ABSTRACT

Title of Dissertation: ENHANCING GENE DELIVERY TO THE MAMMALIAN NUCLEUS FOR APPLICATIONS IN VIRAL REVERSE GENETICS AND HUMAN ARTIFICIAL CHROMOSOME DEVELOPMENT

David M. Brown,
Doctor of Philosophy, 2017

Dissertation directed by: Professor Jonathan D. Dinman,
Cell Biology and Molecular Genetics
Prof. John I. Glass, Synthetic Biology,
J. Craig Venter Institute

Delivery of transgenic DNA into mammalian cells is critical to realizing the potential of synthetic biology in advancing gene therapy, construction of entire chromosomes and production of new vaccines and therapeutics in cultured mammalian cells. New synthetic biology techniques such as rapid, inexpensive DNA synthesis have opened the door to engineering biology. However, now the delivery of these synthetic DNA constructs to the nucleus of a living cell is the limiting step in the development of these applications. Living cells possess numerous cellular barriers that a synthetic DNA construct needs to cross to be successfully expressed. In this dissertation, I explore two methods to enhance DNA delivery across the nuclear membrane barrier. First, a

plasmid delivery system was developed involving a papillomavirus scaffolding protein that when expressed by the transfected cell line, more consistently delivered plasmids bearing a specific DNA binding site to the nucleus of mammalian cells. This technique enabled us to produce infectious influenza virus more effectively when transfecting mammalian cells with DNA copies of influenza virus genes. These improvements accelerate production of vaccine against influenza virus. Second, we improved an existing method of transferring large DNA molecules cloned in *Saccharomyces cerevisiae* into cultured cells through polyethylene glycol mediated fusion of the yeast and cultured cells. Creating a reporter yeast strain allowed us to track the percentage of fused cells and the percentage that achieved YCp delivery allowing us to easily optimize the process. By synchronizing recipient cells in mitosis when the nuclear envelope is broken down we increased the delivery efficiency of large YCps ten-fold. This was accomplished by fusing yeast spheroplasts harboring large YCps (up to 1.1 Mb) with cultured cell lines. A statistical design of experiments approach was employed to further boost the vector delivery rate 300-fold to achieve a YCp delivery rate of 1/840 cells. This method was adapted to deliver a 152-kb herpes simplex virus genome cloned in yeast into mammalian cells to produce infectious virus. Finally, we discuss future applications for this technology including the development of human artificial chromosomes and applications in viral reverse genetics for vaccine development.

ENHANCING GENE DELIVERY TO THE MAMMALIAN NUCLEUS FOR
APPLICATIONS IN VIRAL REVERSE GENETICS AND HUMAN ARTIFICIAL
CHROMOSOME DEVELOPMENT

by

David M. Brown

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
 Professor Jonathan D. Dinman, Chair
 Professor John I. Glass, Co-Chair
 Professor Iqbal Hamza
 Professor Reed S. Shabman
 Professor Edward Eisenstein

Dedication

To my loving and supportive wife Lauren

Acknowledgements

None of this work would have been possible without the support from my family and friends. Thank you to my parents for supporting my interest in science and all your love and support.

I would like to thank my past mentors and teachers including my high school biology teacher Mary Jane Roethlin for first inspiring a passion for biology and excellence. Thanks to Dr. Robin Kinnel for introducing me to scientific research and giving me the opportunity to learn and work in his lab. Thanks to my undergraduate thesis advisor Dr. Mike McCormick for his engagement and enthusiasm and who further inspired me to work in research.

I want to thank committee members, Dr. John Glass, Dr. Jonathan Dinman, Dr. Iqbal Hamza, Dr. Edward Eisenstein, and Dr. Reed Shabman, for their advice and help along the way. Their suggestions have always been very constructive and have helped shape my academic and scientific thinking. I would also like to thank Dr. David Wentworth and Dr. Daniel Perez for serving on my committee in the past. I would like to specifically thank Jonathan Dinman for chairing my committee and helping me navigate my graduate career and education.

I would also like to thank the other numerous people who helped me with my scientific, academic and career development. Specifically: Alina Chan, Pamela

Table of Contents

List of Abbreviations

ADA	Adenosine Deaminase deficiency
BAC	Bacterial artificial chromosome
BARDA	US Biomedical Advanced Research and Development Authority
BPV	Bovine papillomavirus
BS	Binding site
CCU	color changing units
CDC	Centers for Disease Control
ChIP	Chromatin immunoprecipitation
CHO	Chinese hamster ovary
CMV	Cytomegalovirus
CPE	Cytopathic effect
DMEM	Dulbecco's modified Eagle's medium
DOE	Design of experiments
DTS	DNA targeting sequences
FBS	Fetal bovine serum
FRAP	Fluorescence recovery after photobleaching
GFP	Green fluorescent protein
GMP	Good manufacturing practice
HA	Hemagglutinin
HAC	Human Artificial Chromosome
HEK	Human embryonic kidney
HF	High Fidelity
HMG	high mobility group protein
HSV-1	Herpes simplex virus type 1
JCVI	J. Craig Venter Institute
MDA	Multiple Displacement Amplification
MDCK	Madin-Darby canine kidney
MEM	Minimum essential medium
MLS	mitochondrial localization signal
MMCT	Microcell mediated chromosome transfer
mtDNA	mitochondrial DNA
NA	Neuraminidase
NLSs	Nuclear localization sequence
NP	Nucleocapsid
OFAT	One factor at a time
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pDNA	plasmid DNA
PEG	Polyethylene glycol
PEI	Polyethylenimine
PFA	Paraformaldehyde
PR8X	Influenza strain A/Puerto Rico/8/34
RCA	Rolling circle amplification
RSM	Response surface methodology
RT-PCR	Reverse transcriptase-polymerase chain reaction
SCID	Severe combined immunodeficiency

SSR	Site-specific recombinase
STLC	S-trityl-L-cysteine
SV40	simian virus 40
TAR	Transformation-associated recombination
$TCID_{50}$	Tissue culture Infective Dose
tetO	Tetracycline operator
tetR	Tet Repressor
TNFα	Tumor necrosis factor alpha
TPA	2-O-tetradecanoyl-phorbol-13-acetate
tTS	transcriptional silencer
YACs	Yeast artificial chromosomes
YCp	Yeast centromeric plasmid

List of Figures

List of Tables

Chapter 1 Introduction

Introduction

The emerging field of synthetic biology has resulted from a dramatic and rapid drop in the costs of both DNA sequencing and DNA synthesis (**Figure 1-1**). Researchers can now order small custom genes for a few hundred dollars and receive them in just a few weeks [1, 2]. The synthesis of DNA has become a commodity with numerous different vendors capable of producing highly complex sequences that were impossible to synthesize just 15 years ago. Additionally, the construction of megabase sized chromosomes has been demonstrated [3, 4]. There are numerous and enormous potential applications for synthetic biology constructs: creation of complete metabolic pathways [2, 5], construction and use of synthetic genomes [6, 7], vaccine and antibody production [8], construction of industrial enzymes, as well as many other applications. Although there has been a rapid improvement in the synthesis of genetic constructs, DNA delivery methodologies have lagged other synthetic biology techniques, particularly regarding the delivery of large DNA constructs. The current state-of-the-art system to deliver human artificial chromosomes was first developed more than 30 years ago [9], and has since only improved 10-fold in delivery efficiency. This thesis describes two different projects aimed at improving DNA delivery methods so that the potential of other synthetic biology tools can be realized.
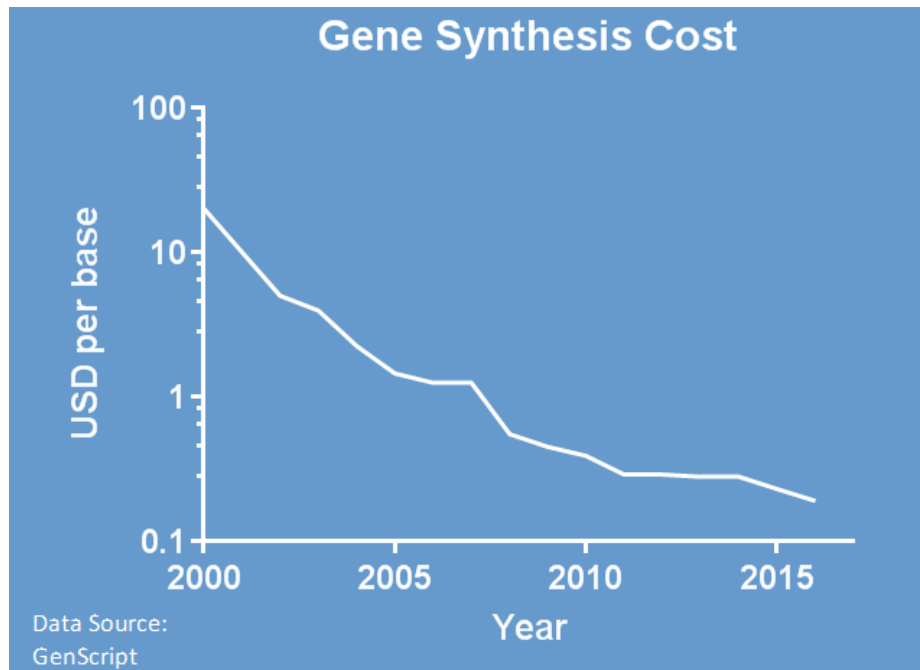
Figure 1-1 Gene synthesis cost.

