An In Vitro Selected Sequence Capable of Ultrahigh Transgene Expression in Vaccinia Virus Infected Cells

by

Julia Anne Flores

A Thesis Presented in Partial Fulfillment
Of the Requirements for the Degree
Master of Science

Approved April 2012 by the
Graduate Supervisory Committee:

John Chaput, Chair
Joshua LaBaer
Bertram Jacobs

ARIZONA STATE UNIVERSITY

May 2012

ABSTRACT

Recombinant protein expression is essential to biotechnology and molecular medicine, but facile methods for obtaining significant quantities of folded and functional protein in mammalian cell culture have been lacking. Here I describe a novel 37-nucleotide in vitro selected sequence that promotes unusually high transgene expression in a vaccinia driven cytoplasmic expression system. Vectors carrying this sequence in a monocistronic reporter plasmid produce >1,000-fold more protein than equivalent vectors with conventional vaccinia promoters. Initial mechanistic studies indicate that high protein expression results from dual activity that impacts both transcription and translation. I suggest that this motif represents a powerful new tool in vaccinia-based protein expression and vaccine development technology.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor John Chaput and also Brian Wellensiek for advising and assisting me along every step of this project as well as all members of the Chaput lab both past and present. They have been a great group of people to work with and I appreciate all of the help and encouragement that I have received. I would also like to acknowledge members of the Jacobs lab, particularly James Jancovich and Stacy White for indispensible training, Karen Kibler for obtaining the vTF7-3 strain, and Karen Denzler and Trung Huynh for being so helpful with all of my questions. I would like to thank the members of my committee, Drs. John Chaput, Bertram Jacobs and Joshua LaBaer for guiding me through this experience. Finally I would like to thank my wonderful family, especially my patient husband, for their love and support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 EUKARYOTIC TRANSLATION INITIATION

Overview of Eukaryotic Translation

The translation of mRNA transcripts into protein in eukaryotes is a multi-step process consisting of four main stages: (1) initiation, in which the ribosome, methionyl tRNA and other necessary factors assemble onto the mRNA start codon, (2) elongation of the peptide strand via the natural peptidyl transferase activity of the ribosome, (3) termination of the translated message and release of the protein, and (4) recycling of the translation machinery by disassembly of the mRNA-ribosome complex.[1] Of the four steps, most translational regulation occurs at initiation, underscoring the importance of understanding the factors that control ribosomal recruitment and translation initiation.[1]

Gene regulation via translation control is advantageous because it can provide a faster phenotypic change than regulation at the transcription level because the mRNA is already made and ready to be translated.[1] Genome-wide translation regulation can be achieved by adjusting the availability of eukaryotic initiation factors (eIFs), poly(A) binding protein (PABP) or other species required for initiation.[1,2] The translation of individual mRNAs can be controlled through interaction with microRNAs

(miRNAs), secondary structures in the 5' UTR, open reading frames (ORF) upstream of the actual start codon and the context of the AUG start site (i.e. Kozak sequence).[1,2]

Efforts to understand eukaryotic translation have revealed that translation initiation requires the coordination of many factors. The form and function of initiation factors continue to be discovered. Furthermore, recent additions to this growing body of knowledge expose a gap in our understanding of this complex process and all of the dynamic components that contribute to initiation. Here, I provide a review of canonical eukaryotic translation initiation and highlight some nontraditional mechanisms and unanswered questions.

Translation Initiation

The overall process of eukaryotic translation initiation is depicted in Figure 1. The current understanding of the canonical process of eukaryotic translation initiation begins with eukaryotic initiation factor (eIF) 2-GTP joining with a charged initiator tRNA (Met-tRNAi) to form the ternary complex (TC).[3] Each round of translation initiation results in the release of eIF2-GDP that must be recycled to eIF2-GTP for the next round of initiation. eIF2 has a higher affinity for GDP than GTP so the exchange is assisted by eIF2B.[4] The TC and the 40S ribosomal subunit (to which is bound eIFs 1, 1A, 3 and possibly 5) are brought together by eIF3 (a multi-

domain protein responsible for binding several key players in translation initiation) to form the 43S pre-initiation complex (PIC)[2,5] A capped, polyadenylated mRNA strand is prepared to receive the 43S PIC by the eIF4F complex (consisting of the ATP dependant RNA helicase eIF4A, the cap binding protein eIF4E and the scaffold protein eIF4G).[6] eIF4E binds to the 5' cap, bringing the eIF4F complex and the mRNA strand together.[6] EIF4B also binds the mRNA strand and plays a non-essential role assisting the RNA binding and helicase activity of eIF4A.[7] The poly(A) binding protein (PABP) binds to the 3' poly(A) tail of the mRNA strand and interacts with eIF4G as well, causing the mRNA to form a loop. This interaction increases the affinity of eIF4F for the cap and prevents mRNAs with truncated or degraded 3' ends from being translated.[8,9] The 43S complex loads onto the mRNA sequence near the 5'cap through interactions between eIF3 and eIF4G.[10] The ribosome then begins to scan the mRNA searching nucleotide triplets for the AUG start codon with the help of several initiation factors.[11,12] EIF4A unwinds RNA secondary structure[13] and eIFs 1 and 1A aid scanning and correct initiation codon selection by blocking the A site of the ribosome until the correct codon in a favorable context is reached.[14] Translation generally initiates at the first AUG codon in the 5' region of a mRNA message with a few established exceptions: 1) when the first AUG is too close to the cap,[15] 2) when the AUG is in an unfavorable (non-Kozak) sequence context and is bypassed

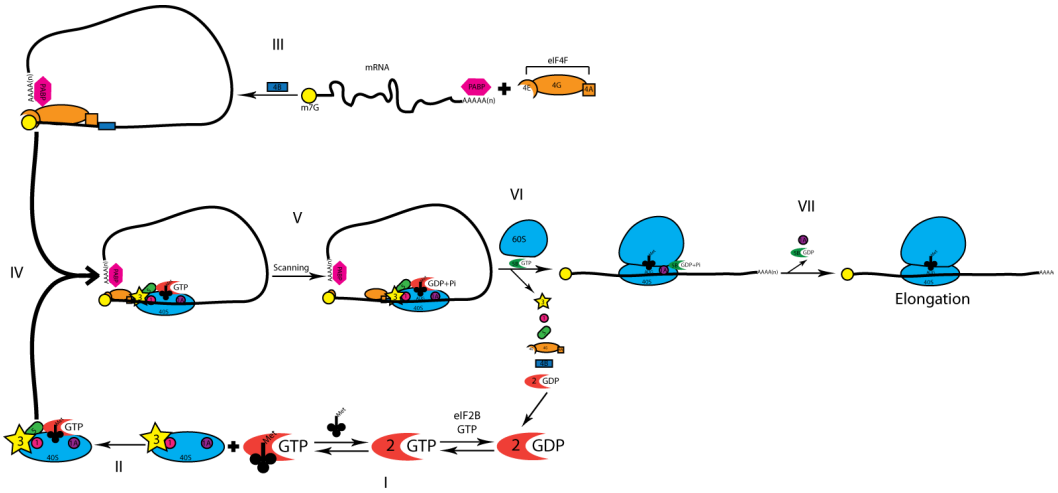

Figure 1: Eukaryotic Translation Initiation. **I.** eIF2B mediates the exchange of GDP for GTP in eIF2.[4] eIF2-GTP binds Met-tRNA$_i^{Met}$ to form the ternary complex (TC).[3] **II.** The 40S ribosomal subunit, with eIFs 1, 1A, 3 and possibly 5 bind the TC forming the 43S complex.[5] **III.** Meanwhile, eIF4F complex composed of the subunits eIF4E, eIF4A and eIF4G binds to the 5' m^{7}G cap of an mRNA message with the help of the RNA binding protein eIF4B. eIF4G also interacts with Poly A Binding Protein (PABP) bound to the 3' poly A tail, causing the mRNA to form a circle.[6] **IV.** The 43S complex loads onto the mRNA near the 5' cap through interactions between eIF3 and eIF4G.[6] **V.** The complex scans the mRNA checking successive nucleotide triplets against the Met-tRNA$_i^{Met}$ anticodon for the complementary AUG start codon.[11, 12] Upon reaching the start codon, eIF2-GTP hydrolyzes to eIF2-GDP. **VI.** Hydrolysis of eIF2-GTP and binding of the 60S subunit, assisted by eIF5B-GTP, triggers the release of eIFs 1, 2, 3, 4B, 4F and probably 5. eIF2-GDP is recycled to start another round.[17, 18] **VII.** eIF5B-GTP hydrolyzes and is released with eIF1A.[17] The 80S complex is ready for elongation.

through leaky scanning, and 3) when the ribosome translates a short open reading frame, remains associated with the mRNA and reinitiates translation at a downstream AUG .[11] Once an AUG pairs correctly with the anti-codon on the methionine tRNA the GTPase activating protein eIF5 accelerates the hydrolysis of eIF2-GTP to eIF2-GDP.[16] At this point eIFs 1, 3, 4B, 4F, 5, and 2-GDP dissociate from the mRNA-ribosome complex while eIF5B-GTP binds along with the 60S ribosomal subunit.[17,18] The binding of the larger subunit causes hydrolysis of eIF5B-GTP to eIF5B-GDP, which then dissociates along with eIF1A.[19] The ribosome is now positioned to receive charged tRNA and begin the process of elongation.

