HisMBP tag.

Isolating the chimera in this manner can be challenging. Purifications with a Tris based buffer and using the reducing agent BME tended to produce a more focused UV peak in the chromatogram (Figure 11-1) with the concentrated cleaved protein visible by SDS PAGE (Figure 11-2). In earlier trials with a PBS based buffer and the reducing agent TCEP the peak was either absent or less focused (data not shown), yet the fairly pure and concentrated cleaved target was present and was observed by SDS PAGE. Since the cleaved target is present in solution and does not bind to the column theoretically a peak should not necessarily be anticipated as the target simply passes through and is not focused in any manner.

In general, when performing FPLC purification one should collect as many fractions throughout the process as possible regardless of the visible absorption recorded by the

chromatogram. Each fraction should be analyzed by SDS PAGE before assessing the purification or attempting to concentrate the target protein. Proper selection of the fractions will help to reduce dilution of the final sample, and prevent contamination by any undesirable co-purified proteins.

11.1. Results

Figure 11-1 is the FPLC chromatogram for the purification of the cleavage reaction solution with a HisTrap HP column. The peak at 15 ml was collected in fractions A2 and A3 and contained the concentrated CycT1m-Tat. An SDS PAGE analysis (Figure 11-2) of the cleavage reaction demonstrates nearly complete cleavage of the chimera in lane 2 of the gel. The A2 and A3 FPLC fractions corresponding to the UV peak in the chromatogram in Figure 11-1 appear to be highly concentrated and of greater than ~80% purity. In the final yield approximately 6.5 mg of CycT1m-Tat were recovered in 8 ml after purification from a 1 liter culture, 17 g of wet cell pellet weight, and 40 mg of the full chimera. The chromatogram peak at 35 ml corresponds to the elution of the HisMBP tag and the His_6-TEV protease from the HisTrap HP column in the full strength 500 mM imidazole elution buffer.

Notably, CycT1m-Tat migrates at nearly 25 kDa despite a true molecular weight of 17.5 kDa, and a doublet of CycT1m-Tat was often apparent. Proteins frequently do not migrate to their calculated positions due to preferential SDS loading at hydrophobic regions. It has also been suggested that the SDS loading capacity of a protein may be

related to the protein structure, secondary or tertiary, and here the protein may not have been completely denatured by the presence of SDS (136). Finally, observations of similarly anomalous migratory behavior are well documented in proteins with a high number of basic residues, such as are present in the Tat chimera (137,138).

The appearance of the doublet of CycT1m-Tat could indicate the presence of two species of the target in different oxidation states, as while it might be reasonable to suspect a truncated form of the target as the source of the doublet, a single molecular weight species at 17.5 kDa was later confirmed by matrix-assisted laser desorption-ionization/time-of-flight (MALDI-TOF) experiments on the same sample (data not shown).

All fractions from the final purification of CycT1m-Tat were analyzed by SDS-PAGE to ensure that some fraction of the target was not lost in the waste material nor was it binding to the column. Fractions from the cleaning of the column at the end of the procedure were also analyzed, and no proteins were observed in the gel for these washes (Figure 11-2). Finally, a control containing the His_6-TEV in a concentration equivalent to that of the cleavage solution was run on the gel (Figure 11-2 lane 13) to assess whether this low concentration of the enzyme would be visible, and the enzyme is not apparent in the gel at this concentration.

11.2. Buffer Exchange and Concentration

At the end of the purification procedure the buffer must be exchanged based on either downstream applications, or storage requirements. Removing the imidazole is necessary for accurate measurement of protein concentrations since imidazole absorbs at 280 nM. Downstream binding assays are confounded by the presence of high salt concentrations such as the 0.5 M NaCl in the FPLC binding buffer, and so the buffer was exchanged for a lower salt concentration solution free from the presence of imidazole. If storage is the next step the glycerol concentration will likely be increased, but for binding and other assays glycerol and perhaps BME may need to be removed.

Buffer exchange worked well with the use of a dialysis membrane and where little change the in concentration took place, but was considerably more complicated when both concentration and buffer exchange were attempted simultaneously. When the salt concentration was lowered and the glycerol was removed during concentration with Vivaspin® 5,000 MWCO columns (GE Healthcare Bio-Sciences AB Uppsala, Sweden) the protein appeared to aggregate. Though the precipitation was not visibly apparent, the concentration of the retentate decreased, when measuring absorbance at 280 nm by Nanodrop, rather than increased with each successive spin even when using a refrigerated centrifuge. Measurements of washes from pipetting up and down on the membrane indicated that at least some of the protein had precipitated onto the membrane. The addition of arginine may be helpful toward mitigating this aggregation, but could also interfere with downstream applications such as NMR. The empirical determination of

appropriate conditions for highly concentrated solutions of CycT1m-Tat could prove daunting, and here also high throughput methods would definitely be desirable.

11.3.Storage of CycT1m-Tat

When storing proteins a number of factors can affect the stability and biological activity of the protein and must be considered. Degradation of stored proteins can result from changes in temperature, protease activity, and the presence of some heavy metals. Another frequent concern is damage caused by oxidation. Aggregation and precipitation can result from changes in concentration, pH, ionic composition, and changes in temperature. Freeze thaw damage caused by extremes of temperature can also degrade proteins. During storage important cofactors of the protein can also be lost that can compromise the integrity of the protein such as the zinc atom from each of the zinc fingers. In general, the chimera was stored in 20% glycerol in 20 mM Tris, 0.2 M NaCl, 10 uM $ZnCl_2$, 5 mM BME, and pH 7.4 at -80°C in the interest of circumventing as much of the damage from storage as possible. Samples of the chimera stored in these conditions for several months and then run in SDS PAGE analysis appeared as a single band at the same molecular weight as when freshly purified.
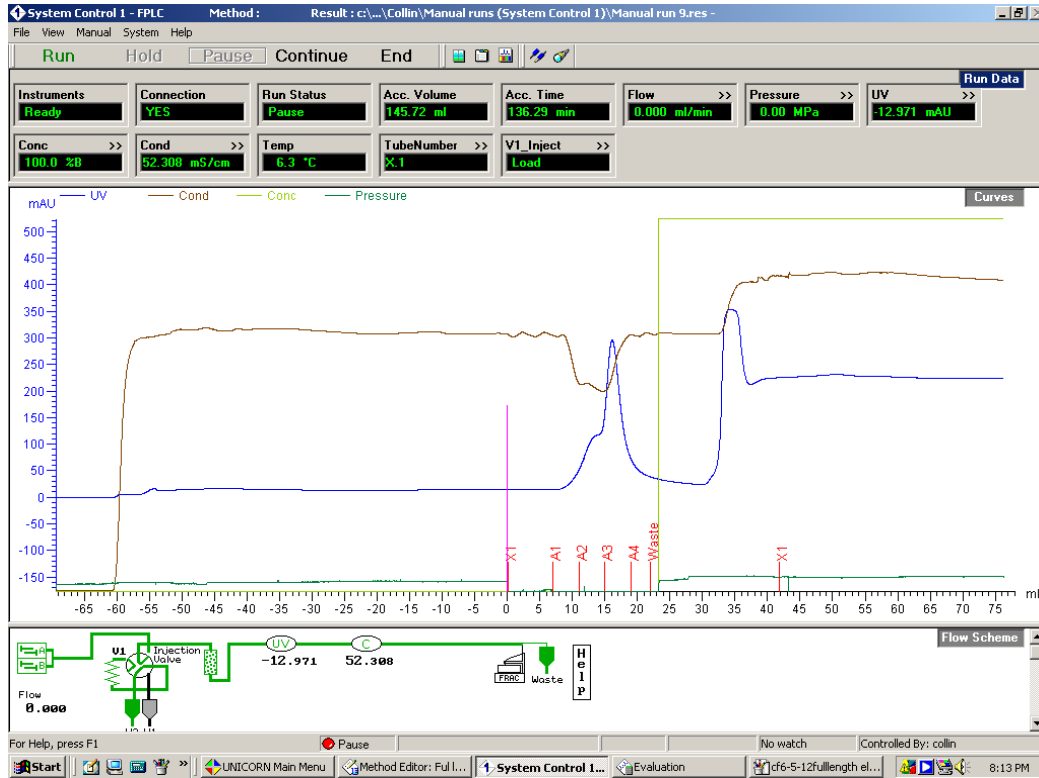
Figure 11-1 FPLC chromatogram of CycT1m-Tat

The FPLC chromatogram of the elution of CycT1m-Tat. The chimera with HisMBP tag removed flows through the HisTrap HP column without binding to it and is collected in the flow through solution. A peak at 15 ml was collected in fractions A2 and A3 and contained the concentrated (from the first round of purification) and purified CycT1m-Tat.
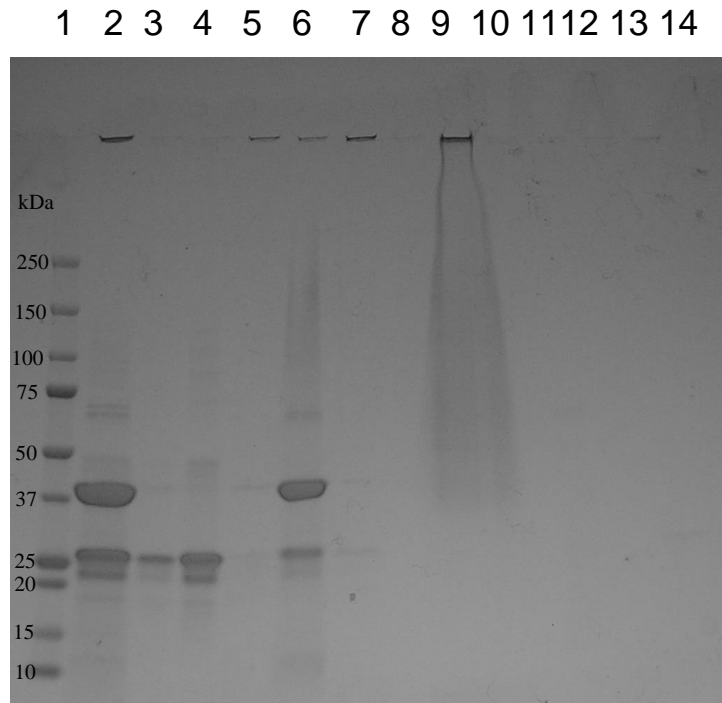
Figure 11-2 SDS PAGE of cleavage reaction and FPLC fractions and washes.

The SDS PAGE analysis of the cleavage reaction solution and the FPLC fractions and

wash solutions from left to right:

1- ladder   2 – cleavage reaction   3 – CycT1m-Tat A2   4 – CycT1m-Tat A3

5 – waste 1   6 – waste 2 (MBP, His_6-TEV)   7 – NaCl wash   8 – water wash

9 – NaOH wash   10 – isopropanol wash   11 – binding buffer wash

12 – second NaOH wash   13 – His_6-TEV 37.5:1 dilution control.

Note: His_6-TEV at 25 kDa in 6 runs at a molecular weight that is nearly indistinguishable

here from that of CycT1m-Tat. Distinguishing the protease from the chimera is addressed

by western blot in the following chapter.

Chapter 12 Non-Quantitative Binding Assays

Once the recombinant protein purification was complete two experimental techniques were used to confirm the correct Tat-chimera sequence. First the presence of the target protein was confirmed by the binding of primary and secondary antibodies in a Western blot, and second the correct molecular weight, corresponding to the appropriate target sequence, was confirmed by matrix-assisted laser desorption-ionization/time-of-flight (MALDI-TOF). With the use of both techniques it was possible to confirm the identity of the chimera with a reasonable degree of certainty before moving on to additional assays.

12.1. Western Blot

Western blotting was developed by W. Neal Burnette who submitted the idea in 1981 in an article to the journal of Analytical Chemistry (139). Ironically, the article was initially rejected. The original technique involves the electrophoretic transfer of proteins from SDS-PAGE (during which the proteins are initially separated vertically on the basis of molecular weight) to nitrocellulose filters from which a specific protein can then be detected by the binding of radioactive iodine-labeled antibodies or "probes" for the protein of interest. Today the technique is still widely used for the detection of proteins, and is available commercially with both radioactive and non-radioactive visualization alternatives including fluorescently and chromogenically labeled antibodies.

Since the chimera contains a myc antibody binding sequence between the hCycT1 and

Tat portions of the protein (Figures 3-1 and 4-11) the presence of the Tat chimera was

confirmed by first binding a primary anti-myc antibody to the sequence, and

subsequently binding a secondary chromogenic antibody to the primary antibody for

visualization by Western blot. The Novex® WesternBreeze™ Chromogenic Anti-mouse

Kit (Life Technologies Grand Island, NY) detects picogram levels of protein, and was

used with the iBlot® western blotting system (Life Technologies Grand Island, NY).

12.2. Western Blot Results

In an unfortunate coincidence the chimera and the His_6-TEV protease were found to run

at nearly the same molecular weight (~25 kDa) in SDS-PAGE. However, while the His_6-

TEV runs true to size the actual molecular weight of the chimera is ~17.5 kDa. Thus an

alternative method of distinguishing between the two proteins is essential. A Western blot

of the Positope™ positive control, full purified chimera, the cleavage reaction, and the

His_6-TEV protease confirmed that the band appearing in (Figure 12-2 SDS PAGE) 5 is

His_6-TEV which is not myc tagged and is therefore not present in the Western blot

(Figure 12-1 Western 5). Comparing the two bands in s four and five of Figure 12-2 SDS

PAGE it is clear that the two proteins migrate with remarkable similarity in SDS PAGE.

A second Western blot of several samples of the full chimera and CycT1m-Tat in a

variety of buffers (Tris based storage buffer, fluorescence titration buffer, and Octet

buffer) confirms the presence and integrity of the full chimera and CycT1m-Tat samples that were used in downstream assays (Figure 12-3 Western) (Figure 12-4 SDS-PAGE).

12.3.MALDI-TOF

The basic principle of mass spectrometry (MS) is that a moving charged particle is accelerated in a magnetic field which causes the deflection of the particle. The degree of deflection of the charged particles is dependent upon the mass/charge ratio. A detector receives the deflected particle, the signal is amplified, and the mass to charge ratio is recorded and plotted on the x axis, while the intensity of the signal is plotted on the y axis in what is called a mass spectrum.

Developed in 1988 by Franz Hillenkamp at University of Műnster in Germany, matrix-assisted laser desorption-ionization/time-of-flight (MALDI-TOF) is a highly sensitive soft ionization technique that is able to accurately analyze intact biomolecules (Figure 12-5). In MALDI-TOF the sample is mixed with excess matrix and is then dried to the MALDI plate. A laser is directed at the sample on the plate surface and ionizes the sample by proton transfer to the sample from the matrix which absorbs the laser light. The time that the ionized particle takes to reach the detector is known as "ion drift" and is inversely proportional to the square root of the mass to charge ratio of the particle. When the particle reaches the detector the signal is amplified and then recorded (140).

This technique is particularly useful in the analysis of biomolecules because MALDI-TOF is a soft ionization technique that produces less fragmentation and is therefore useful in the study of larger intact biomolecules. Other soft ionization techniques include: chemical ionization, fast atom bombardment, and liquid secondary ionization, Over the past 25 years use of MALDI-TOF has grown with the field of proteomics as the technique is highly sensitive and accurate and provides reasonable resolution for biomolecules up to several hundred kilodaltons (141).

12.4.Results: MALDI-TOF of CycT1m-Tat

Subsequent to confirming the presence of the chimera with Western blot, the appropriate molecular weight of the protein was confirmed by matrix-assisted laser desorption-ionization/time-of-flight (MALDI-TOF). The Bruker AutoFlex III MALDI TOF/TOF Mass spectrometer in the Jahn Lab at the State University of New York College of Environmental Science and Forestry (SUNY-ESF) was used to confirm the molecular weight of a 20 uM sample of CycT1m-Tat in a 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) matrix.

The base peaks of the two spectra produced from two samples of CycT1m-Tat (Figure 12-4 MALDI-TOF) indicate molecular weights of 17,485.1 and 17,485.9 daltons respectively and are within 4.4 daltons of the 17481.5 daltons predicted by the Protein Calculator version 3.3 provided online by The Scripps Research Institute (Appendix 3). This relatively small experimental error may be attributable to: protonation, differences in

isotopic abundance calculations of the calculator, and/or to calibration error for the

Bruker AutoFlex III.

12.5. Electrophoretic Mobility Shift Assay

An electrophoretic mobility-shift assay (EMSA), formerly known as a gel-retardation

assay and also known as a gel shift or band shift assay is sensitive and inexpensive

technique that was originally described by both Fried and Crothers, and by Garner and

Revzin, in 1981 (142,143). This non-denaturing electrophoresis technique permits both

the qualitative and quantitative characterization of protein-nucleic acid complexes (144).

In EMSA molecules are separated at near neutral pH on the basis of shape, size, and

charge since SDS is omitted and no negative charge is artificially imparted as it would be

in denaturing electrophoresis (145).

12.6. Characterization of the Protein-Nucleic Acid Complex

In EMSA the formation of complexes between nucleic acid and protein is evidenced by a

reduction in the distance traveled by the higher molecular weight bound complex through

the polyacrylamide gel when compared to the distance traveled by the free nucleic acid.

This so called "retardation" of the distance traveled by the nucleic acid is referred to as

"shifting" or "super-shifting" and is rendered visible by staining. The band produced by

the complex appears vertically higher or closer to the well than that of the lighter

unbound nucleic acid.

The binding affinity of the protein-nucleic acid complex can be characterized, to a limited degree, by the length of time that the complex travels through the gel before separating into its primary components. In some cases the relative intensity of the bands produced by the complexed nucleic acid versus those produced by the free nucleic acid can be measured and used for a more quantitative analysis. However, both the formation and stability of the complex are affected by numerous factors including: binding buffer, and running buffer components, gel concentration, temperature, and competitor molecules. For any particular protein-nucleic acid interaction EMSA conditions must be determined empirically (146).

The EMSA technique is useful for nucleic acids ranging in size from short oligonucleotides to longer and more complicated nucleic acid structures including small circular DNA, but is limited to approximately 5,000 base pairs or less (144). Assays with shorter oligonucleotide segments tend to be hindered by difficulties related to binding sites positioned close to the end of the molecules, while longer nucleic acids tend to exhibit more non-specific binding (144). With respect to protein, the EMSA technique is effective for proteins ranging from small oligopeptides to molecular weights greater than 1,000 kDa and is useful for both crude and purified solutions (144).

12.7. Visualizing Proteins and Nucleic Acids in the EMSA

In a traditional EMSA the protein-nucleic acid interaction is visualized by ^{32}P labeling of the nucleic acid, and provides a high degree of sensitivity detecting 0.1 nM or less of the nucleic acid (144). However, new and highly sensitive staining techniques have been developed that permit the circumvention of radioactive techniques. Chemiluminescent, fluorescent, and immunohistochemical detection methods are commercially available, reasonably sensitive, and stain either nucleic acids or proteins, or both.

For use in protein staining SYPRO Ruby Protein Gel Stain (Molecular Probes, Inc. Eugene, OR) is a highly sensitive luminescent stain with two excitation maxima at approximately 280 nM, and 450 nM, and with an emission maximum of approximately 610 nM (147). This stain can be visualized at 300 nM with a UV transilluminator. The sensitivity of this stain is comparable to that of silver staining with a lower detection limit between 0.25 ng and 1 ng of protein. This highly sensitive protein stain does not stain nucleic acids, and binds to basic amino acids and the peptide backbone of the protein (147).

The nucleic acid stain SYBR Gold (Molecular Probes, Inc. Eugene, OR) has excitation maxima for the dye-nucleic acid complexes at approximately 495 nM, and 300 nM and an emission maximum at approximately 537 nM. This proprietary unsymmetrical cyanine dye exhibits a greater than 1000-fold fluorescence enhancement when bound to nucleic acids and has a high quantum yield (~0.6) upon binding to double- or single-stranded nucleic acid (148). While this stain is highly sensitive and useful for EMSA, it is important to note that in general unsymmetrical cyan dyes bind the minor groove of

nucleic acids (149) and could potentially interfere with the binding interaction of a protein ligand.

12.8. Limitations of Electrophoretic Mobility Shift Assays

In addition to the complications already presented regarding the use of dyes that have the potential to alter or inhibit binding interactions there are other limitations to EMSA. Specifically, little can be gained from the assay regarding the molecular weight of the complex, or of the location of the binding interaction. Moreover, during electrophoresis samples are not at equilibrium, and both the ionic strength of the buffer and the so called "caging effects" of the gel may cause certain interactions to be stabilized, while the rapid dissociation of other interactions may prevent their detection entirely (144).

Glycerol, often added to the sample buffer for the purposes of increasing the density of the sample for loading, often has a stabilizing effect both on the unbound protein, and on the protein-nucleic acid complex (144,150,151) and can confound attempts at more quantitative analysis. Because of these issues, the aforementioned complications experienced with the positioning of small oligonucleotide binding sites, and the potential effects of the temperature of the gel on the affinity of the complex EMSA may be more useful in a semi-quantitative role. Here EMSA was used in the preliminary confirmation of binding interactions between recombinant purified proteins and their native nucleic acid binding partners prior to moving on to the more quantitative methods of surface plasmon resonance, and biolayer interferometry.

An additional complication in this non-denaturing assay is the issue of protein charge. While nucleic acid is always negatively charged and migrates from the negative to the positive electrode, proteins can be positively charged, negatively charged, or can have no net charge at their respective isoelectric points complicating migration in non-denaturing gel electrophoresis. In the case of positively charged proteins in non-denaturing gels the positively charged proteins will migrate in the opposite direction from the nucleic acid toward the negative electrode and in some cases can migrate up out of the wells of the gel and be dispersed into the running buffer.

12.9. Method

In order to qualitatively characterize the interaction between the recombinant Tat chimera and the TAR RNA an EMSA was performed in which both the full Tat chimera, and CycT1m-Tat (from which the tag had been removed) were bound to a 27 nucleotide stem-loop portion of the wild type TAR RNA (Figure 13-1). Two additional proteins were included in the assay: the MBP tag alone in order to determine whether the tag itself bound the RNA, and a 13 amino acid arginine rich peptide which is the minimal portion of the Tat protein found by Weeks et al. 1990 to bind the TAR RNA, and which also contains the nuclear localization signal (NLS) of the Tat protein.

The concentration of the polyacrylamide gel used for EMSA was determined empirically to suit the 27-nt nucleic acid, and the various proteins ranging in molecular weight from

1.7 kDa for the Tat minimal peptide to 60.8 kDa for the full tagged Tat chimera. A 12% acrylamide non-denaturing gel was pre-run for one hour at ~8°C and 100 Volts in 1 X TBE buffer at pH 8.0. Nucleic acid was heated for three minutes at 90°C vortexed lightly and briefly and then cooled on ice briefly before being added to samples. Protein samples in 20 mM Tris, 0.2 M NaCl, 10 uM $ZnCl_2$, 5 mM BME, and pH 7.4 were incubated with the wild type TAR RNA stem loop to a final concentration of 500 nM of the RNA, and 5 and 10 uM of each of three experimental proteins assayed. Glycerol was added to each of the samples to a final concentration of 5% in order to facilitate loading on the gel. Once the samples were mixed they were equilibrated at 4°C for 15 min. Positive and negative control samples of known binding partners HIV-1 SL3 RNA 33-nt with and without HIV-1 Ncp7 (7 kDa), respectively, were loaded in the first and second s of each gel. Negative control samples of the wild type TAR RNA alone were loaded in the third, and negative control samples of each of the proteins in the absence of RNA were loaded in the final three s (Appendix 16).

The EMSA gel was run for 90 minutes at ~8 °C and 100 Volts, removed from the apparatus, incubated in Millipore Milli-Q Biocel ultrapure water (Millipore, Billerica, MA) for 15 minutes, and then stained with SYBR® Gold nucleic acid stain (Molecular Probes, Inc. Eugene, OR) for one hour. After rinsing the gel with water the RNA was visualized and photographed with the Kodak Gel Imager equipped with UV transilluminator. The gel was then incubated in 100 ml of fixing solution comprised of 50% methanol and 7% acetic acid for 30 minutes two times. The gel was then incubated in 60 ml of SYPRO® Ruby gel stain solution (Molecular Probes, Inc. Eugene, OR) overnight, and then

visualized and photographed again with the Kodak Gel Imager. The two images were

then superimposed over one another in order to facilitate the interpretation of the pattern

of protein and nucleic acid migration.

12.10. Results: Electrophoretic Mobility Shift Assay

In Figure 13-2 (EMSA) the negative control in one demonstrates the migration of the

free HIV-1 SL3 RNA at 500 nM, and with the addition of 10 uM of Ncp7 the positive

control in two demonstrates the super-shifting of the SL3 RNA as predicted for this well

characterized interaction. The negative control free wild type TAR RNA in three

migrates slightly less than the SL3 as predicted based on the two nucleotide difference

between the two oligonucleotides. In the experimental four the addition of Tat minimal

peptide at 10 uM binds to TAR 500 nM and produces a small shift consistent with

expectations for the low molecular weight (1.7 kDa) peptide. In experimental five a

super-shift appears where the 10 uM recombinant CycT1m-Tat (17.5 kDa) binds the 500

nM TAR RNA. In this interaction the RNA appears to have been completely prevented

from migration into the gel. In six a super-shift appears where 10 uM recombinant full

MBP tagged chimera (60.8 kDa) binds the 500 nM TAR RNA and is indistinguishable

from the shift produced by CycT1m-Tat in the prior. In seven no shift is observed where

10 uM of the MBP tag has been incubated with the 500 nM TAR RNA indicating that the

MBP tag alone does not bind TAR. In s eight, nine, ten, and eleven the previous four s

are repeated albeit with 5 uM of each of the proteins and the same 500 nM TAR RNA

with similar results except for the binding of the minimal peptide where slightly less

shifting is observed. In the same figure lanes twelve, thirteen, fourteen, and fifteen

contain the proteins alone at 10 uM and demonstrate that none of the recombinant

proteins contains any observable nucleic acid contamination.

12.11.Discussion

The results of the EMSA were as predicted for both positive and negative controls, and

for the experimental samples. Shifting of the TAR RNA was minor for the low molecular

weight Tat minimal peptide, and was absent for MBP which is not known to bind TAR.

Super-shifting of the TAR RNA for both the full Tat chimera, and CycT1m-Tat is readily

apparent. The binding affinity of the two samples was almost indistinguishable from the

EMSA experiment alone. However, CycT1m-Tat did appear to bind with slightly higher

affinity than its tagged counterpart.

In Figure 13-2 s 5 and 6, and 9 and 10 for CycT1m-Tat and the full chimera, at 10 uM

and at 5 uM respectively, the nucleic acid is not visible. The large, high molecular

weight, and high affinity complexes formed in these s appear to have super-shifted the

nucleic acid preventing its migration into the gel. In general, conditions known to prevent

visualization of nucleic acid in EMSA are where the nucleic acid has been degraded,

where protein-nucleic acid complexes are too large for the gel system, where protein

aggregation occurs, or where the ratio of nucleic acid to protein is higher than the gel

system can accommodate (144). The appearance of the nucleic acid in both the positive

and negative controls indicates that degradation was unlikely. Lower concentration of

polyacrylamide in other trials failed to improve resolution (data not shown). Reducing the ratio of nucleic acid to protein, and the addition of non-ionic detergents or higher concentrations of glycerol might improve the visualization of the nucleic acid (144). However, since MBP (43.4 kDa) at identical concentrations did not prevent the migration of the TAR RNA it could reasonably be concluded that neither the ratio of nucleic acid to protein nor the concentration of the gel were particularly unsuitable. While aggregation of the recombinant full chimera and CycT1m-Tat could not be ruled out, the EMSA did prove to be a successful screening tool for the Tat-TAR interaction as specific binding was readily apparent (Figure 13-2).

With regard to the free protein samples, at the pH 8.0 of the 1 X TBE buffer the charges of the Tat minimal peptide, CycT1m-Tat, full MBP tagged chimera, and the MBP tag alone are +7.5, -0.5, -5.9 and -5.9, and their respective isoelectric points are 12.7, 7.9, 6.6, and 5.71 (Appendices 2,3,4, and 5). These characteristics along with the fact that in this non-denaturing system protein migrates based on a combination of size, shape, and charge and are influenced by a multitude of additional factors including pH, and ionic conditions could be responsible for the modest migration of the free protein through the gel.

Although not quantitative, EMSA did prove to be a suitable initial method for screening binding interactions between the recombinant chimeric proteins and the TAR stem-loop nucleic acid.
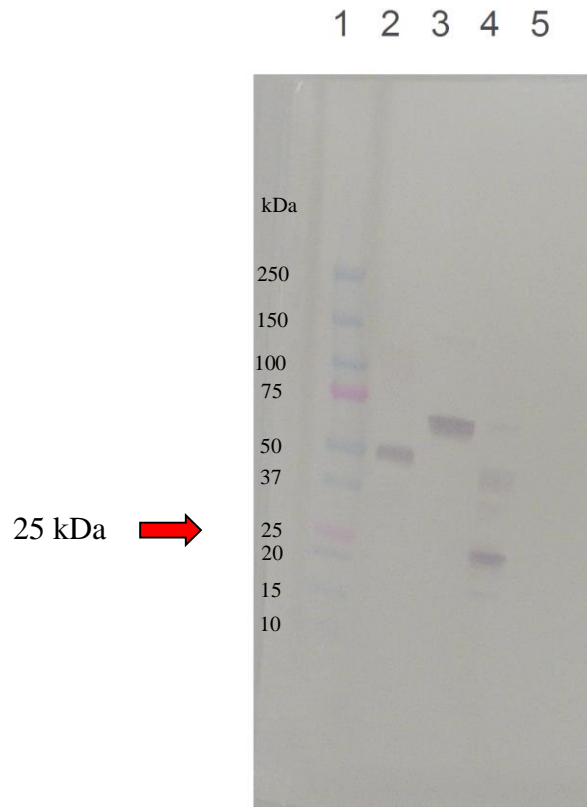
Figure 12-1 Western blot of the Tat chimera and His₆-TEV protease.