Cost to consumers of a sequence verified custom DNA synthesis product, lowest

price each year from GenScript in USD

Gene Delivery

Gene delivery is the process of introducing foreign DNA into a host cell so that it can

be expressed by host polymerases. Most commonly cationic lipids are used to transfect,

the process of deliberately introducing naked DNA by non-viral methods, eukaryotic

cells. The steps of the process are shown in **Figure 1-2.** A transfected gene must cross

multiple cellular barriers to be successfully delivered. First, the cationic lipid-DNA

complex (lipoplex) must be recognized by endocytosis mediating proteins on the cell

membrane and endocytosis must occur. Next the lipoplex must escape the endosome

and disassociate. Finally, the DNA vector must cross the nuclear envelope barrier and be expressed.
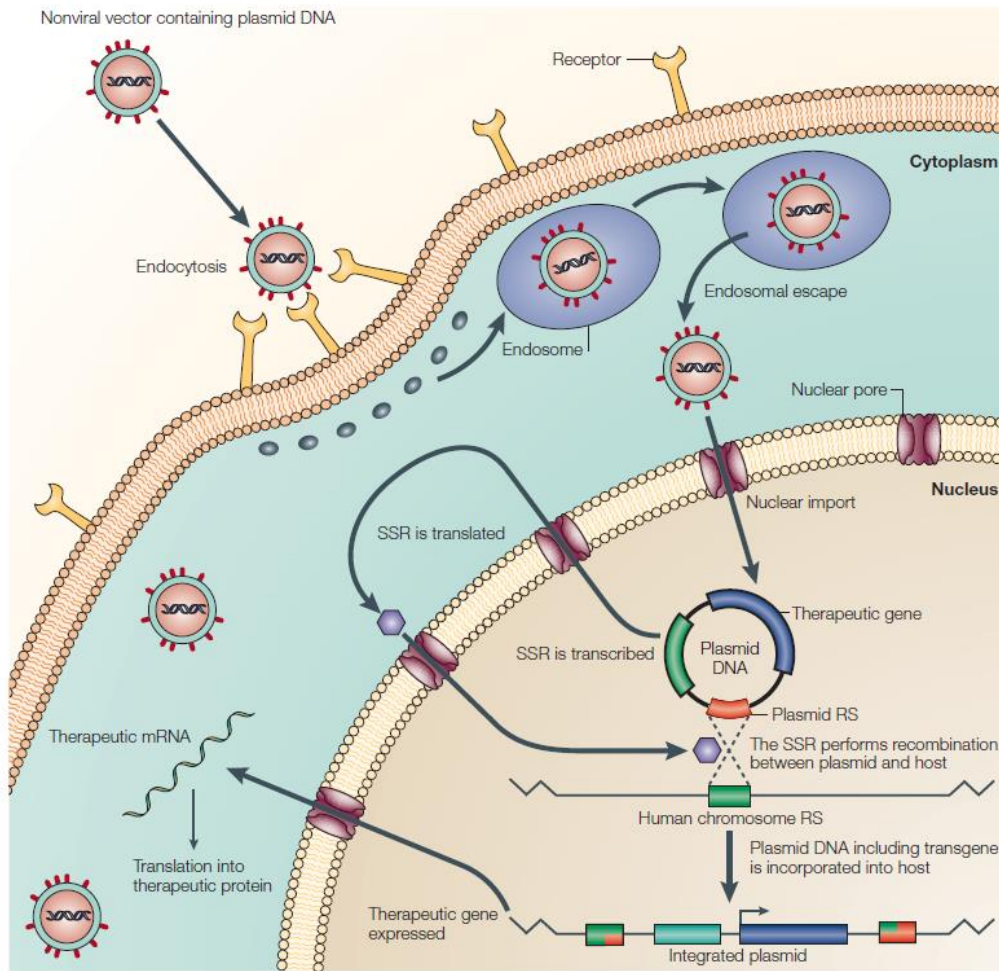
Figure 1-2 Site-directed integration of foreign DNA.

Schematic overview showing how permanent genetic changes can be made to a host cell using DNA that can mediate its own site-directed integration. Non-viral vectors that are capable of receptor-mediated endocytosis, endosomal escape, and nuclear import deliver plasmid DNA that contains a therapeutic gene and a site-specific recombinase (SSR) into the cell nucleus. The SSR is expressed and recognizes unique recombination sites (RS) within the plasmid and the host chromosome. This enables

3

recombination to take place, and consequently integrates the plasmid DNA into the host genome. The therapeutic gene can replicate during cell division and thereby provide sustained expression of the therapeutic gene product. Adapted from [10].

Barriers to Effective Gene Delivery

To have successful gene expression in cultured cells using recombinant DNA, the vector DNA must be installed in a recipient cell by crossing multiple cell barriers. First, DNA vectors must cross the plasma membrane. This must be done rapidly in order to avoid degradation from serum nucleases, cell surface associated nucleases, and to avoid destruction via an immunological response [11]. There are several methods biotechnologists use to drive DNA across cellular membranes. One of the most frequent approaches is complexing the DNA with cationic lipids [12-14]. These chemicals are well studied and numerous adaptations have been developed.

Cationic lipids are positively charged lipids that complex with DNA to form particles with a net positive charge. Extensive evidence [15, 16] shows that a nonspecific endocytosis pathway is triggered by the negatively charged heparin sulfate proteoglycan on the cell surface followed by by endocytosis. Although very efficient at inducing endocytosis, the process of using cationic lipids creates a second cellular barrier to be crossed by entrapping lipoplexes in endosomes. Endosomal escape is required because if the DNA remains in the endosome it is targeted for destruction [17-23].

4

One hypothesis for how endosomal escape occurs is through the 'proton-sponge effect' [24-26]. Briefly, cationic polymers have an abundance of unprotonated amine groups, which are thought to absorb excess protons in the endosome. Excess proton and chlorine accumulation in endosomes through transmembrane ATPase's leads to an increase in osmotic pressure. This ruptures the endosome, releasing the contents into the cytoplasm [27]. Alternatively, some cationic lipids have been designed to fuse with the endosomal membrane or form pores in the membrane [28, 29], leading to the export of DNA. Additionally, there are several commercially available transfection agents that have been optimized for cytoplasmic entry where the cationic lipid fuses with the cell membrane releasing the DNA vector into the cytoplasm directly bypassing the endosome, eliminating the requirement for endosomal escape and the endosomal barrier.

The third barrier is DNA transport into the nucleus/nucleolus. This is a major bottleneck for DNA delivery. Verkman and colleagues microinjected fluorescently labeled DNA into the cytoplasm of Hela cells and used fluorescence recovery after photobleaching (FRAP) to determine diffusion coefficients of the microinjected DNA. Using a focused laser spot the fluorescence of labeled DNA was eliminated and the fluorescence signal recovery was measured as surrounding unbleached DNA diffuse into the bleached area. They found that fluorescently labeled DNA up to 250 bp readily recovered, however DNA larger than 2 kb essentially did not diffuse [11, 30, 31]. Other groups have also observed this finding. Dowtey *et al*. found that plasmid DNA injected into the cytoplasm of rat myotubes did not diffuse away from the site of injection [32]. In contrast, oligonucleotides are readily imported to the nucleus [11].