The study of several non-canonical variations of translation initiation is contributing to a greater understanding of the molecular origin of disease and the regulation of protein expression. One such mechanism is ribosomal “shunting.” First described in viral mRNA transcripts, shunting is a process whereby the 43S complex is able to temporarily disassociate from the mRNA and re-associate and initiate translation at a location further downstream.[20] There is evidence that shunting also occurs in uninfected eukaryotic cells and may contribute to the dysregulation of protein expression in Alzheimer patients.[21] The shunting model offers a possible explanation for the cap-dependent initiation of translation at a start site downstream of an AUG codon in “good” context or when the 5’ UTR of a mRNA strand contains a stable secondary structure that is

bypassed in order to reach a down stream initiation site.[22,23] The shunting activity is mediated by sequences within the 5'UTR that are complementary to and interact with the 18S ribosomal subunit.[24,25] An additional explanation of these phenomena is the "tethering and clustering" model proposed by Mauro and colleagues.[26] In this model, the 43S complex is "tethered" through the cap binding proteins to the mRNA strand that is able to fold over and allow the complex to interact with sequences complementary to the 18S ribosome, thereby bypassing any intervening secondary structure or AUGs.[26]

Transcripts containing an Internal Ribosomal Entry Site (IRES) are able to forego scanning and cap dependant translation by recruiting the 40S ribosomal pre-initiation complex directly to the 5'UTR.[27,28] IRES sequences were also first discovered in viruses that lack proteins responsible for adding a 5' cap to the RNA message.[29] Viral IRESs function by folding into large tertiary structures that mimic some or all of the eukaryotic initiation factors and in doing so are able to recruit the ribosome to the initiation site (see Figure 2).[28] The picornavirus family, including encephalomyocarditis virus (EMCV), poliovirus and foot and mouth disease virus, contain secondary structures that bind eIF4G directly and do not require the cap binding protein eIF4E.[30,31] The Hepatitis C Virus avoids using any part of the eIF4F complex or eIF4B through contacts with the stem loop structures of the IRES and the 40S

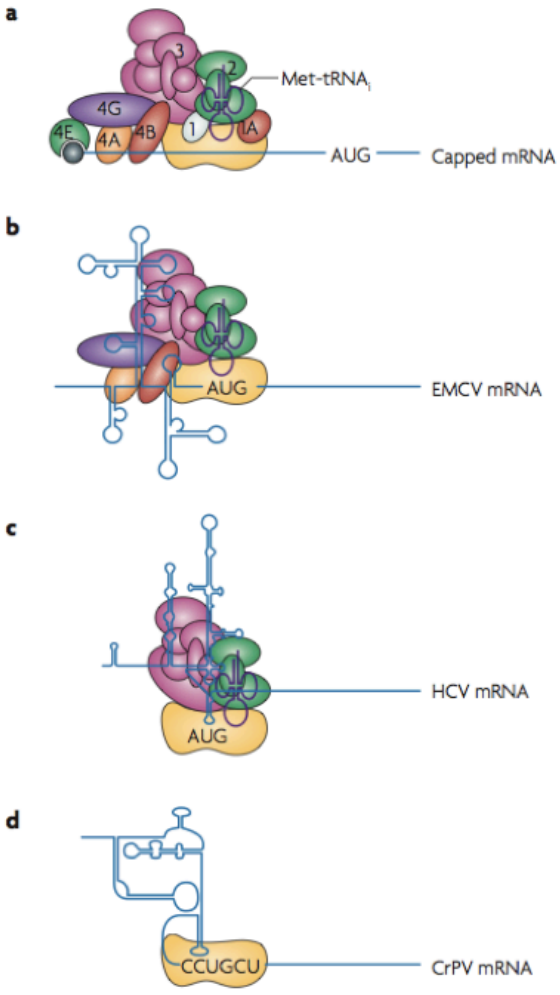

Figure 2: Capped mRNA and viral mRNA ribosome recruitment strategies. Comparison of canonical eukaryotic cap-dependent translation initiation (a) and the internal initiation mechanism used by the picornavirus encephalomyocarditis virus (EMCV) internal ribosome entry site (IRES) (b), hepatitis C virus (HCV) IRES (c) and the cricket paralysis virus (CrPV) IRES (d). Reprinted (adapted in part or whole) with permission from *Nature Reviews Microbiology* (2008).[27]

ribosome.[32] The cricket paralysis virus (CrPV) IRES dispenses with the need for any initiation factors, including eIF2, or for the initiator tRNA by interacting directly with the 40S subunit and using a GCU codon to begin translation.[28] IRESs have also been discovered in eukaryotic cells and

have been shown to direct expression of essential genes during starvation, cellular stress or conditions when regular protein production is shut down through the lack of initiation factors.[33,34]

Unanswered Questions

Despite the advances that have been made in the understanding of eukaryotic translation initiation, there are still many unanswered questions. For instance, it has recently been discovered that the vast majority of the genome is transcribed but only a small percentage (<2%) is annotated as protein coding.[35,36] Among these non-coding transcripts, researchers have identified a new class of RNAs termed long intergenic non-coding RNA (lincRNA). LincRNA are defined as transcripts greater than 200 bases that have no potential to be translated by the canonical translation initiation mechanism.[37] However, the transcription of many of these sequences is highly regulated and there is evidence lincRNAs play important roles in development and cell differentiation.[37] The *polished rice (pri)* transcript from *Drosophila* was considered to be non-coding because it contained only short open reading frames.[38] In fact, *pri* mRNA codes for four very small peptides, between 11 to 32 amino acids, that play a pivotal role in the development of trichomes on *Drosophila* larvae.[38] This discovery raises the possibility that other non-canonical peptides may be hidden in "non-coding" transcripts.

Alternative methods for initiation, such as IRESs, still require further investigation. Though the existence of IRESs in viral genomes is well established, the discovery and characterization of these elements in cells is still in the early stages. Recently, there has been a great deal of conflicting evidence qualifying or completely discrediting supposed cellular IRESs. A common assay to identify IRESs places the potential sequence in the intergenic region of a bicistronic plasmid. If the second gene is able to be expressed when the canonical translation initiation of the first gene is inhibited, it is assumed that the sequence mediated an internal translation initiation event. This assay for identifying IRES sequences was proved to be flawed by a series of studies showing that often the protein produced from the second gene may actually be attributed to cryptic promoter activity or poor experimental design.[39–42] The validity of cellular IRESs needs to be solidified by requiring sequences to pass rigorous, well designed tests before being identified as IRESs. Such tests could include direct transfections of RNA containing the candidate IRES into cells. Direct RNA transfections bypass the nucleus and the potential splicing and mRNA modifications that take place there. These studies should also include structural analysis of the mRNA post-expression to ensure it is fully intact. These tests will help to rule out the possibility of transcriptional or nuclear processing artifacts, such as splicing, contributing to increased protein production.[39,41]

The mechanisms for regulating the translation initiation of individual mRNAs almost invariably inhibit initiation. These inhibitory mechanisms include miRNA, RNA binding proteins, RNA secondary structures in the 5'UTR and, under most conditions, upstream Open Reading Frames.[1,2,43] The many pathways for translational suppression lead to the question: Is there a mechanism to enhance the translation initiation of individual mRNA strands, and if so could this mechanism depend on the presence of an mRNA sequence element? An element dependent mechanism could enable a cell to quickly produce a large amount of protein without having to repeatedly go through the energetically expensive and time-consuming processes of mRNA transcription and maturation. An example of a small sequence capable of enhancing translation when placed in the 5'UTR of a gene was discovered in the mouse genome.[44] This sequence was shown to function through complementary base pairing with the 18S ribosome.[45]

1.2 RECOMBINANT MAMMALIAN PROTEIN EXPRESSION SYSTEMS

Proteins are macromolecular structures in living cells that have diverse roles from enzyme catalysts to receptors of extracellular signals, transducers of intracellular signals, and structural scaffolds. These diverse functions make proteins useful tools in molecular medicine, biotechnology, industry and agriculture. Although many systems have been developed for producing recombinant proteins, the ability to manufacture human proteins

with correct post-translational modifications (PTM) has remained challenging. In general, human proteins with correct patterns of PTMs are produced expensively and in low yields, which is cost prohibitive for many basic and applied research projects.[46] Here, I briefly review the most commonly used expression systems and discuss their advantages and disadvantages.

Bacterial Expression Systems

Bacterial expression systems rely principally on the *E. coli* and related *Bacillus* species for heterologous protein expression. In these systems, the gene of interest is placed downstream of a regulated promoter on a plasmid that is transformed into the bacterial cell for high and controlled expression.[47]

Bacterial systems have the advantage of being inexpensive to maintain, safe to use, easy to manipulate, and producing large quantities of recombinant protein.[47] There are many limitations to bacterial expression systems as well. Disadvantages of this system include an inability to produce large mammalian proteins as well as PTMs such as glycosylation that are often essential for proper function.[48] In some cases, high production yields can become problematic as overproduction can lead to cellular toxicity (an issue which can be addressed with specially engineered strains and growth conditions).[49] Proteins produced in an *E.*

coli system are often packaged in inclusion bodies and require further processing for proper folding and activity.[48] In addition, care must be taken when isolating the protein to remove any toxins produced by the bacteria. Thus, bacterial systems are the system of choice when manufacturing small (<30 kD), unmodified proteins, but are incapable of producing large proteins with complex PTMs.[46]

Yeast Expression Systems

Yeast is engineered to produce recombinant proteins by transformation with a plasmid containing the gene of interest. Once inside the cell, the plasmid can either be maintained as an episome or incorporated into the genome through homologous recombination, depending on the experimental design.[50] The most common yeast species used for protein expression is *S. cerevisiae*, which has a well characterized eukaryotic genome.[50]

Relative to *E. coli* systems, yeast systems produce less protein, but can produce large functional proteins with many common PTMs.[48] Yeast can also be grown in defined, protein free media. This facilitates the purification of the recombinant protein and makes yeast culture more affordable than insect or mammalian cell culture.[46]

Yeast systems can produce glycosylated proteins, but their pattern of PTMs is often different than the patterns found in proteins produced in

mammalian systems. For example, *S. cerevisiae* produces O-linked oligosaccharides with mannose while mammalian proteins have sialylated O-linked oligosaccharides. Yeasts also lack proper chaperonins that are required to properly fold some complex mammalian proteins.[46] Though the glycosylation and folding abilities of yeast systems surpass those of bacterial systems, they are still insufficient for producing many mammalian proteins.