A Western blot of the Tat chimera and His₆-TEV protease from left to right:

1 – ladder   2 – Positope™ control

3 – full chimera control   4 – cleavage reaction

5 – His₆-TEV control

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine buffer.

Run condition: 200V, 50 minutes.
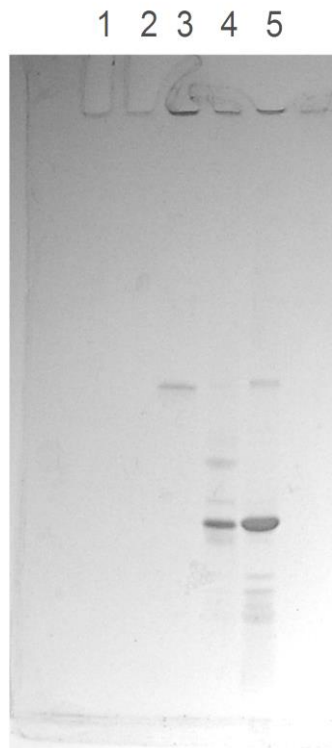
Figure 12-2 SDS-PAGE of the Tat chimera and His$_6$-TEV protease.

The SDS-PAGE analysis of the Tat chimera and His$_6$-TEV protease samples from left to right:

1 – ladder   2 – Positope™ control,

3 – full chimera control   4 – cleavage reaction,

5 – His$_6$-TEV protease control.

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine buffer. Run condition: 200V, 50 minutes. The same gel is used in Figure 12-1 Western blot.
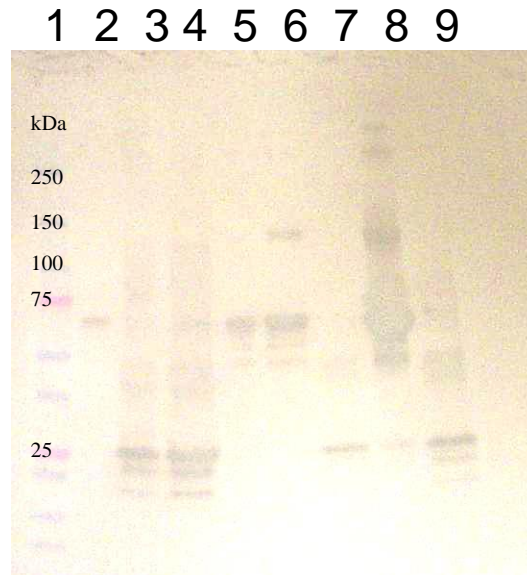
Figure 12-3 Western blot of full chimera and CycT1m-Tat.

A Western blot of the full chimera and CycT1m-Tat samples in various concentrations and buffers from left to right:

1 – ladder

2 – Positope™

3 – CycT1m-Tat 0.5 mg/ml

4 – CycT1m-Tat 1.0 mg/ml

5 – full chimera

6 – full chimera concentrated

7 – CycT1m-Tat (fluor buffer)

8 – full chimera (Octet buffer)

9 – CycT1m-Tat (Octet buffer)

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine buffer. Run condition: 200V, 50 minutes.
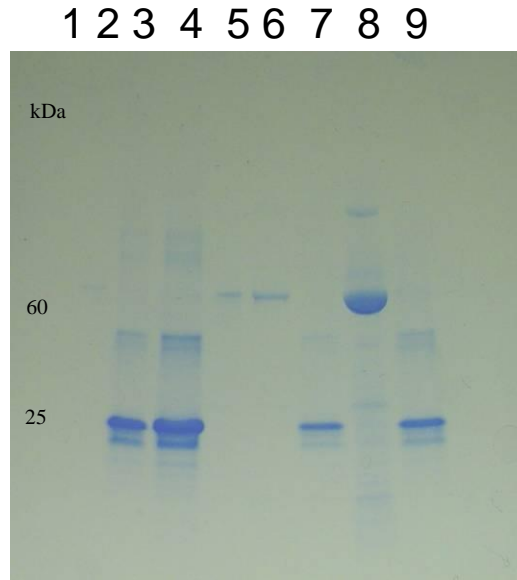
Figure 12-4 SDS-PAGE of full and CycT1m-Tat.

SDS-PAGE analysis of the full chimera and CycT1m-Tat samples from left to right:

1 – ladder   2 – Positope™   3 – CycT1m-Tat 0.5 mg/ml

4 – CycT1m-Tat 1.0 mg/ml   5 – full chimera,

6 – full chimera (fluor buffer)   7 – CycT1m-Tat (fluor buffer)

8 – full chimera concentrated (Octet buffer)   9 – CycT1m-Tat (Octet buffer)

Precast Gradient Gel 4-20% (Bio-Rad, Hercules, CA), running buffer: 1X Tris-Glycine

buffer. Run condition: 200V, 50 minutes. The same gel is used in Figure 12-3.
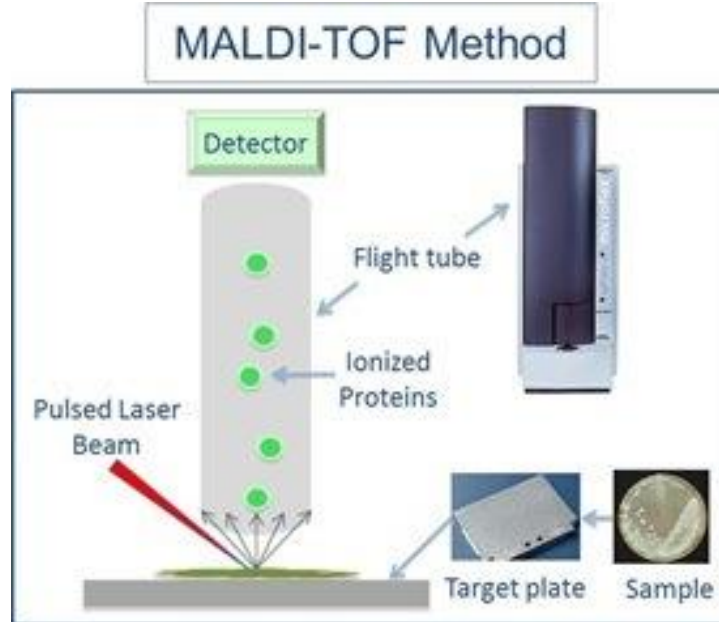
Figure 12-5 Diagram of the MALDI-TOF Method of Mass Spectrometry.

In MALDI-TOF a pulsed laser beam ionizes proteins affixed to the surface of the target

plate (bottom right). The mass to charge ratio of the ionized proteins causes the particle to

be deflected. A detector receives the deflected particle, the signal is amplified, and the

mass to charge ratio is recorded and plotted on the x axis, while the intensity of the signal

is plotted on the y axis in what is called a mass spectrum. Reprinted from: "The Next

Generation Technology for Microbial Identification: MALDI-TOF", Copyright 2011

Accugenix, Inc. with permission from Accugenix, Inc. Accessed online at:

http://www.accugenix.com/microbial-identification-bacteria-fungus-knowledge-

center/micro-id-basics/maldi-tof-method/

Comment 1
Comment 2
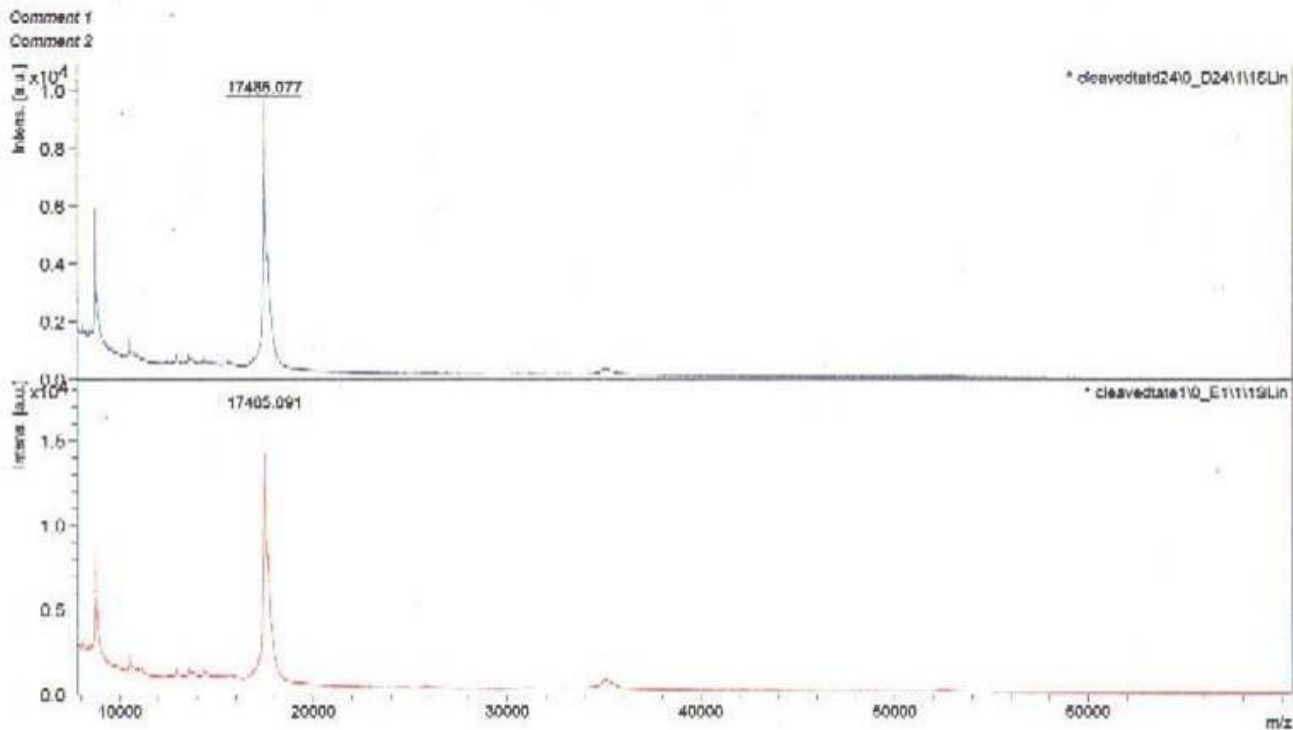
Figure 12-6 Bruker Daltonics flexAnalysis of CycT1m-Tat

The Bruker AutoFlex III MALDI TOF/TOF Mass spectrometer in the Jahn Lab at the

State University of New York College of Environmental Science and Forestry (SUNY-

ESF) was set for the range of 8000 to 50000 Daltons. The molecular weight determined

by two trials is in close agreement with the theoretically calculated molecular weight of

17,481.5 Daltons.
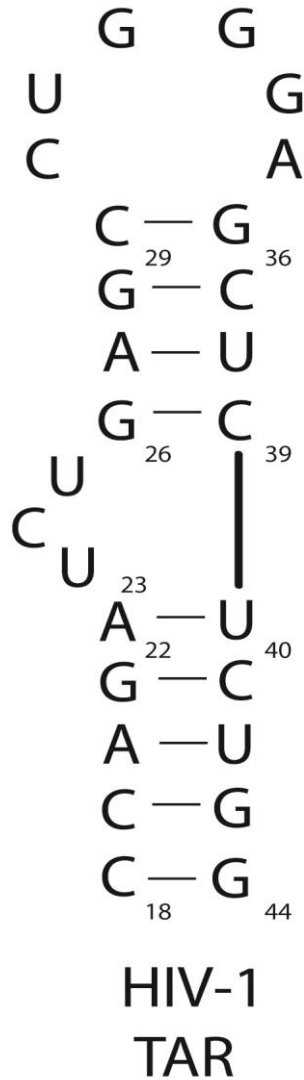
Figure 12-7 Wild type HIV-1 TAR (27-nt) stem loop structure for EMSA

The wild type HIV-1 TAR (27-nt) stem-loop nucleotide sequence and structure used in the electrophoretic mobility shift assay (EMSA) for the characterization of the binding complexes formed with the full MBP tagged Tat chimera, cleaved tag free Tat chimera, and the Tat minimal peptide.
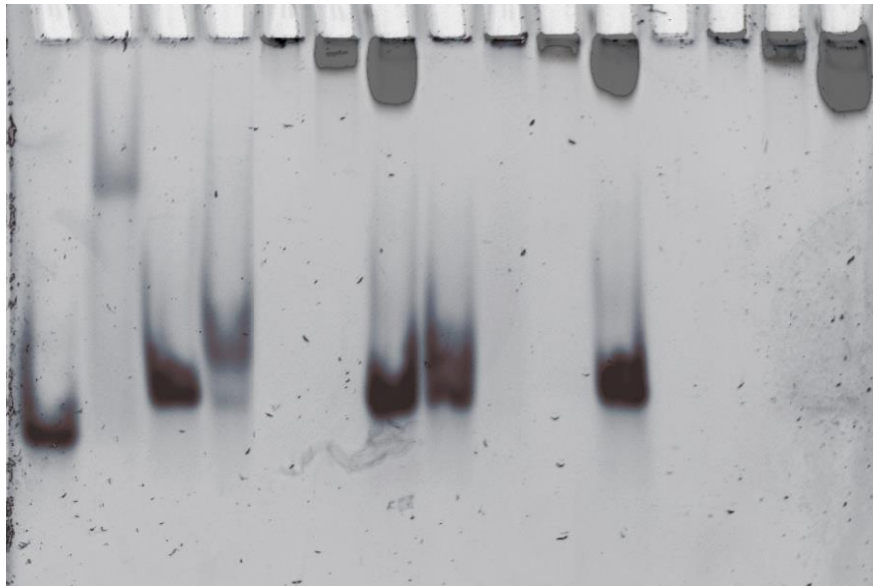
Figure 12-8 Electrophoretic mobility shift assay (EMSA)

EMSA of wild type HIV-1 TAR (27-nt) RNA complexed with full-length MBP tagged Tat chimera, CycT1m-Tat, and the Tat minimal peptide (see also, Appendix 16). An image of gel stained with SYBR Gold (RNA) was superimposed with an image from the same gel stained with SYPRO Ruby (protein). [SL3] = [TAR] = 500 nM when present. Lanes are color-coded for the most important analytes, numbered left to right:

1 – SL3 RNA

2 – SL3 + NCp7 protein 10 uM

3 – TAR RNA

4 – TAR + Tat minimal peptide 10 uM

5 – TAR + CycT1m-Tat 10 uM

6 – TAR + MBP tagged chimera 10 uM

7 - TAR + MBP 10 uM

8 - TAR + Tat minimal peptide 5 uM

9 - TAR + CycT1m-Tat 5 uM

10 - TAR + MBP tagged chimera 5 uM

11 - TAR + MBP 5 uM

12 – Tat minimal peptide 10 uM

13 – CycT1m-Tat 10 uM

14 – MBP tagged chimera 10 uM

15 – MBP 10 uM

Chapter 13 Label-Free Binding Analysis

From its introduction in the 1960's (152-154), biosensor technology has developed into a multibillion dollar market utilized by a large and diverse group of industries as well as by the general public (152). Though many different definitions exist, in general a biosensor is a device that is used to detect a signal that is produced when a target molecule interacts with a biological component in close proximity to a transducer (152,155). This label-free technology is now widely employed, and is the basis of several home diagnostic devices available to the general public such as: the ClearBlue pregnancy test, and the electronic blood glucose monitors commonly used by diabetics (152).

For biochemists, biosensors permit the detection and quantitation of interactions between unlabeled biological components, and facilitate a wide variety of experiments by obviating the tedious and problematic process of labeling the biological components under investigation. In addition to the difficulties encountered during labeling, the study of labeled components is frequently hindered by inefficient labeling, as well as by interference, or the necessity to rule out interference, caused by the label itself. A wide array of enzymes, antibodies, and nucleic acids are now assayed by label-free methods using biosensors that employ electrode, thermistor, or optical transducers to convert the interaction between biomolecules into a quantifiable signal (152,156).

13.1. Optical Biosensors

Among optical biosensors surface plasmon resonance (SPR) is at present the most commonly used affinity-based biosensor technique. Several different manufacturers have developed equipment that varies widely in price, ease of use, and popularity among researchers. BIAcore® is currently the market leader, and most of the published research protocols detail techniques applicable to the BIAcore system (GE Healthcare Biosciences, Piscataway, NJ). The system in our lab is a GWC Technologies SPRimager®II (Madison, WI) that, while less popular among researchers, is a more moderately priced model that is capable of a comparable range of assays. Another important variant is biolayer interferometry, discussed below in Section 16.6. As of this writing the bulk of published SPR research still tends to pertain to the investigation of protein-protein interaction, and a good deal less published work is to be found regarding the interaction between a protein and a nucleic acid binding partner such as is the subject of this research.

13.2. Surface Plasmon Resonance

The technology behind SPR affinity-based biosensors is based on the detection of changes in the refractive index of a solution with a positive real part of the dielectric constant Re (ε) flowing over an oscillating electromagnetic wave parallel to a metal (Au, Ag, Cu, Ti, Cr) sensor surface with a negative ε (157). Surface plasmons at the metal surface produce a self-propagating electromagnetic wave called surface plasmon

polaritons (SPP) when infrared polarized light couples (through a prism-coupling arrangement) with free oscillating electrons in the metal (156-159).

Coupling must occur at a specific angle known as the angle of incidence. When biomolecules bind within the sensitive region of 300 nm from the metal surface, incident light is lost to the metal at the surface and the reduction in light is detected as the angle of reflection is shifted by the binding event (Figure 14-1) (159). The difference between the angle of reflection prior to adsorption of the analyte and after binding has occurred is plotted as response versus time and is displayed in real time on the system's computer monitor (Figure 14-2) (156,158,159).

In the typical SPR assay a "bait" ligand, either a protein or nucleic acid, is affixed to a gold surface over which a solution containing the analyte binding partner is passed. The interaction between the analyte and the ligand is recorded as intensity of response versus time in a real-time curve generated by the system software (Figure 14-3). The kinetics of the reaction, in terms of the on-rate, off-rate, and dissociation constant of the interaction, are then calculated by fitting the curves of the association and dissociation phases of the response. The height of the SPR response is related to the mass of material bound to the sensor surface.

13.3. Surface Plasmon Resonance Surface Chemistry

When performing binding assays with the GWC SPRimager®II a self-assembled monolayer (SAM) must be created on the gold spots on the surface of a glass chip prior to beginning the experiment. The SAM provides the surface chemistry for the attachment of the ligand to the metal on the chip surface. Several different types of surface chemistry are available including: amine, streptavidin, or polyethylene glycol (PEG). Deciding which surface chemistry to use for a specific interaction is most frequently an empirical process. The preparation of the SAM is performed over the course of two days. Preparing the chip fresh prior to each experiment is advisable, although regeneration is theoretically possible results are frequently not reproducible.

13.4. Limitations of Surface Plasmon Resonance

Set-up and execution of the SPR experiment with the GWC SPRimager®II is a relatively long procedure that is complicated to assemble, and it is generally not possible to recover sample after the experiment. Accurate determination of the ligand concentration on the surface is not possible due to the potential for artifacts related to the application of the surface chemistry to the chip. Ligand may or may not have successfully bound to the surface chemistry in some regions of the chip surface producing inter-spot or intra-spot binding heterogeneity (160). Consequently, the concentration of bound analyte cannot be accurately known if the concentration of bound ligand is unknown.

13.5.Mass Transport Limitation

In addition to the SPR limitations related to the efficiency of the surface chemistry, are limitations related to analyte delivery. When attempting to measure the binding affinity of an interaction with a fast on-rate it is common for the measured rate to be slower than the true rate according to chemical kinetics. This can occur when the density of the affixed ligand is so great as to sterically hinder the delivery of the analyte to the binding site, and also when the on-rate of the interaction is faster than the rate of analyte delivery that the system can accommodate (160). Further, where the on-rate of the interaction is fast a local depletion of analyte concentration can occur preventing an accurate determination of on-rate (160).

Interference with dissociation can occur when the density of the ligand prevents the free dissociation and diffusion of the analyte. The size and concentration of the analyte must also be considered, as large proteins can hinder binding at other sites. Thus the potential effects of mass transport limitation on the determination of binding kinetics in SPR are not trivial. However, in the evolution of SPR protocols researchers have found that determining and employing a minimal working ligand concentration, and keeping the concentration of the analyte well below that of the ligand are prudent approaches to minimizing the potential for mass transport limitation problems (personal conversation with Dr. Thomas Duncan, Upstate Medical University, Syracuse NY) (160).

13.6. Biolayer Interferometry

Improving upon SPR biosensor technology, biolayer interferometry (BLI) is another label-free optical biosensor detection platform used to study the interaction of biomolecules. In BLI (Figure 14-4 and 14-5) the interaction of light waves in constructive and destructive interference produces a wave pattern that is the basis of optical interferometry technology. Glass fiber-based biosensors, with a variety of available surface chemistries, bind the ligand and are then dipped into 96 well plates containing the analyte in solution.

White light travels down the sensor and is reflected back to the detector from two surfaces within the tip: a reference layer surface, and the interface surface where the ligand meets the analyte solution (Figure 14-6) (161). At the spectrometer individual reflected wavelengths that are either in phase or out of phase combine to form an interference pattern. When binding occurs at the second interface surface of the tip, the change in the interference pattern of the reflected wavelengths is detected as the intensity variation by wavelength. The change in intensity is then plotted as response (in nanometers) versus time (161). The kinetics of the reaction can then be calculated from the association and dissociation phases of the resulting response versus time curve, as in SPR.

However, unlike SPR where microfluidics govern the delivery of the analyte through a flow cell, BLI systems such as the ForteBio's Octet optical biosensor provide "microfluidic-free" delivery of the analyte by dipping the fiber optic sensor tip loaded

with the affixed ligand into an open shaking 96 well microplate rather than flowing the

analyte solution over the chip as in SPR (161). In this manner many of the complications

of microfluidics and analyte delivery are avoided. The ForteBio Dip and Read™

Biosensors are available in a wide variety of surface chemistries and can also be

customized for a specific application. Assays can be performed in crude or purified

analyte solutions, and in many cases it is possible to recover the sample after the

experiment.

13.7. Biolayer Interferometry ForteBio Octet Method

Many attempts to obtain kinetic data for the interaction between the recombinant Tat

chimera and the TAR stem-loop using the GWC SPRimager®II failed to produce

consistent and reproducible results. At the time of the experiments manufacturer

suggestions for ligand density of 1 to 2 mM on the chip surface proved to be too dense

and appeared to contribute to steric hindrance. Results were somewhat more reproducible

when ligand concentrations were reduced, however the three dimensional nature of the

surface chemistry on the chip was also a likely contributor to steric hindrance (personal

conversation with Dr. Thomas Duncan, Upstate Medical University). Hence, BLI was an

attractive alternative for acquiring these data. The Octet Red 96 (Pall FortéBio Corp,

Menlo Park, CA) provides real time data for protein kinetics and quantitation and was

generously made available by Dr. Thomas Duncan of Upstate Medical University in

Syracuse, New York. The streptavidin biosensor tips used with the Octet system affix the

ligand with a surface chemistry that is more two dimensional in nature and less likely to

contribute to steric hindrance (personal conversation with Dr. Thomas Duncan, Upstate Medical University).

To begin each experiment the instrument was set at 22°C so that it could be heated to the experimental temperature of 25°C rather than cooled as this requires more time. Streptavidin-coated fiber optic tips were incubated in buffer at pH 7.4, containing 20 mM Tris, 0.2 M NaCl, 10 µM $ZnCl_2$, 0.02% Sodium Azide, 5 mM BME, and 0.5 mg/ml fat free bovine serum albumin (BSA) for a minimum of 10 minutes prior to beginning the assay.

The first trial compared the response for ligand concentrations of TAR RNA-27 nt at 0.3 µM, 0.6 µM, 1.25 µM, 2.5 µM, and 5.0 µM in a BLI binding assay with a static CycT1m-Tat analyte concentration of 1.5 µM (data not shown). From the results of this assay a 0.5 µM ligand concentration was chosen as the working concentration for subsequent trials. The 0.5 µM ligand concentration was low enough to minimize mass transport limitations, while high enough to produce the appropriate and characteristic response curve for BLI.

13.8. Non-specific Binding

The second experiment was designed to assess the level of nonspecific binding of the protein to the streptavidin tips utilizing negative controls. Raw aligned data in figures 14-7 and 14-8 demonstrates very low level response for the non-specific binding of the CycT1m-Tat (17.5 kDa) to the streptavidin sensor tips in the absence of 5' biotinylated

TAR RNA ligand at analyte concentrations of 1.5, 0.5, and 0.2 μM. This response was similar even at concentration of up to 4.0 μM of CycT1m-Tat (data not shown). The nonspecific response is approximately 0.05 nm above the background response for the negative control sensor in the absence of both the RNA ligand and CycT1m-Tat analyte, and well below the positive control response of 0.5 nm for 1.5 μM CycT1m-Tat bound to the 0.5 μM 5' biotinylated TAR stem-loop.

13.9. Tat-TAR Binding Interaction

Eight disposable fiber-optic sensors were used in the third experimental assay. The single use Dip and Read™ Streptavidin (ForteBio Menlo Park, CA) sensors were incubated in a solution of 0.5 μM 5' biotinylated TAR 27-nt stem-loop RNA ligand (10.6 kDa) purchased from Integrated DNA Technologies, Coralville, IA. Protein analyte solutions used for the assay were: full Tat chimera (60.8 kDa) 4 μM, CycT1m-Tat (17.5 kDa) 4 μM, 2 μM, 1 μM, 0.5 μM, 0.25 μM, and MBP (42.0 kDa) 4 μM (MBP purchased from GenWay Biotech, Inc. San Diego, CA.). All samples were prepared in the same buffer solution containing 0.5 mg/ml fat free bovine serum albumin (ffBSA) to inhibit nonspecific binding.

A 96 well plate was prepared with buffer solution in rows 1, 3, and 5 (Figure 14-9) (Table 14-1). The 5' biotinylated TAR RNA 0.5 μM ligand (in buffer solution) was loaded in 2. In 4 various analyte solutions (in the same buffer solution) were loaded as listed in Table 14-1. For the first 300 seconds (step 1) (Figure 14-10) of the experiment

the streptavidin biosensor tips were incubated in 1 (buffer 1), followed by a 400 second

incubation (step 2) in 2 with 0.5 µM 5' biotinylated TAR RNA 27-nt affixing the ligand

to the tip. At 700 seconds the tips were incubated in 3 (a fresh buffer solution of the same

composition) (step 3) for 600 seconds to remove unbound ligand. In step 4 the tips were

incubated for 60 seconds in a fresh buffer solution in 5 in order to achieve a stable

baseline in fresh buffer unadulterated by the presence of unbound RNA. In step 5 the

biosensor tips were incubated in the various concentrations of analyte solutions for 900

seconds which was sufficient to achieve saturation. In step 6 the tips were returned to the

unadulterated buffer in 5 for 1800 seconds for the dissociation step. Data were displayed

in real time using the ForteBio Data Acquisition 6.4 software program (ForteBio Menlo

Park, CA) (Figure 14.9).

13.10. Determining the Dissociation Constant of the Tat-TAR Interaction

In the formation of the bound Tat-TAR complex the binding event can be represented as

a two-state process by:
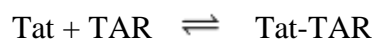
$$\text{Tat} + \text{TAR} \rightleftharpoons \text{Tat-TAR}$$

Thus the corresponding equilibrium dissociation constant K_d is defined as:

$$K_d = \frac{[\text{Tat}] \cdot [\text{TAR}]}{[\text{Tat-TAR}]}$$

where [Tat], [TAR] and [Tat-TAR] represent molar concentrations of the Tat protein, the

TAR RNA ligand and the bound [Tat-TAR] complex, respectively.

The equilibrium of this reaction will be reached when

$[\text{Tat}] \cdot [\text{TAR}] \, k_{on} = [\text{Tat-TAR}] \, k_{off}$

where k_{on} and k_{off} are the on and off rates of the reaction in $M^{-1} s^{-1}$, and s^{-1} respectively.

The equation can then be rearranged

$$K_d = \frac{k_{off}}{k_{on}} = \frac{[\text{Tat}] \cdot [\text{TAR}]}{[\text{Tat-TAR}]}$$

The actual concentrations of Tat, TAR, and the Tat-TAR complex cannot be known

during BLI as discussed previously. Therefore the equilibrium constant K_d must be

calculated from the ratio of the off rate to the on rate for the Tat-TAR interaction.

The data calculated by the ForteBio Data Analysis 6.4 reports a response of k_{obs} for the

binding event. The equation for a trend line of the linear regression of k_{obs} versus

concentration is:

$$k_{obs} = k_a[L] + k_d$$

where k_a is equal to the on rate of the interaction, L is the ligand concentration, and k_d is

the off rate of the interaction in units of per nanomolar per second when reported by Data

Analysis 6.4.

13.11.Biolayer Interferometry ForteBio Octet Results: Fitting the Data

The measured response was minimal and consistent with expectations for the negative control MBP protein (the solubility enhancing tag alone) (Figure 14-10, B7). A larger response, with a rapid onset and early approach to saturation was observed for the full MBP-Tat chimera (Figure 14-10 A7). The samples of CycT1m-Tat demonstrated the largest response that increased in proportion to concentration, saturating at about 2-4 uM for the present conditions. CycT1m-Tat comprises only 29% of the full chimera, so each binding event for the full protein should carry 3.4 times more mass than for CycT1m-Tat (60/17.5 = 3.4) if both proteins had the same binding constant and an equivalent number of binding sites. Figure 14-10 portrays the opposite situation, where the final A7-signal is less than half of that for C7. While it is possible that the MBP portion of the full chimera renders its CycT1m-Tat domain less effective in binding TAR, the most likely explanation is that the MBP portion of the full chimera occludes some of the TAR RNA sites on the Octet tip. The data are consistent with a high-affinity interaction between CycT1m-Tat and surface-immobilized TAR RNA.