The reason large DNA fragments do not diffuse in cells is because of the high density of eukaryotic cytoplasm and because cytoskeletal elements create a meshwork of structures making it difficult for uncompacted DNA to move freely. An additional problem is nuclear pores are far too small to allow DNA to freely diffuse into the nucleus [33-36].

Since transfecting DNA to the nucleus for gene delivery has been used for decades, DNA must be capable of moving into the nucleus by some mechanism. One hypothesis for how this might occur, here referred to as 'diffusion based translocation", is that during the M phase of cell division, the cytoskeleton is reorganized and the nuclear envelope is broken apart thus allowing for the diffusion of large DNA molecules to take place. It has long been understood that the density of dividing cells is dynamic in the range of 1.02 − 1.08 g/ml. HeLa cells have the lowest cytoskeletal density during mitosis, just as the nuclear envelope is disassembled numerous other cells lines have also shown a decrease in density just before a mitotic division [37, 38]. Successful diffusion based translocation requires that during a mitotic event, large DNA plasmids diffuse to the nuclear area to be encapsulated when the nuclear envelope reforms.

Evidence for this potential mechanism is the difficulty of transfecting non-dividing cells. Frequently, non-dividing cells cannot be transfected, or their transfection rate is extremely low. Fasbender *et al.* found that transfection of non-dividing mature airway epithelia cells using the cationic lipid DMRIE-DOPE was 90% less efficient than in younger still dividing epithelia cells [39, 40]. A different mechanism of nuclear entry must operate in nondividing cells because they can be transfected, although at low levels. Akita at al. provided additional evidence for a different mechanism operating

6

in nondividing cells. Using the cationic lipid lipofectamine-PLUS, DNA can be detected in the nucleus 1 h post transfection [17]. One mechanism by which DNA may move to the nucleus is NLS-based nuclear targeting. Numerous proteins that normally localize to the nucleus utilize nuclear localization sequences (NLSs). They have been well characterized and are discussed in **Chapter 2**. These signal sequences activate an active transport pathway capable of shuttling protein–plasmid complexes through nuclear pores [41, 42]. Nuclear pores typically have a 70 kDa or 10 nm particle diameter limit for import [43-45]; however, during NLS-based targeting the nuclear pore complex is expanded to ~30 nm, which allows the import of large molecules. With varying success, a number of groups have attempted to covalently and non-covalently bond NLSs to cationic lipids or DNA vectors themselves for improved transfection efficiency [46-49].

Different lipids almost certainly function via different mechanisms. Zabner *et al.* microinjected PEI polyplexes into a mammalian recipient nucleus, which in contrast to cationic lipids, did not affect transgene expression. The poor gene expression of directly injecting lipoplexes is likely the result of poor disassociation between the DNA and the lipoplex [30]. Some lipid carriers are capable of mediating DNA entry into the nucleus by fusing with the nuclear envelope or even by promoting uptake as a complex [50-53], while others disassociate in the cytoplasm after endosome escape [54, 55]. Some gain entry via the NLS targeting pathway, riding along the cytoskeletal network [56]. Still others simply disassociate leaving naked DNA in the cytoplasm requiring diffusion to the nucleus during mitosis. Three separate groups found strong correlations between cell division and growth on transfection efficiency when

transfections were done using poly-L-lysine, polyethyleneimine, or lipofectamine [53, 57, 58].

Diffusion and Trafficking of DNA

The diffusion of DNA vectors is responsible for a portion of every DNA delivery system. Dynamic light scattering or Brownian motion tracking [59-63] have been used to estimate diffusion coefficients of DNA. Supercoiled DNA molecules typically diffuse on the order of 10^{-8} cm^2/s, which is an order of magnitude slower than nearly all proteins [62]. This is likely because of their large, >0.2 um structure, while most globular proteins measure generally between 2-10 nm [64, 65].

Despite some difficulty in measuring diffusion rates of plasmids inside cells, plasmid diffusion coefficients (D) have been reliably modeled and experimentally verified as being proportional to the molecular weight of the plasmid (M). In general, the approximate relationship observed for plasmid diffusion is $D \propto M^{-0.6}$ [62, 63, 65, 66]; however slightly different relationships have been observed. Experimentally, a 2 kb plasmid has a diffusion coefficient of ~5.5 x 10^{-8} cm^2/s [62], a 20 kb plasmid a diffusion coefficient of ~1.5 x 10^{-8} cm^2/s [63], and a 200 kb plasmid a diffusion coefficient of ~0.25 x 10^{-8} cm^2/s. As the size of a desired DNA vector increases, relying on diffusion as the primary method by which a plasmid is transferred into a cell becomes more tenuous. As mentioned above, the cytoplasm has very high viscosity and is filled with a dense network of cytoskeletal elements, organelles, vesicles, and transational machinery that greatly slow the diffusion of plasmids once they are in the cytoplasm.

Additional evidence of this comes from disruption of the actin cytoskeleton by cytochalasin D, which improves the diffusion coefficient of microinjected DNA [67].

The nuclear envelope is the limiting step in gene delivery

It is likely that the nuclear envelope is the limiting barrier of DNA delivery. Capecchi *et al.* observed that direct microinjection of pDNA into the cytoplasm resulted in no gene expression, but direct microinjection of the same amount of pDNA into nuclei achieved more than 50% gene expression [68]. Other groups observed a similar effect [69, 70]. This observation is supported by the poor rates of diffusion of DNA in the cytoplasm; once DNA escapes from the endosome it will generally be localized to that region. In another study using cationic lipids to deliver pDNA, Zabner *et al.* showed that only 10% of cells expressed a reporter when transcription was driven using the nuclear CMV promoter, but cell expression of pDNA increased to 98% when transcription of the reporter was driven by the cytoplasmic T7 polymerase [30].

A special consideration for delivering large DNA molecules

The first consideration when attempting to deliver large DNA is its susceptibility to shear forces. DNA strands longer than a few hundred kilobases are very susceptible to mechanical shear forces. These fracture DNA along the phosphodiester bond and break it apart into smaller fragments if not handled delicately [71-73]. A few strategies have been employed to gently handle large DNA fragments. One common method to protect DNA breakage from shear forces is to purify DNA from yeast cells entrapped inside low melt agarose plugs [4, 74-79]. The plugs can be treated with zymolase,

digesting the yeast cell wall, followed by treatment with a detergent such as SDS, Proteinase K to digest nuclear proteins, and EDTA to inhibit any DNase activity. Naked DNA remains localized in caverns inside the plug and protected until ready for use. The disadvantage of this method is eventually the plug must be melted, exposing the DNA to some shear forces. This method is also impossible for use in microinjection as large agarose clusters remain even after melting, clogging the injection needle. When DNA is in a condensed form, it becomes far more resistant to mechanical shear forces, similar to a chromosome. This property has been exploited with compaction reagents as another method to protect DNA from shear forces [80, 81]. Most transfection protocols include relatively high salt concentrations to condense DNA to improve transfection [82, 83] because condensed DNA is more resistant to shear forces. Gnirke *et al.* could successfully microinject an intact 500 kb YAC by keeping DNA in 100 mM NaCl as well as using polyamines spermine or spermidine [84]. However, microinjecting large YACs is likely not a practical method of delivering large DNA. Typically, when large YACs are microinjected into a mammalian cell, there is low integration efficiency (<5%) and few fully intact YACs (~20%) [85]. Finally, another method to protect vector DNA is to keep it maintained inside a living vector cell. This can be accomplished through fusion of a donor cell to a recipient cell and is discussed in the methods of gene delivery below.