Baculovirus Expression Systems

Insect cells represent an attractive host for recombinant protein expression because of their capacity to produce proteins with folding and PTMs that are similar to proteins produced in mammalian cells.[51] In the baculovirus system, the insect virus is engineered to contain the desired protein in place of a nonessential, highly expressed native gene. Insect cells (usually fall army worm cell lines Sf9 or Sf21) are then infected with the aculovirus and produce the protein.[52] Recombinant baculoviruses were traditionally produced through homologous recombination with a transfer vector, which occurred at a low rate and required time-consuming plaque purifications. Bacmid technology streamlines this process by putting the baculovirus genome into a linearized bacterial artificial chromosome that is deficient in an essential gene. Homologous recombination with a transfer vector in an *E. coli* cell adds the recombinant

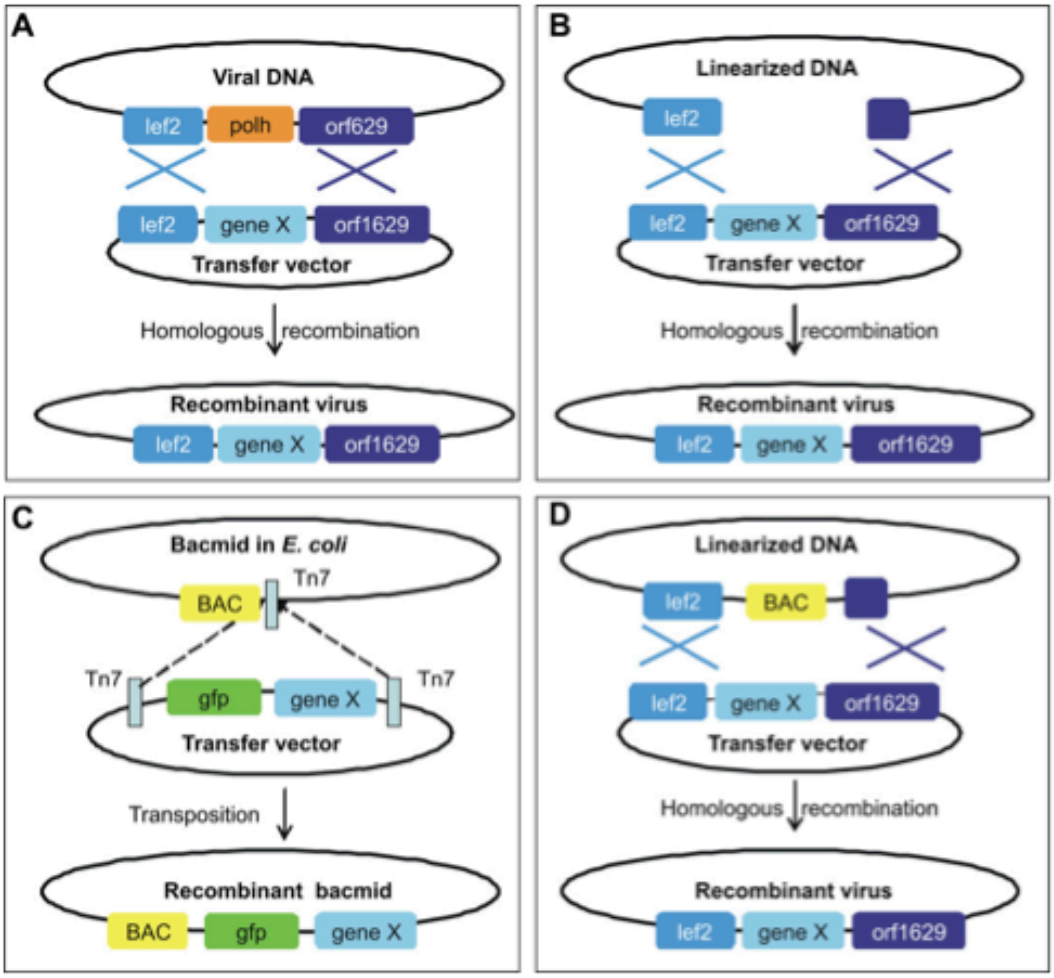

Figure 3: Baculovirus expression vectors over time. Various methods exist to generate recombinant baculoviruses expressing a foreign gene (gene X). Historically, recombinant baculoviruses were generated through homologous recombination (A). Subsequently, linearized vectors were developed to increase the percentage of recombinants, as an essential gene, orf1629, will be restored only upon recombination (B). Bacmid technology is based on transposition of gene X into a bacterial artificial chromosome containing the baculovirus genome and which is amplified and manipulated in E. coli (C). Bacmid technology and the repair of an essential gene were combined to avoid bacterial sequences in the virus genome and further automated (flashback/BacMagic) for high-throughput recombinant virus generation (D). Reprinted with permission (adapted in whole or part from *J. Inv. Path.* (2011).[51]

protein and restores the essential gene resulting in improved recombination frequency and removing the need for plaque purification.[51]

Cultivation of insect cells also has the advantage of being more affordable and producing more product than mammalian cells.[48] The virus is extremely safe to work with because it is incapable of infecting vertebrate cells and as such poses no danger to humans.[51] However, the system does have drawbacks, including the technical challenges of building the recombinant virus. The baculovirus infection can impair protein processing, folding, secretion, and also results in high protease activity at the lytic stage of the viral life cycle.[52,53] While the baculovirus system produces more protein than mammalian cell culture systems it still has relatively low yields compared to bacterial or yeast systems. There are also differences between the PTMs of some baculovirus produced proteins and their mammalian counterparts that cause the baculovirus proteins to be inactive.[54,55]

Mammalian Cell Culture Expression Systems

Mammalian cell culture systems are necessary to produce proteins that require mammalian-specific PTMs in order to achieve functional activity.[48] Mammalian cell based expression systems are either plasmid or virus based.

Plasmid based systems

Under transient transfection conditions, plasmids containing the gene of interest under the control of a strong transcription promoter are transformed into the cells using calcium phosphate transfection, electroporation or lipofection methods.[56] Protein is harvested from the cells 1-10 days post transfection.[57]

Protein can also be expressed by creating a stably transfected cell line, although this process is considerably more difficult and can take several months to achieve. A typical stable transfection protocol requires transforming cells with both the gene of interest and a gene that confers a selective advantage such as antibiotic resistance. A small percentage of foreign DNA will enter the nucleus and integrate into the cell genome. Applying a selective pressure reduces the cell population so that it only contains members that were successfully transformed. Individual cells are then grown and evaluated for protein production. A colony of cells will be selected to found a cell line based on protein production levels and growth characteristics.[56]

Virus based systems

Adenoviruses can infect a diverse range of human cells and replicate with high efficiency, making them good candidates for expression of mammalian genes.[58] Adenoviruses engineered for recombinant protein expression are replication incompetent outside of suitable complementing

cells to increase the safety profile. There are two methods used to generate recombinant adenovirus. The first requires transfecting a shuttle vector carrying the gene of interest into complementing cells that have been infected with adenovirus. Homologous recombination places the gene of interest into the viral genome under control of an early promoter and the complementing host cell enables viral replication. Successive screening and plaque purifications identify and purify the desired recombinant virus. In an alternative method for producing recombinant adenovirus, homologous recombination takes place inside *E. coli* cells between a shuttle vector and a vector containing the adenovirus genome. Once the desired adenovirus is created, it is infected into (non-complementing) host cells where it enters the nucleus but does not integrate into the genome and the cell transiently produces the recombinant protein.[58,59]

Lentivirus is a retrovirus that integrates its genome into that of the host cell and is used for generating cell lines that stably produce recombinant proteins. Lentivirus has the additional advantage of being able to infect non-replicating cells. The lentivirus system is very similar to the adenovirus system in that virus used for protein expression is engineered to be replication incompetent without the help of a packaging system. Homologous recombination between the modified lentivirus and a transfer cassette containing the gene of interest takes place in a

Table 1: Comparison of Mammalian Protein Expression Systems

Expression System	System Construction Time	Expression Time	Advantages	Disadvantages
E. coli	Days	Hours	Simple technology High yields Safe Fast Affordable	Incapable of producing large proteins or PTMs Inclusion bodies, require purification and renaturation
Yeast	Days	≤ 1 day	Simple technology High yields Safe Fast Affordable Glycosylation	Differences in glycosylation patterns Mammalian specific PTMs Improper folding
Baculovirus	Weeks to months	1-several days	Good yields Safe Many mammalian-like PTMs	Advanced technology Time consuming Cannot produce all mammalian PTMs
Mammalian cell: Transient transfection	Weeks	1-several days	Safe Mammalian PTMs	Low yields Expensive
Mammalian cell: Stable transfection	Months	1-several days	Safe Mammalian PTMs	Low yields Time consuming Expensive
Mammalian cell: Viral systems (Adenovirus and Lentivirus)	Weeks to months	1-several days	Mammalian PTMs Better yields than plasmid based systems	Safety Time consuming Relatively low yields Expensive

packaging cell containing vectors that provide the viral proteins necessary for replication.[60] Recombinant lentivirus conveniently integrates the gene of interest along with a promoter sequence into the host genome, generating a stably transfected cell line.

The vaccinia virus system, which is also used to express recombinant proteins, is described elsewhere in this manuscript.

While viral systems are efficient and generally produce more protein than their plasmid based counterparts, they have the disadvantage of being potentially hazardous to the humans who work with them. Care must also be taken to prevent contamination of the final protein product with virus as well.[46] In addition, generating a new recombinant virus for every protein is time consuming and expensive.

Thus, although recombinant proteins from mammalian cell culture systems are correctly folded and modified, their production requires an expensive investment of time and money for a relatively low yield, which discourages their use unless necessary for function.

1.3 VACCINIA VIRUS REVIEW

Basic Vaccinia Biology

Vaccinia viruses (VVs) belong to the family *Poxviridae* and genus *Orthopoxvirus*.[61] This genus also contains variola virus, the causative agent of smallpox. VVs are DNA viruses with large, double stranded linear

genomes ranging in size between 178-192 kb, depending on the strain.[62] Such a large genome is required because VVs carry all the genes necessary for DNA transcription and viral replication to occur in the cytoplasm.[61] Each viral particle also comes packaged with all the enzymes necessary for the transcription and maturation of viral RNA.[63]

Virus can enter the body through a break in the skin and manifests as a localized lesion.[61] Upon viral entry into a cell, early promoters direct the expression of proteins that block the body's extracellular and intracellular antiviral defenses.[64] Proteins for DNA replication and transcription are also made early in the viral life cycle.[64] Genes that are expressed at an intermediate time point in the viral life cycle produce transcription factors for late genes expressed at the end of the viral life cycle prior to cell lysis. Late genes include proteins necessary for virion assembly.[64] VVs are also capable of infecting a wide range of cultured mammalian cells. In cultured cells, DNA replication has been observed as early as 1-2 hours post infection, with new virus produced in 8 hours.[64]

There are four VV virion types that play different roles in the spread of infection. Intracellular mature viruses (IMV) are released after lysis and are responsible for host-to-host transmission. Intracellular enveloped virus (IEV) fuses with the cell membrane to form cell-associated enveloped virus (CEV), which mediate the spread of infection to neighboring cells.