13.12.Association 1:1 Model Fit

The association and dissociation phases of the binding response versus time curves (Figure 14-10 steps 5 and 6) were displayed in real time during the experiment. Excluding the MBP protein and full Tat chimera, the association phase (step 5) for the five remaining samples of CycT1m-Tat chimera (4.0 µM, 2.0 µM, 1.0 µM, 0.5 µM, and

0.25 μM) were fit to a 1:1 model. Using ForteBio Data Analysis 6.4 software program the 1:1 model was selected with settings for global fit and R_{max} unlinked by sensor. As seen visually in Figure 14-11, the data did not fit perfectly to the 1:1 model (the fit is improved by invoking a second binding class – see next paragraph). The regression analysis of the 1:1 model to the data in Figure 14-12 for the five concentrations of CycT1m-Tat gave an R^2 value of 0.98 (Figure 14-12).

13.13. Association 2:1 Model Fit

The same data for the association phase of CycT1m-Tat samples fit well to a 2:1 model (Figure 14-13). A linear regression analysis for the response k_{obs1} was plotted for all data points with an R^2 value of 0.99 (Figure 14-14). The slope of the trend line for concentration (nM) versus response is equal to k_a, the on rate for the primary binding event, which accounted for approximately 88% of the response. For the predominant high affinity species the on rate k_a is equal to 2.0×10^{-5} nM^{-1} s^{-1} or 2.0×10^4 M^{-1} s^{-1}.

In the linear regression analysis k_{obs2} for the second binding event, a minor fraction of the response, a single data point was removed as an artifact of the software calculations (Figure 14-15). From the remaining data points the R^2 value was 0.92 and from the slope of the trend line k_a is equal to 1.0×10^{-6} nM^{-1} s^{-1} or 1.0×10^3 M^{-1} s^{-1}.

13.14. Dissociation Fit 1:1 Model

The majority of data for the dissociation step (step 6) fit well to a 1:1 model. The first 200 seconds indicate the presence of a second species which is fast dissociating and did not fit well to the 1:1 model. These first 200 seconds were excluded from the 1:1 model fit. The predominant species, however, was slow to dissociate. Only approximately 0.1 nm of the approximately 0.8 nm total response or 12% was attributable to a fast dissociation by the presence of a second species (Figure 14-16). The average off rate, k_d, for the five cleaved Tat samples was $8.8 \pm 0.6 \times 10^{-4}$ s^{-1} with a $R^2 > 0.99$.

13.15. Equilibrium Dissociation Constant

The equilibrium dissociation constant K_D is a measure of the strength of the binding interaction between two molecules and is defined as the ratio of the off rate to the on rate of the interaction:

$$K_D = \frac{k_d}{k_a}$$

where k_d is the off rate of the interaction in units of s^{-1}, and k_a is the on rate of the interaction in units M^{-1} s^{-1}.

Using the $k_a = 2.0 \times 10^4$ M^{-1} s^{-1} for the 2:1 model predominant interaction, and $k_d = 8.8 \times 10^{-4}$ s^{-1}, the K_D for the interaction between CycT1m-Tat and the TAR 27-nt stem-loop is 44×10^{-9} M (44 nM) indicating a high affinity for the protein-RNA interaction.

For these studies, the CycT1m-Tat protein was not purified to absolute homogeneity and it takes time to concentrate the protein with appropriate buffer exchanges, during which time one might expect a small amount of degradation. There could also be a small amount of the full-length chimera present. Thus, a leading explanation for the Octet results is that 85-90% of the CycT1m-Tat fusion binds TAR RNA with low nanomolar affinity, exhibiting fast association and slow dissociation ($k_a \sim 2 \times 10^4$ M^{-1} s^{-1} and $k_d \sim 9 \times 10^{-4}$ s^{-1}), while related impurities having reduced affinity for TAR RNA exhibit slower association and faster dissociation ($k_a \sim 1 \times 10^4$ M^{-1} s^{-1} and a larger k_d [value not determined] and accounting for ~12% of the Octet response).

13.16. Biolayer Interferometry Discussion: Non-Ideal Behavior

For the sake of completion, it is useful to formally consider the effects that may occur in Biolayer interferometry due to imperfect samples. Non-ideal behavior is typically observed under three circumstances: (1) heterogeneity of the ligand (2) heterogeneity of the analyte (3) mass transport limitation (4) interactions more complicated than a simple 1:1 binding ratio (ForteBio Product Literature).

13.17. Heterogeneity of the Ligand or Analyte

From Figure 14-10 step 5 the five Tat samples appear to reach equilibrium in the association phase, but fail to reach a complete dissociation in step 6. Fitting the data to a 1:1 model demonstrates that a small portion of the dissociation is fast in nature and deviates substantially from the predominant interaction that dissociates slowly. This may indicate heterogeneity of the ligand, the analyte, or possibly both.

Heterogeneity of the RNA ligand can result from a variety of causes, among them: artifacts related to the surface chemistry, dimerization, contamination, and degradation. Similarly, heterogeneity of the analyte can be caused by a multitude of factors such as: impurities in the sample, incomplete or improper folding of some portion of the analyte, contamination, the absence of cofactors, "bridging" (binding more than one ligand molecule), and degradation. Determining the factors or combination of factors responsible for the observed heterogeneity could require many additional assays, and unless the non-ideal behavior is resolved it is possible that the source of the heterogeneity may remain obscure.

13.18. Non-ideal Behavior and Mass Transport Limitation

Non-ideal behavior in the association phase is often associated with a mass transport limitation. Since the association phase of the experiment did reach equilibrium a mass transport limitation is unlikely to be the cause of the observed deviation from ideal behavior.

13.19.Non-ideal Behavior with Multiple Binding Sites

Non-ideal behavior is frequently observed when binding events have a stoichiometry more complicated than a simple one to one ratio. Since the Tat chimera has two zinc fingers which purportedly bind at two separate regions of the TAR stem-loop (the apical loop and the tri-nucleotide bulge) it is possible that the binding of the two zinc fingers may contribute in some manner, as yet unanticipated, to the observed analyte heterogeneity although in theory, binding of either zinc finger individually should produce a similar response (personal conversation with Dr. Thomas Duncan). Moreover, partially or incorrectly folded proteins may be present, and the zinc co-factor may or may not be present in all zinc finger sites. Either of these conditions present in some fraction of the analyte solution has the potential to inhibit, mitigate, or even to enhance binding of the protein resulting in the observance of heterogeneous behavior. Finally, several of these conditions could be contributing simultaneously to the observed non-ideal behavior.

13.20.Reversibility of the Binding Interaction

Since the dissociation step of the experiment did not continue longer than 1800 seconds, and complete dissociation was not observed it is not possible to confirm that the interaction is reversible. Additional experiments with such additives as EDTA (to remove the zinc ions from the zinc fingers), or mutation experiments could be employed to disrupt the binding interaction and demonstrate reversibility in future work.
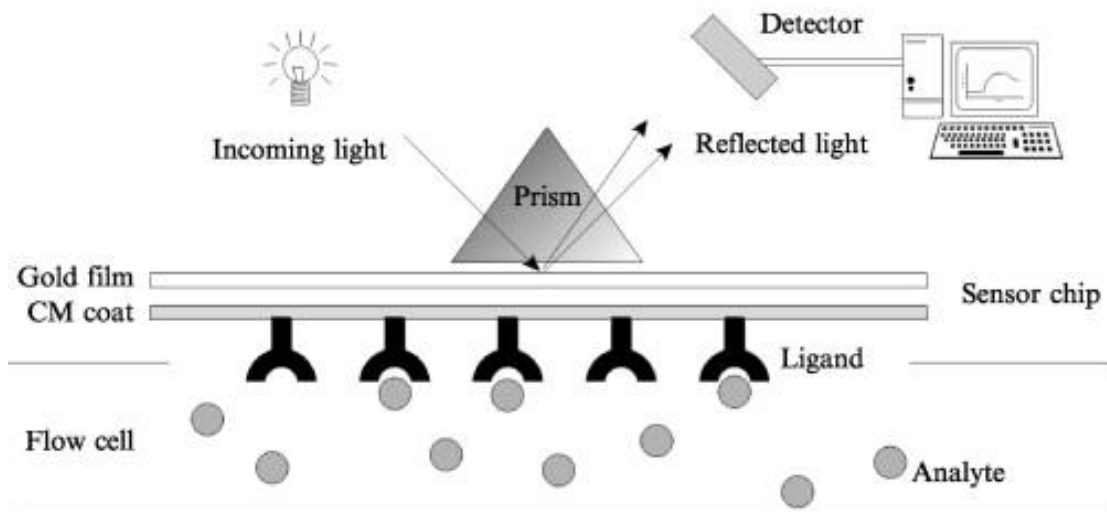
Figure 13-1 Surface plasmon resonance Kretchmann configuration

The surface plasmon resonance (SPR) Kretchmann configuration is the standard design

for most SPR instruments. A soluble analyte solution is delivered to the chip surface

through a flow cell. The interaction between the ligand affixed to the chip surface and the

analyte in solution is detected by an increase in mass at the surface that corresponds to a

shift in the angle of the reflected light. This shift is detected and displayed in real time on

the system's computer monitor. Reprinted from: *Methods in Enzymology,* vol. 399

Hartmann-Petersen, R., and Gordon, C., Pages 164-177, Copyright 2005 (158) with

permission from R. Hartmann-Petersen.
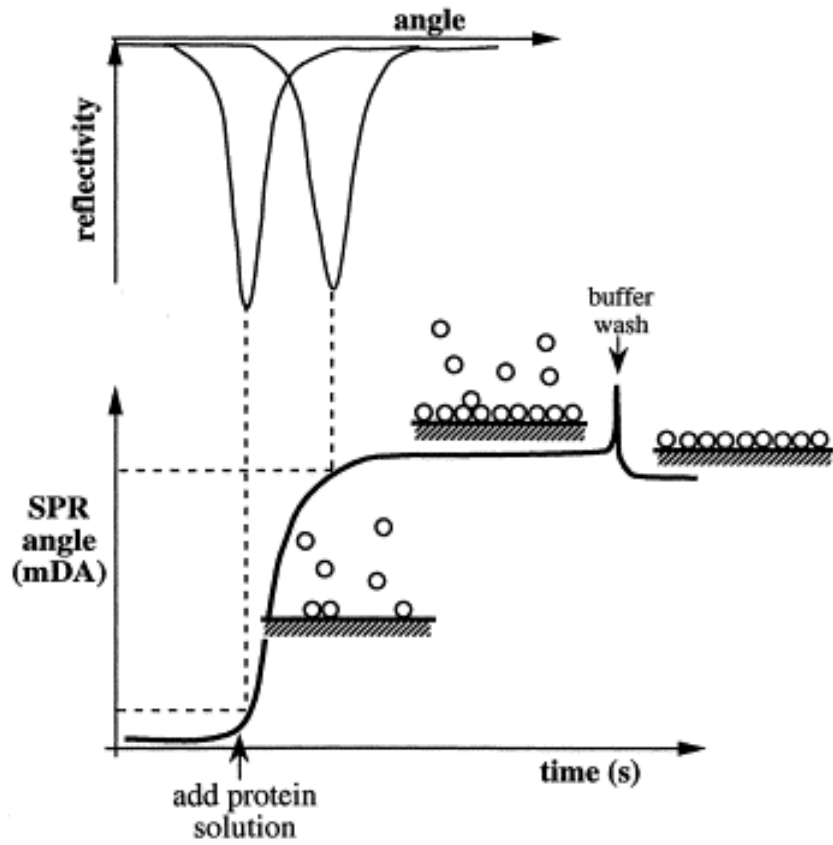
Figure 13-2 Adsorption profile for SPR

A diagram of the adsorption profile for the surface plasmon resonance (SPR) assay. A shift in the angle of reflection is produced when an analyte in solution binds to a ligand affixed to the chip surface. The shift in the angle of reflection is detected and plotted as response versus time in a real time display on the system monitor. Reprinted from: *Methods in Enzymology,* vol. 399 Hartmann-Petersen, R., and Gordon, C., Pages 164-177, Copyright 2005 (158) with permission from R. Hartmann-Petersen.
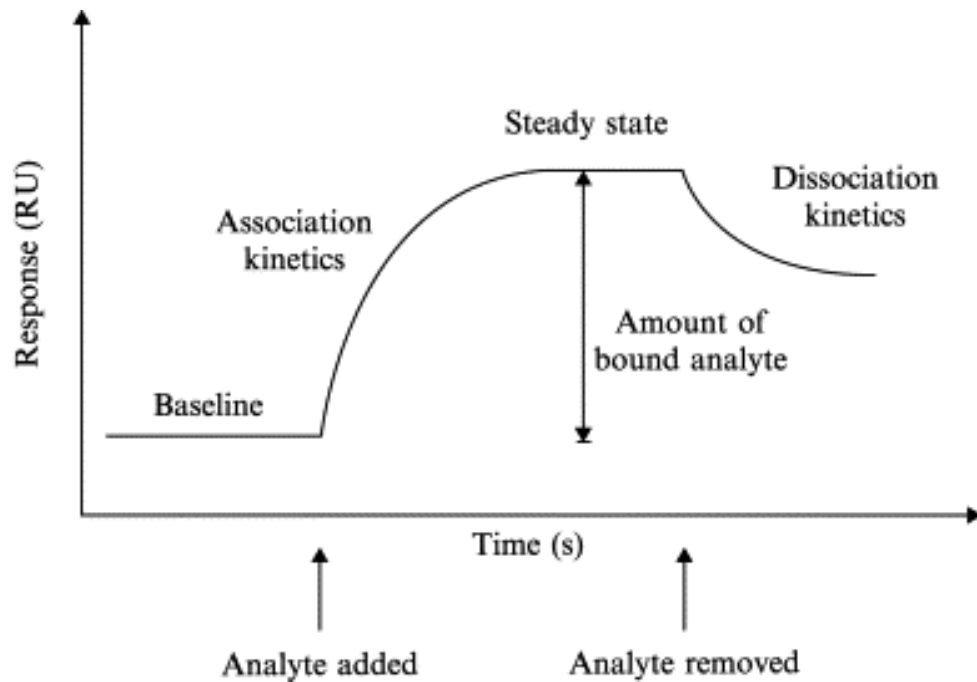
Figure 13-3 A surface plasmon resonance sensorgram.

A surface plasmon resonance sensorgram in which real time data is plotted as response

versus time and displayed on the system monitor. Reprinted from: *Methods in*

Enzymology, vol. 399 Hartmann-Petersen, R., and Gordon, C., Pages 164-177, Copyright

2005 (158) with permission from R. Hartmann-Petersen.
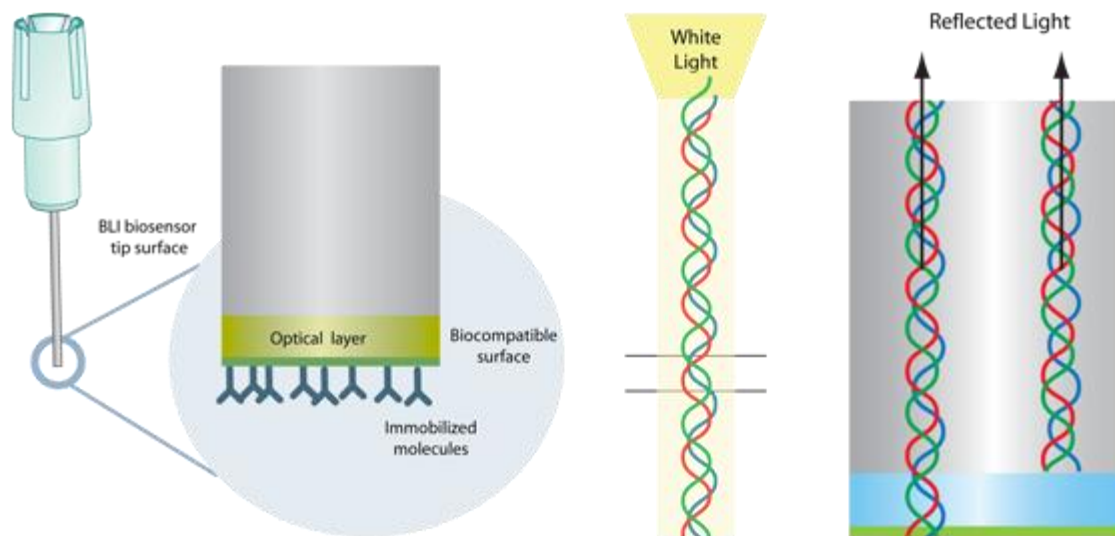
Figure 13-4 Biolayer interferometry sensor tip

A diagram of a biolayer interferometry glass fiber-based sensor tip. White light traveling

down the tip is reflected from two points at the tip surface. Changes in the wave patterns

of the reflected light are used to detect binding at the tip surface. Reprinted with

permission from ForteBio.
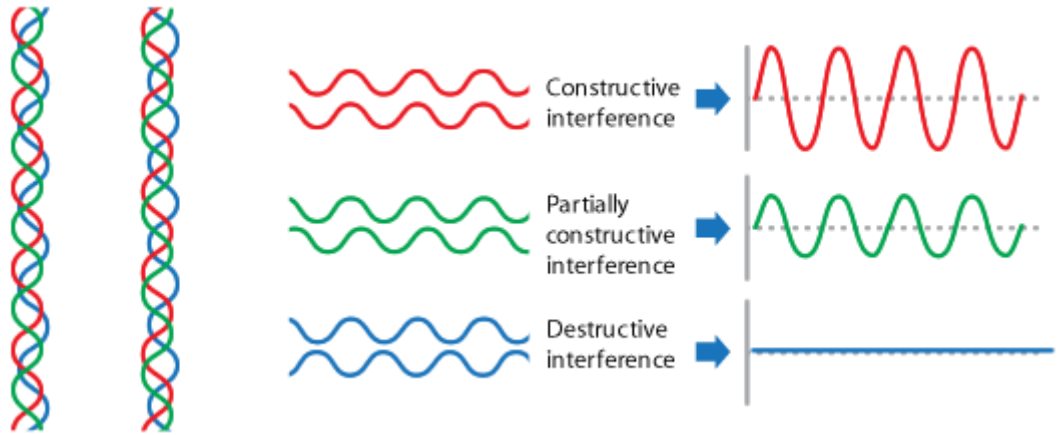
Figure 13-5 Constructive, partially constructive and destructive interference

Constructive, partially constructive and destructive interference patterns of white light.

Reprinted with permission from ForteBio.
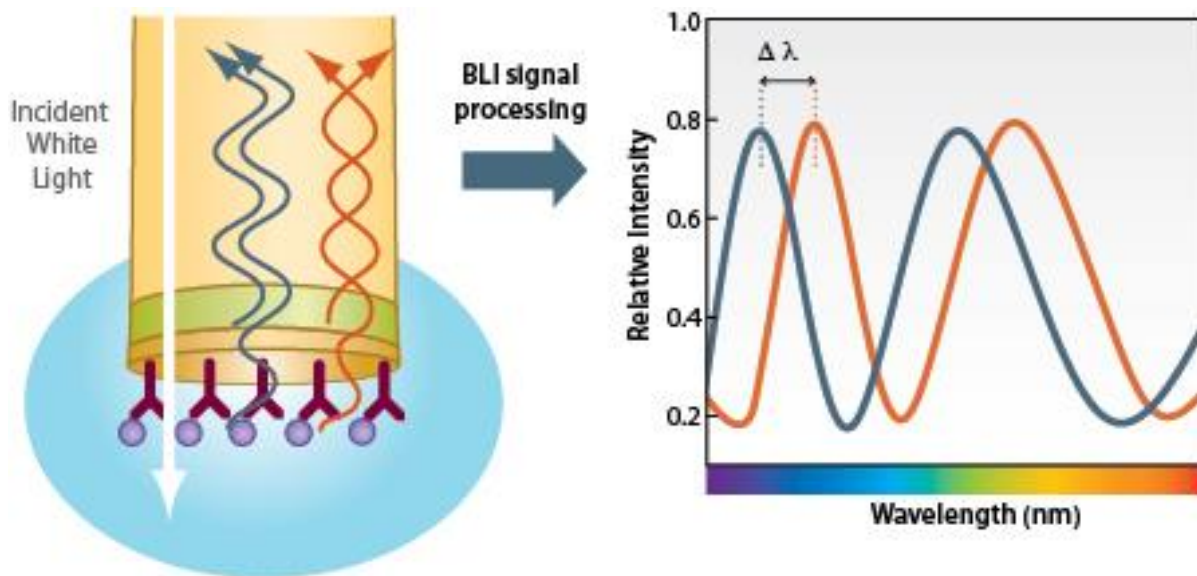
Figure 13-6 Interference captured by the spectrometer

Interference captured by the spectrometer reported as relative intensity in nanometers.

Reflected wavelengths are altered by binding at the surface and the thickness of the

optical layer. Reprinted with permission from ForteBio.
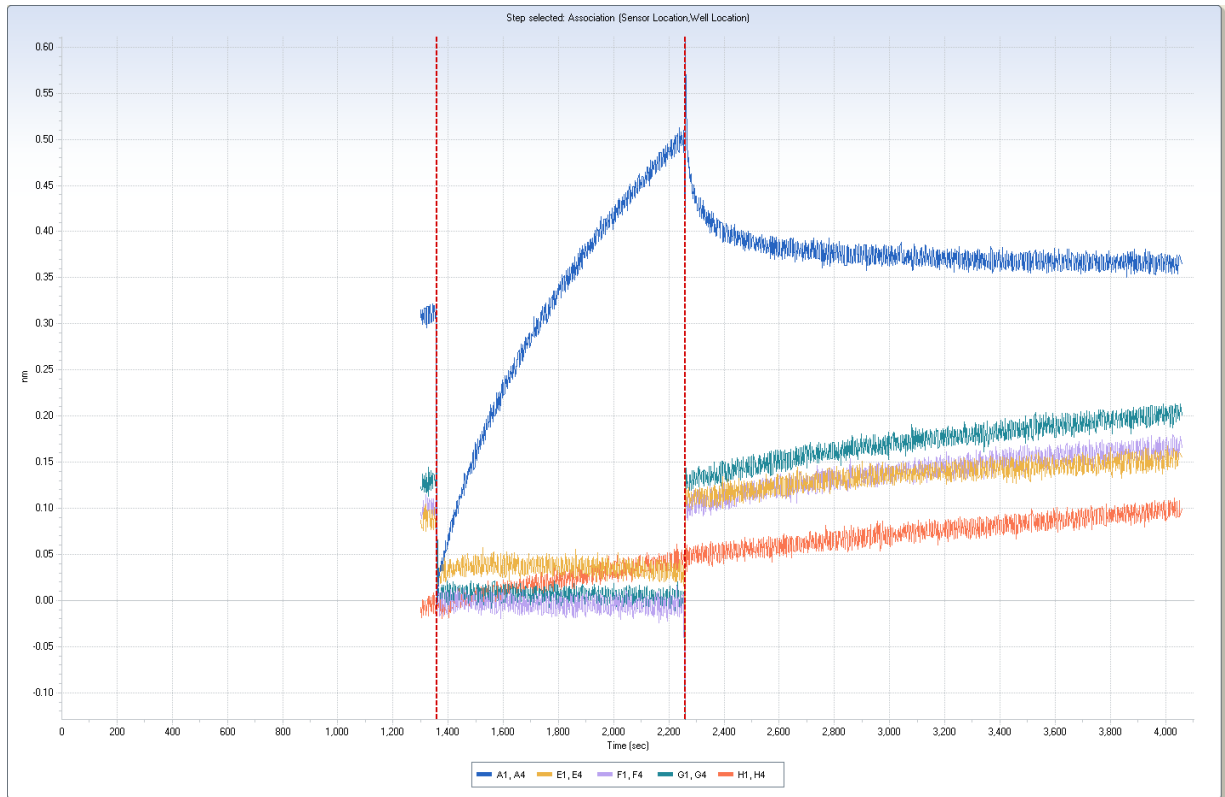
Figure 13-7 Nonspecific binding to Octet biosensor tips

Assessment of the nonspecific binding of CycT1m-Tat to the streptavidin biosensor tips.

Raw aligned data demonstrates a low level of Tat chimera nonspecifically bound to the

streptavidin biosensor tips in the absence of TAR RNA. From left to right:

A (dark blue) – Positive control - 5 uM TAR RNA with 1.5 M cleaved Tat,

E (yellow) – No RNA bound to sensor with 1.5 uM CycT1m-Tat,

F (violet) – No RNA bound to sensor with 0.5 uM CycT1m-Tat,

G (turquoise) – No RNA bound to sensor with 0.2 uM cleaved Tat,

H (red-orange) – Negative control - No RNA bound to sensor and buffer only (No

Protein) analyte solution
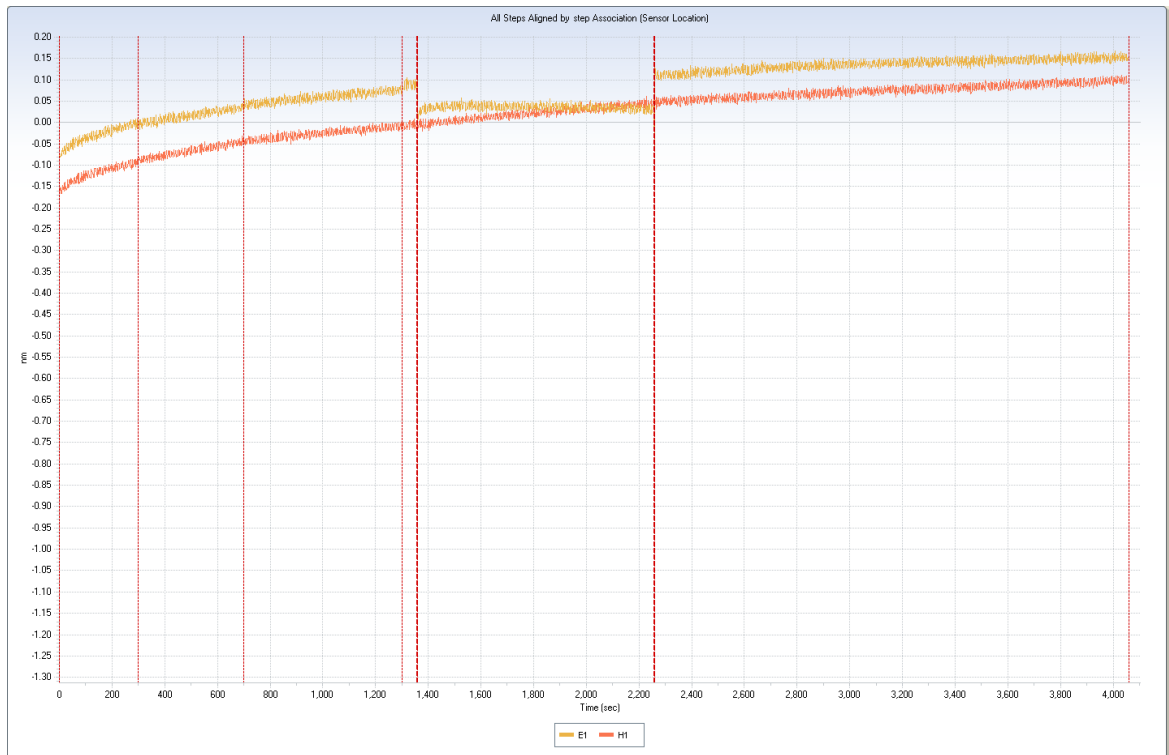
Figure 13-8 Raw aligned nonspecific protein response

Raw aligned nonspecific protein response for streptavidin biosensor tips in the absence of

the RNA ligand:

E (yellow) – No RNA 1.5 uM Cleaved Tat,

H (red-orange) – No protein No RNA.
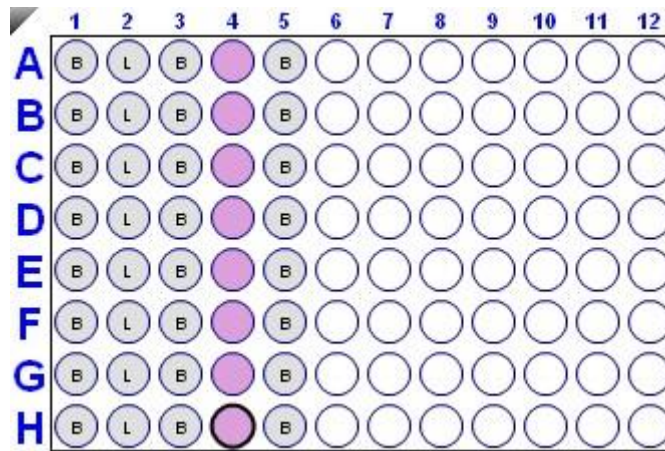
Figure 13-9 The experimental set-up of Octet Red 96 well plate

A diagram of the experimental set-up of Octet Red 96 well plate arrangement for the

loading of samples in Table 14-1.