Methods of Gene delivery

In this review, we have divided gene delivery methods mechanistically into four broad methods (**Table 1-1**). The most common method used, and the method primarily

discussed so far in this review has been endocytosis-mediated methods using a cationic lipid. Although cationic lipids are the most commonly used reagents used for this method, other proteins or chemicals have also been successful at inducing endosome formation. The second method is pore generating methods; these are primarily physical methods, such as microinjection or ballistic approaches that fire tiny particles at high velocity at cells using a 'gene gun' [86, 87] to puncture one or multiple membranes resulting in DNA delivery via diffusion through the transient pores generated. Certain chemicals may also function via pore generating methods such as calcium phosphate. The third method is cellular fusion methods. Cellular fusion is when DNA from a donor cell is transferred into a recipient cell. Cell fusion is particularly useful when dealing with large DNA vectors because it protects them from shear forces. The fourth method is viral methods, i.e. transduction. These methods use native viral infection mechanisms to deliver DNA to a transduced cell.

Table 1-1 Characteristic features of major classes of DNA delivery systems.

	Principle	Advantages	Disadvantages	Examples
Viral methods	Transfer of DNA through viral infection pathway	• High transduction efficiency • Can be used with nondividing cells as well as dividing cells • in vivo trials successful	• Strong induction of immune response • Most can delivery less than 10 kb; size restricted to less than 150 kb • Careful planning needed • Oncogenesis and insertional mutagenesis possible	retroviruses, lentiviruses, HSV-1, Adenoviruses, Adeno-associated virus
Endocytosis mediated methods	Transfer of DNA complexed with polymers through cellular endocytosis pathways	• low cost • common and highly effective in vitro	• transient gene expression • low efficiency in vivo • efficiency decreases for vectors larger than 150 kb	cationic lipids, other polymers
Pore generating methods	Transfer of DNA through pores generated in the plasma membrane	• effective both in vitro and in vivo • dividing and nondividing cells can be transfected • specific tissue transfection	• Specialized equipment required • Optimization required • local damage at application site • efficiency decreases for vectors larger than 150 kb	Electroporation, microinjection, ultrasound, calcium phosphate, gene gun
Cell Fusion methods	Transfer of DNA through the fusion of cell membranes	• Large DNA vectors more than 1 Mb can be accommodated • Delivery of functional protein also possible	• only can be used in vitro • Low frequency of recipient cell transfection	MMCT, Yeast fusion

12

Delivering DNA using cell fusion methods

Membrane fusion inside the cell is essential for cell survival. Generally, the mechanism by which two membranes fuse involves two steps: aggregation and membrane merging. Two membranes will naturally aggregate via Van der Waals attraction without the hydration electrostatic forces created by water molecules which creates a strong repulsive barrier keeping the phospholipid membrane intact. Lowering the number of water binding sites or lowering the number of charges on the membrane will lead to aggregation of nearby membranes [88, 89]. Once membranes aggregate, the process still remains partially unknown, but one hypothesis is that the membranes will fuse when at least two regions of transient and sufficient bilayer disorder temporarily come into close contact [88]. Meaning that defects are required for successful cell fusion, and that membrane proteins provide these defects. The outer bilayer leaflets then merge while the inner leaflets remain distinct and no mixing of the aqueous layer is observed [90].

One of the methods of choice to induce cell fusion is to use polyethylene glycol (PEG); however, calcium, diacylglycerol, peptides, or high membrane curvature, can also induce fusion [90, 91]. Saez *et al* and Boni *et al.* showed by electron microscopy and X-ray scattering respectively that PEG-based fusion primarily functions by volume exclusion, resulting in an osmotic force that aggregates membranes together in a dehydrated region [91, 92]. PEG-mediated cell fusion has been widely adopted for a wide range of cell types, although optimization is often needed for effective delivery and to avoid toxicity and cell lysis. The process has been shown to result in transfer of

13

DNA to and from bacterial protoplasts to bacterial protoplasts [79, 93-97], from bacterial mini cells to yeast spheroplasts [98, 99] from bacterial protoplasts to mammalian cells [76, 78, 100], from yeast spheroplasts to mammalian cells [101-104], (**Chapter 3**), from yeast spheroplasts to insect cells (**Chapter 3**), and from yeast spheroplasts to avian cells (**Chapter 3**). Additionally, very large DNA molecules have been transferred this way. A 1.1 Mb mycoplasma genome was transplanted from one mycoplasma to another [79], and vectors in excess of 1 Mb have been transferred into mammalian cells from yeast spheroplasts [105](**Chapter 3**)

Fusing yeast spheroplasts

In **Chapter 3**, we developed an improved method of fusing yeast spheroplasts to cultured cell lines via PEG mediated cellular fusion for applications in DNA delivery (**Figure 1-3**). The advantage of this method is that yeast can be used as a construction platform to build large DNA molecules [3, 106, 107] that can then be delivered to mammalian cells. Yeast is an organism with extensive genetic tools thanks to it's recombination capacity [75, 108]. This means that yeast is an excellent organism in which to construct large DNA molecules as linear yeast artificial chromosomes (YACs) or as circular yeast centromeric plasmids (YCPs), and then to deliver those molecules into mammalian cells.

The spheroplast cellular fusion DNA method was first developed in the 1990s and has since been used to transfer yeast artificial chromosomes to some cell types including goat fibroblasts [109], mouse A-9 cells [105], mouse ES cells [110, 111], mouse L cells

[111] human RCC-1 cells [112], and HEK293 cells [103, 113]. In addition, yeast DNA

in excess of 500 kb has been used to create transgenic mice through spheroplast fusion

with mouse embryonic stem cells [101]. However, DNA delivery from yeast has

suffered from low efficiency generally in the range of 1 transfected cell per 250,000

recipient cells [110, 112, 114, 115]. The advantages of using YCp delivery is discussed

in **Chapter 4**.
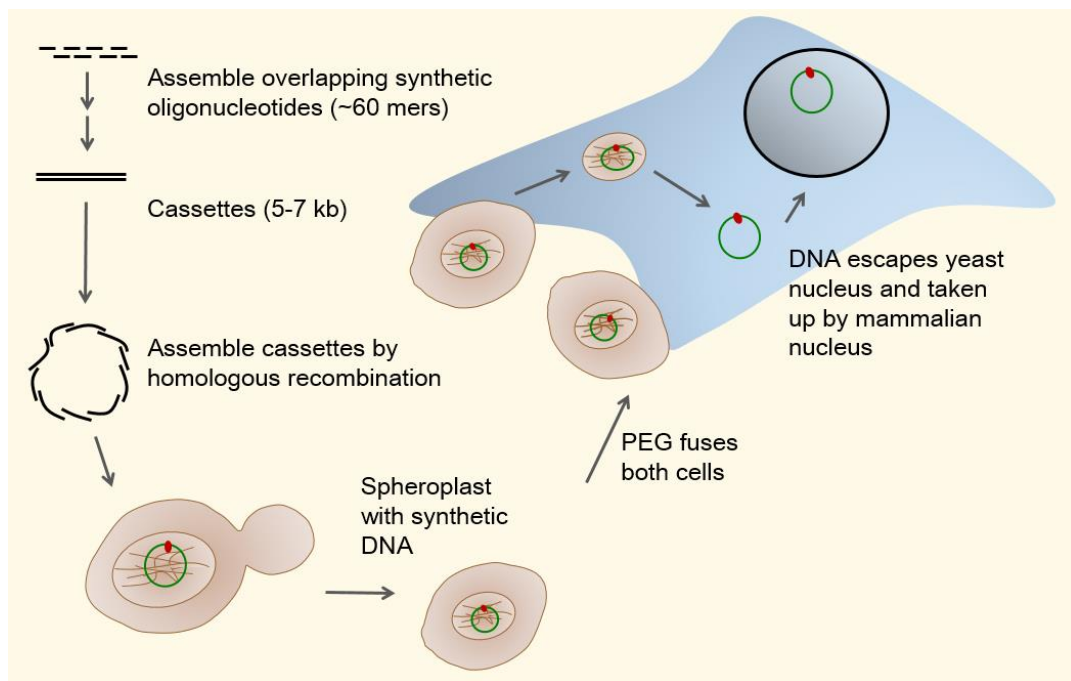
Figure 1-3 Schematic showing how large DNA vectors can be assembled and delivered via yeast fusion.

First synthetic overlapping oligonucleotides can be assembled into cassettes using

Gibson assembly (**Appendix A**). Cassettes can then be cloned into yeast via TAR

cloning. Yeast can then be spheroplasted and fused with mammalian cells resulting

in delivery of its DNA cargo (**Chapter 3**).