The virus is spread to distant cells in the same host or cell culture by extracellular envelope virus (EEV).[64]

History of the VV Smallpox Vaccine

Vaccine technology was pioneered by Edward Jenner over 200 years ago when he began exposing people to a milder member of the poxvirus family to prevent infection with the more virulent and potentially lethal smallpox.[65] Because of the high conservation between all members of the orthopoxvirus genus, exposure to a VV also serves to vaccinate against variola virus.[65] Jenner originally used cowpox virus as a vaccine, but over years of use, experimentation and poor record keeping a related but genetically distinct virus emerged and consequently the exact origin of VV is unknown.[61] The effectiveness of the vaccine, coupled with the fact that humans are the only known reservoir and do not carry an asymptomatic form of smallpox, encouraged the worldwide campaign to eradicate this disease.[61] Global elimination of smallpox was achieved in 1980 through widespread vaccination, strict surveillance, and quarantine programs.[61] Interest in VVs remains high due to their potential use as vaccines for other diseases, their utilization in mammalian expression systems, and due to the fear of accidental or malicious reintroduction of smallpox.[66]

Potential for Recombinant VV Vacines

Given the success of vaccinia in ridding the world of smallpox, researchers are developing the technology to use VV as a vehicle to introduce antigens of other diseases to the immune system.[65] The large genome and high productivity of VVs make them an ideal recombinant system for incorporating and expressing foreign proteins (Figure 4).[57] VV vectors have been reported to incorporate at least 25 kb of foreign DNA.[67] However, there are still concerns about the safety of VV. There are rare cases of severe and even lethal reactions to the vaccine, mostly among immunocompromised recipients.[65] In an effort to attenuate VV, viral strains that are incapable of replicating in human cells have been developed.[68] These include the modified vaccinia virus Ankara (MVA) strain, which can infect but not replicate in human cells. MVA is also made less virulent by the loss of genes involved in evading the host immune system.[62] Unfortunately, these attenuated strains tend to make less effective vaccines that cannot generate an immune response strong enough for vaccination.[65] Strategies to create vaccines that are both safe and effective include co-expressing immune modulators with attenuated vaccines and deletion of virulent genes from more dangerous viruses.[69,70]

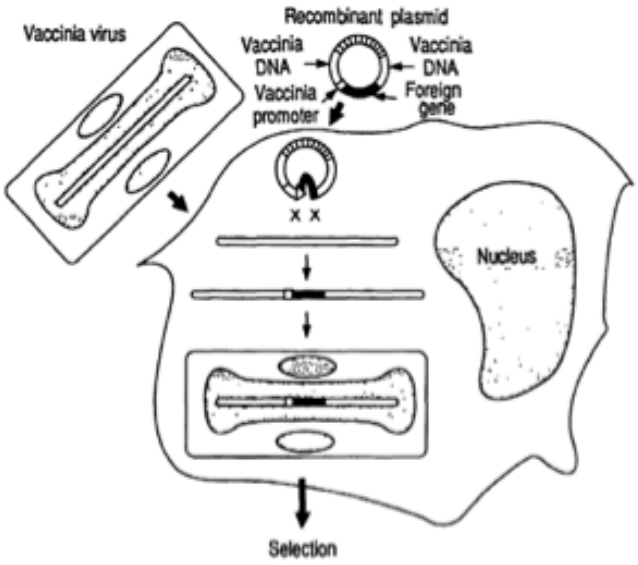

Figure 4: Formation of a recombinant vaccinia virus by homologous recombination. Reprinted with permission from *Science* (1991).[63]

VV Expression Systems

In order to produce proteins with proper folding and post translational modifications that are suitable for human therapeutics, a mammalian cell based system is required.[71] VV is particularly well suited for heterologous protein expression because transcription occurs in the cytoplasm, thereby eliminating the possibility of interference from nuclear splicing or mRNA transport. VV also has a wide host range and can infect most cultured mammalian cells.[71] The vTF7-3 strain has been modified to contain the bacteriophage T7 RNA polymerase specifically for recombinant protein expression.[72] The gene of interest, along with a T7 promoter, can be incorporated into the viral genome through homologous recombination or engineered onto a plasmid and transfected into cells that

are subsequently infected with the virus.[72–74] VV expression systems can produce more protein than standard stable transfections of mammalian cells and have the potential to be used in large-scale bioreactors.[75] Bleckwenn et al reported a yield of ~12 µg EGFP/million infected cells in a vaccinia bioreactor system.[75] Because it is a safety concern to work with VV, most vaccinia strains require Biosafety Level 2 (BSL2) procedures and vaccination is recommended for personnel at risk for exposure.[71] The potential for infection can be lessened or removed by using attenuated strains such as MVA, which can be used under BSL1 conditions.[68,73]

Chapter 2

CHARACTERIZIATION OF A SEQUENCE THAT FUNCTIONS AS A TRANSLATION ENHANCER AND VV PROMOTER

2.1 CONTRIBUTIONS

The following chapter describes the discovery and characterization of a sequence element that works as both a VV promoter and a translation enhancer. The project was conceived by Professor John Chaput. The reported experiments were performed by Julia Flores, Brian Wellensiek and Brett Stephens under the guidance of John Chaput. Brian Wellensiek performed the mRNA display selection and Brett Stephens did the initial characterization and preliminary end mapping deletion analysis of the Best sequence. Julia Flores finished the end mapping deletion analysis and ran the western blots, realtime PCR, DNA analysis, RACE, and time course experiments with assistance from Brian Wellensiek. John Chaput wrote the manuscript with comments from Julia Flores and Brian Wellensiek.

2.2 INTRODUCTION

Vaccinia virus (VV), a member of the poxvirus family, received worldwide attention when it was used to eradicate variola virus, the causative agent of smallpox.[65] Since then, poxviruses have been engineered for other biomedical and biotechnology purposes, which includes their use as a vehicle for protein expression in mammalian cell culture and as a vector for therapeutic vaccines directed against infectious agents and cancer.[63] The interest in poxvirus technology is due to a number of unique properties that are not found in other DNA viruses. Principal among these is the ability for poxviruses to replicate in the cytoplasm of their host cell, which avoids many of the complications associated with RNA splicing and export.[76] Poxviruses can infect a wide range of host cells and produce gene products with mammalian patterns of post-translational modifications (PTMs).[73] Non-mammalian systems can produce proteins with alternative modifications, which is problematic if these proteins are to be used as therapeutics, as targets in drug-based screens or as antigens for antibody production.[77,78]

It is known that recombinant vaccinia viruses are capable of producing milligram quantities of protein in mammalian cells.[73] However, these viruses require special expertise to construct and are therefore not suitable as a general platform for routine expression needs. Whether similar expression levels can be achieved in a transient transfect-infect

assay is an interesting question with significant practical implications in many areas of basic and applied research.[46]

We postulated that it might be possible to improve protein expression levels in mammalian cells if TEEs could be identified that effectively engaged the translation machinery. From a screen of ~250 *in vitro* selected sequences, we identified a short (37-nt) motif that is capable of achieving unusually strong transgene expression in a wide range of VV infected mammalian cells. Relative to the recombinant T7 vaccinia expression system, our system can produce >100-fold more protein after 6-12 hours of transient expression and >10,000-fold more protein when compared to other expression systems.[72]

2.3 MATERIALS AND METHODS

Cell Culture

All cells used in this study were obtained from the American Type Culture Collection (ATCC). HeLa and HEK293 cells were maintained in DMEM (Invitrogen), while BHK cells were maintained in MEM (Invitrogen). All media was supplemented with 5% fetal bovine serum (FBS, HyClone) and 5 mg/ml gentamicin (Invitrogen). Cells were kept at 37°C in a humidified atmosphere containing 5% CO_2.

VV Strains

The *Copenhagen* and vTF7-3 viral strains used in this study were obtained from ATCC. The modified vaccinia virus Ankara (MVA) was from Sanofi Pasteur. The *Copenhagen* strain (VC2) is a wild type vaccinia virus,[65] MVA is an attenuated vaccinia virus strain that is non-pathogenic in humans[62] and vTF7-3 is a recombinant vaccinia virus strain that has been engineered to express T7 RNA polymerase.[72] Viral stocks were stored in MEM with 2% FBS.

Transfection/Infection

Cells were seeded at a density of 15,000 cells per well in white 96-well plates 18 hours prior to transfection (Figure 5). Plasmid transfections were carried out using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. In brief, complexes containing 200 ng of plasmid and 0.5 μl of Lipofectamine 2000 were formed in Opti-MEM (Invitrogen). During complex formation, DMEM was discarded from the 96-well plate, and 50 μl of fresh Opti-MEM were added to the cells. Complexes (50 μl) were then carefully overlaid onto the cells (total volume 100 μl). Immediately following DNA transfections, cells were infected with VC-2,

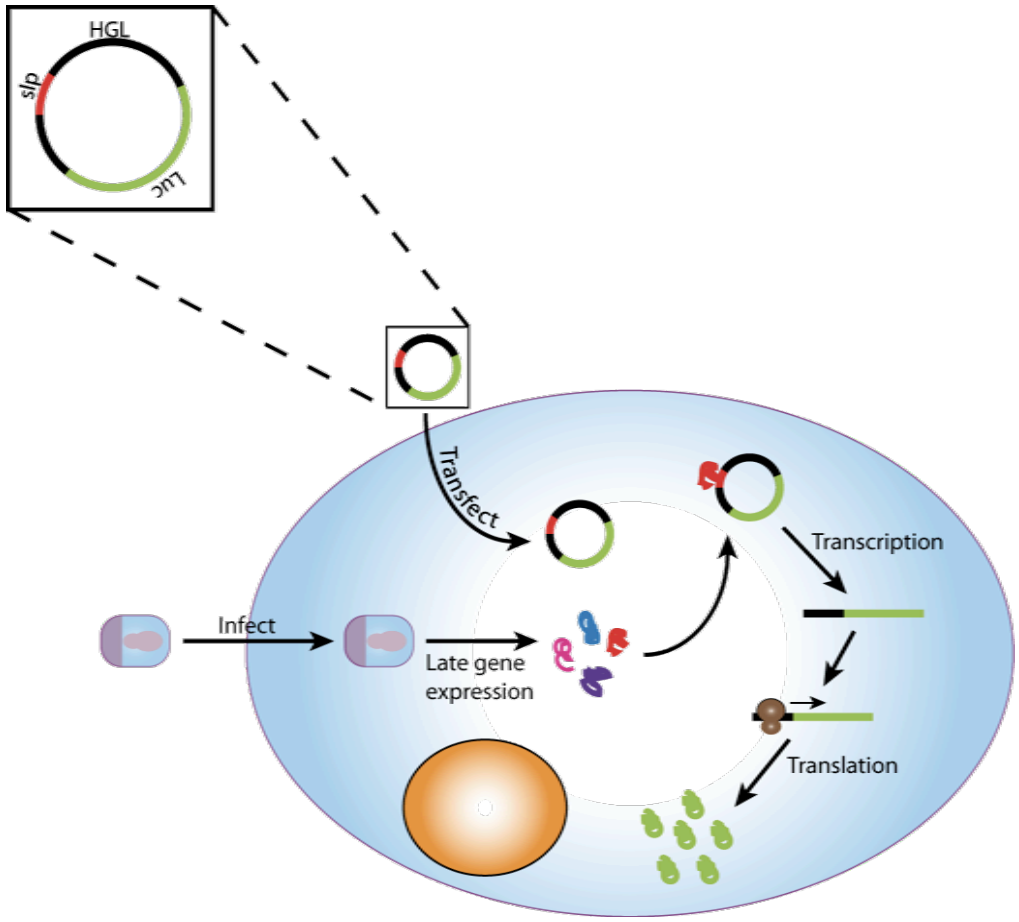

Figure 5: VV Transfect/Infect system. The luciferase gene was placed under the control of VV promoter SLP and transfected into cells that were infected with VV. Figure courtesy of Brian Wellensiek.