Experiement Set Up					
	1	**2**	**3**	**4**	**5**
A	Buffer	RNA 0.5 uM	Buffer	**FL Chimera 4 uM**	Buffer
B	Buffer	RNA 0.5 uM	Buffer	**MBP 4uM**	Buffer
C	Buffer	RNA 0.5 uM	Buffer	**Cleaved Tat 4 uM**	Buffer
D	Buffer	RNA 0.5 uM	Buffer	**Cleaved Tat 2 uM**	Buffer
E	Buffer	RNA 0.5 uM	Buffer	**Cleaved Tat 1 uM**	Buffer
F	Buffer	RNA 0.5 uM	Buffer	**Cleaved Tat 0.5 uM**	Buffer
G	Buffer	RNA 0.5 uM	Buffer	**Cleaved Tat 0.25 uM**	Buffer
H	Buffer	RNA 0.5 uM	Buffer	**NO PROTEIN Control**	Buffer

Table 3 The experimental set up for Octet 96 well plate

The experimental set up for Octet 96 well plate from left to right: 1 - buffer at pH 7.4 containing 20 mM Tris, 0.2 M NaCl, 10 uM $ZnCl_2$, 0.02% Sodium Azide, 5 mM BME, and 0.5 mg/ml fat free bovine serum albumin (BSA), 2 – wild type TAR 27-nt RNA at 0.5 μM in same buffer as 1, 3 – buffer as in 1, 4 – various concentrations of recombinant proteins, 5 – buffer as in 1, and 3.
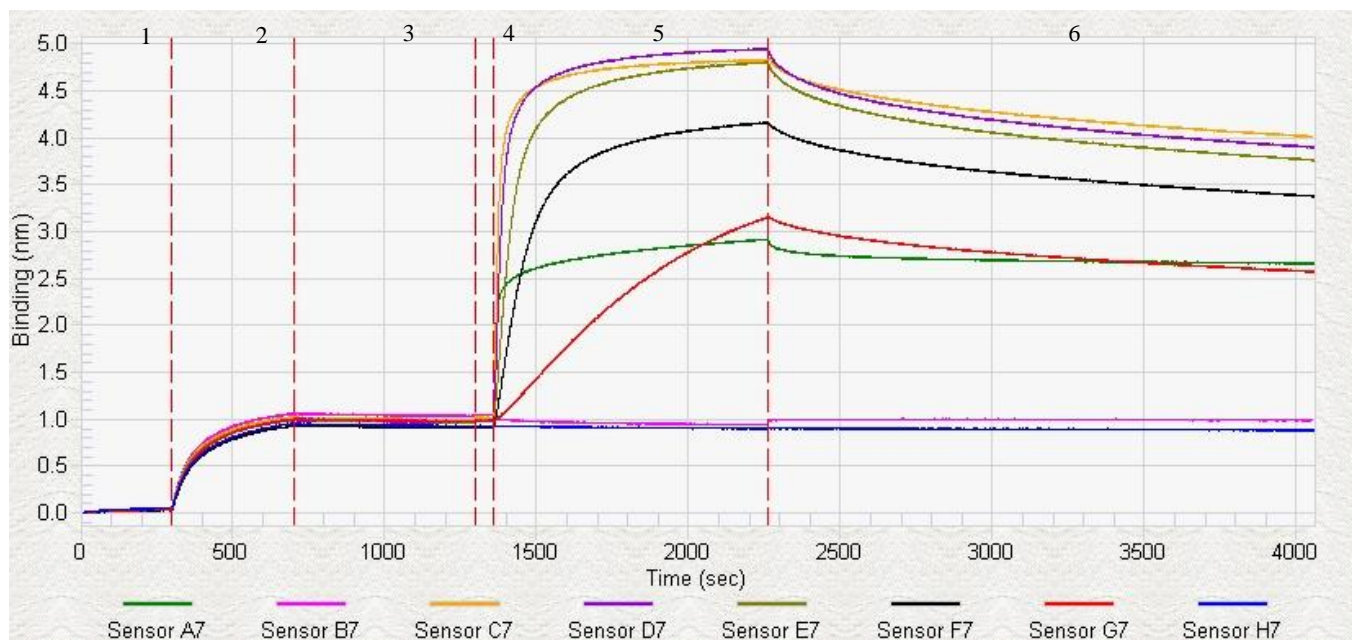
Figure 13-10 ForteBio Octet Sensogram for TAR stem-loop bound to Tat chimera.

A ForteBio Octet Sensogram for the TAR RNA stem-loop bound to the full Tat chimera, CycT1m-Tat, and MBP. Steps 1 through 6 are delineated by dashed red lines. Samples from left to right:

A7 (green) – full chimera 4 μM, B7 (pink) – MBP 4 μM,

C7 (yellow) – CycT1m-Tat 4 μM, D7 (purple) – CycT1m-Tat 2 μM,

E7 (dingy green) - CycT1m-Tat 1 μM, F7 (black) – CycT1m-Tat 0.5 μM,

G7 (red) – CycT1m-Tat 0.25 μM, H7 (blue) – negative control (no protein)
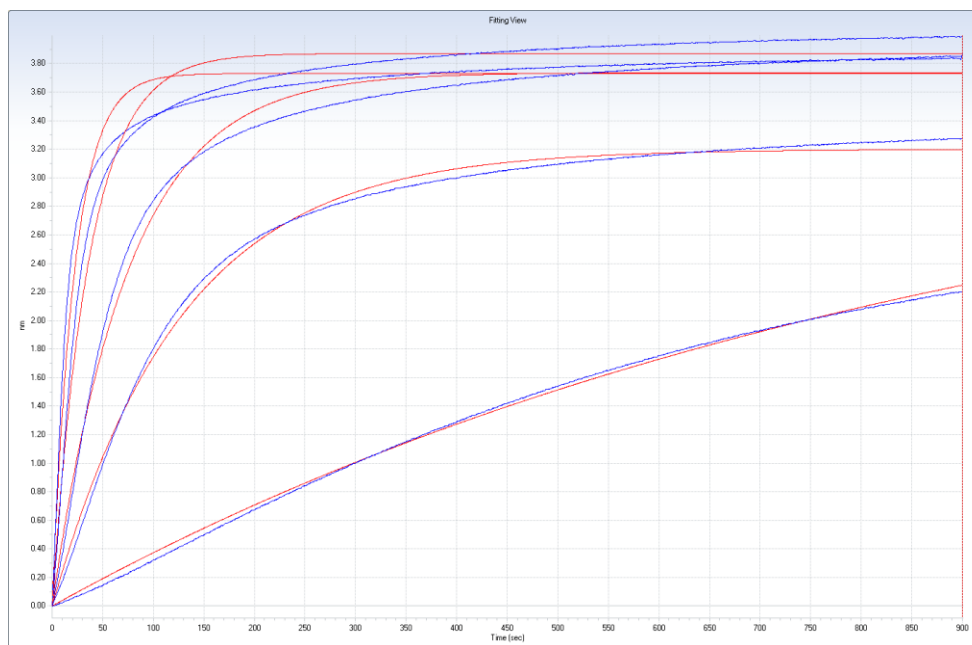
Figure 13-11 Association phase 1:1 data fit

The association phase data for TAR [0.5 µM] RNA bound to CycT1m-Tat at 4, 2, 1, 0.5, and 0.25 µM fit to a 1:1 model. The blue line is the actual binding data and indicates a response in proportion to concentration. The red line is the curve fitting model. The actual data deviated from the 1:1 model considerably indicating a multiphasic response and the possibility of analyte heterogeneity.
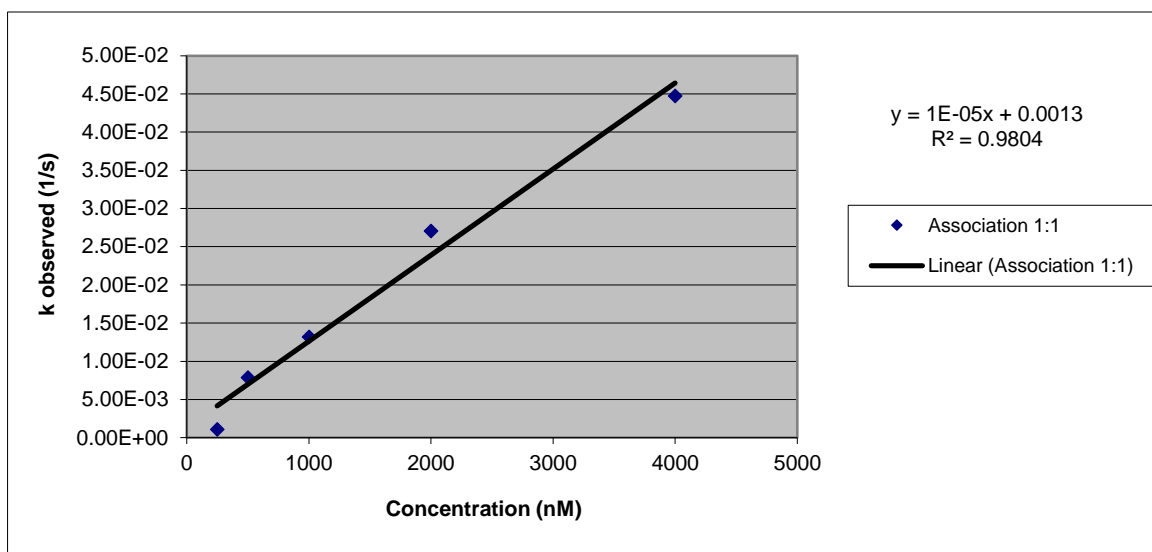
Figure 13-12 Association phase 1:1 data fit regression analysis.

The association phase 1:1 data fit regression analysis for the TAR-Tat interaction in

Figure 14-11. The slope of the trend line for concentration versus k_{obs} is $k_a = 1.0 \times 10^{-5}$

nM^{-1} s^{-1} the on rate of the reaction.
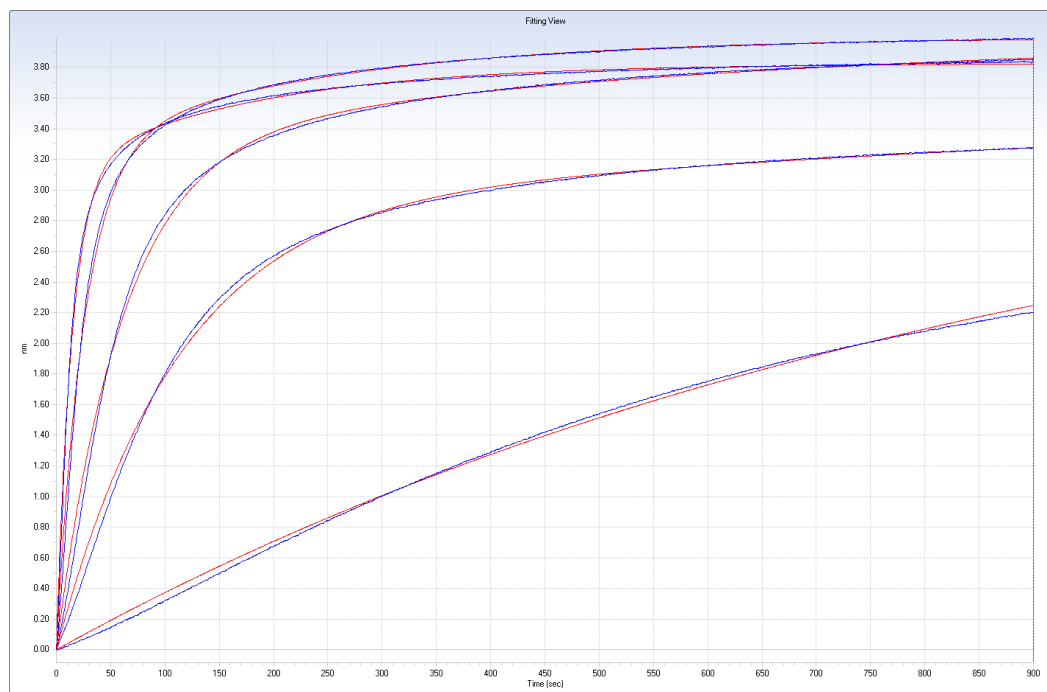
Figure 13-13 Association phase 2:1 data fit.

The association phase data for TAR [0.5 **µM**] RNA bound to CycT1m-Tat at 4, 2, 1, 0.5, and 0.25 **µM** fit to a 2:1 model. The blue line is the actual binding data and indicates a response in proportion to concentration. The red line is the curve fitting model. The actual data fit well to a 2:1 model consistent with predictions for the multiphasic response and analyte heterogeneity.
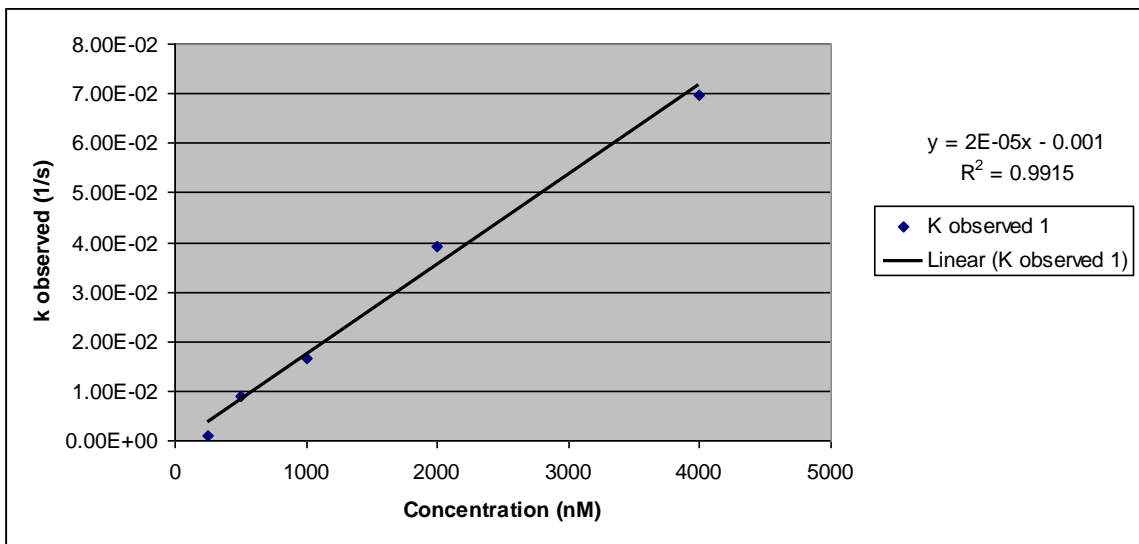
Figure 13-14 – Association phase 2:1 data fit regression analysis

The Association phase 2:1 data fit regression analysis *for the predominant binding event* k_{obs1}. The slope of the trend line for concentration versus k observed (1/s) is $k_a = 2.0 \times 10^{-5}$ nM^{-1} s^{-1} the on rate of the primary reaction.
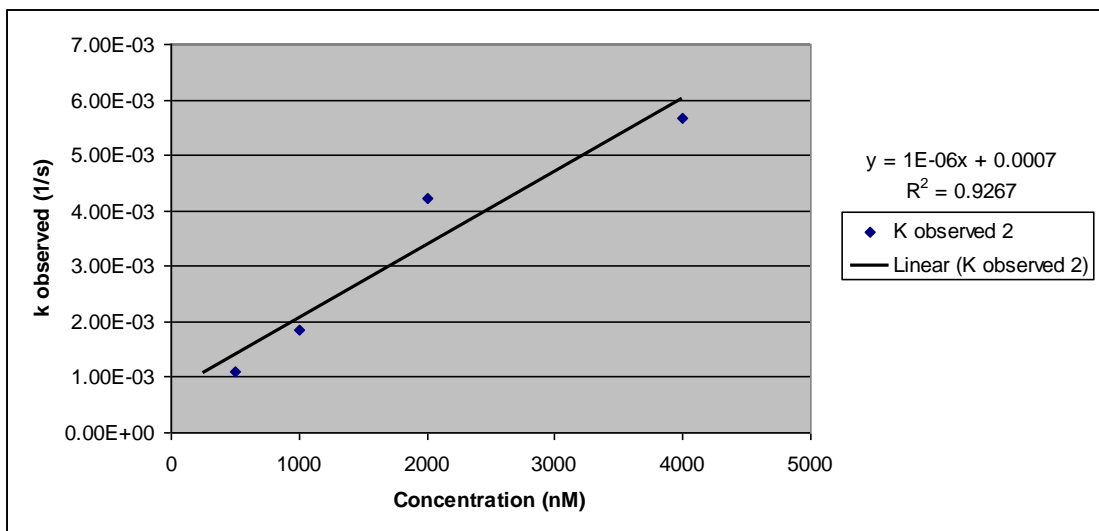
Figure 13-15 Association phase 2:1 data fit regression analysis

The Association phase 2:1 data fit regression analysis *for the minor binding event* k_{obs2}. The slope of the trend line for concentration versus k observed (1/s) is $k_a = 1.0 \times 10^{-6}$ nM^{-1} s^{-1} the on rate of the reaction.
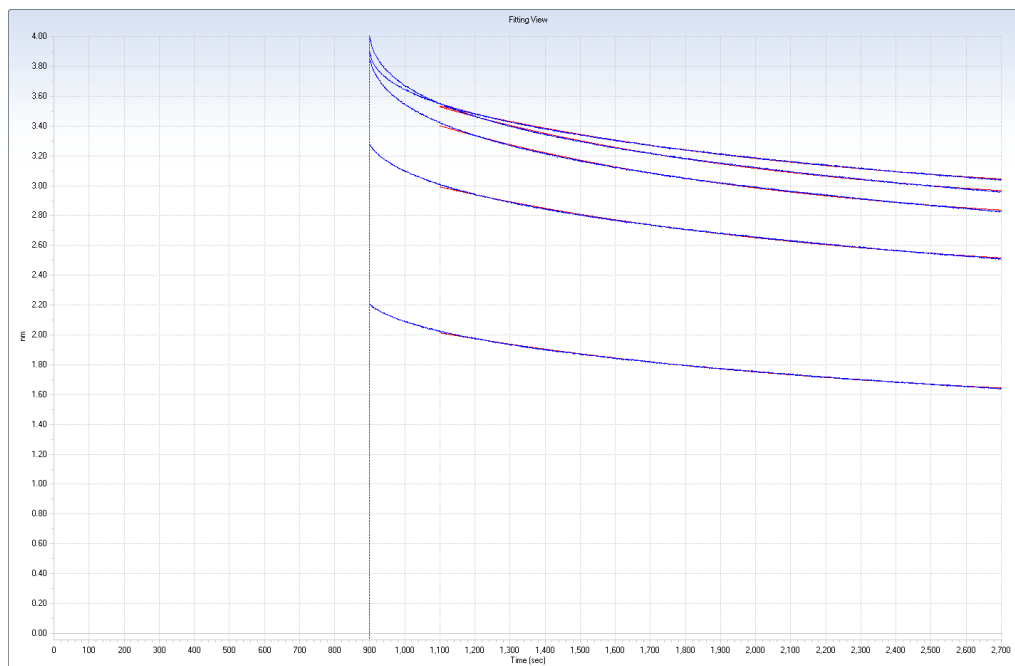
Figure 13-16 Dissociation 1:1 fit minus

The dissociation phase for TAR [0.5 μM] RNA bound to CycT1m-Tat at 4, 2, 1, 0.5, and

0.25 μM fit to a 1:1 model. The first 200 seconds have been omitted from the curve

fitting to remove the data for the fast dissociating species. Less than 15% of the response

appears to be attributable to the low affinity fast dissociating species.

Sensor	Sensor Type	Sample ID	Conc. (nM)	Response	kdis(1/s)	kdis Error	Dissoc X^2	Dissoc R^2
C7	SA (Streptavidin)	Cleaved TAT	4000	3.838	8.58E-04	3.62E-06	0.026	0.999
D7	SA (Streptavidin)	Cleaved TAT	2000	3.988	9.43E-04	4.28E-06	0.045	0.999
E7	SA (Streptavidin)	Cleaved TAT	1000	3.851	9.64E-04	4.29E-06	0.046	0.999
F7	SA (Streptavidin)	Cleaved TAT	500	3.273	8.37E-04	4.09E-06	0.031	0.999
G7	SA (Streptavidin)	Cleaved TAT	250	2.193	7.98E-04	3.59E-06	0.014	0.999

Table 4 Dissociation phase responses

A table of Octet dissociation data indicating the off rates k_d for the dissociation step of TAR RNA [0.5 µM] bound to five concentrations of CycT1m-Tat (4.0 µM, 2.0 µM, 1.0 µM, 0.5 µM, and 0.25 µM). The average k_d was $8.8 \pm 0.6 \times 10^{-4}$ s^{-1}.

Chapter 14 Labeling the Tat Chimera with Stable Isotopes for NMR

As discussed in Chapter 3, the first NMR structures of a TAR-arginine complex were reported by Puglisi, Tan et al. 1992. The structure of TAR alone was reported by Aboul-ela, Karn et al. in 1995, and a 24 residue Tat peptide bound to TAR was reported by Long and Crothers in 1999 (21,22,40,162). However, these structures did not consider or contain any portion of the human Cyclin T1 protein, the essential TAR interacting component of PTEF-b. At present there is still no structural NMR data of the purportedly essential 261 Cys residue of Cyclin T1 that is believed to complete zinc finger 2 in concert with three additional cysteine residues Cys22, Cys25, and Cys27 of Tat (Figure 1-5) (15) , and the active domain of Tat remains obscure.

14.1. Structural Detail of the Tat-TAR Interaction

Initially, the Tat-TAR interaction was predicted to occur at the bulge region of the TAR stem-loop where an arginine rich region of Tat has been shown to interact (21). However, Garber et al. 1998, and Wei et al. 1998 found, by Western blot, that an additional binding interaction between zinc finger 2 and the apical stem-loop contributed substantially to the high affinity of the interaction (15,17). While the crystal structure solved by Tahirov and Babayeva in 2010 of PTEF-b and HIV-1 Tat was groundbreaking the reported structure was disordered in this crucial region of hCycT1 and also failed to elucidate the active

domain of Tat (7). The opportunity then remains to gain insight into this important region from 2D ^{1}H-^{15}N HSQC spectra that could either support or refute the findings of Garber and Wei, while providing structural detail to facilitate rational drug design.

Structural information for the folded state of a protein is reported by 2D ^{1}H-^{15}N HSQC NMR in the form the dispersion pattern, intensity, and number of cross peaks observed, with most of the cross peaks corresponding to the amide ^{1}H-^{15}N nuclei of individual residues in the backbone of the protein. The pattern of cross peaks then provides a diagnostic fingerprint specific to the protein of interest (163). These peaks can then be assigned to begin the process of determining the structure of the protein.

14.2. Determining the Suitability of the Tat-TAR Interaction for NMR

The results of biophysical techniques discussed previously indicated that recombinant CycT1m-Tat was comprised of the appropriate sequence and molecular weight, and that the unlabeled protein could be sufficiently purified and concentrated for structural work. When expressed in Turbo Broth™ a concentration approaching the requisite 0.6 ml of 0.2 mM protein for 2D ^{1}H-^{15}N HSQC NMR was obtained. However, it is still possible that heterogeneity of the sample could prove problematic in structural experiments, for example: by the presence of more cross peaks than residues in the HSQC NMR spectrum. In addition, the presence of multiple domains, as is the case with Tat, has been cited as a potential impediment to NMR structural determination (93,164).

Most of the structures solved by 2D ^{1}H-^{15}N HSQC NMR are less than 30 kDa (164).

Proteins larger than 30 kDa may be candidates for the more recently developed TROSY

technique (165). In general, proteins with multiple domains tend to be more troublesome

subjects for structural study due to difficulties with expression in *E. coli*, conformational

heterogeneity, and the frequently large size of these proteins (93). Initially however, the

wild type TAR RNA at 10.2 kDa, and CycT1m-Tat 17.5 kDa appear to be within the

acceptable range of size necessary for structural study. Although the CycT1m-Tat

chimera does exceed, by four residues, the approximate 150 residue recommended limit

cited by Edwards, Arrowsmith et al. 2000. It is most likely that fully ^{15}N,^{13}C-labeled

protein and RNA will be used to determine the structure by combination of 2D, 3D, and

4D NMR experiments as was done for determining the structure of the complex between

the HIV-1 nucleocapsid protein and the SL3 RNA stem-loop (166).

14.3. The Expression of ^{15}N Labeled Proteins for NMR

The first step toward obtaining 2D ^{1}H-^{15}N HSQC NMR structural data for CycT1m-Tat is

the production of recombinant protein which is uniformly ^{15}N-labeled. Typically an M9

minimal media formula (Appendix 19) containing the ^{15}N isotope in the form of ^{15}NH$_4$Cl

is used to grow the culture and all other sources of nitrogen are removed thereby forcing

the incorporation of heavy nitrogen. In addition to ^{15}NH$_4$Cl, trace metals, and vitamins

are added to the media to aid the bacterial growth under the sub-optimal and taxing

growth conditions afforded by the M9 minimal media. Frequently, the target protein can

be otherwise expressed with the same protocol as would be used for rich media, but a

lower yield of the recombinant protein may be reasonably predicted (rich media contain amino acids with 100% ^{14}N, which would dilute ^{15}N-crosspeaks to an extent that makes the HSQC spectrum useless).

Since the amount of protein required for NMR experiments is relatively large, optimization of the growth conditions, cell density, and protein expression is highly recommended. When expressing ^{15}N labeled proteins, optimization of the growth conditions is limited considerably by the use of minimal media. High cell density cultures can be employed to improve protein yield, but efforts in this regard are often hampered by resulting plasmid loss, reduced pH, and low levels of dissolved oxygen in the growth medium (165). One attractive approach to increasing the yield of the labeled protein without altering the medium, or increasing the culture density is to perform a double colony selection of the expression strain (165). The double colony selection method of selection can be used for normal or high density cell cultures.

14.4. Double Colony Selection

It is always advisable when embarking on a recombinant protein expression to begin with a fresh transformation of the plasmid of interest into its intended host. Plasmid loss occurs over time and during storage due to ampicillin instability and will reduce the overall yield of recombinant protein expression. In addition to a fresh transformation, it is also advisable to screen the transformed colonies as genetic differences will cause some colonies to demonstrate a higher level of target protein production than other colonies.

Exploiting this fact, in a double colony selection several colonies are selected from a fresh transformation plate and separately screened for the highest level of protein expression and, where possible, the lowest level of background expression. The colony demonstrating the most efficient expression in this first round is then grown and plated, and several colonies are selected from this second plate and separately screened in a second round of selection. The colony demonstrating the most efficient expression in the second round of selection is then employed in a large scale expression and purification (165). In this way it is possible to substantially improve protein expression yield and purification without otherwise altering the protocol.

14.5. Method

Following the work of Sivashanmugam, Murray et al. 2009 with minor modifications, a double colony selection was performed in accordance with their published protocol. A fresh transformation of the pDEST His-MBP-hCycT1-Tat plasmid into Rosetta Gami B (Novagen) was performed (Appendix 18). Four colonies were selected from the transformation plate and grown overnight to an OD_{600} of approximately 2-3 in 2 ml of LB media with 100 ug/ml of ampicillin and 35 ug/ml of chloramphenicol at 37 °C. The culture was then spun down, the cell pellet was resuspended in 5 ml of M9 minimal media (Appendix 19) with 100 µg/ml of ampicillin and 35 µg/ml of chloramphenicol to an $OD_{600} = 0.1$, and the culture was grown to $OD_{600} = 1.0$ at 28 °C. Then the culture was induced with a final concentration of 1 mM IPTG, and grown overnight at 28 °C.

14.6. Results: Protein Expression in Minimal Media

Predictably, protein expression in minimal media was reduced when compared with expression levels in rich media. Despite the lower overall expression however, the third colony from the first round of selection demonstrated the highest level of expression from among the four colonies selected (Figure 15-1 lane 3). A 2 ml culture of LB with 100 ug/ml ampicillin and 35 µg/ml of chloramphenicol was then inoculated with colony three and grown overnight. An aliquot of 100 µl of the overnight culture was plated on LB with 100 µg/ml of ampicillin and 35 µg/ml of chloramphenicol, and four colonies were selected from the plate after overnight incubation at 37 °C. The expression protocol was repeated as for the first round of selection above and the expression levels of all colonies are compared in Figure 15-1. After two rounds of selection the fourth colony (Figure 15-1 lane 8 - DCS 3.4) from the second round demonstrated the highest level expression of the target and the lowest level of background expression of all other cellular proteins. A low level of background expression is desirable when attempting to achieve a high degree of purity after chromatography. Hence, colony 3.4 was selected for ^{15}N labeled protein expression.

To assess the effects of IPTG concentration on protein expression a comparison was made of the effects of increasing concentrations of IPTG from 0.2 mM to 1.0 mM on both the original glycerol stock Rosetta Gami B strain (RGB), and the DCS 3.4 freshly transformed high expression mutant. The culture was incubated for 49 hours at 37°C and reached an OD_{600} of 0.5 for RGB, and 0.3 for DCS. At 20 minutes prior to induction 1%

ethanol was added to the minimal media with 100 µg/ml ampicillin and 35 µg/ml of chloramphenicol and the culture was grown for 42 hours at 28°C.

Comparing the original glycerol stocks of the pDEST HisMBP-hCycT1-Tat plasmid in Rosetta Gami B to the DCS 3.4 mutant the increased expression of the recombinant protein by the DCS 3.4 mutant is clear (Figure 15-2). For both the original transformed colony and the DCS 3.4 1.0 mM IPTG produced the highest level of recombinant protein production (Figure 15-2 lane 8 and lane 15) when compared to IPTG concentrations of 0.2 mM, 0.4 mM, 0.6 mM, and 0.8 mM. Notably, the DCS mutant produced considerably more protein despite a lower OD_{600} at induction.

Increasing the concentration of IPTG above 1.0 mM did not appear to further enhance recombinant protein expression (Figure 15-3), nor did decreasing the temperature post-induction to 14°C improve yield. In fact, at 14°C the recombinant protein expression was decreased markedly. While decreasing the post-induction incubation temperature does tend enhance protein solubility, purportedly by allowing increased time to fold, in this case the detrimental effects of the lower growth rate at 14°C eclipsed any gain in yield of the soluble form.

Attempts to improve the expression level of the recombinant protein further by transferring high cell density starter cultures from rich media to minimal media prior to induction were also unsuccessful, despite careful attention to the maintenance of a neutral pH. The cultures required an extraordinarily lengthy incubation to reach an adequate

OD$_{600}$ for induction yet increasing the incubation time after induction from 42 hours to 88 hours actually appeared to reduce the level of expressed protein rather than enhance it both at 28°C and at 14°C (Figure 15-4). This observation is likely due to degradation, and the buildup of toxic waste by-products that accumulated during the lengthy incubation.

14.7.Protein Expression in BioExpress Cell Growth Media

Since most of the more commonly employed techniques for improving recombinant protein expression levels in minimal media had been exhausted it became apparent that a nutrient deficiency could possibly be causing the low level of expression observed for the recombinant chimera in minimal media. Researchers have observed that some proteins are particularly "stubborn" or do not express well in minimal media (University of Connecticut Health Center). In these cases The Gregory P. Mullen NMR Structural Biology Facility at the University of Connecticut Health Center has succeeded at increasing recombinant expression levels of ^{15}N labeled protein with the use of BioExpress cell growth media (Cambridge Isotope Laboratories, Inc. Andover, MA). However, the downside to the use of BioExpress cell growth media is the high cost of this media when compared to traditional formulas. Some researchers have circumvented this drawback, and found that at little as 10% BioExpress formula in the cell growth media can substantially improve the expression of labeled protein for difficult targets **(167).**

Trials with 100% BioExpress cell growth media were conducted by inoculating a 25 ml culture of BioExpress cell growth media with lightly spun down cells from 2 ml of an overnight culture of either freshly transformed RGB cells or the DCS mutant in LB with 100 ug/ml ampicillin and 35 ug/ml of chloramphenicol. After 4 ½ hours of growth the cultures were induced at OD_{600} ~0.7 with 1 mM IPTG. Samples of the post-induction culture were taken after 4 hours and after 18 hours. The target protein expression levels of samples of the DCS mutant in BioExpress cell growth media were compared to the target protein expression levels of the RGB strain in LB, Turbo, and BioExpress cell growth media (Figure 15-5).