Microcell mediated chromosome transfer

Microcell mediated chromosome transfer (MMCT) is a technique by which chromosomes are transferred from one mammalian cell to another by cellular fusion. The process begins with arresting donor cells in metaphase using colcemid or colchicine. After prolonged exposure, certain murine cell types begin to micronucleate [116, 117], a process by which a membrane forms around individual chromosomes. The process requires a donor cell that can form microcells. Actin microfilaments are then disrupted with cytochalasin B and microcells are centrifuged through a Percoll gradient, which prevents cells from entering interphase. The purified microcells are then filtered through a 3 µm filter to select for microcells containing just one chromosome. Microcells can then be fused to the recipient cell through PEG-mediated fusion. After selection, desired cell hybrids can be identified (**Figure 1-4**).

The technique was first developed by Fournier and Ruddle in 1977 [118] when mouse microcells were fused with CHO cells. MMCT has been used as a delivery tool in engineering human and mammalian artificial chromosomes, genetic mapping, identification of tumor suppressor genes, analysis of genomic imprinting, and production of transgenic animal models of disease [119-122]. For 30 years, few improvements were made to the process, a yield frequency of microcell hybrids by PEG-induced fusion was no more than $1 \times 10^{-6} - 1 \times 10^{-5}$ [115]. However, recent advancements using Measles Virus (MV) fusogenic proteins [123] instead of polyethylene glycol (PEG) have enhanced the efficiency 10-fold. For years, d*e novo* kinetochore formation only occurred in a very limited number of cell lines, HT1080

16

cells being the most used. Using common cell lines such as HeLa, results in 100% integration of the delivered vector. Recently, Ohzeki *et al.* described a method of increasing the number of cell lines by regulating the histone H3K9 acetyl/methyl balance [124].

Donor cell line, transfected with selectable dominant markers in red. E.g *neo* or *bsr* resistance

Exposure to colcemid to block cells in mitosis

Eventually, nuclear envelope forms around individual chromosomes.

Centrifuging in the presence of cytochalasin B, to collect micronucleated cells. Micro-cells are then filtered through membranes to remove microcells with more than one chromosome.

Recipient mammalian cell line.

Generated micro-cell hybrid cell lines are then grown under antibiotic selection. FISH analysis is performed on resulting clonal cell lines.

Figure 1-4 Diagram of microcell-mediated chromosome transfer protocol.
Adapted from: [117].

Pore generating methods of gene delivery

Pore generating methods to deliver DNA are conceptually very simple. These are methods that utilize physical, electrical, or chemical means to create temporary weak points in a cellular barrier, usually the cell membrane, of a target cell to allow a DNA vector to be delivered into the cell via diffusion. Briefly we will review two of the most common methods.

Needle microinjection

Conceptually, the simplest method of DNA delivery is direct injection of DNA into the nucleus of cells. This technique utilizes a thin glass microneedle to puncture both the cell membrane as well as the nuclear envelope. The needle remains inside the cell long enough for the DNA vector to diffuse across the length of the needle into the nucleus. Microinjection has been common practice since the 1980s [68]. The technique has been used to produce recombinant cell lines [84], transgenic animals [125], and for mitochondrial DNA transfer [126]. However, the practicality of the process is limited due to requiring specialized equipment or an automated control system. Moreover, few labs possess the expertise required to execute the process correctly, and since only one cell can be transformed at one time, obtaining a clone is laborious and time consuming. Although not regarded as an efficient method of transfecting cells, important mechanistic characteristics of DNA delivery can be identified when performing microinjections into specific regions of a cell.

18

Electroporation

Electroporation is a physical technique that induces the uptake of DNA into a recipient cell by increasing the permeability of the cell membrane though exposure to a controlled electromagnetic field [127]. Electroporation was first used by Neumann *et al* in mammalian cells. DNA delivery occurs by transiently destabilizing the cell membrane, thus allowing the entry of both DNA and other macromolecules into the cytoplasm [128]. Since the invention of the technique in 1982 [128], it has been used to deliver DNA to numerous cell types including yeast and bacteria. Electroporation can be highly efficient at delivering plasmids; up to 80% of surviving cells will be transformed. The efficacy of gene transfer with this method is affected by different factors that are both physical (pulse duration, field intensity and electrode geometry) and biological (cell size, shape, and density) [129, 130].

Other Endocytosis mediated methods

Cationic lipids are the primary reagent used in endocytosis mediated methods and have already been discussed. Extensive variations of cationic lipids and cationic proteins have been developed. Some chemicals, such as DEAE-dextran, have been abandoned due to low levels of gene transfer, cellular toxicity, and lack of biodegradability [131]. In contrast, the polycation polyethylenimine (PEI) approach has shown promise in enhancing transfection efficiency. In the presence of PEI, pDNA condenses to form a positively charged DNA–PEI complex that protects the DNA from serum nucleases. This complex then interacts with the negatively charged proteoglycans present on the cell membrane and promotes the non-specific endocytosis of the complex [132].

19

Disadvantages of viral methods

Several extensive reviews of viral delivery methods are published [129, 133-135] and the topic is beyond the scope of this project. Viral delivery methods can be extremely efficient at transducing DNA into a target cell, however there are several disadvantages relative to nonviral delivery systems. First, viral delivery methods can trigger innate immune responses in cells resulting in destruction of the DNA being transduced or other major alterations of cell physiology that compromise the DNA transfer or activity of the transferred DNA. Second, viral methods have been shown to have significant safety concerns due to insertional mutagenesis. Third, viral vectors have restrictive limits for DNA that can be incorporated into a vector. The most commonly used viruses have very limited vector capacities, sometimes as small as a few thousand base pairs (adeno-associated virus or adenovirus). The safety of using viral vectors in gene therapy is discussed in **Chapter 4**.

Significance of the project: Enhancing DNA delivery by targeting the nuclear membrane barrier.

This thesis focused on two different projects to enhance DNA delivery by eliminating the nuclear membrane barrier. **Chapter 2** focuses on strengthening the transfection efficiency and consistency of small plasmids. The goal of the project was to understand and enhance influenza virus reverse genetics (the process of producing infectious virions by transfecting cells with DNA copies of viral genomes in vectors that enable viral gene expression and genome replication) to enable a more rapid production of

vaccines in response to pandemics and tangentially assist in other areas of gene delivery, including other viral rescue systems or gene therapy. Generation of vaccines that protect against influenza virus infection is still a lengthy process. Pandemic influenza infections occur periodically, afflicting millions of people. With the possibility of a highly pathogenic H5N1 pandemic, it is important to enhance the capacity of vaccine manufacturers to create large quantities of vaccine before the peak of viral infection. The 2009 H1N1 pandemic demonstrated critical inadequacies in our influenza virus vaccine production pipeline. That year, even though vaccine manufacturers produced a vaccine against the H1N1 strain in record time after the announcement of the pandemic by the World Health Organization, useful amounts of vaccine were not available until after the pandemic had peaked [136]. The vaccine was developed and manufactured fundamentally using the same processes that have been in use for the last 70 years. In **Appendix A** we used synthetic DNA based on viral genome sequences and reverse genetics methods to install synthetic DNA in mammalian cells to produce the exact strain of virus needed to produce the desired vaccine. To assist US public health agencies in preparation for potentially dangerous influenza strain in 2013, using viral genome sequence data only, our team produced a vaccine strains in less than 5 days as opposed to the 35 days that could have been needed using conventional approaches [137]. This suggests that reverse genetics will be an important tool to generate pandemic influenza virus vaccines more rapidly. Further improvements to the system is detailed in **Appendix B utilizing *in vitro* cloning**. Unfortunately, there are still problems associated with reverse genetics due in part to the limited number of cell lines approved for vaccine production and the

21

occasional difficulties faced when producing a difficult-to-rescue strain [136, 138-140]. Increasing the likelihood of rescuing a difficult to rescue influenza virus strain would make the vaccine development process more reliable. In **Chapter 2**, we hypothesized that targeting the nuclear membrane would enhance gene delivery. To test this, we examined four different nuclear targeting systems utilizing two different mechanisms of action designed to enhance transfection efficiency and consistency to produce virus from plasmids encoding the 8 influenza genome segments. DNA for influenza gene segments was created through a rapid assembly technique seen in **Appendix A and B**.