MVA, or vTF7-3 at a multiplicity of infection (m.o.i.) of 5 plaque forming units (PFU)/cell for six hour assays or 30 PFU/cell for 24 hour assays. For protein isolation, HeLa cells were plated at a density of 200,000 cells per well in a 24 well plate and 800 ng of plasmid were combined with 2 μl of Lipofectamine 2000 to form the transfection complexes. Media was removed from the wells and 400 μl of Opti-MEM media were added to the cells, which were then overlaid with 100 μl of transfection complexes.

Luciferase Activity Assay

Cells were lysed in the 96-well plate by discarding the growth-transfection media and adding passive lysis buffer (Promega). Luciferase activity was measured using the Promega Luciferase Assay System according to the manufacturer's protocol with a Glomax microplate luminometer (Promega).

RNA Characterization

RNA was isolated from transfected HeLa cells 6 hours post infection with VC2. Lysate from 2 wells of a 96 well plate were pooled and RNA isolation was performed using the PerfectPure RNA cultured cell kit (5 Prime) according to manufacturer's protocol. Isolated RNA was reverse transcribed with an oligo (dT) primer and Superscript II (Invitrogen). Realtime PCR (iQ^{TM} SYBR® Green Supermix, Bio-Rad) was used to determine the mRNA levels of luciferase as well as the housekeeping gene hypoxanthine-guanine phospho-ribosyltransferase (HPRT). Using the ΔΔCt method, the amount of luciferase mRNA was normalized to HPRT mRNA levels. Luminescence values were adjusted according to normalized luciferase mRNA levels.

End Mapping Deletion Analysis

Deletion analysis of HGL-Best was performed by Klenow extension of a short DNA primer annealed to the synthetic DNA templates containing BamHI and NcoI restriction sites. The double-stranded DNA was restriction digested and ligated into a monocistronic firefly luciferase reporter plasmid (F-luc-mono) carrying a vaccinia virus synthetic later promoter upstream of the insert (Wellensiek *et al.*, manuscript in preparation). Reporter plasmids containing truncated variants of the Best sequence were assayed for activity in the vaccinia virus transfect-infect assay as described above.

Western Blot

Western blot analysis was performed using F-luc-mono engineered with six different leader sequences in the 5' UTR (Best-core, SLP, SLP-Best-core, I1L, I1L-Best-core, and a random filler sequence). A second set of plasmids were generated by replacing the luciferase gene with the gene for HIV-1 Gag. In all cases, proteins were expressed in a standard transfect-infect assay in HeLa cells for 6 hours and lysed with 35 µl of Passive Lysis Buffer (Promega). Cellular debris was removed by centrifugation at 10,000 rcf for 10 min and the supernatant containing the protein of interest was removed and stored at -80°C. For protein analysis, the supernant was diluted with NuPage 4x LDS sample buffer (Invitrogen)

and proteins were denatured by heating at 95°C for 10 min before being run on a NuPage 4-12% Bis-Tris gel (Invitrogen) at 200V for 30 min. Proteins were transferred to a nitrocellulose membrane using the iBlot Gel Transfer system (Invitrogen) using the manufacturers instructions. After blocking for one hour at room temperature in TBS-Tween (20 mM Tris, 125 mM NaCl, pH 7.5, and 0.05% Tween) with 3% milk the membrane was cut along the 50 kDa band of a pre-stained protein ladder so that the protein of interest, either luciferase (66 kDa) or HIV-1 Gag (55 kDa), and the loading control, GAPDH (37 kDa) could be detected separately. The membrane pieces were incubated with the appropriate primary antibodies in TBS-Tween with 3% milk overnight at 4°C. The Firefly Luciferase and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) antibodies were obtained from Abcam while the HIV-1 Gag antibody was obtained from Professor Dr. Hohne at the Charite Institute for Biochemie in Berlin, Germany. Goat-anti-mouse or goat-anti-rabbit HRP conjugated secondary antibodies (Bethyl Laboratories) were then incubated with the membranes for one hour at room temperature. Membranes were visualized with SuperSignal West Pico or Dura Chemiluminescent Substrate (Pierce Biotechnology).

RACE

RNA was isolated using the PerfectPure RNA cultured cell kit (5 Prime) according to manufacturer's protocol. Rapid amplification of cDNA ends (RACE) was performed with the 5' RLM-RACE kit (Invitrogen) using total RNA following the small reaction protocol provided by the manufacturer with primers specific to the luciferase gene. RACE sequences were ligated into pJET 1.2 (Fermentas), cloned, and sequenced at the ASU DNA Sequencing Facility.

DNA Isolation and Real Time PCR

Cellular and plasmid DNA was isolated from transfected HeLa cells six hours post infection with VC2 using the Trizol Reagent (Invitrogen) according to the manufacturer's protocol. Following isolation, DNA was ethanol precipitated and re-suspended in water. Quantitative realtime PCR (iQ^{TM} SYBR® Green Supermix, Bio-Rad) was used to determine the levels of plasmid DNA as well as the housekeeping gene Ribonuclease P (RNase P).[79] Using the ΔΔCt method, the amount of plasmid DNA was normalized to RNase P DNA levels.

Cell Free System

Cell-free characterization was performed using a Human *In vitro* Protein Expression Kit (Pierce). Luciferase expression was achieved

following manufacturer's protocols using 300 ng of linear template for a two-hour transcription at 32°C followed by a 90 min translation at 30°C.

Primers

Real Time PCR Primers

Luciferase:

RTluc.F: 5' GCTGGGCGTTAATCAGAGAG

RTluc.R: 5' GTGTTCGTCTTCGTCCCAGT

HPRT:

RThprt.F: 5'TGCTGAGGATTTGGAAAGGGTG

RThprt.R: 5' CCTTGAGCACACAGAGGGCTAC

RNase P:

hRNaseP.F: 5'CCCCGTTCTCTGGGAACTC

hRNaseP.R: 5'TGTATGAGACCACTCTTTCCCATA

RACE Primers

Luc30.R (Outer primer): 5'GTACGTGATGTTCACCTCGATATGTGCATC

Dlpseq.R (Inner primer): 5'AGGAACCAGGGCGTATCTCT

2.4 RESULTS

We have previously used messenger RNA display to isolate RNA sequences from the human genome that function with TEE activity (Wellensiek *et al.*, manuscript in preparation). After 6 rounds of *in vitro*

selection and amplification, an initial set of 1000 human genomic elements were cloned and sequenced. Of these, 227 mapped with perfect identity (length and sequence) to the human reference genome. Initial experimental characterization revealed that many of these sequences could enhance protein translation levels in a human cell-free expression system. This observation suggested that it might be possible to increase protein production levels in a mammalian cell culture using our selected TEEs to promote ribosomal initiation at the translation start site.

To investigate this possibility, our set of 227 *in vitro* selected sequences were inserted into a firefly luciferase reporter plasmid (F-luc-mono) containing a vaccinia-specific promoter (Fig 1a). Transfected HeLa cells were infected with VV (*Copenhagen* strain, 5 moi) and luciferase activity was measured in 96-well format after 6-hours of cell-based expression (Figure 5). Vectors carrying either a no-insert control or one of ten randomly chosen sequences from the starting library provided a basal level of translation enhancement and no infection controls were used to demonstrate that luciferase activity was the result of cytoplasmic translation and not a product of nuclear expression. Consistent with earlier in vitro assays, plasmids carrying the selected TEEs provided luciferase values that were up to 100-fold stronger than the basal translation level observed for the set of eleven control vectors (Figure 6). However, one sequence, HGL-Best, routinely yielded luciferase values that were as

much as 5,000-fold higher than the basal level, indicating that this sequence exhibited remarkable translation enhancement activity in a vaccinia driven cytoplasmic expression system (Figure 6b).

To determine whether the high activity of HGL-Best was due to greater RNA expression or increased translation enhancement, quantitative real-time PCR was used to measure mRNA levels in cells expressing the HGL-Best containing vector. This analysis revealed that high luciferase activity was principally due to increased levels of expression; however, increased translation levels were also detected by normalizing luciferase activity values for cellular mRNA. When compared to the next highest activity sequences, HGL-Best produces ~10-fold

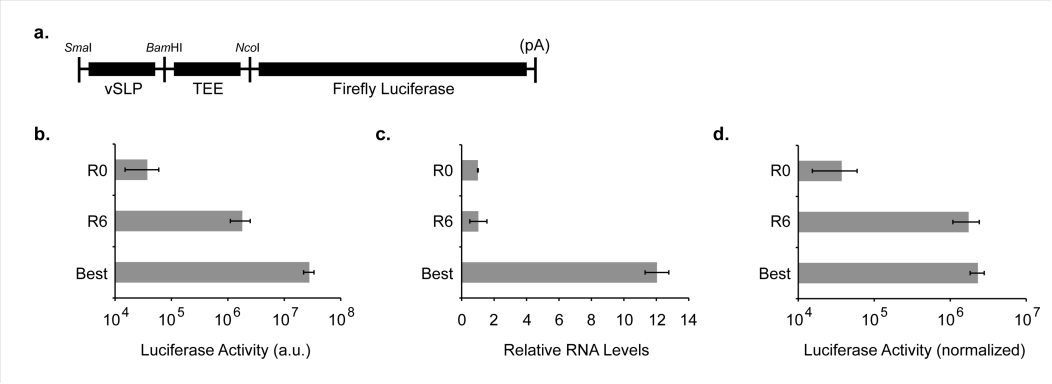

Figure 6: Functional analysis of HGL-Best. (a) Schematic view of the luciferase reporter plasmid used to evaluate in vitro selected TEEs. (b-d) Firefly luciferase gene expression and translation of Best relative to four randomly chosen sequences from the original naïve pool (labeled round 0), and the nine most active sequences discovered after six rounds of in vitro selection and amplification (labeled round 6) in vaccinia infected HeLa cells. Realtime quantitative PCR measurements coupled with chemiluminescence measurement indicate that Best functions as both a vaccinia promoter and translation enhancing element. This dual activity leads to a significant increase in luciferase production in vaccinia infected cells.

more RNA and ~5-fold more protein than all other plasmids tested (Figure 6c,d). We confirmed by quantitative real-time PCR that plasmid copy number was not altered in cells transfected with the HGL-Best plasmid (Figure 7b), indicating that HGL-Best impacts gene expression at the levels of transcription and translation, but not replication. This unusual feature is not without precedent, and at least one other RNA element is known to function in this capacity albeit less efficiently.[80]

We determined the minimal region required to achieve strong gene expression using end-mapping deletion analysis. Variants of HGL-Best were generated by primer extension using templates that contained incremental deletions from the 5' and 3' ends of the full-length 90-nt parent sequence. We compared luciferase activity values for each deletion construct to the parent sequence, which defined a core functional region of 37-nts spanning a boundary from residues 6-42 (Figure 7a). The core region is ~2-fold more active than the full-length sequence and substantial drops in activity were observed with additional deletions that extend into either end of this region. Because of its small size and high functional activity, we focused the remainder of our study on this 37-nt region of HGL-Best, which greatly simplified the engineering of recombinant expression vectors.