The expression levels of the ^{15}N labeled chimera in the DCS mutant (Figure 15-5 lanes 12 and 13) were comparable to the expression levels of unlabeled protein expressed by the RGB strain in rich media (Figure 15-5 lanes 3,4, 6,7,9 and 10). Of note is the clear decrease in target protein expression level observed for the DCS mutant in BioExpress cell growth media after the extended post-induction incubation time of 18 hours when compared to the expression level after 4 hours (Figure 15-5 lane 13). While a similar decrease in target protein expression level is apparent for the RGB strain in LB after 18 hours when compared to the expression level after 4 hours (Figure 15-5 lanes 3, and 4), when expressed in Turbo Media™ the expression level was actually increased after 18 hours. For the RGB strain in BioExpress cell growth media the difference in expression level between the two time intervals is barely detectible (Figure 15-5 lanes 9 and10).

Based on these results, and the yield of the cleaved recombinant protein achieved for unlabeled cultures, it is reasonable to assume that the 2 mg/ml ^{15}N labeled sample concentration required for NMR HSQC would be obtainable by expressing a 1 or 2 liter culture of the DCS mutant in BioExpress cell growth media for approximately 4 or slightly more hours. While typically protein samples might be concentrated for NMR structural study by such methods as the use of concentrating spin columns, CycT1m-Tat did not concentrate well by this method and often appeared to either precipitate or bind to the column material. This loss of protein was evidenced by a concentration which reduced both counter intuitively and unpredictably after multiple rounds of attempted concentration. The decrease in concentration was measured by UV detector at 280 nM with a NanoDrop spectrophotometer (Thermo Scientific Wilmington, DE).
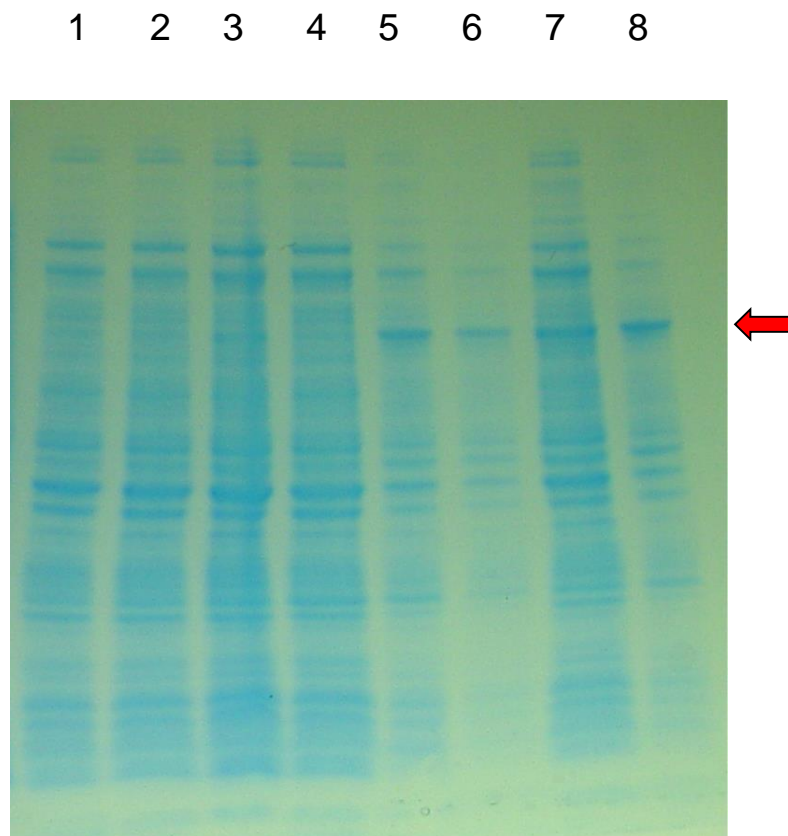
Figure 14-1 Double colony selection and expression in M9 minimal media

The SDS PAGE comparison of the double colony selection (DCS) and expression in M9 minimal media for the selection of a colony expressing high levels of the full Tat chimera. The four colonies from the first round of selection appear in s 1-4. The third colony was selected (3) as the highest expressing colony from the first round and then employed in a second round of selection with colonies from the second round appearing in s 5-8. From left to right:

Round 1:

1 – colony 1   2 – colony 2   3 – colony 3   4 – colony 4

Round 2:

5 – colony 3.1   6 – colony 3.2   7 – colony 3.3   8 – colony 3.4
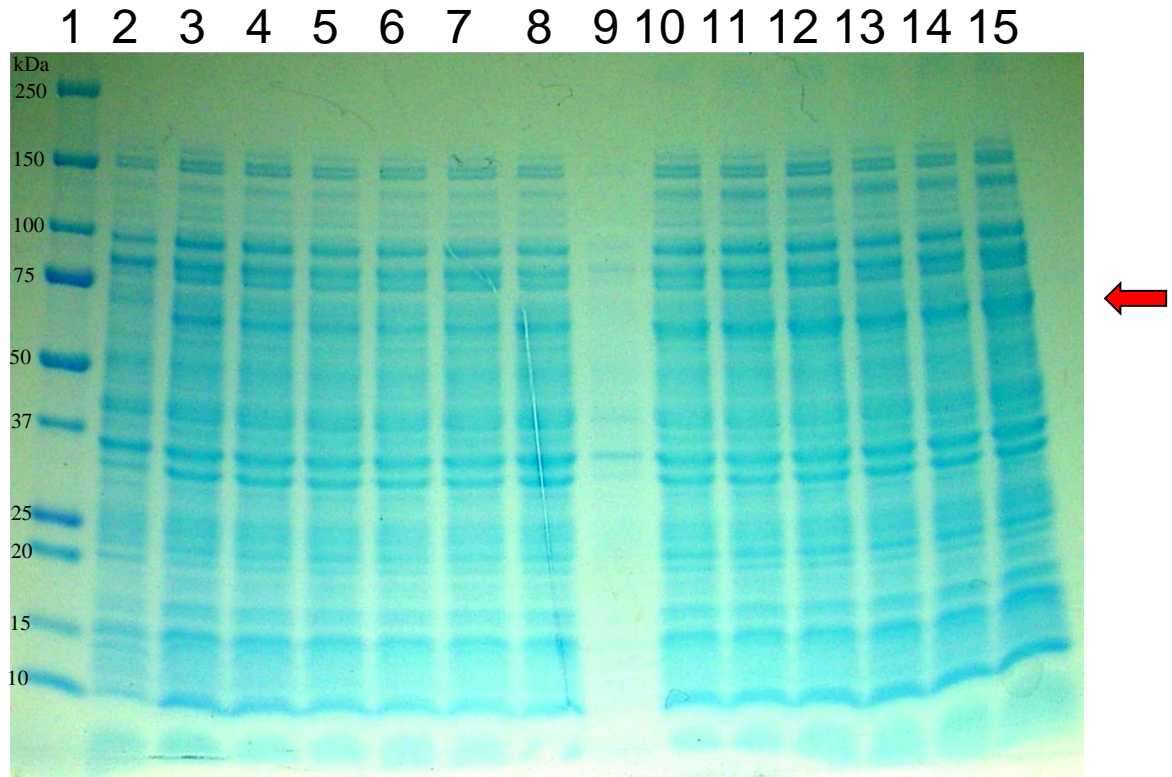
Figure 14-2 Minimal media expression comparison

The SDS PAGE comparison of minimal media expression of the full Tat chimera in the wild type and double colony selection (DCS) mutant (3.4) with variable IPTG concentrations after 42 hours of post-induction growth at 28°C. From left to right:

1 – ladder, 2 – uninduced wild type (wt), 3 – induced wt 0 mM IPTG

4 – induced wt 0.2 mM IPTG 5 – induced wt 0.4 mM IPTG

6 – induced wt 0.6 mM IPTG 7 – induced wt 0.8 mM IPTG

8 – induced wt 1.0 mM IPTG 9 – uninduced DCS colony 3.4

10 – induced DCS 3.4 0 mM IPTG 11 – induced DCS 3.4 0.2 mM IPTG

12 - induced DCS 0.4 mM IPTG 13 - induced DCS 3.4 0.6 mM IPTG

14 - induced DCS 3.4 0.8 mM IPTG 15 – induced DCS 3.4 1.0 mM IPTG
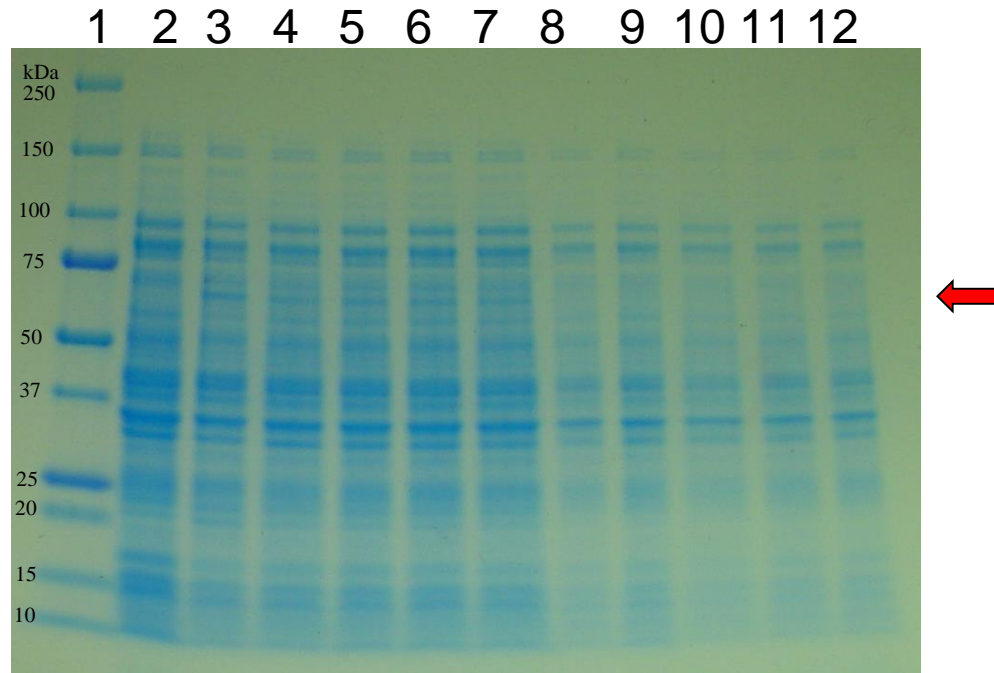
Figure 14-3 An IPTG and temperature comparison for the DCS (42 hours)

An SDS PAGE IPTG induction concentration and post-induction temperature comparison for the double colony selection (DCS) mutant expression of the full Tat chimera in minimal media with increasing concentration of IPTG 1.0-5.0 mM at both 28°C and 14°C for 42 hours post-induction. From left to right:

1 – ladder   2 – uninduced   3 – 1.0 mM IPTG 28°C,

4 - 2.0 mM IPTG 28°C   5 – 3.0 mM IPTG 28°C   6 - 4.0 mM IPTG 28°C,

7 – 5.0 mM IPTG 28°C   8 – 1.0 mM IPTG 14°C   9 – 2.0 mM IPTG 14°C,

10 - 3.0 mM IPTG 14°C   11 - 4.0 mM IPTG 14°C   12 - 5.0 mM IPTG 14°C
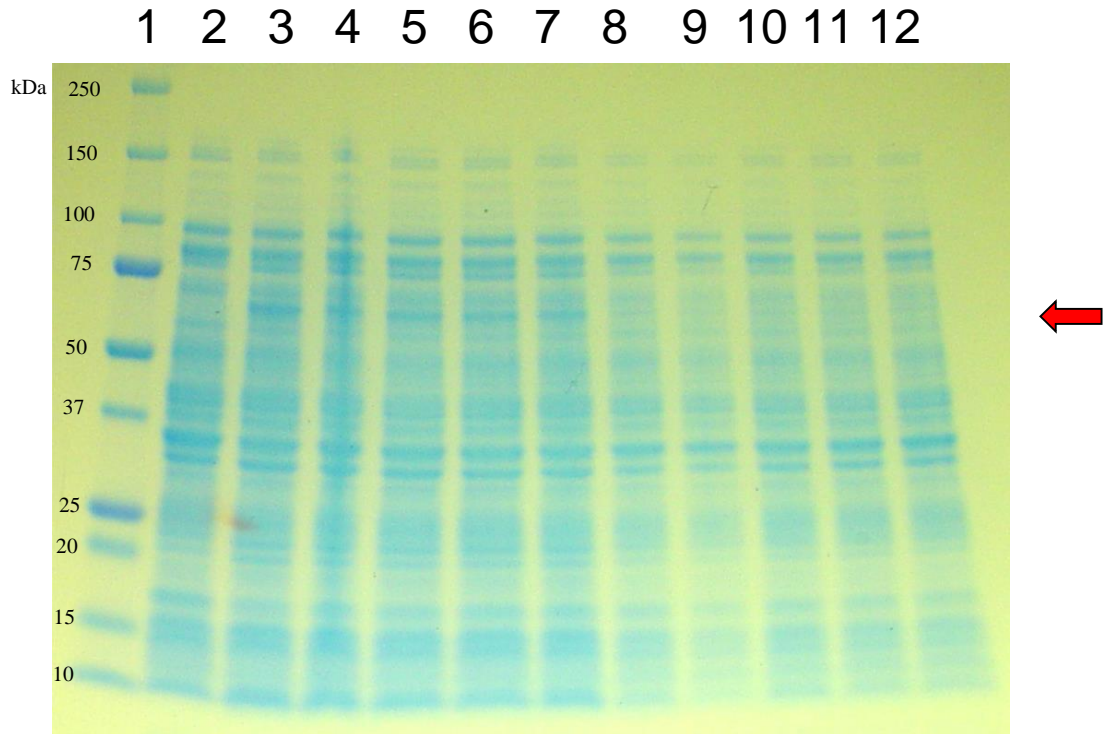
Figure 14-4 An IPTG and temperature comparison for the DCS (88 hours)

An SDS PAGE IPTG induction concentration and post-induction temperature comparison for the double colony selection (DCS) mutant expression of the full Tat chimera in minimal media with increasing concentration of IPTG 1.0-5.0 mM at both 28°C and 14°C for 88 hours post-induction. From left to right:

1 – ladder

2 – uninduced

3 – 1.0 mM IPTG 28°C,

4 - 2.0 mM IPTG 28°C

5 – 3.0 mM IPTG 28°C

6 - 4.0 mM IPTG 28°C,

7 – 5.0 mM IPTG 28°C

8 – 1.0 mM IPTG 14°C

9 – 2.0 mM IPTG 14°C,

10 - 3.0 mM IPTG 14°C

11 - 4.0 mM IPTG 14°C

12 - 5.0 mM IPTG 14°C
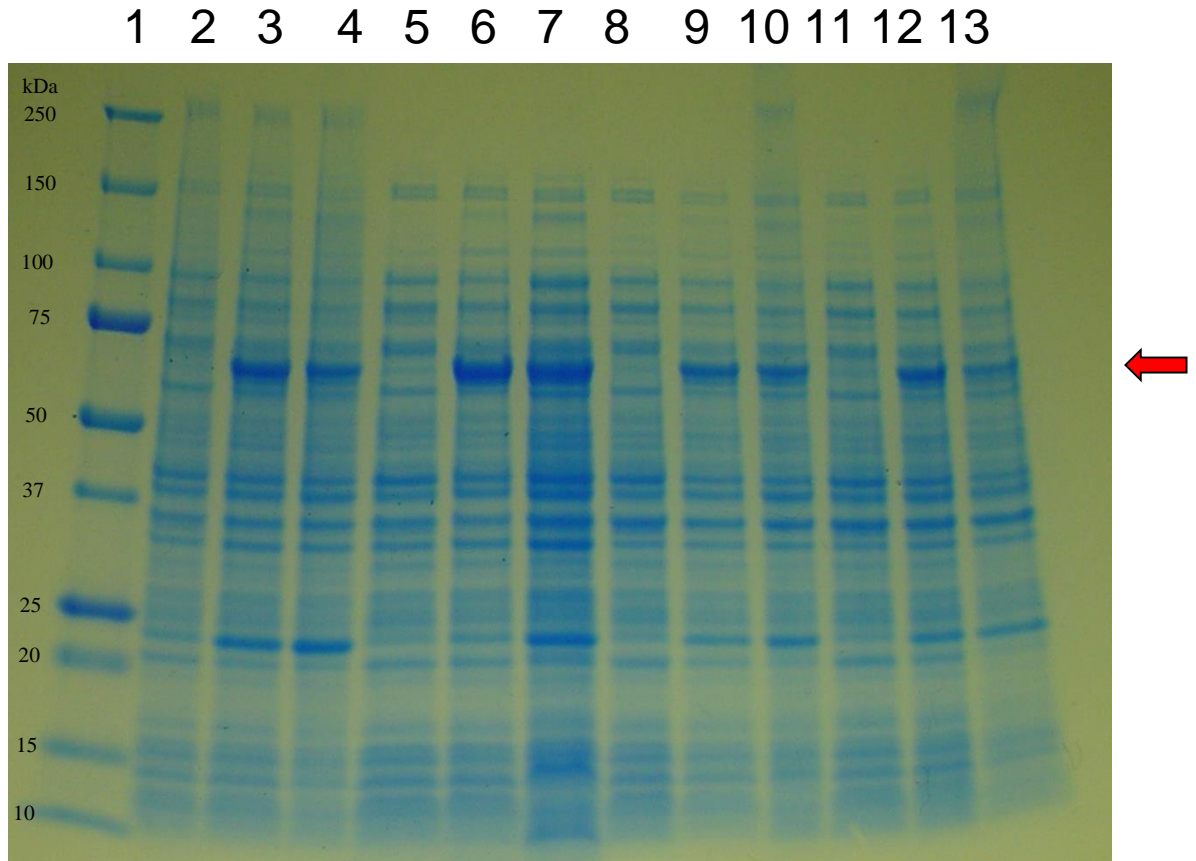
Figure 14-5 Media comparison for expression of the full Tat chimera

The SDS PAGE media comparison of protein expression of the full Tat chimera from

Rosetta Gami B (RGB) wild type and double colony selection mutant (DCS) in LB media, Turbo

Media™, and BioExpress media. All cultures were induced at $OD_{600} = 0.7$ with 1 mM IPTG and

expressed for 4 hours at 28°C. From left to right:

1 – ladder   2 – uninduced RGB in LB   3 – induced RGB in LB, 4 hrs

4 - induced RGB in LB, 18 hrs   5 – uninduced RGB in Turbo Media

6 – induced RGB in Turbo Media, 4 hrs   7 – induced RGB in Turbo Media, 18 hrs

8 – uninduced RGB in BioExpress   9 – induced RGB in BioExpress, 4 hrs

10 - induced RGB in BioExpress, 18 hrs   11 – uninduced DCS in BioExpress

12 – induced DCS in BioExpress, 4 hrs   13 – induced DCS in BioExpress, 18 hrs

Chapter 15 Conclusions and Suggestions for Further Work

The expression and purification of CycT1m-Tat, at a concentration sufficient for structural study, proved to be considerably more challenging than was initially anticipated. The sheer number of options available for modification of expression and purification protocols requires researchers not working with high throughput screening methods to make a series of educated guesses in order to achieve reasonable results in an efficient manner. While such "best guess" conditions may prove to be sufficient to produce adequate protein concentration and purity for a particular downstream application it is unlikely that these will be the optimal conditions for the chosen target. If a substantial quantity of the chimera is to be produced frequently the expression and purification protocols presented in this work should continue to be tested and optimized to the greatest extent necessary for the intended application.

In addition to optimizing expression and purification, the activity of the 257-280 hCycT1-Tat chimera must be assessed. The high affinity binding observed between the re-engineered chimera and TAR is promising, especially with respect to is application toward the production of a diagnostic tool for drug discovery. However, assessing the ability of this chimera to produce transactivation is essential. Moreover, future mutation and deletion experiments modifying key residues of the zinc fingers as well as modifications to arginine residues suspected of playing important stabilizing roles should lend much to our understanding of the key features necessary for recognition, binding, and transactivation

In this section suggestions will be made for: future improvement to the expression and purification of the chimera, activity assays, confirming the zinc content of the chimera, confirming the folded state of the chimera, mutation and deletion experiments, and for the future structure elucidation of the Cyclin T1-Tat-TAR interaction by NMR

15.1. Alternative Expression Strains

An educated guess was made when selecting the expression strain Rosetta Gami B for the recombinant production of the MBP-His$_6$-CycT1m-Tat chimera. The hope initially was not only that the rare codons would be accommodated by Rosetta Gami B but that target solubility might be enhanced by the "tunable" feature of the strain which permits the modulation of expression levels with induction by IPTG. Since no increase in expression level was observed when the concentration of IPTG was varied this feature appears unnecessary for expression of the MBP-His$_6$-CycT1m-Tat chimera.

An additional feature of the Rosetta Gami B expression strain is the formation of disulfide bonds in the cytoplasm as opposed to in the periplasm of the cell. It is possible that the folding of the target protein might in some way be improved by this feature, but no information could be found on the effect of this feature on zinc finger proteins. It is also conceivable, however, that this feature could have no effect or could even hinder the expression of zinc finger proteins which are sensitive to the redox state of the cell (168).

The choice of this strain imposed limitations on the expression of the Tat chimera in that it is not possible to use the Rosetta Gami B strain with an auto-induction media. A comparison could be made between expression and purification in the Rosetta Gami B strain and the Rosetta strain without the "Tuner" lacZY mutation. If a similar yield is obtained expression in auto-induction media may improve yield and facilitate expression by optimizing the induction. Furthermore, it is possible that other strains not considered here may prove superior to the Rosetta strains in the expression of the Tat chimera. If large scale expression of the Tat chimera is undertaken it would be worthwhile to evaluate the performance of several other strains. Ideally, such an evaluation would be done in a high-throughput manner.

15.2. Alternate Solubility Enhancing Fusion Tags

The MBP fusion tag is highly effective at enhancing the solubility of many recombinant proteins. However, from the perspective of structural study, MBP is large (43 kDa) and likely to interfere with many downstream applications; thus its removal is generally required. As has been demonstrated previously in this work, removal of the tag can be time consuming and difficult and will inevitably result in a decrease in the final yield. In addition, removing the fusion tag from less tractable targets frequently and predictably renders the target insoluble once again. For these reasons it might be advantageous to explore other smaller but similarly effective fusion tags such as Small Ubiquitin-like Modifier (SUMO) (12 kDa) which is available with a His_6 modification for affinity

purification. With a fusion tag of this size it may be possible to obtain Tat chimera structural data by NMR while circumventing removal of the fusion tag entirely.

15.3. Optimizing Reduction of the Tat Chimera

Reducing agents used in the expression of the Tat chimera have been discussed at length previously. However, in the literature reviewed there was a conspicuous absence of discussion on the determination of the optimal concentration of reducing agent employed. While insufficient reduction may permit aberrant disulfide bond formation, potentially leading to aggregation and insolubility, conversely it is also possible that the excessive use of reducing agents may alter the secondary structure of the protein and confound downstream applications and structural determination. Therefore, it would be worthwhile to examine the effects of a series of reducing agents and concentrations on the binding affinity of the Tat chimera for the HIV-1 TAR RNA.

15.4. Confirming the Presence of Zinc

Since the native structure of the Tat chimera is dependent on the presence of the two zinc atoms complexed in the zinc fingers of the chimera it is necessary to confirm the presence of zinc in the purified recombinant protein. Several attempts were made to determine the concentration of zinc in the purified samples by a direct spectrophotometric method. Since Zn^{2+} itself is spectroscopically silent 4-(2-pyridylazo)resorcinol (PAR)resorcinol and iodoacetamide were used to measure the presence of Zn^{2+} (169). In this reaction 2-iodoacetamide is used as an alkylating agent to covalently bind to the thiol

group of the cysteine residues in the two zinc fingers of the Tat chimera and release Zn^{2+}.

Free Zn^{2+} is then complexed to form $Zn(PAR)_2$ (170-172) and the solution changes color

from yellow to orange. A change in absorbance at 490-500 nm can then be measured

spectroscopically and used to calculate the approximate concentration of zinc complexed

in the zinc fingers of the recombinantly produced Tat chimera (171,172).

Method

First a standard curve was calculated using the absorbance measured at 500 nm of $ZnCl_2$

and PAR added to a solution to form $Zn(PAR)_2$ for a range of $ZnCl_2$ from 10 uM to 100

uM. The response for this range was fairly linear (Figure 16-1). Next, following a

protocol by Pfister et al. 2000, a 48.75 ul sample of 48.6 uM Tat chimera was incubated

with 1.25 ul of Proteinase K (2.0 ug/ul) and incubated for 30 minutes at 56 °C (172).

Subsequently, Iodoacetamide (5 mM) and PAR (0.2 mM) were added to a total volume of

100 ul, and the absorbance of the solution at 500 nM was measured by NanoDrop

spectrophotometer (Thermo Scientific Waltham, MA).

Results

The absorbance of two redundant samples ranged from .063 to .151 placing the estimated

concentration of Zn^{2+} in the Tat chimera samples at between approximately 13 and 30

uM according to the standard curve (Figure 16-1). Since the final concentration of the Tat

chimera should theoretically have been approximately 24.3 uM, but two atoms of Zn^{2+}

should be present for each molecule of the chimera the predicted concentration of $Zn(PAR)_2$ would be 48.6 uM with, according to the standard curve, a predicted absorbance at 500 nm of approximately 0.280. Allowing for approximately 20% contamination in the sample the actual absorbance measurements of $Zn(PAR)_2$ still fall short of the predicted value. This deviation from prediction may be attributable to a variety of factors including: inaccuracies inherent in the nature of the calculation of extinction coefficients using NMR data, limitations of the NanoDrop spectrophotometer equipment (especially for absorbance values close to background levels), and pipetting error. Furthermore, some of this error may be due to misfolded proteins that are missing either or both Zn^{2+} cofactors.

15.5. Inductively Coupled Plasma Mass Spectrometry

The determination of Zn^{2+} concentration by spectrophotometric means continued to be problematic after many trials. Hence an alternative method of determining the Zn^{2+} concentration of the Tat chimera, such as might be accomplished by inductively coupled plasma mass spectrometry (ICP–MS), is desirable in order to confirm the presence of the zinc cofactor at the appropriate concentration. A generally quantitative instrument, ICP-MS has low detection limits in the parts per thousand range and is used for determining the concentration of a large group of elements including zinc. Such equipment is readily available for use by our lab. Assessment of the Zn^{2+} content of the Tat chimera should be undertaken regularly in this native recombinant expression and purification protocol, and may offer valuable insight into the efficacy of any future modifications of same.

15.6. Activity Assay

In order to assess the activity of the recombinant CycT1m-Tat chimera an activity assay should be conducted. Such an assay might be similar to that performed by Fujinaga et al. 2002 which measured the levels of transactivation produced by the Tat chimera from the HIV-1 LTR in NIH 3T3 cells (28). However, this assay requires working with mammalian cells, and as our laboratory is not currently equipped to handle this work collaboration with another laboratory may be required.

15.7. Confirming the Native Structure and Assessing Folding

Since the data acquired from biolayer interferometry indicates the presence of a (minor) fast dissociating species it is reasonable to suspect the presence of some partially or incorrectly folded proteins. One interesting and high-throughput method that has been useful in assessing protein folding involves the interaction of the promiscuous and high affinity chaperonin protein GroEL with a purified target protein in biolayer interferometry experiments (173-176). Here chaperonin is employed as a "kinetic trap" which binds promiscuously to hydrophobic regions of partially folded proteins. Due to the fairly large size of chaperonin (802 kDa) a large BLI signal is observed when the protein binds preferentially to a partially folded target protein (176). Such an assay might reasonably be used to determine the degree to which the purified Tat chimera sample may be present in a non-native confirmation, as well as to identify optimal conditions for target solubility and correct protein folding.

15.8. Analytical Ultracentrifugation

In addition to determining the binding affinity of the interaction between the CycT1m-Tat chimera and TAR RNA, accurate quantitative characterization requires a determination of the stoichiometry of the unbound recombinant protein as well as the stoichiometry of the protein-RNA complex. A number of matrix-based and matrix-free techniques are available for the identification of protein aggregates. Among the matrix-based techniques SDS-PAGE, and size exclusion chromatography (SEC), and among the matrix-free sedimentation velocity (SV), dynamic light scattering (DLS), and field flow fractionation (FFF) are most common (177).

Analytical ultracentrifugation (AUC) is one of the most efficacious techniques currently available for the quantitative characterization of macromolecular associations in solution (178). The Beckman Coulter ProteomLab XL-A AUC, in the laboratory of Dr. Michael Cosgrove at Upstate Medical University, is equipped with absorbance optics and calculates the sedimentation velocity of macromolecules in solution by measuring the absorbance of monochromatic light passed through a sample cell during high speed centrifugation (177). Sedimentation velocity data provides information about size distribution that can be extracting using SEDFIT computer software developed by Peter Schuck (177,179). The SEDFIT software utilizes Lamm equation solutions to describe the sedimentation data and from these data oligomeric state can reasonably be determined (177,179).

The use of the AUC has generously been extended to us by the Cosgrove lab for our work on the CycT1m-Tat chimera. A gel filtration polishing step will be added to the purification protocol to remove residual contaminants that might confound the data from this sensitive technique. In addition to confirming the monomeric state of the recombinant chimera, this assay may also yield data about the affinity of the interaction between CycT1m-Tat and TAR RNA.