Chapter 3 focuses on enhancing the delivery of large YCps using yeast spheroplast fusion with the goal of using the YCps as a HAC. DNA vectors for mammalian cells are decades behind the development of other organisms. In contrast to easy-to-manipulate *S. cerevisiae* or *E. coli,* researchers generally rely on transient or integrating vectors as their only genetic tools for mammalian cells. But for decades researchers have strived to generate HACs to address these limitations of mammalian vectors [141]. HACs are anticipated to allow mega base scale cloning capacities, copy number control, long-term gene expression, avoidance of anti-viral immune responses, and reduced mutagenesis from insertion into host chromosomes. To date HAC development has consistently relied on excising HACs by joining centromeric fragments (alphoid satellite DNA) from natural human chromosomes into bacterial, phage and yeast artificial chromosomes (BACs, PACs and YACs) [142-145]. Multimerization of the satellite DNA leads to a lack of fidelity and bacterial sequences

can lead to increased mitotic instability, rearrangement and concatenation in HACs [146].

Manipulation and design of HACs is currently only possible in Chinese hamster ovary (CHO) cells using Cre-LoxP recombination [143, 147], and these HACs have several difficulties. The HACs are too large to manipulate in bacteria or yeast therefore cloning and moving the HACs must occur via microcell-mediated chromosome transfer (MMCT). This takes months and requires an expertise rarely found outside the labs that invented the HACs [143, 147]. Further, existing HACs lack an engineered design and often consist of megabases of random rearrangements and long strings of multimerization [148]. To date there are no HACs in distribution that can compare to BACs and YACs in terms of ease-of-use and design precision capabilities. In contrast, YACs containing well-defined basic functional units have been available since the 1980s [149, 150] and are convenient to manipulate using conventional cloning protocols [150]. YACs have been used by hundreds of studies and different researchers, extrapolating its utility so far as to study human genetic elements in YAC transgenic mice [105], and to perform multiplex assembly of exogenous biosynthetic pathways to produce diverse compounds in yeast [151]. Generating a HAC with simple methods of manipulation and transfection analogous to YACs holds the same potential of transforming current practices and widening the breadth of possible experiments and synthetic biology designs in mammalian cell lines. Recent advances in understanding the chromatin requirements for centromere formation may lead to the possibility of engineering a HAC [124, 152-156]. However, extensive engineering of large DNA

23

constructs is currently not possible in mammalian cells and existing delivery methods are insufficient.

In **Chapter 3** we have developed a much simpler and more consistent approach for putting potential HACs into mammalian cell lines via yeast spheroplast fusion. The advantage of this method is that it does not require isolation of the HAC, which could result in chromosome shearing. It has long been known that spheroplasted yeast cells can be fused with mammalian cells in the presence of polyethylene glycol [110]. *S. cerevisiae* is the best choice for both HAC construction and the delivery of HACs into mammalian cells for three main reasons. First, putting the HAC in yeast allows us to build millions of base pairs of DNA into the HAC, which can be moved directly from yeast cells into mammalian cells using cell fusion. The largest reported YAC is 2.3 Mb [157], however it is likely that larger YACs can be made since native yeast chromosomes have been mapped to 5.8 Mb. Second, as discussed earlier, there is great difficulty in manipulating large segments of DNA without shearing it when it is unprotected. DNA constructs delivered from fused living cells are protected from shear. Third, extensive genetic tools are available to engineer large DNA molecules in yeast.   Indeed large DNA molecules, in excess of 1 Mb, have been synthetically engineered in yeast [158] using both low G+C% DNA [7] and high G+C% DNA content [107]. In **Chapter 3** we hypothesized that the nuclear envelope cellular barrier was the limiting step in the YCp delivery process and that enhancing YCp delivery could occur if the nuclear envelope cell barrier was eliminated. We synchronized

recipient cells in M-phase of the cell cycle when the nuclear envelope is broken down, and observed a 10x increase in YCp delivery rates.

Chapter 2 Enhancing influenza virus reverse genetics using a nuclear targeting mechanism

BACKGROUND

Influenza Biology and Disease

It is estimated that influenza infections are associated with an annual average of 25,000 deaths worldwide, reaching over 1,000,000 deaths during pandemic years [159, 160]. As a member of the orthomixovirus family, influenza is a negative sense, segmented, RNA virus containing 8 different viral segments that encode 10-12 different proteins.[161] Infections are characterized by the sudden appearance of respiratory distress as well as other symptoms, including fever, muscle or body aches and fatigue [162-164]. Transmissible though the air by aerosol, influenza A viruses are able to infect a number of different animal hosts, including swine, horses, ferrets, seals, humans, and a large variety of birds [164].

An influenza infection begins with the viral hemagglutinin (HA) surface protein attaching to a cellular sialic acid receptor and inducing receptor mediated endocytosis [165]. The acidification of endosomes induces uncoating of the virions and allows the release of viral RNA into the cytoplasm. The viral RNA, complexed with the nucleocapsid (NP) protein, is transported to the nucleus via NLS based targeting

nuclear transport. The NP protein contains two different nuclear localization signals (NLSs) responsible for targeting to the nucleus [166]. Once in the nucleus, replication and transcription of viral RNA can take place.

Vaccination

Presently, the most effective way to combat influenza is through vaccination. The two surface glycoproteins (HA and NA) are the primary antigens that a host is able to recognize and therefore are purified and used in vaccines [162]. Influenza A is currently categorized into 17 HA types and 9 NA types [167]. Combinations of the two proteins result in virus subtypes that are commonly phrased as H1N1, H3N2, etc. Due to the segmented nature of the virus, coinfection of a single host cell with two different influenza virus strains can result in different combinations of viruses being produced. For example, the HA from one virus could be replaced with an HA from another virus, resulting in a distinct strain. Theoretically, coinfection with two influenza strains could result in 254 different progeny viruses [168]. In addition, because of the rapid rate of mutation, new seasonal flu vaccines must be developed each year.

The first step to develop a vaccine is the production of a seed virus. Conventional influenza vaccine seeds are produced via reassortment from a coinfection, also known as "viral mating", of a circulating strain and a strain adapted to growth in fertile chicken eggs or cultured cells. This lab adapted strain is often called a backbone strain. Typically, the aim of the coinfection is to obtain a reassortant strain with the HA and

NA encoding genome segments donated from the circulating strain and the 6 remaining segments are donating from a backbone strain, such as the high growth strain A/Puerto Rico/8/34 (PR8X) virus. The reassortant should contain the 2 glycoproteins from the circulating strain and the 6 housekeeping genes from PR8X, creating a 6+2 reassortant, also called a vaccine seed. This requires screening numerous samples because the desired 6+2 reassortant is one of many strains produced in the coinfection. Isolating this strain is a bottleneck in producing the initial virus from which large preparations are produced [168, 169]. This is how vaccine seeds were first made in the 1950s. Fortunately, reverse genetics technologies utilizing molecular biology have enabled much more rapid production of the same 6+2 virus starting from DNA copies of influenza virus RNA genomes

Reverse Genetics

Reverse genetics enables the generation of influenza virus entirely from DNA vectors transfected into a mammalian cell. This technique has the added advantage of not needing a helper virus or to screen for 6 + 2 reasortants [161, 168]. To produce an infectious virus, 12 different RNA molecules must be produced from these DNA templates. The 8 negative sense viral RNA segments (HA, NA, PA, PB1, PB2, PA, NP, NS, and M) and 4 positive sense mRNA molecules (PA, PB1, PB2 and NP) capable of producing the viral polymerase and nucleoprotein must be produced (**Figure 2-1**). Reverse genetics is a technique involving 8-12 different plasmids transfected into a host cell culture with each plasmid delivered to the nucleus and nucleolus producing one or two RNA molecules [161, 168, 170]. One significant advancement is the use of