Given the unusual ability for HGL-Best to enhance expression and translation, we decided to determine which nucleotides within the core

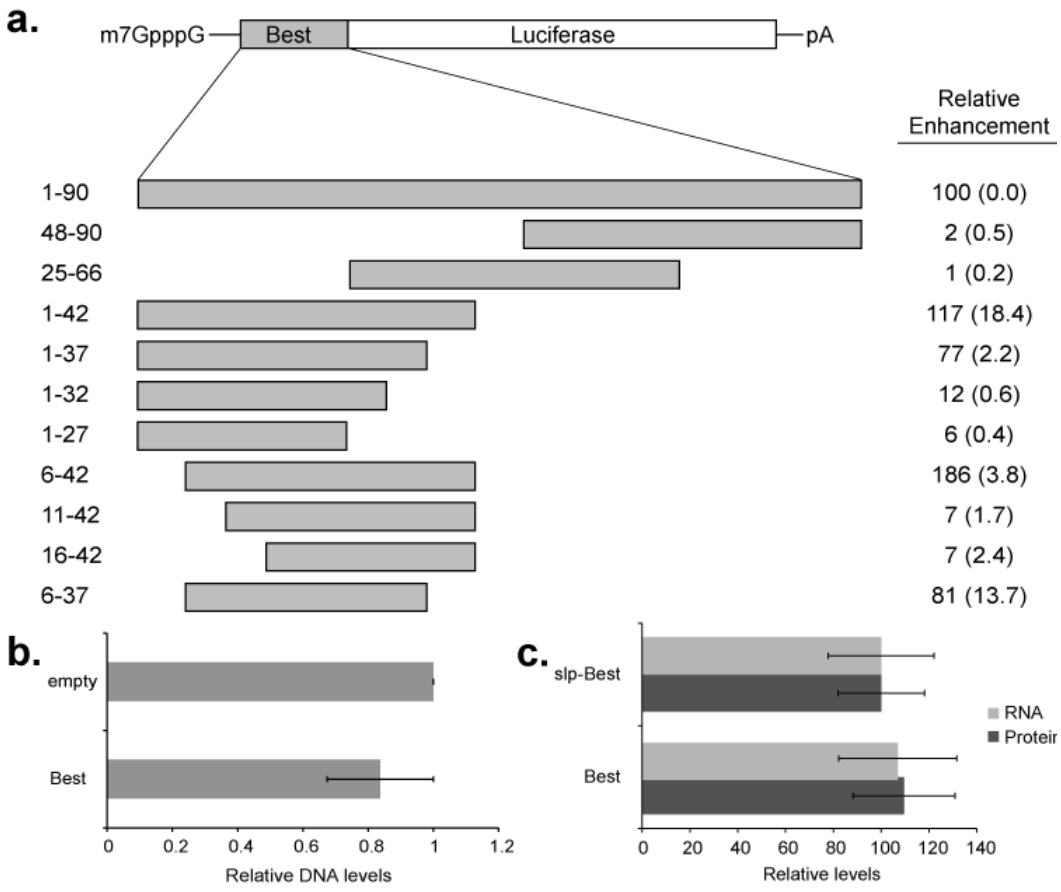

Figure 7: End Mapping Deletion Analysis. (a) Monocistronic constructs of the full-length HGL-Best (Best 1-90) and variations of this sequence were analyzed in a vaccinia transfect-infect luciferase reporter assay to identify the core functional domain of HGL-Best. Subscript labels indicate the nucleotide fragment analyzed for each HGL-Best construct. Translation activity is represented as a percentage relative to the full length HGL-Best sequence. The normalized percent error is shown in parenthesis. (b) Real time PCR analysis of plasmid DNA isolated from cells transfected with a HGL-Best or empty vector. Values were normalized to genomic DNA. (c) Luciferase RNA and protein levels produced by HGL-Best vectors with and without the SLP promoter.

region were responsible for these two activities. Rapid amplification of cDNA ends (RACE) was performed on the 5' end of luciferase mRNA isolated from HeLa cells transfected with the HGL-Best plasmid and

infected with VV. DNA sequencing of the 5' RACE product revealed the presence of 11-nts from the 3' end of the HGL-Best core preceded by a short polyA tail (Data not shown). The short polyA tail is a post-transcriptional modification that occurs as a result of VV expression. This result indicates that the first 26-nts of HGL-Best function as a VV promoter, while the last 11-nts function as a TEE. This prediction is supported by our deletion analysis study, which showed that removing 5-nts from the 3' end of the core HGL-Best sequence reduced luciferase activity ~50%, while a further deletion of 10 or more nts abolished activity altogether.

To confirm VV promoter activity of HGL-Best, we constructed a modified luciferase plasmid (F-luc-mono-Best) that removed the VV synthetic late promoter (SLP) from the vector. We then compared luciferase activities of HGL-Best to two well-established VV promoters, SLP and I1L. Comparisons were made for HGL-Best, SLP and I1L alone, and for HGL-Best in tandem with SLP and I1L in VV infected HeLa cells. After 6 hours of cytoplasmic expression, Western blot analysis indicated that vectors carrying HGL-Best, either alone or in tandem with SLP and I1L, produced substantial amounts of luciferase when compared to traditional SLP and I1L promoters (Figure 8). Prolonged exposure times with a more sensitive chemiluminescent substrate made it possible to visualize protein from the SLP construct, which appeared as a faint band

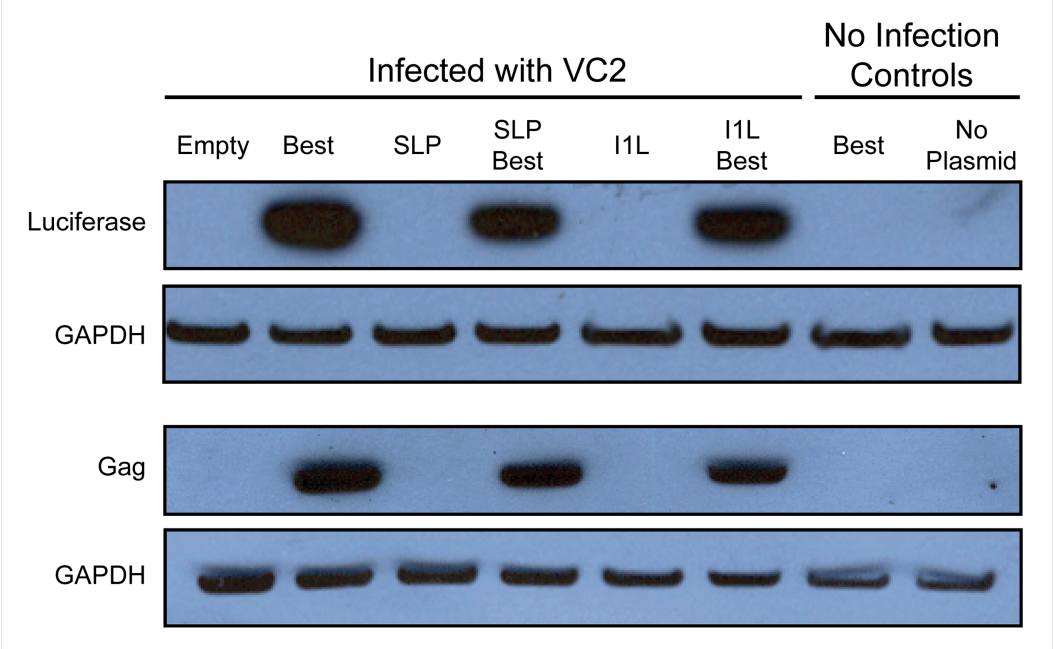

Figure 8: Comparative western blot analysis. The ability of the HGL-Best core region to enhance protein expression when placed in the 5'UTR of luciferase or the HIV gag protein was compared to the protein expression levels of vaccinia promoters SLP and I1L both alone and in combination with Best. HeLa cells were transfected with reporter plasmids and subsequently infected with vaccinia virus. Protein for western blots was harvested 6 hours post infection. Controls include protein from cells transfected with plasmid but not infected with virus and protein from untreated cells.

relative to the HGL-Best generated product. This result is consist with our original screen, which indicated that HGL-Best produced ~5,000-fold more protein than an SLP containing vector lacking a human TEE. We verified that HGL-Best could promote high expression of biologically relevant proteins by replacing the *luciferase* gene in the F-luc-mono-Best vector with *HIV Gag*. Western blot analysis closely mirrored the luciferase expression profile, indicating that HGL-Best was likely capable of broad antigen tolerance (Figure 8).

Recognizing the potential for HGL-Best as a general tool for mammalian protein expression, we compared luciferase expression profiles for HGL-Best to current state-of-the-art technology for vaccinia-based expression. In this regard, the vaccinia T7-EMCV expression system is recognized as the most efficient protein production system for VV-based expression in mammalian cell culture.[63] We therefore designed a control vector that contained an internal ribosomal entry site (IRES) from the encephalomyocarditis virus (EMCV) in place of HGL-Best. Two version of this vector were constructed; one that contained an SLP promoter for expression with the VV strains VC2 and MVA, and a second that contained the T7 promoter for expression in a recombinant VV expression system (vTF7-3) engineered to co-express T7 RNA polymerase. To facilitate a direct comparison of HGL-Best and EMCV in the recombinant T7 expression system, a modified version of F-luc-mono-Best was constructed that contained the T7 promoter upstream of HGL-Best.