15.9. Structural Study

The ultimate goal of this work is to produce a sample of Tat chimera protein of sufficient concentration and purity to obtain structural NMR data. The yield and purity obtained from the small scale minimal media experiments indicated that, theoretically, a large scale expression and purification of the Tat chimera protocol herein should yield adequate protein for structural work. A few additional trials with fresh solutions, and perhaps different formulae, of trace minerals and with minimal solutions of BioExpress media would be helpful toward optimizing the protocol and minimizing expenditures.

15.10. Mutation and Deletion Experiments

Toward improving our understanding of the essential features of the hCycT1-Tat interaction mutation and deletion experiments substituting or removing key residues are likely to yield valuable insight into the nature of interaction with TAR. Specifically, mutation or deletion of residues of the two zinc fingers, mutation and deletion of R251

and R254 (which have been mentioned as possible stabilizers of this interaction (26)), and mutation or deletion of key arginine residues in the arginine-rich region of Tat could offer this much needed structural insight. By minimizing the interaction between the chimera and TAR to that afforded solely by the two zinc fingers the impact of the zinc fingers contribution to binding and transactivation can be more clearly demonstrated.

15.11. HIV-1 Drug Discovery

Having produced a high affinity Tat chimera containing the appropriate zinc cofactors, and after confirming the appropriate biological activity, the production of a Tat-TAR indicator (Figure 2-1) can then commence. Many small molecule inhibitors with potential for the treatement of HIV-1 can be simultaneously screened for efficacy by means of high-throughput methods that detect the compounds ability to disrupt the interaction between the Tat chimera and the HIV-1 TAR RNA. The Tat chimera recombinantly produced here offers the great advantage of a secondary structure more closely resembling the *in vivo* structure encountered in the disease process where such mimicry is otherwise prevented by the dissordered nature of the Tat protein independent of the essential portion of hCycT1.
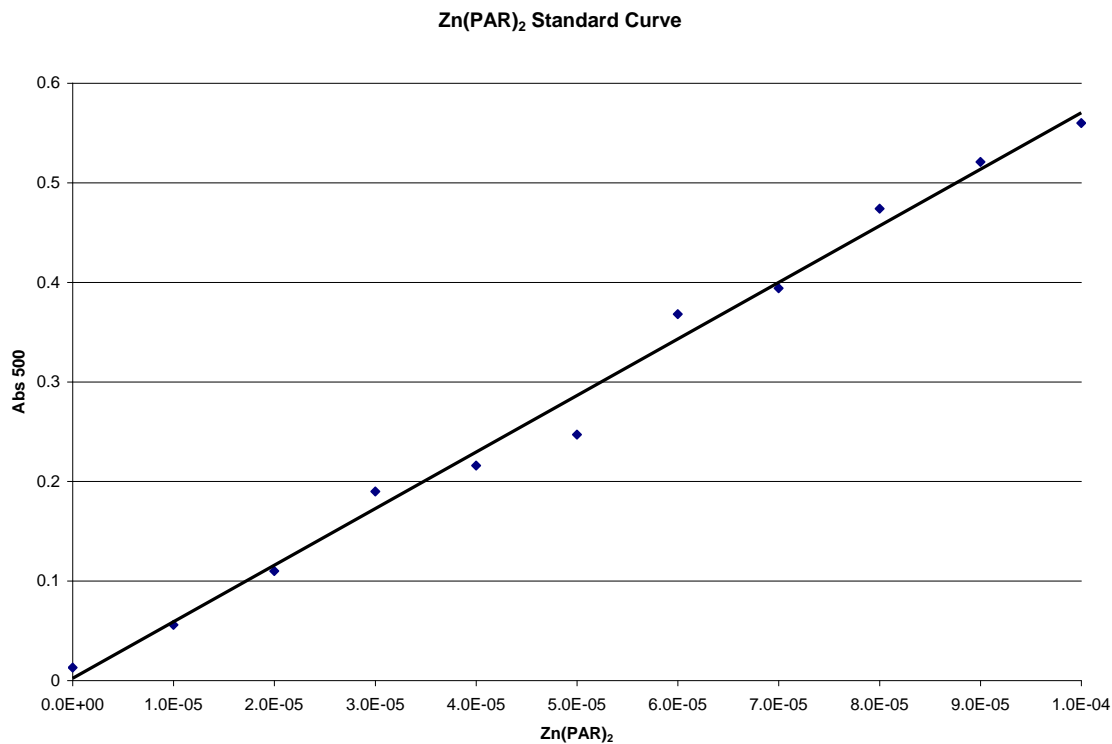
Figure 15-1 Standard curve for the formation of Zn(PAR)2 measured at 500 nm.

A standard curve was calculated by measuring the absorbance at 500 nm of $ZnCl_2$ and

PAR added to a solution to form $Zn(PAR)_2$ for a range of $ZnCl_2$ from 10 uM to 100 uM.

The response measured by a Nanodrop spectrophotometer for this range was reasonably

linear.

Appendix

Appendix 1 - Full length GST-hCycT1-Tat Chimera Sequence Details

```
MEGAVLDIRY GVSRIAYSKD FETLKVDFLS KLPEMLKMFE DRLCHKTYLN GDHVTHPDFM  60
LYDALDVVLY MDPMCLDAFP KLVCFKKRIE AIPQIDKYLK SSKYIAWPLQ GWQATFGGGD  120
HPPKSDLVPR GSPNRLKRIW NWRACEAAKK TKADDRGTDE KTSEQTMPEQ KLISEEDLAM  180
EFLEIDPVDM EPVDPNLEPW KHPGSQPRTA CNNCYCKKCC FHCYACFTRK GLGISYGRKK  240
RRQRRRAPQD SQTHQASLSK QPASQSRGDP TGPTESKKKV ERETETDPFD L          291
```

Isotopically Averaged Molecular Weight = 33491.1992

Estimated pI = 8.10

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the

isolated residues. For a folded protein this is **not** valid. However, this rough value can be

useful for planning protein purifications. pKa values for the individual amino acids from

Stryer Biochemistry, 3rd edition.

Estimated charge at pH 7.00 = 5.3

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the

isolated residues. For a folded protein this is not valid. However, this rough value can be

useful for planning protein purifications. pKa values for the individual amino acids from

Stryer Biochemistry, 3rd edition.

Estimated charge over pH range

pH	Charge
4.00	41.8
4.50	30.0
5.00	19.4
5.50	13.5
6.00	10.3
6.50	7.7
7.00	5.3
7.50	3.3
8.00	0.7
8.50	-3.4
9.00	-8.7
9.50	-16.2
10.00	-26.8

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

PROTEIN CALCULATOR v3.3

Chris Putnam
cdputnam@scripps.edu
The Scripps Research Institute
Last Updated: March 28, 2006

Appendix 2 - Full length MBP-hCycT1-Tat Chimera Sequence Details

```
MKIKTGARIL ALSALTTMMF SASALAKIEE GKLVIWINGD KGYNGLAEVG KKFEKDTGIK
VTVEHPDKLE EKFPQVAATG DGPDIIFWAH DRFGGYAQSG LLAEITPDKA FQDKLYPFTW
DAVRYNGKLI AYPIAVEALS LIYNKDLLPN PPKTWEEIPA LDKELKAKGK SALMFNLQEP
YFTWPLIAAD GGYAFKYENG KYDIKDVGVD NAGAKAGLTF LVDLIKNKHM NADTDYSIAE
AAFNKGETAM TINGPWAWSN IDTSKVNYGV TVLPTFKGQP SKPFVGVLSA GINAASPNKE
LAKEFLENYL LTDEGLEAVN KDKPLGAVAL KSYEEELAKD PRIAATMENA QKGEIMPNIP
QMSAFWYAVR TAVINAASGR QTVDEALKDA QTRITKGANW RACEAAKKTK ADDRGTDEKT
SEQTMPEQKL ISEEDLAMEF LEIDPVDMEP VDPNLEPWKH PGSQPRTACN NCYCKKCCFH
CYACFTRKGL GISYGRKKRR QRRRAPQDSQ THQASLSKQP ASQSRGDPTG PTESKKKVER
ETETDPFDLS
```

Isotopically Averaged Molecular Weight = 60850.9727

Estimated pI = 6.60

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Estimated charge at pH 7.00 = -1.8

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is **not** valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Estimated charge over pH range

pH	Charge
4.00	57.5
4.50	36.6
5.00	18.1
5.50	8.0
6.00	3.4
6.50	0.5
7.00	-1.8
7.50	-3.6
8.00	-5.9
8.50	-9.9
9.00	-16.5
9.50	-28.6
10.00	-47.9

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

PROTEIN CALCULATOR v3.3

Chris Putnam
cdputnam@scripps.edu
The Scripps Research Institute
Last Updated: March 28, 2006

Appendix 3 - Cleaved hCycT1-Tat Chimera Sequence Details

```
GANWRACEAA KKTKADDRGT DEKTSEQTMP EQKLISEEDL AMEFLEIDPV DMEPVDPNLE  60
PWKHPGSQPR TACNNCYCKK CCFHCYACFT RKGLGISYGR KKRRQRRRAP QDSQTHQASL  120
SKQPASQSRG DPTGPTESKK KVERETETDP FDLS                             154
```

Isotopically Averaged Molecular Weight = 17481.4746

Estimated pI = 7.89

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is **not** valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Estimated charge at pH 7.00 = 2.4

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Estimated charge over pH range

pH	Charge
4.00	23.0
4.50	16.1
5.00	9.9
5.50	6.6
6.00	4.9
6.50	3.6
7.00	2.4
7.50	1.3
8.00	-0.5
8.50	-3.3
9.00	-6.6
9.50	-10.6
10.00	-15.9

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

PROTEIN CALCULATOR v3.3

Chris Putnam
cdputnam@scripps.edu
The Scripps Research Institute
Last Updated: March 28, 2006

Appendix 4 – Tat Minimal Peptide Sequence Details

```
RKKRRQRRRP PQG                                                          13
```

Isotopically Averaged Molecular Weight = 1719.0298

Estimated pI = 12.72

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Estimated charge at pH 7.00 = 7.9

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Estimated charge over pH range

pH	Charge
4.00	8.1
4.50	8.0
5.00	8.0
5.50	8.0
6.00	8.0
6.50	8.0
7.00	7.9
7.50	7.8
8.00	7.5
8.50	7.2
9.00	6.9
9.50	6.5
10.00	6.0

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

PROTEIN CALCULATOR v3.3

Chris Putnam
cdputnam@scripps.edu
The Scripps Research Institute
Last Updated: March 28, 2006

Appendix 5 - MBP (Maltose Binding Protein) Sequence

```
MKIKTGARIL ALSALTTMMF SASALAKIEE GKLVIWINGD KGYNGLAEVG KKFEKDTGIK
VTVEHPDKLE EKFPQVAATG DGPDIIFWAH DRFGGYAQSG LLAEITPDKA FQDKLYPFTW
DAVRYNGKLI AYPIAVEALS LIYNKDLLPN PPKTWEEIPA LDKELKAKGK SALMFNLQEP
YFTWPLIAAD GGYAFKYENG KYDIKDVGVD NAGAKAGLTF LVDLIKNKHM NADTDYSIAE
AAFNKGETAM TINGPWAWSN IDTSKVNYGV TVLPTFKGQP SKPFVGVLSA GINAASPNKE
LAKEFLENYL LTDEGLEAVN KDKPLGAVAL KSYEEELAKD PRIAATMENA QKGEIMPNIP
QMSAFWYAVR TAVINAASGR QTVDEALKDA QTRITK
```

Isotopically Averaged Molecular Weight = 43387.5195

Estimated pI = 5.71

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Estimated charge at pH 7.00 = -4.3

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

Estimated charge over pH range

pH	Charge
4.00	34.6
4.50	20.6
5.00	8.2
5.50	1.5
6.00	-1.5
6.50	-3.1
7.00	-4.3
7.50	-5.1
8.00	-5.9
8.50	-7.4
9.00	-10.8
9.50	-19.0
10.00	-33.1

WARNING: pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purifications. pKa values for the individual amino acids from Stryer Biochemistry, 3rd edition.

PROTEIN CALCULATOR v3.3

Chris Putnam
cdputnam@scripps.edu
The Scripps Research Institute
Last Updated: March 28, 2006

Appendix 6 - hCycT1-Tat Site-Directed Mutagenesis Protocol

1. Get template (plasmid with hCycT1-Tat from isolated from XL1)

Start 5 ml LB/Amp liquid culture of XL1-Blue cells containing the plasmids

Do plasmid miniprep using QIAprep Spin Miniprep kit. Elute plasmid with 50ul of H_2O.

A 16-hr 5-ml XL1-Blue culture typically gives a plasmid yield of 7.5ug (~150ng/ul).

2. Site-directed mutagenesis (Day 1)

Material:

- Stratagene QuikChangeII Site-Directed Mutagenesis Kit (Stored @ -20°C)
- Template: dilute to 10ng/μl
- Primers: dilute to 10μM (100~150ng/μl)

Primer design: Use QuikChange Primer Design Tool:

https://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Tool&SubPageType=ToolQCPD&PageID=15

PCR Mix:

5 μl of 10× reaction buffer

10 μl of dsDNA template (1 ng/ul)

1.25 μl of forward primer (100 ng/ul)

1.25 μl of reverse primer (100 ng/ul)

1.0 μl of 10mM dNTP mix

1.0 μl of PfuUltra polymerase (2.5 U/μl)

30.5 μl of nuclease free H_2O

final volume: 50 μl

PCR Cycle:

1) 95°C 30s

2) 95°C 30s

3) 55°C 1min

4) 68°C 5 min

5) go to 2) for a total of 18 cycles

6) hold at 4°C

Note: Make a negative control with no template (add 0.5 µl of nuclease free H_2O instead of template).

3. DpnI digestion (Day 1)

Add DpnI 1.0 µl into each PCR reaction

Incubate @ 37°C for 1hr

4. Check product size on agarose gel (Day 1)

1% gel, 160V, 40min.

Use 5 µl of each sample/control per well, 2.5 µl of 1kb DNA ladder (Promega Cat# G5711).

Note: Do not proceed to next step until you see successful amplification (single band of size ~ 5kb).

5. Transform into XL-1 Blue Competent Cells (Day 1)

Material:

- XL-1 Blue competent cells(Stratagene), stored @ -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 won't be accurate because of dNTPs and free primers.) If you loaded 2.5 µl of 1kb DNA ladder onto your gel earlier, the four brighter bands are 30ng each, other bands are 10ng each. Compare your sample band to the ladder and make your best guess.

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. 1 µl of nuclease free H_2O as negative control.

3) Thaw a tube of XL-1 Blue competent cells on ice. Add 20 µl of competent cells into each tube. If you have competent cells left, make 20 µl aliquots of them and freeze them on dry ice immediately. These aliquots should be good to use for next time, but their transformation efficiency will be lower.

4) Incubate the mixture on ice for 30min. In the meantime, warm up a tube of LB+0.4% glucose @ 42°C.

5) Put the tubes into a thermal cycler. Heat shock @ 42°C for 45s.

6) Immediately put tubes back on ice, incubate for 2min.

7) Add 100 µl of pre-warmed LB+0.4% glucose (42°C), gently mix.

8) Pool the tubes in a falcon tube, shake @ 250rpm in a 37°C incubator for 1hr. Warm up a few LB/AMP plates during this time.

9) Spread 50 µl of the recovered cells onto each LB/AMP plate. Grow @ 37°C overnight.

6. Pick single colony from plate (Day 2)

Pick single colony from plate, inoculate a 5ml LB/AMP liquid culture, grow @ 37°C overnight, shaking @ 220~250rpm.

7. Make glycerol frozen stock and Miniprep (Day 3)

Make permanent glycerol storage culture by adding 350 µl of overnight LB/AMP culture into 150 µl of 50% glycerol (final glycerol concentration 15%). Freeze immediately on dry ice. Store @ -80°C.

Do plasmid miniprep using QIAprep Spin Miniprep kit. Elute plasmid with 50ul of H_2O.

Measure UV using Nanodrop, calculate concentration.

Take 200~800ng of each plasmid miniprep as sequencing sample. Store the rest @ -20°C or -80°C.

8. Sanger Sequencing

Sequencing facility at Upstate Medical University. University Hospital Room 4840

Contact: Vicki Lyle lylev@upstate.edu

Sequencing primers for pGEX-2TK plasmid)

Sample dropped off every day by 11AM will be sequenced on the same day. Results usually get sent via email the second day.

9. Check sequence for mutation (Day 4)

Use BLAST or ClustalW to check mutated sequence against wildtype.

  BLAST: http://blast.ncbi.nlm.nih.gov/ (use bl2seq)

  ClustalW: http://www.ebi.ac.uk/Tools/clustalw2/index.html

Note: Do not proceed to next step until you confirm the plasmid has the right mutation and nothing else.

Appendix 7 - Invitrogen Protocol for Gateway® Reactions

- **BP Reaction**

  Creating a Gateway® entry clone from an *att*B-flanked PCR product is an easy 1

  hour reaction. See below for an overview of the set-up. For more detailed

  information, refer to the manual.

  1. Add the following components to a 1.5 ml tube at room temperature and

     mix:

     attB-PCR product (=10 ng/µl; final amount ~15-150 ng) 1-7 µl

     Donor vector (150 ng/µl) 1 µl

     TE buffer, pH 8.0 to 8 µl

  2. Thaw on ice the BP Clonase™ II enzyme mix for about 2 minutes. Vortex

     the BP Clonase™ II enzyme mix briefly twice (2 seconds each time).

  3. To each sample (Step 1, above), add 2 µl of BP Clonase™ II enzyme mix

     to the reaction and mix well by vortexing briefly twice. Microcentrifuge

     briefly.

  4. Return BP Clonase™ II enzyme mix to -20°C or -80°C storage.

  5. Incubate reactions at 25°C for 1 hour.

  6. Add 1 µl of the Proteinase K solution to each sample to terminate the

     reaction. Vortex briefly. Incubate samples at 37°C for 10 minutes.

Transformation

7. Transform 1 µl of each BP reaction into 50 µl of One Shot ® OmniMAX ™ 2 T1 Phage-Resistant Cells (Catalog no. C8540-03). Incubate on ice for 30 minutes. Heat-shock cells by incubating at 42°C for 30 seconds. Add 250 µl of S.O.C. Medium and incubate at 37°C for 1 hour with shaking. Plate 20 µl and 100 µl of each transformation onto selective plates. Note: Any competent cells with a transformation efficiency of $>1.0 \times 10 8$ transformants/µg may be used.

8. Transform 1 µl of pUC19 DNA (10 ng/ml) into 50 µl of One Shot ® OmniMAX ™ 2 T1 Phage-Resistant Cells as described above. Plate 20 µl and 100 µl on LB plates containing 100 µg/ml kanamycin, or the appropriate selection marker for your donor vector.

Expected Results

An efficient BP recombination reaction will produce >1500 colonies if the entire BP reaction is transformed and plated.

- **LR Reaction**

Transferring your gene from a Gateway® entry clone to destination vector is an easy 1 hour reaction. See below for an overview of the set-up. For more detailed information, refer to the manual.

1. Add the following components to a 1.5 ml tube at room temperature and mix:

   Entry clone (50-150 ng) 1-7 µl

   Destination vector (150 ng/µl) 1 µl

   TE buffer, pH 8.0 to 8 µl

2. Thaw on ice the LR Clonase ™ II enzyme mix for about 2 minutes. Vortex the LR Clonase ™ II enzyme mix briefly twice (2 seconds each time).

3. To each sample (Step 1, above), add 2 µl of LR Clonase ™II enzyme mix to the reaction and mix well by vortexing briefly twice. Microcentrifuge briefly.

4. Return LR Clonase ™ II enzyme mix to -20°C or -80°C storage.

5. Incubate reactions at 25°C for 1 hour.

6. Add 1 µl of the Proteinase K solution to each sample to terminate the reaction. Vortex briefly. Incubate samples at 37°C for 10 minutes.

Transformation

Follow the protocol as indicated for the BP reaction, except use the appropriate selection marker for the LB plates suited to your destination vector (typically 100 µg/ml ampicillin).

Expected Results

An efficient LR recombination reaction will produce >5000 colonies if the entire

LR reaction is transformed and plated.

One Tube Format

If you want to transfer your attB-flanked PCR product directly into an expression clone,

you can easily combine the BP and LR reactions using the following protocol. This will

potentially eliminate the transformation and DNA isolation of the Gateway® entry clone.

1. In a 1.5 ml microcentrifuge tube, prepare the following 15 μl BP reaction:

   attB DNA (50-100 ng) 1.0-5.0 μl

   attP DNA (pDONR™ vector, 150 ng/μl) 1.3 μl

   BP Clonase™ II enzyme mix 3.0 μl

   TE Buffer, pH 8.0 add to a final volume of 15 μl

2. Mix well by vortexing briefly and incubate at 25°C for 4 hours.

   Note: Depending on your needs, the length of the recombination reaction can be

   extended up to 20 hours. An overnight incubation typically yields 5 times more

   colonies than a 1 hour incubation. Longer incubation times are recommended for

   large plasmids (=10 kb) and PCR products (=5 kb).

3. Remove 5 μl of the reaction to a separate tube and use this aliquot to assess the

   efficiency of the BP reaction (see below).

4. To the remaining 10 µl reaction, add:

    Destination vector (150 ng/µl) 2.0 µl

    LR Clonase™ II enzyme mix 3.0 µl

    Final volume 15 µl

5. Mix well by vortexing briefly and incubate at 25°C for 2 hours.

    Note: Depending on your needs, the length of the recombination reaction can be

    extended up to 18 hours.

6. Add 2 µl of proteinase K solution. Incubate at 37°C for 10 minutes.

7. Transform 50 µl of the appropriate competent E. coli with 1 µl of the reaction.

8. Plate on LB plates containing the appropriate antibiotic to select for expression

    clones.

Assessing the Efficiency of the BP Reaction

1. To the 5µl aliquot obtained from "One-Tube" Protocol, Step 3, above, add 0.5 µl

    of proteinase K solution. Incubate at 37°C for 10 minutes.

2. Transform 50 µl of the appropriate competent E. coli with 1 µl of the reaction.

    Plate on LB plates containing the appropriate antibiotic to select for entry clones.

Appendix 8 - pRK793 Plasmid Map for TEV Protease S219V Mutant
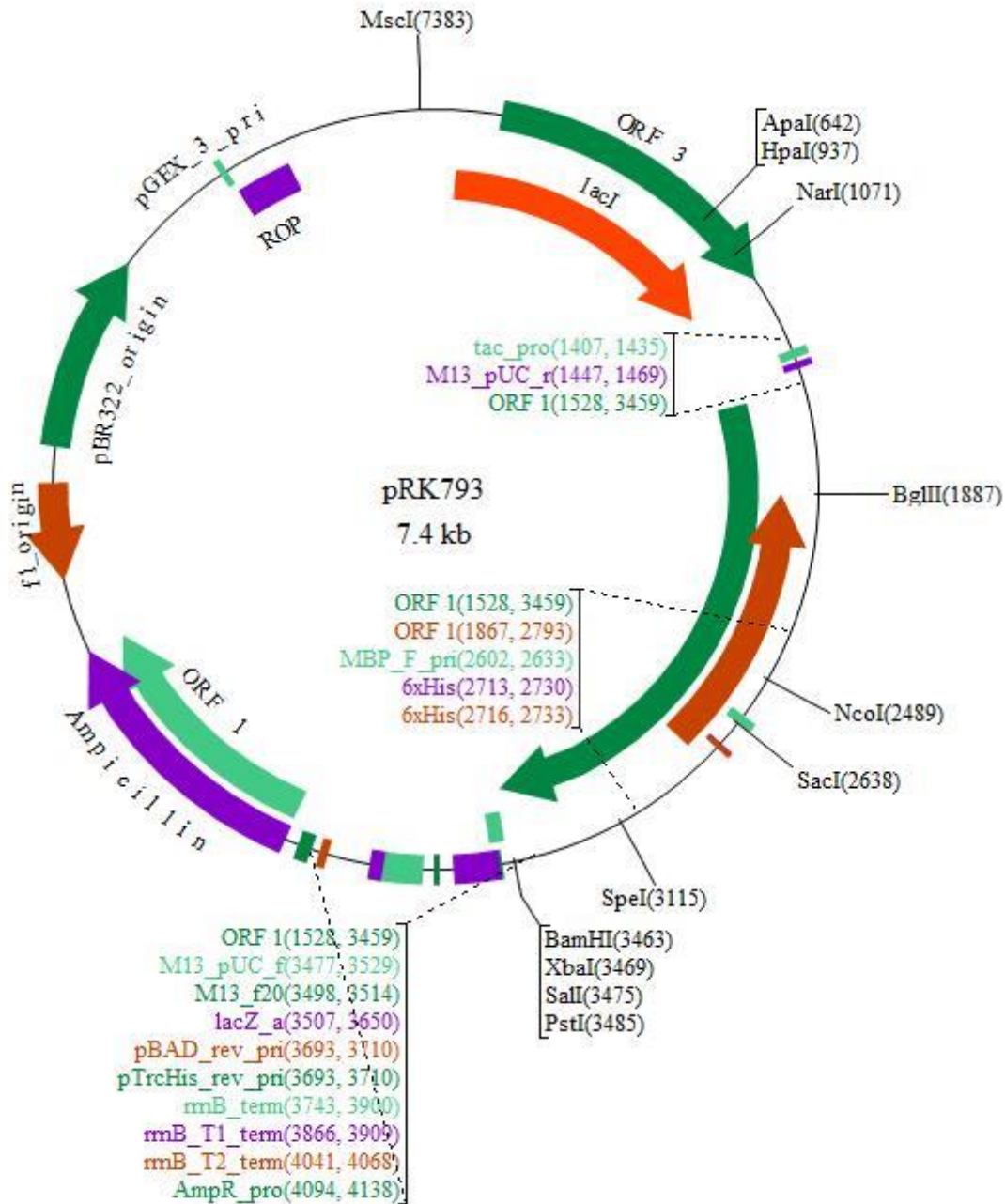
Map of the pRK793 plasmid for recombinant expression of the Tobacco Etch Virus (TEV) protease S219V mutant, Addgene #8827, deposited by Principal Investigator David Waugh, National Cancer Institute, Frederick, MD.

Appendix 9 - M9 Minimal Media Recipe

M9 Minimal Media 1L

$Na_2HPO_4 \cdot 7H_2O$	12.8g

Or

Na_2HPO_4 (anhydrous)	6 g
KH_2PO_4	3 g
NaCl	0.5 g
NH_4Cl	1 g
Carbon Source	0.2% (v/v)

(glucose, sodium gluconate, or glycerol)

Adjust pH to 7.4 with NaOH

Autoclave prior to adding sterile micronutrients:

Stock [C]	Micronutrient	[C] Final
1 M	$MgSO_4$	1 mM
1 M	$CaCl_2$	100 μM
3 mM	$(NH_4)_6Mo_7O_{24} \cdot 4H_2O$	3×10^{-9} M
400 mM	H_3BO_3	4×10^{-7} M
30 mM	$CoCl_2 \cdot 6H_2O$	3×10^{-8} M
10 mM	$CuSO_4 \cdot 5H_2O$	1×10^{-8} M
80 mM	$MnCl_2 \cdot 4H_2O$	8×10^{-8} M
10 mM	$ZnSO_4 \cdot 7H_2O$	1×10^{-8} M
5 mM	$FeSO_4 \cdot 7H_2O$§	1×10^{-6} M

Filter sterilize all micronutrients. Make $FeSO_4$ fresh immediately prior to use

Appendix 10 - Protocol for Expression of the GST Tat Chimera K. Fujinaga.

(handwritten annotations at top) Amp only

GST-PTEF Purification Protocol
(12/97)

1. Inoculate 10 ml of a 50 ml O/N culture into 1L of LB. Expression is in *E. coli* strain BL21p(lysS). *(handwritten: 1 ml Chlorophenicol 0.5ml Amp.)*

2. Grow at 37 °C until A600–0.6, induce w/ 1 mL 1 M IPTG per 1 L culture (1 mM IPTG final).

3. Induce for 4 hr at 30 °C. *(handwritten: keep induced + uninduced 100 µl)*

4. Collect cells by centrifugation at 2,500 rpm for 25 min.

5. Wash cells in 100 mL cold PBS (no Ca^{+2}, no Mg^{+2}) per L culture. Collect cells by centrifugation. *(handwritten: 8000 rpm in Beckman - tube containers)*

6. Resuspend the cells in 15 ml lysis buffer (see below). *(handwritten: add lysozyme.)*

7. ~~Freeze/thaw 2X to aid in cell lysis.~~ (Use dry ice and water). Can store cell lysates at -80 °C at this point if needed. *(handwritten: ~2 g/20 ml)*

8. ~~Thaw tubes quickly in water.~~

9. Sonicate lysate 2-3X until lysate is no longer viscous (keep tubes on ice to keep cold).

10. Spin in pre-cooled ultracentrifuge tubes (Vti 50.2) at 35,000 rpm (100,000g) for 30 min. (4°C)

11. Remove supernatent and keep supernatent. *(handwritten: keep sample of sup.)*

12. Resuspend the pellet in 8 ml of lysis buffer + 2% sarkosyl. Incubate on the nutator at 4°C for one hour.

13. Spin in pre-cooled ultracentrifuge tubes (Vti 50.2) at 35,000 rpm (100,000g) for 30 min. (4°C).

14. ~~Combine supernatent with the supernatent from step # 11.~~ *(handwritten: combine / don't combine sup)*

15. Add 500 µl packed Glutathione Sepharose beads (Pharmacia) which were pre-equilibrated with lysis buffer (no Sarkosyl). Incubate on Nutator for 2 hr @ 4 °C.