bidirectional promoters to reduce the plasmid number and increased viral generation efficiency [171]. This technique involves using a strong pol II promoter (such as CMV) to produce positive sense RNA (mRNA) and a pol I promoter to produce negative sense RNA. When positioned on opposite ends of a viral segment, both positive sense and negative sense RNA can be transcribed from a single region of plasmid DNA (pDNA). To generate influenza virus successfully, a process called "rescuing the virus," a single mammalian nucleus must obtain at least one copy of each of the 8 plasmids. The mechanism by which pDNA is delivered to the nucleus is reviewed in **Chapter 1**.
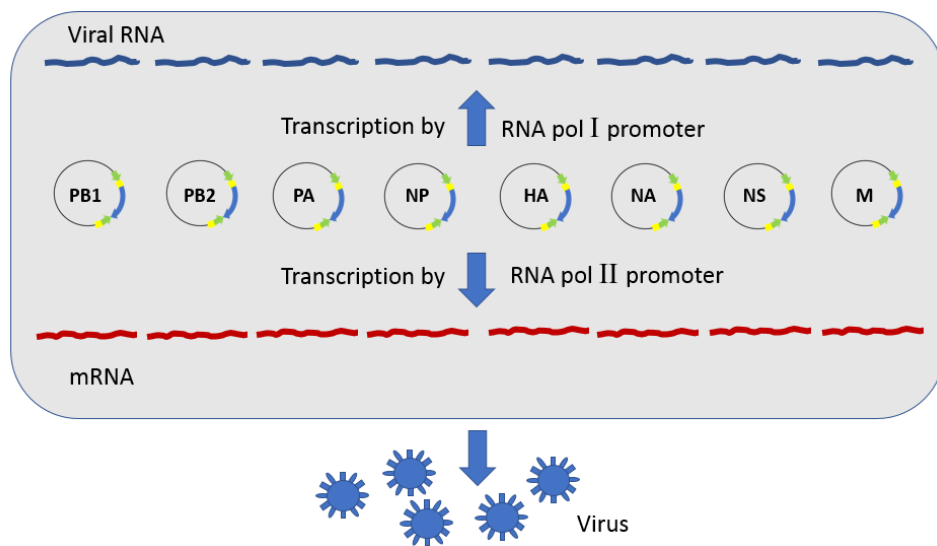
Figure 2-1 Schematic of the reverse genetics of influenza.

Based on figure from [161], green designates promoters and yellow terminators.

INTRODUCTION

Described in **Appendix 1**, a collaborative team at Novartis Vaccines and the JCVI endeavored to produce a manufacturing process to rapidly produce vaccine seeds in response to pandemics. Starting with oligonucleotides synthesized from sequence data, influenza vaccine seeds could be generated in just 5 days, trimming one month or more from the current 5-month process going from identification of a new pandemic virus to delivering an influenza vaccine to the public. Although there has been tremendous improvement in both DNA synthesis technology and vaccine seed development, some influenza strains lacked rescue consistency, meaning they failed to be rescued every time a transfection was performed. Some difficult-to-rescue strains could be rescued more consistently when an alternative backbone strain was used, and every strain attempted could eventually be rescued if enough replicates were performed. However, improvements to viral rescue efficiency and consistency could improve the vaccine development platform. Herein we demonstrated that the rescue consistency of transfected plasmids coding for the difficult-to-rescue strain, A/Perth/16/2009, and bearing a specific DNA binding site, can be enhanced if the transfected cell line expresses a papillomavirus scaffolding protein.

Enhancing Rescue Efficiency

The overall aim of this chapter is to enhance the viral rescue and transfection efficiency of influenza gene segments. Because plasmid delivery to the nucleus is the likely limiting step in the reverse genetics process, at least for difficult-to-rescue viruses, we

formed two general hypotheses for how to enhance gene delivery to the nucleus defined here:

1) *Diffusion based nuclear targeting* – Enhancing gene delivery using a scaffolding protein that binds both to the host chromosome and to the delivered gene during a cellular division resulting in co-localization of a transfected plasmid and the newly reformed nucleus.

2) *NLS based nuclear targeting* - Enhancing gene delivery using proteins bearing nuclear localization signals (NLSs) bound to a transfected plasmid.

To test these two hypotheses four plasmid systems (E2, TFAM, NF-KB, SV40) were developed using different binding proteins and binding sites each described below. Each system was evaluated by three different metrics: transfection efficiency, rescue efficiency and rescue consistency. Transfection efficiency was determined by the minigenome assay and is a measure of the percentage of cells successfully producing viral polymerase expression. Rescue efficiency was determined by viral titers in $TCID_{50}$ assays and is a measure of viral production efficiency. We determined rescue consistency by performing multiple transfections of the same DNA template and evaluating the percentage of times virus was rescued and is defined as the likelihood of obtaining successful virus generation from a single transfection experiment.

Diffusion based nuclear targeting

Diffusion based nuclear targeting during cellular division is designed to assist simple diffusion based translocation discussed in **Chapter 1**. In this hypothesized mechanism

to enhance gene delivery, a nuclear shuttle protein that binds both to a specific DNA sequence and to a host chromosome packaging protein, a scaffold can form between the pDNA and the mammalian chromosome, allowing for the co-localization of the plasmid with the host chromosome. The scaffolding ensures that the plasmid is retained in the nucleus once a nuclear envelope reforms. Nuclear transport is still only possible during a cellular division, but during a division process, there is a greater chance pDNA will localize to the nucleus once it reforms. To date almost no studies have attempted to enhance gene delivery by utilizing a diffusion based targeting mechanism. Those studies that did primarily used polycationic or polyanionic domains [172]. Enhancing gene delivery through diffusion based nuclear targeting using a scaffolding protein appears to never have been attempted prior to this study.
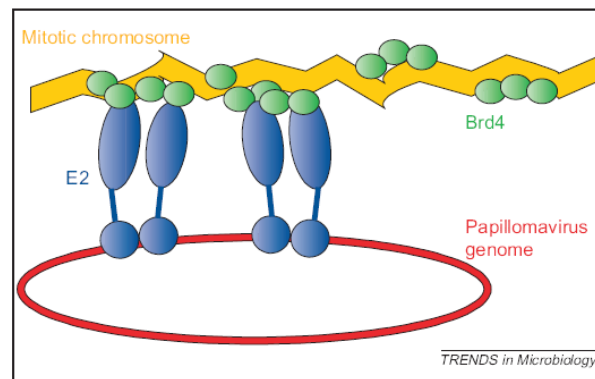
Figure 2-2 Tethering of papillomavirus genomes to mitotic chromosomes.
Figure from [173].

Testing if the E2 system enhances gene delivery

Papillomaviruses are extrachromosomal dsDNA viruses containing a circular genome approximately 8 kbp in size, encoding 8 major proteins [174]. The E2 protein is

31

transcribed early in the papillomavirus life cycle and is responsible for a number of functions including persistence, replication, transcription, and apoptosis regulation [175]. Importantly, E2 binds to an $ACCN_6GGT$ motif that is found in the replication origin and transcriptional regulatory regions of the viral genome. Consecutive repeats of the motif result in strong binding [173]. Bovine papillomavirus (BPV) uses E2 protein to bind both a sequence specific region of papillomavirus DNA as well as BRD4 (eukaryotic chromosomal packaging protein). This scaffolding action allows BPV to persist in daughter cell nuclei following a cellular division seen in **Figure 2-2**. McBride *et al.* tested whether the papillomavirus E2 protein could maintain plasmid persistence in *Saccharomyces cerevisiae*. Plasmids containing the yeast autonomous replication sequence replication element but with the centromeric element replaced by E2 binding sites could be maintained in yeast cells expressing both the E2 protein and the BRD4 element [176].

The E2 protein has 2 binding sites, one of which binds to BRD4 and another that binds to a specific DNA sequence motif. BPV E2 protein will be used to develop a diffusion based nuclear targeting mechanism [174].

To test if diffusion based nuclear targeting can enhance gene delivery we constructed plasmids containing an E2 binding site and a cell line expressing the BPV E2 protein. Cells expressing the transgene BPV E2 should be able to bind pDNA carrying the sequence specific DNA binding site, referred to as the E2 system. This protein-DNA

complex can then diffuse freely during a cellular division. Once in contact with a mammalian chromosome the complex can bind, and therefore co-localize, to the newly formed nucleus of the daughter cell, potentially enhancing gene delivery. We hypothesize that plasmids containing influenza gene segments and E2 binding sites (E2 BS) utilizing the E2 system will scaffold to the host chromosome and enhance gene delivery and further, enhance influenza rescue.