An initial comparison was performed in HeLa cells for vectors carrying HGL-Best alone and in tandem with SLP and T7 to T7 EMCV and SLP EMCV. Expression profiles were monitored by time course analysis for cells infected with VC2, vTF7-3, and MVA (Figure 9). Relative to EMCV, the HGL-Best containing vectors each exhibit a rapid burst in protein production that saturates at 6 hours post-infection for VC2 and 12

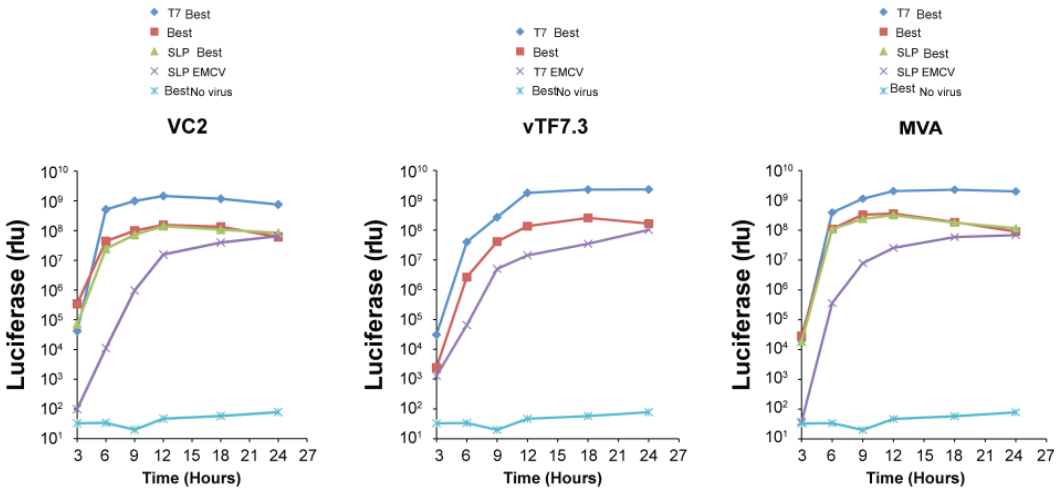

Figure 9: VV time course in HeLa cells. HeLa cells were transfected with plasmid and infected with one of three VV strains (VC2, vTF7-3 or MVA) and we obtained a time course of their expression profiles. Luciferase levels were read at 3 hour intervals for the first 12 hours and at 6 hour intervals until 24 hours. Relative Light Units are reported.

hours post-infection for vTF7-3 and MVA. Of the various HGL-Best vectors, the T7 modified version consistently produced ≥10-fold more luciferase after 24 hours of expression than HGL-Best either alone or in tandem with SLP. In comparison to EMCV, the HGL-Best vectors produce 100-1,000 fold more protein after 6-12 hours of expression. Only after 24 hours of expression is EMCV equivalent to HGL-Best alone and with SLP, which is still less efficient than the T7 HGL-Best vector.

2.5 DISCUSSION

The increasing need for recombinant mammalian proteins in medicine, agriculture and industry drives a continuous search for ways to improve the yield, quality and safety of recombinant expression systems.

Small, unmodified proteins can be produced quickly and cheaply using bacterial or yeast based systems but these organisms are incapable of generating mammalian PTMs that are sometimes crucial for activity.[46] The baculovirus insect cell system is capable of producing many but not all mammalian PTMs and has the additional drawback of being a technically difficult system to develop.[51] Only in mammalian cells can the complete set of correct PTMs be manufactured.[48] However, these systems are notorious for producing low yields and being expensive to maintain. Transient transfection of DNA into mammalian cells is quick and simple but produces the lowest yields. While stably transfected cells can have higher yields, generating a stably transfected cell line is a complicated process that can take over a year.[56] Very high protein levels have been reported from infecting cells with recombinant viruses engineered to produce the desired protein but once again, creating the recombinant virus can be a complex, time consuming process.

A simpler variation of a virus based mammalian expression system involves transfecting cells with a plasmid and then infecting them with a VV that drives the transcription and translation of the protein in the cytoplasm. To use this system, all that is required is the creation of a plasmid with the gene of interest under the control of the appropriate promoter. Here, we describe the identification and characterization of sequence HGL-Best, which can improve protein yields from the VV

transfect/infect system by 10-fold when placed in the 5' UTR of a gene. We determined that HGL-Best achieved this increase in protein production by acting both as a VV promoter (transcribing 10-fold more protein than standard promoters) and translation enhancer (producing 5-fold more protein than other plasmids tested when normalized to RNA levels). The promoter activity of HGL-Best was confirmed by its ability to perform equally well without a VV promoter.

Translation enhancing elements often contain a shorter core sequence responsible for activity. We found through end mapping deletion analysis that the 90 nt HGL-Best has a core region only 37 nt long. RACE experiments revealed that the last 11 bases of the sequence are present on the 5' end of the mRNA and suggests they are involved in translation enhancement, implying that the first 26 bases are responsible for promoter activity. The small size of the HGL-Best sequence is another advantage. Unlike EMCV and other IRES sequences used to promote translation initiation, which can be hundreds of bases long, HGL-Best is less than 50 bases and can be easily inserted into plasmids.

We considered the possibility that the high activity of HGL-Best is a phenomenon unique to luciferase, but western blot analysis of HIV Gag protein showed that the sequence could generate increased amounts of this protein as well. We isolated protein produced by plasmids containing the *luciferase* or *HIV Gag* gene under the control of HGL-Best or VV

promoters alone and in tandem with HGL-Best and found that only in the presence of HGL-Best were high levels of product detected. This demonstrates that the activity of HGL-Best is not protein specific and could be used to produce many different proteins, including those with complex PTMs. It has previously been shown that recombinant proteins produced in the VV system receive the appropriate mammalian PTMs, including glycosylation, phosphorylation, myristalation, folding and proteolytic cleavage.[63]

In direct comparisons, HGL-Best is shown to outperform known VV promoters in all tested VV strains (VC2, MVA and vTF7-3) by producing higher titers of protein at earlier time points. A time course of HGL-Best protein levels revealed the expression profile and optimum expression time of the sequence. High levels of protein can be seen as early as 6 hours post infection with the HGL-Best sequence alone whereas it takes 24 hours for the standard system plasmid (T7/SLP-EMCV) to produce similar levels. Combining our sequence with the T7 promoter creates the most effective protein expression vector. T7 Best vector was ~50,000-fold more efficient with VC2 and ~1,000 fold more efficient with vTF-3 and MVA than the conventional plasmid at 6 hours. At 24 hours the T7-Best plasmid was 10-fold higher than the standard in all viruses. Including the T7 promoter has the additional advantage of creating a more versatile plasmid that could be used in cell free or bacterial expression systems as

well. It is unclear why the addition of the T7 promoter amplifies the activity of HGL-Best when coupled with viruses such as VC2 and MVA that do not express the T7 polymerase. RACE data of mRNA isolated from VC2 and vTF7-3 infected cells that were transfected with the T7-Best plasmid show the identical sequence in the 5' UTR as seen with Best alone. This indicates that the enhancement happens at transcription, though previous studies have shown that the T7 promoter is not a significant VV promoter and the T7 sequence does not change the site of transcription initiation.[72] These results lead us to speculate that the T7 promoter works synergistically with the HGL-Best mechanism of action to increase mRNA production.

Chapter 3

CONCLUSIONS

Developing cost effective technologies that can be used to generate large quantities of human proteins in their native biological form could revolutionize biomedical research.[46] Major advantages of our technology over existing mammalian expression systems are: (i) a method for rapidly producing large quantities of biologically relevant protein in a timeframe that can effectively compete with existing *E. coli* expression systems; (ii) a readily available source of viral vectors that are easy to engineer and can be used immediately without further processing; (iii) a straightforward system that bypasses the need for stably transfected cell lines; and (iv) an expression system based on commercially available reagents that is compatible with a wide range of cell lines and vaccinia viruses, including the modified vaccinia Ankara (MVA) strain, which is non-pathogenic in humans and approved for use under standard biosafety level 1 conditions.[73] We suggest that this technology could accelerate protein production in mammalian cells for routine laboratory analysis.

In addition to improving protein yields from VV expression systems, HGL-Best also has the potential to advance VV based vaccine development. Due to thorough characterization and its success in eradicating smallpox, VV is an attractive candidate for the development of

recombinant, live vaccines.[63] Attenuated strains that are safer than the wild type VV have been developed for this purpose but unfortunately, these strains tend to make less effective vaccines.[65] Incorporating HGL-Best into these genetically modified strains could boost antigen expression levels and help to generate the immune response necessary for a successful vaccination.

REFERENCES

1. Sonenberg, N. & Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731–745 (2009).

2. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113–127 (2010).

3. Kapp, L. D. & Lorsch, J. R. GTP-dependent recognition of the methionine moiety on initiator tRNA by translation factor eIF2. *J. Mol. Biol* **335**, 923–936 (2004).

4. Konieczny, A. & Safer, B. Purification of the eukaryotic initiation factor 2-eukaryotic initiation factor 2B complex and characterization of its guanine nucleotide exchange activity during protein synthesis initiation. *J. Biol. Chem* **258**, 3402–3408 (1983).

5. Kapp, L. D. & Lorsch, J. R. The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem* **73**, 657–704 (2004).

6. Gingras, A.-C., Raught, B. & Sonenberg, N. eIF4 INITIATION FACTORS: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation. *Annu. Rev. Biochem.* **68**, 913–963 (1999).

7. Rogers, G. W., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F. *Journal of Biological Chemistry* **276**, 30914 –30922 (2001).

8. Kahvejian, A., Svitkin, Y. V., Sukarieh, R., M'Boutchou, M.-N. & Sonenberg, N. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev* **19**, 104–113 (2005).

9. Sachs, A. B. & Davis, R. W. The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. *Cell* **58**, 857–867 (1989).

10. LeFebvre, A. K. *et al.* Translation Initiation Factor eIF4G-1 Binds to eIF3 through the eIF3e Subunit. *Journal of Biological Chemistry* **281**, 22917 – 22932 (2006).

11. The scanning model for translation: an update. *J Cell Biol* **108**, 229–241

(1989).

12. Spirin, A. S. How does a scanning ribosomal particle move along the 5'-untranslated region of eukaryotic mRNA? Brownian Ratchet model. *Biochemistry* **48**, 10688–10692 (2009).

13. Schütz, P. *et al.* Crystal structure of the yeast eIF4A-eIF4G complex: An RNA-helicase controlled by protein–protein interactions. *Proc Natl Acad Sci U S A* **105**, 9564–9569 (2008).

14. Passmore, L. A. *et al.* The Eukaryotic Translation Initiation Factors eIF1 and eIF1A Induce an Open Conformation of the 40S Ribosome. *Molecular Cell* **26**, 41–50 (2007).

15. Sedman, S. A., Gelembiuk, G. W. & Mertz, J. E. Translation initiation at a downstream AUG occurs with increased efficiency when the upstream AUG is located very close to the 5' cap. *J Virol* **64**, 453–457 (1990).

16. Paulin, F. E. ., Campbell, L. E., O'Brien, K., Loughlin, J. & Proud, C. G. Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. *Current Biology* **11**, 55–59 (2001).