16. Collect beads by centrifugation (1,200 rpm, 2 min.) and wash 3x20 mL lysis buffer + 0.075 % SDS.

- 243 -

17. Do a final wash with PBS (no Ca^{+2}, no Mg^{+2}) containing 10 mM DTT.

18. Aliquot and store at - 80 ^oC.

Lysis Buffer: 50 ml

50 mM Hepes-KOH ph7.8 ....5 ml (0.5 M)
100 mM KCl (MW 75)...............2.5 ml (2M)
Triton X-100.....................5 ml (10%)
2 mM EDTA........................200 µl (0.5M)
0.1 mM PMSF.....................25 µl (0.2M; 0.35g/10 ml)
Benzamidine......................0.1 g
0.1 mM NaF........................500 µl (0.5M; 0.2g/10ml)
0.1 mM NaVO4.....................50 µl (0.1M; 0.183g/10ml)
Aprotinin........................100 µl (10 mg/ml)
Leupeptin........................10 µl (5 mg/ml)
1 mM Nametabi....................50 µl (1 M)

WATER............................33.4 ml

Dialysis Buffer:

            2000 ml

25 mM Hepes(MW 238)................ 100 ml (0.5M)
20% glycerol .... 500 g
100 mM KCl .... 14.9 g of (MW 74.5)
0.1mM EDTA ....400 µl of 0.5 M
Just before use add:-
10mM DTT..........................3.08g/2L
1mM NaMetabis....................0.4g /2L

Appendix 11 - Tat Chimera Expression Protocol

HISMBP-TEV hCycT1/Tat Chimera in Rosetta Gami B Cells

non-DE3 non-p(lysS) strain

TRIS Buffer

1. Inoculate a 250 ml starter culture of LB media with RGB Mut 5 with 100 ug/ml ampicillin and 35 ug/ml chloramphenicol and grow overnight at 37°C

2. Autoclave 1 L baffle bottom flasks of Turbo media

3. Inoculate 150 ml of overnight culture into 1 L flask of Turbo media with 1.0 ml of [100mg/ml] ampicillin.

4. Grow at 37°C until A_{600} = 0.6-0.7 (~3 hours)

5. At 20 min prior to induction add ethanol to Turbo media to final concentration of 1% for expression of heat shock and chaperone proteins. Add 10 ml of EtOH per L culture.

6. Take 1 mL uninduced sample (for running gel later) spin down and freeze. Resuspend in 200 ul 1 X SDS loading buffer

7. Induce culture with 1.0 ml 1M IPTG per 1 L to final concentration of 1.0 mM.

8. Incubate overnight shaking at 200-240 rpm at room temp - 28°C.

9. Take 1 mL induced sample (for running gel later); spin down and freeze. Resuspend in 100 ul 1 x SDS

10. Collect cells by centrifugation at 2,500 rpm for 25 min.

11. Weight pellets determine weight of cell paste.

12. Pellet can be frozen at this point.

13. Resuspend the cells (in lysis buffer) <u>completely</u> with no clumps by pipetting in ~46 ml per L cell culture.

14. Incubate on ice for 15-20 min.

15. Optional: Add 4 ml BPER per L and incubate additional 10 min on ice

16. Sonicate (<u>without inducing foaming</u>) at moderate level in burst of 25 sec followed by 25 sec on ice four times or until solution changes color and transparency.

17. Precipitate nucleic acids with 2 ml 4% PEI /L (polyethyleneimine) (per Liter of culture). Add and mix well by shaking or vortexing lightly. This will produce a visible white precipitate. (Additional guideline is 20-80 ul of 5% per ml crude).

18. Spin in pre-cooled ultracentrifuge tubes (Vti 50.2) at <u>21,000 rpm</u> for 30 min at 4ºC

19. Remove supernatant and keep supernatant.

20. Syringe filter supernatant.

21. Take a 20 uL sample, add 20 uL 5x SDS sample buffer, and 60 ul dH20; vortex heat at 90℃ for 10 min. vortex and reserve. As <u>soluble extract sample</u>.

22. Resuspend pellet in 5 ml Insoluble Extraction Solution for Insoluble and prepare sample as per supernatant above. This is <u>insoluble extract</u> gel sample.

Lysis Buffer pH 7.4 (s/b 7.5) 100 ml

Tris 20 mM (MW 121.14)   0.242 g

NaCl (MW 58.44) 0.500 M   2.92 g

$ZnCl_2$ (MW 136.31) 10 uM   0.2 ml of 5 mM

Imidazole 20 mM (68.077 g/mol) .136 g

BME 5 mM

Ethylene Glycol 20 %   20.0 ml

HALT protease inhibitor (EDTA free) (10 ul/ml) 1.0 ml

Arginine (MW 174.2) 50 mM   0.87 g

Lysozyme (Cf ~150 ug/ml in H20)

Insoluble Extract Solution (pH 8.0) 20 ml

Urea (mw 60.06) 8 M   9.61 grams

Na_2HPO_4 (mw 141.96) 100 mM   283.9 mg

Tris-Cl (121.14) 10 mM   24.2 mg

DTT 10 mM

Appendix 12 – Tat Chimera FPLC Purificiation Protocol with TEV Cleavage

Full Length MBP Tagged Protein Recovery

- Turn on UV and warm up 15 min

- Pump wash A and B with dH2O

- Attach 2 - 5 ml HisTrap column columns in series.

- Wash with dH2O 1 ml/min.

- Equilibrate column with Tris binding buffer

- Clean superloop by removing it from the system and disassembling it. Be sure to refill the loop and push plunger to top before re-assembling.

- Apply Sample: Load supernatant in lysis buffer into superloop and then inject at 0.2 ml/min (2x5 ml column) during sample application switch to binding buffer. (Make sure to stop the flow after injecting and returning to load)

- Wash with 5 column volumes of binding buffer to remove unbound proteins until no material appears in the effluent at flow rate of 0.4-0.6 ml/min (2x5 ml column)

- Gradient elution with elution buffer (IMAC-B) for 30 min to 100% elution buffer

- Analyze fractions by SDS PAGE and/or Nanodrop and combine purest fractions for cleavage.

- Perform a buffer exchange on the full chimera back into binding buffer

<u>Pro-TEV Cleavage of HIS MBP Tag</u>

1. Use a ratio of 1/50 grams of TEV per gram of protein to be cleaved.

2. Incubate the cleavage reaction overnight at ~14 ºC

CLEAN THE COLUMN AS PER DIRECTIONS BELOW BEFORE APPLYING

CLEAVAGE REACTION BACK ONTO COLUMN

<u>Re-Application of Cleaved Protein to HisTrap Column for Removal of TEV and MBP</u>

<u>Tag</u>

1. Apply the cleaved supernatant in binding buffer to the superloop and inject onto column at a flow rate of 0.2 to 0.4 ml

2. Collect Fractions after 7 column volumes of buffer have passed.

3. Determine concentration with Nanodrop

4. Analyze the fractions by SDS PAGE

5. Store Fractions in 20% glycerol at -80 ºC

Solutions:

Binding Buffer pH 7.4 1 L (1 X)

Tris 20 mM (MW 121.14)	2.42 g
NaCl 0.500 M (MW 58.44 g/mol)	29.22 g
$ZnCl_2$ (MW 136.31) 10 uM	2.0 ml of 5 mM
Imidazole 20 mM (68.077 g/mol)	1.36 g
Sodium Azide 0.02%	0.20 g
BME 5 mM	
Glycerol 5%	50 ml

Wash Buffer pH 7.4 1 L (1 X)

Tris 20 mM (MW 121.14)	2.42 g
NaCl 0.500 M (MW 58.44 g/mol)	29.22 g
$ZnCl_2$ (MW 136.31) 10 uM	2.0 ml of 5 mM
Imidazole 60 mM (68.077 g/mol)	4.08 g
Sodium Azide 0.02%	0.20 g
BME 5 mM	
Glycerol 5%	50 ml

Elution Buffer pH 7.4 1 L 1 X

Tris 20 mM (MW 121.14)	2.42 g
NaCl 0.500 M (MW 58.44 g/mol)	29.22 g
$ZnCl_2$ (MW 136.31) 1 uM	2.0 ml of 5 mM
Imidazole 500 mM (68.077 g/mol)	34.0 g
Sodium Azide 0.02%	0.20 g
BME 5 mM	
Glycerol 5%	50 ml

Short Term Storage Buffer pH 7.4 1 L (1 X)

Tris 20 mM (MW 121.14)	2.42 g
Glycerol 5%	50 ml
NaCl 0.200 M (58.44 g/mol)	11.68 g
$ZnCl_2$ (MW 136.31) 10 uM	2.0 ml of 5 mM
Sodium Azide 0.02%	0.20 g
BME 5 mM	

Octet Buffer pH 7.4 1 L (1 X)

Tris 20 mM (MW 121.14)	2.42 g

NaCl 0.200 M (58.44 g/mol)   11.68 g

ZnCl$_2$ (MW 136.31) 10 uM   2.0 ml of 5 mM

Sodium Azide 0.02%   0.20 g

BME 5 mM

Column Maintenance:

Stripping and Recharging should be performed after 5 to 7 purifications. It should not be

necessary to perform this procedure after batch purification of the same protein

Stripping Buffer (GE) pH 7.4 100 ml

Sodium Phosphate   20 mM

Monosodium Phosphate, monohydrate   0.0623%   0623 g

Disodium Phosphate, heptahydrate   0.4149%   0.415 g

NaCl (MW 58.44) 0.5 M   2.922 g

EDTA 50 mM

1. Wash with 5 to 10 volumes of stripping buffer

2. Wash with 5 to 10 volumes of binding buffer

3. Wash with 5 to 10 volumes of dH$_2$O

4. Recharge column with 0.5 ml or 2.5 ml of 0.1 M $NiSO_4$ in dH_2O on on HisTrap HP 1 ml and 5 ml column respectively (other salts may also be used).

5. Wash with 5 column volumes of dH_2O

6. Wash with 5 column volumes of binding buffer (to adjust pH before storage)

Column can be cleaned or stored at this point

<u>Cleaning</u>

1. Remove ionically bound proteins by washing the column with several column volumes of 1.5 M NaCl; then wash the column with approximately 10 column volumes of dH_2O.

2. Remove precipitated proteins, hydrophobically bound proteins, and lipoproteins by washing the column with 1M NaOH, contact time usually 1-2 hours (12 hours or more for endotoxin removal). Then wash the column with approximately 10 column volumes of binding buffer, followed by 5-10 column volumes of dH_2O.

3. Remove hydrophobically bound proteins, lipoproteins, and lipids by washing the column with 5-10 column volumes 30% isopropanol for about 15-20 minutes. Then wash the column with approximately 10 column volumes of dH_2O.

   Alternatively, wash the column with 2 column volumes of detergent in a basic or acidic solution. Use, for example 0.1-0.5% nonionic detergent in 0.1 M acetic acid, contact time 1-2 hours. After treatment, always remove residual detergent by washing with at least 5 column volumes of 70% ethanol. Then wash the column with approximately 10 column volumes of dH_2O.

<u>Storage</u>

Ethanol 20%

Appendix 13 - Expression and Purification of MBP-His$_6$-TEV (S219V)-Arg$_5$

Plasmid: pRK793

Extinction Coefficient: 32,290 M^{-1} cm^{-1}

Adapted from (134):

Expression:

1. **<u>Inoculate 50-150 ml of LB</u>** broth containing 100 ug/ml ampicillin and 30 ug/ml

   chloramphenicol in a 500 ml bafflebottom shake flask from a glycerol stock of

   pRK793 transformed *E. coli* BL21 (DE3) CodonPlus-RIL cells (here I will

   substitute **Rosetta 2** (non-DE3 as per Novagen: do not produce T7 polymerase,

   provide more tRNA and possess OmpT and Lon mutations eliminating proteases).

   Place in an incubator and shake overnight at 250 rpm and 37ºC.

2. **<u>Add 25 ml of the saturated overnight culture to each 1 L</u>** of fresh LB broth

   containing 100 ug/ml ampicillin, 30 ug/ml chloramphenicol and **0.2% glucose**

   (glucose should be added separately and not autoclaved in LB because it will

   caramelize). Glucose (a/k/a dextrose) should be mixed slowly into a water

   solution to prevent clumping. Filter sterilize the glucose solution before adding it

   to LB. To ensure that there will be an adequate yield of pure protein at the end of

   the process, Tropea et al. grow 4-6 L of cells at a time. (Here we attempt 3

   separate 1 L cultures).

3. Shake the flasks at 250 rpm and 37ºC until the cells reach mid-log phase (OD$_{600nm}$

   ~0.5); approximately 2 h.

4. Take a 1 ml uninduced sample, spin down, and resuspend in 100 ul 1 x SDS loading buffer.

   Heat sample at 95 ºC for 10 minutes, then spin down at 14,000 rpm for 5 minutes.

5. **<u>Shift the temperature to 30ºC</u>** and induce the culture(s) with IPTG at a final concentration of 1 mM (5 ml of 200 mM IPTG stock solution per liter of culture). Continue shaking at 250 rpm for 4-6 h. Place cultures at 4ºC.

6. Take a 1 ml induced sample, spin down, and resuspend in 100 ul 1 x SDS loading buffer.

7. Recover the cells by **<u>centrifugation at 5,000 x g for 15 min at 4 ºC</u>**, and store at -80ºC. A 6 L preparation yields 30-40 g of cell paste.

Purification:

His$_6$-TEV(S219V)-Arg$_5$ protease can be purified to homogeneity in two steps: immobilized metal affinity chromatography (IMAC) using Ni-NTA resin followed by size exclusion chromatography.

1. All procedures are performed at 4-8ºC. Thaw the cell paste from 6 L of culture on ice and suspend in ice-cold cell lysis/IMAC equilibration buffer (10 ml/g cell paste).

2 L Cell Lysis/IMAC Equilibration buffer pH 8.0:

Sodium Phosphate Dibasic 50 mM   14.2 g

NaCl 200 mM   23.38 g

Glycerol 10 %   200 ml

Imidazole 25 mM   3.41 g

Add H2O to 1980 ml

Adjust pH with HCl

Adjust Volume to 2 L with H2O

Re-check pH

Filter through 0.22 um polyethersulfone membrane or equivalent

Store at 4ºC.

2. Lyse the cell suspension and measure the volume using a graduated cylinder. Add polyethylenimine (PEI) to a final concentration of 0.1% (a 1:50 dilution of the 5% stock solution at pH 8) and mix gently by inversion. Immediately centrifuge at 15,000 x g for 30 min and filter.

3. Apply the supernatant to two tandem 5 ml HisTrap HP columns equilibrated in cell lysis/IMAC equilibration buffer. Wash the columns with equilibration buffer until a stabile baseline is reached and then elute the bound His_6-TEV(S219V)-Arg_5 with a linear gradient to 100% elution buffer over ten column volumes.

Note: It may be helpful to split up the supernatant into several batches to accommodate the 10 ml column volume. Also, make sure to sample flow through to monitor efficiency, and run flow through over column a second time (after eluting from it) if necessary to recover unbound TEV.

1 L IMAC Elution Buffer pH 8.0:

Sodium phosphate dibasic 50 mM	7.1 g
NaCl 200 mM	11.69 g
Glycerol 10%	100 ml
Imidazole 250 mM	17.02 g

Adjust volume to 750 ml with H_2O

Adjust pH to 8.0 using HCl.

Adjust volume to 1 L with H_2O

Let cool to room temperature

Re-check pH.

Filter through 0.22 um polyethersulfone membrane or equivalent.

Store at 4ºC.

4. Pool the peak fractions containing the protease and measure the volume. Add EDTA to a final concentration of 2 mM (a 1:250 dilution of the 0.5 M EDTA, pH 8 stock solution) and mix well. Add DTT to a final concentration of 5 mM (1:200 dilution of the 1M DTT stock solution) and mix well.

5. Concentrate the sample approximately tenfold using an Amicon ultrafiltration YM10 membrane. Remove the precipitation by centrifugation at 5,000 x g for 10 min. Estimate the concentration of the partially pure protein solution spectrophotometrically at 280 nm using a molar extinction coefficient of 32,290 M^{-1} cm^{-1}. The desired concentration is between 5 and 10 mg/ml.

6. Filter through a 0.2 uM syringe filter, aliquot and flash freeze with liquid nitrogen. **Store at -80ºC**.

Appendix 14 - Mass Spectrometry Sample Preparation Protocol

Solutions required for MALDI-TOF:

Acetonitrile

Trifluoroacetic Acid 0.1%

Sinapinic Acid 10 mg

TA solution – 1 Part Acetonitrile to 2 parts Trifluoroacetic Acid 0.1%

1. Prepare a 1.5 ml solution of 0.1% aqueous Trifluoroacetic acid from the ampules (located in the crisper of the fridge in 215) by adding 1.5 ul of TFA to a total of 1.5 ml in H_2O

2. Prepare a solution with 10 mg of Sinapinic Acid in 1 ml of TA solution

3. Allow the solution to dissolve for ~ 30 min then spin down the undissolved SA briefly

4. Dilute the protein samples for your experiment to ~ 20 uM using TFA 0.1% (if possible the target sample should be dialyzed into H_2O PRIOR to the dilution with TFA. However, due to precipitation Tris buffer was used in lieu of $H_2 0$ PRIOR to dilution. Do not use PBS with MALDI-TOF.)

5. Mix the protein sample 1:1 with TA solution (ex.: 10 ul of ~20 uM sample + 10 ul TA solution)

6. Spot 1 ul of each sample on to MALDI Sample Plate

Appendix 15 - Western Blot Protocol

For use with the WesternBreeze® Chromogenic Immunodetection Kit

Anti-myc FITC antibody

1. Run SDS page gel with appropriate concentrations of protein, and 10 ul of the Positope Control Protein in a control .

2. Take gel to iBlot with bottom, top, filter, and sponge.

3. Set up iBlot and run for 7-8 min.

4. Remove the gel and discard all but the membrane

5. Place membrane in a plastic container, cover, and add 10 ml of Invitrogen NC Blocking Solution.

`

   Blocking Solution:

   Ultra filtered Water 14 ml

   Blocker/Diluent (Part A) 4 ml

   Blocker/Diluent (Part B) 2 ml

   Total Volume   20 ml

6. Rock the membrane slowly for 30 min (~1 rev/sec).

7. Decant Blocking Solution

8. Rinse membrane with 20 ml dH2O and incubate in dH2O for 5 min; decant and repeat.

9. Add 10 ml of Primary Antibody Solution with 2 ul of FITC Myc Antibody (1:5000 dilution). For nitrocellulose membranes the primary antibody solution is the blocking solution above. Use remaining 10 ml. Incubate for one hour.

10. Wash membrane in 20 ml of Antibody Wash (**Anti-Body Wash provided in kit is 16 X be sure to dilute before use**) for 5 minutes; decant and repeat 3 times.

11. Add 10 ml of Secondary Antibody Solution incubate for 30 min then decant.

12. Wash membrane in 20 ml of Antibody Wash for 5 minutes; decant and repeat 3 times.

13. Wash the membrane with 20 ml dH2O for 2 minutes, decant and repeat 2 times.

14. Add 5 ml of Chromogenic Substrate and incubate until purple bands appear on the membrane (1-60 minutes).

15. Wash the membrane with 20 ml dH2O for 2 minutes, decant and repeat 2 times.

16. Allow the membrane to dry on filter paper in the open air.

Appendix 16 – EMSA Sample Preparation

PAGE	Non-Denaturing	Tat minimal: 1.7 kDa
Date:	6/9/2012	hCycT1-Tat: 17.481 kDa
Stain:	SYBR Gold RNA SYPRO Ruby Protein	Full Length Chimera 60.8 kDa
Pre-Run	100 V > 1 hr	Ncp7 ~7 kDa
Gel %:	Native PAGE 12%	SNAP COOL RNA
Voltage:	100 V	85 ºC 5 min then Ice 5 min
Run Time	90 min	C initial
	Spin down 2 sec then Incubate Samples 4 ºC for 15	V initial
Running Buffer:	1 X TBE	
Sample Buffer:	Tris 20 mM Binding Buffer 5 mM BME 20 mM Imidazole 10 uM ZnCl2 5% Glycerol	
Stock Protein Date:		**Date of Gel:**

Setup:	Component	Volume (ul)
Lane 1	**SL3 Negative Control 500 nM**	20
Lane 2	**SL3-Ncp7 Positive Control 500 nM/10 uM**	20
Lane 3	**TAR WT 500 nM**	20
Lane 4	**TAR WT 500 nM Positive Control Tat Minimal 10 uM**	20
Lane 5	**TAR WT 500 nM hCyctT1-Tat no tag 10 uM**	20
Lane 6	**TAR WT 500 nM Full Length MBP tagged chimera 10 uM**	20
Lane 7	**TAR WT 500 nM MBP 10 uM**	20
Lane 8	**TAR WT 500 nM Positive Control Tat Minimal 5 uM**	20
Lane 9	**TAR WT 500 nM hCyctT1-Tat no tag 5 uM**	20
Lane 10	**TAR WT 500 nM Full Length MBP tagged chimera 5 uM**	20
Lane 11	**TAR WT 500 nM MBP 5 uM**	20
Lane 12	**Tat Minimal 10 uM**	20
Lane 13	**hCycT1 Tat no tag 10 uM**	20
Lane 14	**Full Length His MBP tagged chimera 10 uM**	20
Lane 15	**MBP 10 uM**	20

Sample/Lane # 1	SL3 Negative Control 500 nM				
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)	Ck
RNA	**SL3**	4.00E-06	2.5	5.00E-07	
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10x	2.0	1x	
Glycerol	50%		2.0		
dH2O			11.5		
Gel Pilot loading dye			1.0		
BME		0.1	1.0		
		Total Volume	20.0		

Sample/Lane # 2	SL3-Ncp7 Positive Control 500 nM/10 uM				
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)	Ck
RNA	**SL3**	4.00E-06	2.5	5.00E-07	
Protein	**Ncp7**	5.50E-05	3.6	1.00E-05	
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10x	2.0	1x	
Glycerol	50%		2.0		
dH2O			8.9		
BME		0.1	1.0		
		Total Volume	20.0		

Sample/Lane # 3	TAR WT 500 nM				
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)	Ck
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07	
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x	
Glycerol	50%		2.0		
dH2O			12.5		
BME		0.1	1.0		
		Total Volume	20.0		

Sample/Lane # 4	TAR WT 500 nM Positive Control Tat Minimal 10 uM				
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)	Ck
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07	
Protein	**Tat Minimal**	3.00E-05	6.7	1.00E-05	
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x	
Glycerol	50%		2.0		
dH2O			5.8		
BME		0.1	1.0		
		Total Volume	20.0		

Sample/Lane # 5	TAR WT 500 nM hCyctT1-Tat no tag 10 uM				
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)	Ck
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07	
Protein	**hCycT1-Tat no tag**	3.00E-05	6.7	1.00E-05	
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x	
Glycerol	50%		2.0		
dH2O			5.8		
BME		0.1	1.0		
		Total Volume	20.0		

Sample/Lane # 6	TAR WT 500 nM Full Length MBP tagged chimera 10 uM				
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)	Ck
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07	
Protein	**Full Length MBP tagged chimera**	3.00E-05	6.7	1.00E-05	
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x	
Glycerol	50%		2.0		
dH2O			5.8		
BME		0.1	1.0		
		Total Volume	20.0		

Sample/Lane # 7	TAR WT 500 nM MBP 10 uM				
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)	Ck
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07	
Protein	**MBP**	2.38E-05	4.2	5.00E-06	
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x	
Glycerol	50%		2.0		
dH2O			8.3		
BME		0.1	1.0		
		Total Volume	20.0		

Sample/Lane # 8	TAR WT 500 nM Positive Control Tat Minimal 5 uM			
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07
Protein	**Tat Minimal**	3.00E-05	3.3	5.00E-06
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x
Glycerol	50%		2.0	
dH2O			9.2	
BME		0.1	1.0	
		Total Volume	20.0	

Sample/Lane # 9	TAR WT 500 nM hCyctT1-Tat no tag 5 uM			
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07
Protein	**hCycT1-Tat no tag**	3.00E-05	3.3	5.00E-06
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x
Glycerol	50%		2.0	
dH2O			9.2	
BME		0.1	1.0	
		Total Volume	20.0	

Sample/Lane # 10	TAR WT 500 nM Full Length MBP tagged chimera 5 uM			
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07
Protein	**TAR WT 500 nM Full Lengt**	3.00E-05	6.7	1.00E-05
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x
Glycerol	50%		2.0	
dH2O			5.8	
BME		0.1	1.0	
		Total Volume	20.0	

Sample/Lane # 11	TAR WT 500 nM MBP 5 uM			
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)
RNA	**TAR WT**	4.00E-06	2.5	5.00E-07
Protein	**MBP**	2.38E-05	8.4	1.00E-05
Buffer	**Tris 20 mM + 10 uM ZnCl2**	10X	2.0	1x
Glycerol	50%		2.0	
dH2O			4.1	
BME		0.1	1.0	
		Total Volume	20.0	

Sample/Lane # 12	Tat Minimal 10 uM			
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)
Protein	Tat Minimal 10 uM	3.00E-05	6.7	1.00E-05
Buffer	Tris 20 mM + 10 uM ZnCl2	10X	2.0	1x
Glycerol	50%		2.0	
dH2O			8.3	
BME		0.1	1.0	
		Total Volume	20.0	

Sample/Lane # 13	hCycT1 Tat no tag 10 uM			
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)
Protein	hCycT1-Tat no tag	3.00E-05	6.7	1.00E-05
Buffer	Tris 20 mM + 10 uM ZnCl2	10X	2.0	1x
Glycerol	50%		2.0	
dH2O			8.3	
BME		0.1	1.0	
		Total Volume	20.0	

Sample/Lane # 14	Full Length His MBP tagged chimera 10 uM			
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)
Protein	Full Length MBP	3.00E-05	6.7	1.00E-05
Buffer	Tris 20 mM + 10 uM ZnCl2	10X	2.0	1x
Glycerol	50%		2.0	
dH2O			8.3	
BME		0.1	1.0	
		Total Volume	20.0	

Sample/Lane # 15	MBP 10 uM			
	ID	Initial Conc.(M)	Vol (uL)	Final Conc.(M)
Protein	MBP 10 uM	2.38E-05	8.4	1.00E-05
Buffer	Tris 20 mM + 10 uM ZnCl2	10X	2.0	1x
Glycerol	50%		2.0	
dH2O			6.6	
BME		0.1	1.0	
		Total Volume	20.0	

Appendix 17 - EMSA Protocol

Sample Preparation:

Snap Cool RNA: ~3 min @ 90° C, vortex, incubate on ice.

Pre-run gel @ 100 V for ~ 30 min to 1 hour 30 min 4 $^{\circ}$ C.

TBE running buffer (1 x) or other

Run gel @ 100 V

Actual running time: ~1:30 min @ 4 $^{\circ}$ C

12% Acrylamide Native Gel

30 ml Total Volume (~ 2 gels)

       9 ml Accugel 40%

       3 ml 10X TBE

       18 ml dH$_2$O

       12 µl TEMED

       300 µl APS 20% (.2 g/ml)

10 X TBE electrophoresis buffer

Tris Base 108g [Cf] 890 mM

Boric Acid 55 g[Cf] 890 mM

dH2O 960 ml

EDTA [0.5 M]   40 ml

pH 8.0[Cf] 20 mM

300 µl APS 20% (.2 g/ml)

Nucleic Acid Stain:

Prepare 1X SYBR Gold gel stain:

5 µl SYBR Gold (50,000X) in

50 ml TBE buffer

Remove gel from cast and stain with 1 X staining solution, and gentle agitation for 10-40

min. in a plastic container protected from light.

Wash gel two times in 150 ml dH$_2$O for 10 sec.

Visualize nucleic acid by UV transillumination, take a picture.

Protein Stain:

Quick Protocol			
	Reagent	Basic Protocol	Rapid Protocol
Fix	50% Methanol, 7% Acetic acid	100 mL 30 min 100 mL 30 min	100 ml, 15 min 100 ml, 15 min
Stain	SYPRO Ruby gel Stain	60 mL overnight	60 ml Microwave 30 sec, agitate 30 sec,

			microwave 30 sec, agitate 23 min
Wash	10% Methanol, 7% Acetic acid	100 mL 30 min	100 mL, 30 min
Hands-on Time		10 min	15 min
Total time		~18 hours	90 minutes

<u>Fix</u>

After electrophoresis, place the gel into a clean container with 100 mL of fix solution and agitate on an orbital shaker for 30 minutes. Repeat once more with fresh fix solution. Pour off the used fix solution.

<u>Stain</u>

Add 60 mL of SYPRO Ruby gel stain. Agitate on an orbital shaker overnight.

<u>Wash</u>

Transfer the gel to a clean container and wash in 100 mL of wash solution for 30 minutes. Transfer step helps minimize background staining irregularities and stain speckles on the gel. Before imaging rinse the gel in ultrapure water a minimum of two times for 5 minutes to prevent possible corrosive damage to the imager. Visualize protein with the UV transilluminator; take a picture.

Overlay nucleic acid and protein images with Photoshop.

Appendix 18 - Novagen Rosetta B Gami Transformation Protocol

Materials:

1. Novagen Rosetta B Gami non-DE3 non-pLys competent cells stored at -80°C

2. Chimera plasmid DNA, dilute to 10ng/μl

3. LB Media

4. 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 competent cells on ice. Add 20ul of Rosetta Gami B competent cells to each tube

3. Incubate the mixture on ice for 5 minutes

4. Heat-shock the cell mixture at 42°C for 30 seconds

5. Immediately put tubes back on ice, incubate for 2 minutes

6. Add 80 μl of room temperature SOC, gently mix

7. Recover for 1 hour with light shaking at 37°C incubator for 1 hour

8. Spread 50 μl of the recovered cells onto each pre-warmed LB/AMP/Chl plate

9.   Grow plates at 37°C overnight

10. Pick a single colony from the plate and inoculate a 5-ml LB/AMP/Chl liquid culture

11. Grow at 37°C overnight, shaking at 220~250rpm

12. Add 50 % glycerol to make a permanent glycerol storage culture

13. Store the frozen culture at -80°C.

Appendix 19 - Double Colony Selection Expression Level Analysis

1. Streak fresh LB+Amp+Chl plate with old glycerol stock for control

2. Select and mark 3 to 4 different freshly transformed colonies from LB plates and one colony from glycerol stock plate to inoculate two falcon tubes in 2 ml LB (can also try turbo) for each colony making a total of 8 tubes

3. Incubate tubes at 37°C until OD_{600} = 2-3

4. Spin down tubes at 1500 g for 5 min

5. Resuspend the pellets in 5 mL of minimal M9 medium to OD_{600} between 0.07 and 0.1

6. Save 100 ul from control tube prior to induction

7. Induce each tube at OD_{600} = 1.0 Induce one tube for each colony with 1.0 mM IPTG

8. Incubate overnight at 28°C

9. Take OD_{600} of final culture

10. Collect 250 uL from each tube and spin at 3300g for 5-10 minutes

11. Prepare SDS PAGE sample by resuspending the pellets in 50 ul of 2X SDS loading buffer

12. Incubate samples for 20 minutes at 70°C and spin down at maximum speed for 20 minutes prior to loading

13. Analyze gel and choose highest protein yielding colony

14. Grow high yield colony O/N in LB or Turbo

15. Plate high yield culture on LB + amp + chl plate

16. Repeat steps 1 through 13

17. Prepare Glycerol Stock for second high yielding double-colony selected culture and store at -80°C

Sivashanmugam et al. Optimized high-cell density IPTG-Induction Minimal Medium 1 L

Na$_2$HPO$_4$·7H$_2$O 50 mM (MW 268.07)   13.4 g

KH$_2$PO$_4$ (pH 8.0-8.2) 25 mM (MW 136.1 g)   3.4 g

NaCl 10 mM (MW 58.44)   0.6 g

NH$_4$Cl 0.1 % (MW 53.49)   1.0 g

AUTOCLAVE HERE BEFORE ADDITION OF REMAINING REAGENTS

MgSO$_4$ 5 mM (MW 120.36)   0.6 g or 5 ml of 1 M

CaCl$_2$ 0.2 mM (MW 110.98)   0.2 g or 1 ml of 0.2 M

Glucose 1.0 %   25.0 ml of 40 %

ZnCl$_2$ 0.1 mM   0.5 ml of 0.2 M

Trace Metals   2.5 ml of 1000x Trace Metals

Vitamin Mix   10.0 ml of 100x Vitamin Mix

Appendix 20 - Expression of Double Colony Selected Mutant

1. Inoculate 1L of turbo media in baffle bottom flask with highest yielding colony

2. Grow at 37°C at 240 rpm until OD_{600} ~5

3. Save 1 ml uninduced sample and prepare with approximately 500 ul 1X SDS

4. Collect cells by spinning down at 1500 g for 5 min

5. Resuspend cells in 1L of minimal media prepared as above

6. Allow cells to recuperate for 1-1.5 hours at 37°C

7. Add 1% EtOH 20-30 min prior to induction

8. Induce with 1 mM IPTG

9. Grow 24-48 hours

10. Purify as per hCycT1-Tat chimera in rich media

References

1. Richter, S. N., Frasson, I., and Palu, G. (2009) *Curr Med Chem* **16**, 267-286
2. Addo, M. M., Altfeld, M., Rosenberg, E. S., Eldridge, R. L., Philips, M. N., Habeeb, K., Khatri, A., Brander, C., Robbins, G. K., Mazzara, G. P., Goulder, P. J. R., Walker, B. D., and Collaboration, t. H. C. S. (2001) *Proceedings of the National Academy of Sciences* **98**, 1781-1786
3. Cullen, B. R. (1991) *FASEB J* **5**, 2361-2368
4. Efthymiadis, A., Briggs, L. J., and Jans, D. A. (1998) *J Biol Chem* **273**, 1623-1628
5. Selby, M. J., Bain, E. S., Luciw, P. A., and Peterlin, B. M. (1989) *Genes Dev* **3**, 547-558
6. Kim, Y. S., Kim, J. M., Jung, D. L., Kang, J. E., Lee, S., Kim, J. S., Seol, W., Shin, H. C., Kwon, H. S., Van Lint, C., Hernandez, N., and Hur, M. W. (2005) *J Biol Chem* **280**, 21545-21552
7. Tahirov, T. H., Babayeva, N. D., Varzavand, K., Cooper, J. J., Sedore, S. C., and Price, D. H. (2010) *Nature* **465**, 747-751
8. D'Orso, I., and Frankel, A. D. (2010) *Viruses-Basel* **2**, 2226-2234
9. Amendt, B. A., Hesslein, D., Chang, L. J., and Stoltzfus, C. M. (1994) *Molecular and Cellular Biology* **14**, 3960-3970
10. Mancebo, H. S., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D., and Flores, O. (1997) *Genes Dev* **11**, 2633-2644
11. Richter, S., Ping, Y.-H., and Rana, T. M. (2002) *Proceedings of the National Academy of Sciences* **99**, 7928-7933
12. Peterlin, B. M., and Price, D. H. (2006) *Mol Cell* **23**, 297-305
13. Jadlowsky, J. K., Nojima, M., Schulte, A., Geyer, M., Okamoto, T., and Fujinaga, K. (2008) *Retrovirology* **5**, 63
14. Asamitsu, K., Hibi, Y., Imai, K., Victoriano, A. F. B., Kurimoto, E., Kato, K., and Okamoto, T. (2011) *Journal of Molecular Biology* **410**, 887-895
15. Garber, M. E., Wei, P., KewalRamani, V.N., Mayall, T.P., Herrmann, C.H., Rice, A.P., Littman, D. R. and Jones, K.A. (1998) *Genes & Development* **12**, 3512-3527
16. Bres, V., Tagami, H., Peloponese, J. M., Loret, E., Jeang, K. T., Nakatani, Y., Emiliani, S., Benkirane, M., and Kiernan, R. E. (2002) *EMBO J* **21**, 6811-6819
17. Wei, P., Garber, M.E., Fang, S., Fischer, W.H., and Jones, K.A. (1998) *Cell Press* **92**, 451-462
18. Anand, K., Schulte, A., Vogel-Bachmayr, K., Scheffzek, K., and Geyer, M. (2008) *Nat Struct Mol Biol* **15**, 1287-1292
19. Xie, B., Calabro, V., Wainberg, M. A., and Frankel, A. D. (2004) *Journal of Virology* **78**, 1456-1463
20. Davidson, A., Leeper, T. C., Athanassiou, Z., Patora-Komisarska, K., Karn, J., Robinson, J. A., and Varani, G. (2009) *Proc Natl Acad Sci U S A* **106**, 11931-11936
21. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., and Williamson, J. R. (1992) *Science* **257**, 76-80

22. Aboul-ela, F., Karn, J., and Varani, G. (1996) *Nucleic Acids Research* **24**, 3974-3981

23. Mu, Y., and Stock, G. (2006) *Biophysical journal* **90**, 391-399

24. Churcher, M. J., Lowe, A. D., Gait, M. J., and Karn, J. (1995) *Proceedings of the National Academy of Sciences* **92**, 2408-2412

25. Dethoff, E. A., Hansen, A. L., Musselman, C., Watt, E. D., Andricioaei, I., and Al-Hashimi, H. M. (2008) *Biophysical journal* **95**, 3906-3915

26. Richter, S., Cao, H., and Rana, T. M. (2002) *Biochemistry* **41**, 6391-6397

27. Das, C., Edgcomb, S. P., Peteranderl, R., Chen, L., and Frankel, A. D. (2004) *Virology* **318**, 306-317

28. Fujinaga, K., Irwin, D., Taube, R., Zhang, F., Geyer, M., and Peterlin, B.M. (2002) *American Society for Microbiology* **76**, 12934-12939

29. Karn, J. (1999) *J Mol Biol* **293**, 235-254

30. Kehn-Hall, K., Guendel, I., Carpio, L., Skaltsounis, L., Meijer, L., Al-Harthi, L., Steiner, J. P., Nath, A., Kutsch, O., and Kashanchi, F. (2011) *Virology* **415**, 56-68

31. Magnuson, D. S. K., Knudsen, B. E., Geiger, J. D., Brownstone, R. M., and Nath, A. (1995) *Annals of Neurology* **37**, 373-380

32. Richter, S. N., and Palu, G. (2006) *Curr Med Chem* **13**, 1305-1315

33. Stelzer, A. C., Frank, A. T., Kratz, J. D., Swanson, M. D., Gonzalez-hernandez, M. J., Lee, J., Andricioaei, I., Markovitz, D. M., and Al-hashimi, H. M. (2011) *Nature Chemical Biology* **7**, 553-559

34. Mousseau, G., Clementz, M. A., Bakeman, W. N., Nagarsheth, N., Cameron, M., Shi, J., Baran, P., Fromentin, R., Chomont, N., and Valente, S. T. (2012) *Cell Host Microbe* **12**, 97-108

35. Borer, P. N., and Hudson, B. S. (2009) Branched and Multi-Chain Nucleic Acid Switches for Sensing and Screening. U.S.A.

36. Borer, P. N., and Hudson, B. H. (2010) Method of Generating Branched and Multi-Chain Nucleic Acid Switches for Ligand Detection. U.S.A.

37. Borer, P. N., and Hudson, B. S. . (2011) Branched and Multi-Chain Nucleic Acid Switches: Methods for Screening and Sensing. U.S.A.

38. DeCiantis, C. L., Jensen, D. K., Hudson, B. S., and Borer, P. N. (2007) *Biochemistry* **46**, 9164-9173

39. Borer, P. N., Hudson, B. S., and DeCiantis, C. L. . (2010) Switchable Nucleic Acids for Diagnostics, Screening and Molecular Electronics. U.S.A.

40. Aboul-ela, F., Karn, J., and Varani, G. (1995) *J Mol Biol* **253**, 313-332

41. Palmer, I., and Wingfield, P. T. (2004) *Curr Protoc Protein Sci* **Chapter 6**, Unit 6 3

42. Zhang, S.-m., Fan, R., Yang, T.-y., Sun, Y., Li, J.-y., Xu, Q.-z., and Zhou, P.-k. (2009) *Virologica Sinica* **24**, 518-528

43. Frankel, A. D., Bredt, D. S., and Pabo, C. O. (1988) *Science* **240**, 70-73

44. Kirsch, T., Boehm, M., Schuckert, O., Metzger, A. U., Willbold, D., Frank, R. W., and Rösch, P. (1996) *Protein Expression and Purification* **8**, 75-84

45. de Marco, A., Vigh, L., Diamant, S., and Goloubinoff, P. (2005) *Cell Stress Chaperones* **10**, 329-339

46. McKenna, M. C., Muchardt, C., Gaynor, R., and Eisenberg, D. (1994) *Protein Expr Purif* **5**, 105-111

47.   Romani, B., Engelbrecht, S., and Glashoff, R. H. (2010) *Journal of General Virology* **91**, 1-12

48.   Li, L., Dahiya, S., Kortagere, S., Aiamkitsumrit, B., Cunningham, D., Pirrone, V., Nonnemacher, M. R., and Wigdahl, B. (2012) *Advances in Virology* **2012**, 28

49.   Aldovini, A., Debouck, C., Feinberg, M. B., Rosenberg, M., Arya, S. K., and Wong-Staal, F. (1986) *Proc Natl Acad Sci U S A* **83**, 6672-6676

50.   Ensoli, B., Buonaguro, L., Barillari, G., Fiorelli, V., Gendelman, R., Morgan, R. A., Wingfield, P., and Gallo, R. C. (1993) *J Virol* **67**, 277-287

51.   Park, J., Hangyu, L., Lee, Y., Kang, Y.H., Rhim, H., and Choi, S.Y. (2000) *Journal of Biochemistry and Molecular Biology* **33**, 337-343

52.   Bigalke, J. M., Czudnochowski, N., Vollmuth, F., Vogel-Bachmayr, K., Anand, K., and Geyer, M. (2011) *Methods* **53**, 78-84

53.   Freund, J., Vértesy, L., Koller, K.-P., Wolber, V., Heintz, D., and Kalbitzer, H. R. (1995) *Journal of Molecular Biology* **250**, 672-688

54.   Löhr, M., Kibler, K. V., Zachary, I., Jeang, K.-T., and L. Selwood, D. (2003) *Biochemical and Biophysical Research Communications* **300**, 609-613

55.   Zhang, J., Tamilarasu, N., Hwang, S., Garber, M. E., Huq, I., Jones, K. A., and Rana, T. M. (2000) *J Biol Chem* **275**, 34314-34319

56.   Ivanov, D., Kwak, Y. T., Nee, E., Guo, J., García-Martínez, L. F., and Gaynor, R. B. (1999) *Journal of Molecular Biology* **288**, 41-56

57.   Fong, Y. W., and Zhou, Q. (2000) *Mol Cell Biol* **20**, 5897-5907

58.   Preston, A. (2011) Plasmid Vectors. Genetics Institute Inc

59.   Gallo, R. C. (1999) *Proc Natl Acad Sci U S A* **96**, 8324-8326

60.   de Mareuil, J., Carre, M., Barbier, P., Campbell, G., Lancelot, S., Opi, S., Esquieu, D., Watkins, J., Prevot, C., Braguer, D., Peyrot, V., and Loret, E. (2005) *Retrovirology* **2**, 5

61.   Nygren, P.-Å., Stefan, S., and Uhlén, M. (1994) *Trends in Biotechnology* **12**, 184-188

62.   Smith, D. B., and Johnson, K. S. (1988) *Gene* **67**, 31-40

63.   Mercado-Pimentel, M. E., Jordan, N. C., and Aisemberg, G. O. (2002) *Protein Expression and Purification* **26**, 260-265

64.   Braun, P., and LaBaer, J. (2003) *Trends Biotechnol* **21**, 383-388

65.   Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H., and Hard, T. (2002) *Protein Sci* **11**, 313-321

66.   Shih, Y.-P., Kung, W.-M., Chen, J.-C., Yeh, C.-H., Wang, A. H. J., and Wang, T.-F. (2002) *Protein Science* **11**, 1714-1719

67.   Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X., and Butt, T. R. (2006) *Protein Science* **15**, 182-189

68.   Kapust, R. B., and Waugh, D. S. (1999) *Protein Science* **8**, 1668-1674

69.   Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B. (2003) *Protein Sci* **12**, 1313-1322

70.   Nallamsetty, S., Austin, B. P., Penrose, K. J., and Waugh, D. S. (2005) *Protein Science* **14**, 2964-2971

71.   Pryor, K. D., and Leiting, B. (1997) *Protein Expression and Purification* **10**, 309-319

72.   Routzahn, K. M., and Waugh, D. S. (2002) *J Struct Funct Genomics* **2**, 83-92

73. Invitrogen. (2009) Gateway(R) Technology with Clonase(TM) II: A universal technology to clone DNA sequences for functional analysis and expression in multiple systems., Manual part no. 25-0749 Ed., Life Technologies

74. Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., and Waugh, D. S. (2001) *Protein Eng* **14**, 993-1000

75. Austin, B. P., Nallamsetty, S., and Waugh, D. S. (2009) *Methods Mol Biol* **498**, 157-172

76. Biomol, G. (2004) Ampliqon IIII Competent Cells. Biomol GmbH, Hamburg

77. Novy, R., Drott, D., Yaeger, K., and Robert Mierendorf. (2001) Overcoming the codon bias of *E. coli* for enhanced protein expression. in *inNovations*, Novagen, Inc.

78. Stewart, E. J., Åslund, F., and Beckwith, J. (1998) *EMBO J* **17**, 5543-5550

79. Prinz, W. A., Åslund, F., Holmgren, A., and Beckwith, J. (1997) *Journal of Biological Chemistry* **272**, 15661-15667

80. Koken, S., Greijer, A., Verhoef, K., van Wamel, J., Bukrinskaya, A., and Berkhout, B. (1994) *J. Biol. Chem.* **269**, 8366-8375

81. Kalantari, P., Narayan, V., Natarajan, S. K., Muralidhar, K., Gandhi, U. H., Vunta, H., Henderson, A. J., and Prabhu, K. S. (2008) *J Biol Chem* **283**, 33183-33190

82. Ortiz de Orue Lucana, D., Wedderhoff, I., and Groves, M. R. (2012) *J Signal Transduct* **2012**, 605905

83. Stathopoulos, C. (1998) *Membr Cell Biol* **12**, 1-8

84. Kukkonen, M., and Korhonen, T. K. (2004) *International Journal of Medical Microbiology* **294**, 7-14

85. Luo, S., McNeill, M., Myers, T., Hohman, R., and Levine, R. (2008) *Archives of Microbiology* **189**, 181-185

86. Dekel, E., and Alon, U. (2005) *Nature* **436**, 588-592

87. Expression Technologies, I. (2003) Bacterial *E. coli* Growth Media.

88. Marr, A. G. (1991) *Microbiol Rev* **55**, 316-333

89. Sezonov, G., Joseleau-Petit, D., and D'Ari, Richard. (2007) *Journal of Bacteriology* **189**, 8746-8749

90. Grabski, A., Mehler, M., and Drott, D. (2005) *Nature Methods* **2**, 233-235

91. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A., and Lian, L.-Y. (2004) *Journal of the American Chemical Society* **126**, 8933-8939

92. Oganesyan, N., Ankoudinova, I., Kim, S. H., and Kim, R. (2007) *Protein Expr Purif* **52**, 280-285

93. Edwards, A. M., Arrowsmith, C. H., Christendat, D., Dharamsi, A., Friesen, J. D., Greenblatt, J. F., and Vedadi, M. (2000) *Nature Structural Biology* **7 Suppl 1**, 970-972

94. Thomas, J., and Baneyx, F. (1996) *J Biol Chem* **271**, 11141 - 11147 Record 11125 of 11152 - MEDLINE (R) 11141/11196-11112/11196

95. Georgiou, G., and Valax, P. (1996) *Current Opinion in Biotechnology* **7**, 190-197

96. Kusano, K., Waterman, M. R., Sakaguchi, M., Omura, T., and Kagawa, N. (1999) *Archives of Biochemistry and Biophysics* **367**, 129-136

97. Winter, J., Neubauer, P., Glockshuber, R., and Rudolph, R. (2000) *Journal of Biotechnology* **84**, 175-185

98. Van Dyk, T. K., Reed, T. R., Vollmer, A. C., and LaRossa, R. A. (1995) *J Bacteriol* **177**, 6001-6004

99. Barroso, J. F., Elholm, M., and Flatmark, T. (2003) *Biochemistry* **42**, 15158-15169

100. Gragerov, A., Nudler, E., Komissarova, N., Gaitanaris, G. A., Gottesman, M. E., and Nikiforov, V. (1992) *Proc Natl Acad Sci U S A* **89**, 10341-10344

101. Neidhardt, F. C., VanBogelen, R. A., and Vaughn, V. (1984) *Annual Review of Genetics* **18**, 295-329

102. Chaudhuri, S., Jana, B., and Basu, T. (2006) *Cell Biology and Toxicology* **22**, 29-37

103. Bolen, D. W. (2004) *Methods* **34**, 312-322

104. Hochachka, P. W. (2002) *Biochemical adaptation : mechanism and process in physiological evolution*, Oxford University Press, Oxford ;

105. Liu, Y., and Bolen, D. W. (1995) *Biochemistry* **34**, 12884-12891

106. Bolen, D. W., and Baskakov, I. V. (2001) *Journal of Molecular Biology* **310**, 955-963

107. Kashanian, S., Paknejad, M., Ghobadi, S., Omidfar, K., and Ravan, H. (2008) *Hybridoma (Larchmt)* **27**, 99-106

108. Harries, D., and Rösgen, J. (2008) A Practical Guide on How Osmolytes Modulate Macromolecular Properties. in *Methods in Cell Biology* (Dr. John, J. C., and Dr. H. William Detrich, III eds.), Academic Press. pp 679-735

109. Alzahrani, Z. (2009) Salting in, salting out, and dialysis of proteins. in *Physical Biochemistry Lab*, Kansas State University

110. Wilson, E. K. (2012) *Chemical & Engineering News* **90**, 42-43

111. Arakawa, T., and Timasheff, S. N. (1984a) *J Biol Chem* **259**, 4979-4986

112. Arakawa, T., and Timasheff, S. N. (1984b) *Biochemistry* **23**, 5912-5923

113. Arakawa, T., Bhat, R., and Timasheff, S. N. (1990) *Biochemistry* **29**, 1924-1931

114. Arakawa, T., and Timasheff, S. N. (1983) *Arch Biochem Biophys* **224**, 169-177

115. Arakawa, T., and Timasheff, S. N. (1984c) *Biochemistry* **23**, 5924-5929

116. Arakawa, T., and Timasheff, S. N. (1985) *Biophys J* **47**, 411-414

117. Lee, J. C., and Timasheff, S. N. (1981) *J Biol Chem* **256**, 7193-7201

118. Lee, L. L., and Lee, J. C. (1987) *Biochemistry* **26**, 7813-7819

119. Schellman, J. A., and Somero, G. N. (1996) *Biophys J* **71**, 1985-1993

120. Koolman, J., and Röhm, K.-H. (2005) *Color atlas of biochemistry*, Thieme, Stuttgart ; New York

121. Das, U., Hariprasad, G., Ethayathulla, A. S., Manral, P., Das, T. K., Pasha, S., Mann, A., Ganguli, M., Verma, A. K., Bhat, R., Chandrayan, S. K., Ahmed, S., Sharma, S., Kaur, P., Singh, T. P., and Srinivasan, A. (2007) *PLoS ONE* **2**, e1176

122. Varughese, M. M., and Newman, J. (2012) *Journal of Biophysics* **2012**, 7

123. Tsumoto, K., Ejima, D., Kita, Y., and Arakawa, T. (2005) *Protein Pept Lett* **12**, 613-619

124. Runge, F. F. (1855) *Der Bildungstrieb der Stoffe, veranschaulicht in selbstständig gewachsenen Bildern.*

, Author's publication, Oranienburg

125. Tswett, M. S. (1905) *Proceedings of the Warsaw Society of Naturalists* **14**, 20-39

126. Bussemas, H. H. a. E., Leslie S. (2004) *LCGC North America* **22**, 262-270
127. Dian, C., Baráth, P., Knaust, R., McSweeney, S., Moss, E., Hristova, M., Ambros, V., and Birse, D. (2002) *Life Science News* **10**
128. Qiagen. (2008) Compatibility of reagents with Ni-NTA. Qiagen
129. Nevin Gerek, Z., Kumar, S., and Banu Ozkan, S. (2013) *Evol Appl* **6**, 423-433
130. Rajalingam, D., Kathir, K. M., Ananthamurthy, K., Adams, P. D., and Kumar, T. K. (2008) *Anal Biochem* **375**, 361-363
131. Albini, A., Benelli, R., Presta, M., Rusnati, M., Ziche, M., Rubartelli, A., Paglialunga, G., Bussolino, F., and Noonan, D. (1996) *Oncogene* **12**, 289-297
132. Rusnati, M., Coltrini, D., Oreste, P., Zoppetti, G., Albini, A., Noonan, D., di Fagagna, F. d. A., Giacca, M., and Presta, M. (1997) *Journal of Biological Chemistry* **272**, 11313-11320
133. Parks, T. D., Howard, E. D., Wolpert, T. J., Arp, D. J., and Dougherty, W. G. (1995) *Virology* **210**, 194-201
134. Tropea, J., Cherry, S., and Waugh, D. (2009) Expression and Purification of Soluble His6-Tagged TEV Protease High Throughput Protein Expression and Purification. (Doyle, S., and Walker, J. eds.), Humana Press. pp 297-307
135. Tolia, N. H., and Joshua-Tor, L. (2006) *Nat Methods* **3**, 55-64
136. Rath, A., Gilbowicka, M., Nadeau, V.G., Chen, G. and Charles M. Deber. (2007) *PNAS* **106**, 1760-1765
137. Marenchino, M., Armbruster, D. W., and Hennig, M. (2009) *Protein Expression and Purification* **63**, 112-119
138. Sharadamma, N., Khan, K., Kumar, S., Neelakanteshwar Patil, K., Hasnain, S. E., and Muniyappa, K. (2011) *FEBS Journal* **278**, 3447-3462
139. Burnette, W. N. (1981) *Analytical Biochemistry* **112**, 195-203
140. Spectrometry, S. M. (2003) Introduction to MALDI-TOF MS. (Francisco, U. s. G. U. o. C. S. ed.
141. Marvin, L. F., Roberts, M. A., and Fay, L. B. (2003) *Clinica Chimica Acta* **337**, 11-21
142. Fried, M., and Crothers, D. M. (1981) *Nucleic Acids Res* **9**, 6505-6525
143. Garner, M. M., and Revzin, A. (1981) *Nucleic Acids Res* **9**, 3047-3060
144. Hellman, L. M., and Fried, M. G. (2007) *Nat Protoc* **2**, 1849-1861
145. Gagnon, K. T., and Maxwell, E. S. (2011) *Methods Mol Biol* **703**, 275-291
146. Roder, K., and Schweizer, M. (2001) *Biotechnol Appl Biochem* **33**, 209-214
147. Probes(TM), M. (2005) SYPRO(R) Ruby Protein Gel Stain. in *Product Information* (Technologies, I. D. ed., Eugene, OR
148. Probes(TM), M. (2006) SYBR® Gold Nucleic Acid Gel Stain. in *Invitogen Detection Technologies*, Invitrogen, Eugene, OR
149. Karlsson, H. J., Bergqvist, M. H., Lincoln, P., and Westman, G. (2004) *Bioorg Med Chem* **12**, 2369-2384
150. Garner, M. M., and Rau, D. C. (1995) *EMBO J* **14**, 1257-1263
151. Vossen, K. M., Wolz, R., Daugherty, M. A., and Fried, M. G. (1997) *Biochemistry* **36**, 11640-11647
152. Davis, F., and Higson, S. P. J. (2009) *Seminars in Cell & Developmental Biology* **20**, 1-2

153. Clark Jr, L. C., and Lyons, C. (1962) *Annals of the New York Academy of Sciences* **102**, 29-45
154. Updike, S. J., and Hicks, G. P. (1967) *Nature* **214**, 986-988
155. Abdiche, Y. N. (2010) Complementary Use of Label-Free Real-Time Biosensors in Drug Discovery of Monoclonal Antibodies. in *Biosensors* (Serra, P. A. ed., InTech
156. Fang, Y. (2011) *International Journal of Electrochemistry* **2011**
157. Brockman, J. M., Nelson, B. P., and Corn, R. M. (2000) *Annu Rev Phys Chem* **51**, 41-63
158. Hartmann-Petersen, R., and Gordon, C. (2005) Quantifying Protein–Protein Interactions in the Ubiquitin Pathway by Surface Plasmon Resonance. in *Methods in Enzymology* (Raymond, J. D. ed.), Academic Press. pp 164-177
159. Green, R. J., Frazier, R. A., Shakesheff, K. M., Davies, M. C., Roberts, C. J., and Tendler, S. J. B. (2000) *Biomaterials* **21**, 1823-1835
160. Zhao, M., Wang, X., and Nolte, D. (2010) *Biomed. Opt. Express* **1**, 983-997
161. Do, T., Ho, F., Heidecker, B., Witte, K., Chang, L., and Lerner, L. (2008) *Protein Expression and Purification* **60**, 147-150
162. Long, K. S., and Crothers, D. M. (1999) *Biochemistry* **38**, 10059-10069
163. Hill, J. M. (2008) *Methods Mol Biol* **426**, 437-446
164. Koth, C. M., and Edwards, A. M. (2003) From Clone to Crystal: Maximizing the Amount of Protein Samples for Structure Determination. in *Advances in Protein Chemistry* (Richard, D. S., and Timothy, D. V. eds.), Academic Press. pp 343-352
165. Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., and Li, Q. (2009) *Protein Sci* **18**, 936-948
166. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science* **279**, 384-388
167. Holdeman, T. C., and Gardner, K. H. (2001) *J Biomol NMR* **21**, 383-384
168. Webster, K. A., Prentice, H., and Bishopric, N. H. (2001) *Antioxidants & redox signaling* **3**, 535-548
169. Sabel, C. E., Shepherd, J. L., and Siemann, S. (2009) *Anal Biochem* **391**, 74-76
170. Lin, Y., Robbins, J. B., Nyannor, E. K., Chen, Y. H., and Cann, I. K. (2005) *J Bacteriol* **187**, 7881-7889
171. Zhou, Z. S., Peariso, K., Penner-Hahn, J. E., and Matthews, R. G. (1999) *Biochemistry* **38**, 15915-15926
172. Pfister, T., Jones, K. W., and Wimmer, E. (2000) *J Virol* **74**, 334-343
173. Viitanen, P. V., Donaldson, G. K., Lorimer, G. H., Lubben, T. H., and Gatenby, A. A. (1991) *Biochemistry* **30**, 9716-9723
174. Naik, S., Haque, I., Degner, N., Kornilayev, B., Bomhoff, G., Hodges, J., Khorassani, A. A., Katayama, H., Morris, J., Kelly, J., Seed, J., and Fisher, M. T. (2010) *Biopolymers* **93**, 237-251
175. Voziyan, P. A., Johnston, M., Chao, A., Bomhoff, G., and Fisher, M. T. (2005) *Journal of Structural and Functional Genomics* **6**, 183-188
176. Fisher, M. T. (2012) Use of Chaperone Proteins and Octet Bio-Layer InterferometryTo Detect Protein Folding /Unfolding Events in a High-Throughput Format. (Center, U. o. K. M. ed.

177.   Arthur, K. K., Gabrielson, J. P., Kendrick, B. S., and Stoner, M. R. (2009) *J Pharm Sci* **98**, 3522-3539

178.   Howlett, G. J., Minton, A. P., and Rivas, G. (2006) *Current Opinion in Chemical Biology* **10**, 430-436

179.   Schuck, P. (2000) *Biophys J* **78**, 1606-1619

Biographical Data

NAME OF AUTHOR: Collin L. Fischer

PLACE OF BIRTH: New York, New York, U.S.A.

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Syracuse University, Syracuse, New York

Oswego State University, Oswego, New York

Queens College CUNY, Queens, New York

DEGREES AWARDED:

Bachelor of Arts in Biology Cum Laude, 2003, Oswego State University

AWARDS AND HONORS:

Syracuse University CASE Scholarship 2011-2012

Syracuse University Graduate Assistantship 2005-2010

Howard Hughes Summer Program for Undergraduate Research 2000

PROFESSIONAL EXPERIENCE:

Teaching Assistant Dept. of Chemistry, Syracuse University 2007-2012