17. Pestova, T. V. *et al.* The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* **403**, 332–335 (2000).

18. Unbehaun, A., Borukhov, S. I., Hellen, C. U. T. & Pestova, T. V. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev* **18**, 3078–3093 (2004).

19. Acker, M. G. *et al.* Kinetic analysis of late steps of eukaryotic translation initiation. *J Mol Biol* **385**, 491–506 (2009).

20. Yueh, A. & Schneider, R. J. Selective translation initiation by ribosome jumping in adenovirus-infected and heat-shocked cells. *Genes Dev* **10**, 1557–1567 (1996).

21. Rogers, G. W., Edelman, G. M. & Mauro, V. P. Differential utilization of upstream AUGs in the β-secretase mRNA suggests that a shunting mechanism regulates translation. *Proc Natl Acad Sci U S A* **101**, 2794–2799 (2004).

22. Suzuki, Y. *et al.* Statistical Analysis of the 5' Untranslated Region of Human mRNA Using 'Oligo-Capped' cDNA Libraries. *Genomics* **64**, 286–

297 (2000).

23. Fütterer, J., Kiss-László, Z. & Hohn, T. Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* **73**, 789–802 (1993).

24. Yueh, A. & Schneider, R. J. Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. *Genes Dev* **14**, 414–421 (2000).

25. Chappell, S. A., Dresios, J., Edelman, G. M. & Mauro, V. P. Ribosomal shunting mediated by a translational enhancer element that base pairs to 18S rRNA. *Proc Natl Acad Sci U S A* **103**, 9488–9493 (2006).

26. Chappell, S. A., Edelman, G. M. & Mauro, V. P. Ribosomal tethering and clustering as mechanisms for translation initiation. *Proc. Natl. Acad. Sci. U.S.A* **103**, 18077–18082 (2006).

27. Fraser, C. S. & Doudna, J. A. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nature Reviews Microbiology* **5**, 29 (2006).

28. Hellen, C. U. & Sarnow, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev* **15**, 1593–1612 (2001).

29. Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* **62**, 2636–2643 (1988).

30. Ohlmann, T., Rau, M., Pain, V. M. & Morley, S. J. The C-terminal domain of eukaryotic protein synthesis initiation factor (eIF) 4G is sufficient to support cap-independent translation in the absence of eIF4E. *EMBO J* **15**, 1371–1382 (1996).

31. Balvay, L., Soto Rifo, R., Ricci, E. P., Decimo, D. & Ohlmann, T. Structural and functional diversity of viral IRESes. *Biochim. Biophys. Acta* **1789**, 542–557 (2009).

32. Kieft, J. S., Zhou, K., Jubin, R. & Doudna, J. A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7**, 194–206 (2001).

33. Gilbert, W. V., Zhou, K., Butler, T. K. & Doudna, J. A. Cap-independent translation is required for starvation-induced differentiation in yeast. *Science* **317**, 1224–1227 (2007).

34. Spriggs, K. A., Stoneley, M., Bushell, M. & Willis, A. E. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. *Biol. Cell* **100**, 27–38 (2008).

35. Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet* **8**, 413–423 (2007).

36. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).

37. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics* **10**, 155 (2009).

38. Kondo, T. *et al.* Small Peptides Switch the Transcriptional Activity of Shavenbaby During Drosophila Embryogenesis. *Science* **329**, 336–339 (2010).

39. Baranick, B. T. *et al.* Splicing mediates the activity of four putative cellular internal ribosome entry sites. *Proceedings of the National Academy of Sciences* **105**, 4733 –4738 (2008).

40. Bert, A. G., Grépin, R., Vadas, M. A. & Goodall, G. J. Assessing IRES activity in the HIF-1α and other cellular 5′ UTRs. *RNA* **12**, 1074 –1083 (2006).

41. Van Eden, M. E., Byrd, M. P., Sherrill, K. W. & Lloyd, R. E. Demonstrating internal ribosome entry sites in eukaryotic mRNAs using stringent RNA test procedures. *RNA* **10**, 720–730 (2004).

42. Mancuso, G. & Rugarli, E. I. A cryptic promoter in the first exon of the SPG4 gene directs the synthesis of the 60-kDa spastin isoform. *BMC Biol* **6**, 31 (2008).

43. Mathews, M., Sonenberg, N. & Hershey, J. W. B. *Translational control in biology and medicine*. (CSHL Press: 2007).

44. Chappell, S. A., Edelman, G. M. & Mauro, V. P. A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. *Proc. Natl. Acad. Sci. U.S.A* **97**, 1536–1541 (2000).

45. Dresios, J., Chappell, S. A., Zhou, W. & Mauro, V. P. An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. *Nat. Struct.*

Mol. Biol **13**, 30–34 (2006).

46. Demain, A. L. & Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances* **27**, 297–306 (2009).

47. Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **72**, 211–222 (2006).

48. Schmidt, F. R. Recombinant expression systems in the pharmaceutical industry. *Appl Microbiol Biotechnol* **65**, 363–372 (2004).

49. Chou, C. P. Engineering cell physiology to enhance recombinant protein production in Escherichia coli. *Appl Microbiol Biotechnol* **76**, 521–532 (2007).

50. Romanos, M. A., Scorer, C. A. & Clare, J. J. Foreign gene expression in yeast: a review. *Yeast* **8**, 423–488 (1992).

51. van Oers, M. M. Opportunities and challenges for the baculovirus expression system. *Journal of Invertebrate Pathology* **107, Supplement**, S3–S15 (2011).

52. McCarroll, L. & King, L. A. Stable insect cell cultures for recombinant protein production. *Current Opinion in Biotechnology* **8**, 590–594 (1997).

53. Ikonomou, L., Schneider, Y.-J. & Agathos, S. N. Insect cell culture for industrial production of recombinant proteins. *Applied Microbiology and Biotechnology* **62**, 1–20 (2003).

54. Höss, A. *et al.* Altered Phosphorylation Pattern of Simian Virus 40 T Antigen Expressed in Insect Cells by Using a Baculovirus Vector. *J. Virol.* **64**, 4799–4807 (1990).

55. Tate, C. G. *et al.* Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1610**, 141–153 (2003).

56. Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotech* **22**, 1393–1398 (2004).

57. Wurm, F. & Bernard, A. Large-scale transient expression in mammalian

cells for recombinant protein production. *Current Opinion in Biotechnology* **10**, 156–159 (1999).

58. Russell, W. C. Update on Adenovirus and Its Vectors. *J Gen Virol* **81**, 2573–2604 (2000).

59. Bett, A. J., Haddara, W., Prevec, L. & Graham, F. L. An efficient and flexible system for construction of adenovirus vectors with insertions or deletions in early regions 1 and 3. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 8802 (1994).

60. Delenda, C. Lentiviral vectors: optimization of packaging, transduction and gene expression. *The Journal of Gene Medicine* **6**, S125–S138 (2004).

61. *Sherris Medical Microbiology: An Introduction to Infectious Diseases*. (McGraw-Hill: New York, 2003).

62. Antoine, G., Scheiflinger, F., Dorner, F. & Falkner, F. G. The Complete Genomic Sequence of the Modified Vaccinia Ankara Strain: Comparison with Other Orthopoxviruses. *Virology* **244**, 365–396 (1998).

63. Moss, B. Vaccinia Virus: A Tool for Research and Vaccine Development. *Science* **252**, 1662–1667 (1991).

64. Bahar, M. W. *et al.* How vaccinia virus has evolved to subvert the host immune response. *Journal of Structural Biology* **175**, 127–134 (2011).

65. Jacobs, B. L. *et al.* Vaccinia virus vaccines: past, present and future. *Antiviral Res* **84**, 1–13 (2009).

66. Lefkowitz, E. J., Wang, C. & Upton, C. Poxviruses: past, present and future. *Virus Research* **117**, 105–118 (2006).

67. Smith, G. L. & Moss, B. Infectious poxvirus vectors have capacity for at least 25 000 base pairs of foreign DNA. *Gene* **25**, 21–28 (1983).

68. Carroll, M. W. & Moss, B. Host Range and Cytopathogenicity of the Highly Attenuated MVA Strain of Vaccinia Virus: Propagation and Generation of Recombinant Viruses in a Nonhuman Mammalian Cell Line. *Virology* **238**, 198–211 (1997).

69. Dai, K. *et al.* Pathogenicity and immunogenicity of recombinant Tiantan Vaccinia Virus with deleted C12L and A53R genes. *Vaccine* **26**, 5062–5071 (2008).

70. Perera, L. P. *et al.* Development of Smallpox Vaccine Candidates with Integrated Interleukin-15 That Demonstrate Superior Immunogenicity, Efficacy, and Safety in Mice. *J. Virol.* **81**, 8774–8783 (2007).

71. Bleckwenn, N. A., Bentley, W. E. & Shiloach, J. Exploring Vaccinia Virus as a Tool for Large‐Scale Recombinant Protein Expression. *Biotechnology Progress* **19**, 130–136 (2003).

72. Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 8122 (1986).

73. Hebben, M. *et al.* High level protein expression in mammalian cells using a safe viral vector: Modified vaccinia virus Ankara. *Protein Expression and Purification* **56**, 269–278 (2007).

74. Mackett, M., Smith, G. L. & Moss, B. General Method for Production and Selection of Infectious Vaccinia Virus Recombinants Expressing Foreign Genes. *J. Virol.* **49**, 857–864 (1984).

75. Bleckwenn, N. A., Bentley, W. E. & Shiloach, J. Vaccinia Virus‐Based Expression of gp120 and EGFP: Survey of Mammalian Host Cell Lines. *Biotechnology Progress* **21**, 186–191 (2005).

76. Carroll, M. W. & Moss, B. Poxviruses as expression vectors. *Current Opinion in Biotechnology* **8**, 573–577 (1997).

77. Yin, J., Li, G., Ren, X. & Herrler, G. Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of Biotechnology* **127**, 335–347 (2007).

78. Walsh, G. Post-translational modifications of protein biopharmaceuticals. *Drug Discovery Today* **15**, 773–780 (2010).

79. Malik, A. N., Shahni, R., Rodriguez-de-Ledesma, A., Laftah, A. & Cunningham, P. Mitochondrial DNA as a non-invasive biomarker: Accurate quantification using real time quantitative PCR without co-amplification of pseudogenes and dilution bias. *Biochemical and Biophysical Research Communications* **In Press, Corrected Proof**,

80. Elfakess, R. & Dikstein, R. A Translation Initiation Element Specific to

mRNAs with Very Short 5′UTR that Also Regulates Transcription. *PLoS ONE* **3**, e3094 (2008).