NLS based nuclear targeting

NLS based nuclear targeting employs a nuclear localization signal (NLS) to actively translocate DNA into the nucleus via the proteins importin alpha and beta (Imp α/ β) [41, 177-180] using dynein-mediated transport along microtubules. Proteins containing a NLS are recognized by Imp α, then docked by Imp β and nucleoporins, which allows the complex to be translocated to the nucleus through the nuclear pore complex. NLSs are well understood and are capable of moving DNA fragments of at least 8 kb to the nucleus [41, 181]. The first NLS discovered is a seven amino-acid NLS (PKKKRKV) from simian virus 40 (SV40). A single copy of this NLS present on a peptide is sufficient to localize foreign DNA to the nucleus [41, 177, 181-183]. Numerous other similar NLSs have been discovered, termed classic NLS, these are generally stretches of basic amino acids. Classic NLSs have shown affinity for Imp α/ β resulting in up to 100-fold faster nuclear import than proteins lacking NLS. Classic NLSs and have been shown to translocate proteins to the nucleus in a variety of different cell types [184]. The highly basic residue (RKKRKKK) has been identified as a nucleolar localization signal present on the cellular protein NIK [185, 186]. Behr *et al*. synthetically

conjugated a 3.3 kbp gene to a single NLS peptide and demonstrated that was sufficient to carry the DNA to the nucleus [187]. However, the process of attaching the NLS peptide to pDNA was cumbersome, time consuming, and required special equipment. Coutelle *et al.* demonstrated that a fusion protein comprising the tetracycline repressor protein and the SV40 NLS can be used to enhance transfection efficiency of a DNA vector containing a tetracycline repressor binding site 30 fold [181]. In addition to classical NLSs, non-classical NLSs have been identified. These lack stretches of basic amino acids and are characterized as being largely hydrophobic. Native influenza virus contains two non-classical NLSs on the NP protein [188]. Since NP complexes with viral RNA, NLS based targeting is critical for efficient flu replication during an infection. To test if NLS based nuclear targeting could enhance gene delivery of influenza gene segments and enhance vial rescue, we created three plasmid delivery systems, detailed below, utilizing nuclear shuttle proteins bearing NLSs.

Testing if the TFAM system enhances gene delivery

TFAM is a 204 amino acid protein capable of binding DNA in a sequence specific manner. It contains two high mobility group protein (HMG) domains responsible for DNA binding [189-191]. Native TFAM binds to a site on mitochondrial DNA (mtDNA). Bennett et al. [192] used this information to design a recombinant TFAM molecule engineered with a mitochondrial localization signal (MLS) to deliver mtDNA cargo to the mitochondria of a mammalian cell. The mtDNA rapidly translocated into the mitochondria through the use of this protein. Since this protein demonstrated success in translocating large pieces of DNA, it was selected as a potential nuclear

shuttle protein to test NLS based nuclear targeting. Based on Bennett et al. [192] design we replaced the MLS with a NLS to create a modified TFAM protein, along with plasmids bearing the TFAM binding site to create the TFAM system. We hypothesize that TFAM conjugated to a NLS will bind to plasmids containing the TFAM binding site (TFAM BS) and then be used as a nuclear shuttle protein importing plasmids to the nucleus enhancing gene delivery. Additionally, we hypothesize that gene delivery of plasmids containing influenza gene segments using the TFAM system will be enhanced, and in turn enhance vial rescue.

We created The TFAM system by transfecting cells expressing the NLS-TFAM nuclear shuttle protein, produced from a modified version of the human TFAM coupled to a nuclear/nucleolar localization signal. Plasmids containing the TFAM binding site outside of the influenza coding region of DNA were used in virus rescue. We hypothesized that the NLS-TFAM nuclear shuttle protein will bind to the TFAM binding sites on plasmids and be recognized by the NLS cellular transport machinery shuttling the plasmids to the nucleus.

Testing if the SV40 system enhances gene delivery

Native cellular proteins bearing NLSs will bind sequence specific regions of foreign DNA targeting them for import. Termed DNA targeting sequences (DTS), they are sufficient to localize pDNA to the nucleus containing those specific DNA sequences. The first DTS identified was a region of the SV40 promoter and enhancer [193]. A 72 bp repeated region present on the pDNA was shown to be sufficient to localize the

pDNA to the nucleus [194]. Since then, several different DTSs have been identified [195, 196] including an Epstein Barr virus sequence capable of increasing apparent gene expression by 100 fold compared to plasmids lacking a DTS [184]. These DTSs seem promising; however, their success has been varied. Other groups have not been successful in replicating the increase in transport efficiency using these DTSs [197, 198]. The varied successes of NLS mediated transport may be due to the different conditions used by various groups. We employed the SV40 DTS to create the SV40 system. The SV40 system functions similarly to the TFAM system in that it tests NLS based nuclear targeting, however the process is simplified because only the SV40 DTS is needed on the plasmids for import using the NLS cellular machinery. The SV40 system will test NLS based nuclear targeting having one set of plasmids bearing the SV40 DTA and another set where it is absent. We hypothesize that gene delivery of plasmids bearing a DTS as well as influenza gene segments be enhanced, and in turn enhance vial rescue.

Testing if the NF-KB system enhances gene delivery

An inducible DTS has also been identified, NF-κB is a DNA binding protein that is expressed by almost all mammalian cell types including HEK293 cells and MDCK cells after induction by TNFα or TPA (2-O-tetradecanoyl-phorbol-13-acetate) [199]. In the NF-KB system, NF-κB can function in the same way the TFAM protein would function (NLS based nuclear targeting). NF-κB possess a sequence specific DNA binding site as well as a NLS [200]. DNA vectors that contain a repetitive binding site recognized by NF-κB have been shown to enhance transgene delivery to the nucleus

36

by NLS based targeting after induction with TNFα or TPA [199]. This method has the advantage of better controllability of nuclear shuttle protein expression level. The NF-κB binding site used in this project is listed in the sequences section (**Appendix E**). We employed the NF-κB inducible DTS to create the NF-κB system. The NF-κB system is also similar to both the TFAM and SV40 systems in that it tests NLS based nuclear targeting. However, the process is more controllable because the nuclear shuttle protein can be induced. We hypothesize that gene delivery of plasmids bearing the NF-κB binding site (NF-κB BS) as well as influenza gene segments be enhanced and controlled when induced, and in turn enhance vial rescue.

MATERIALS AND METHODS

Cells.

Human embryonic kidney (HEK293) cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). Madin-Darby canine kidney (MDCK) cells were maintained in minimum essential medium (MEM) supplemented with 5% FBS.

Plasmid construction

To create the proposed DNA transport systems, four different sets of plasmids containing influenza genes as well as a DNA binding region for each system were created, each with a corresponding binding site for each system. These constructs were assembled using Gibson Assembly [201]. We made 11 plasmids for each system. Each system has the corresponding binding site as well as an influenza gene controlled by bidirectional promoters. The genes used for each construction system were from

influenza virus strains A/Puerto Rico/8/1934 (PR8) (PA, PB1, PB2, NP, NS, M, HA, and NA) and from A/Perth/16/2009 (HA and NA) to construct the 11th plasmid pPolI-NS-EGFP a corresponding binding site was added to pPolI-NS-EGFP. A description of each plasmid is shown in **Table 2-1** including a description of the modifications for each engineered protein.

Mini-genome replication assay

The mini-genome assay has been performed previously [202]. Briefly, to construct the pPolI-NS-EGFP plasmid, the EGFP gene, flanked by the NY1682 NS noncoding region, was cloned into the recombination-based influenza reverse-genetics plasmid pG26A12 (modified from pHH21 [203, 204]) between the RNA polymerase I promoter and terminator for expression of vRNA-like negative-sense EGFP RNA, which can be replicated and transcribed by the viral RNA polymerase complex. The transfected pPolI-NS-EGFP expresses GFP when four influenza plasmids (PA, PB1, PB2, and NP) are also expressed.

Virus generation

Influenza virus was generated via reverse genetics. Co-cultured HEK293 cells expressing the gene of interest were transfected with the eight DNA fragments carrying the cDNA of each influenza gene segment, as well as any relevant binding site, using a protocol adapted from Hoffmann and Webster [171]. Briefly, according to the manufacturer's instructions, either 5 µl of X-tremeGENE 9 (Roche Diagnostics) or 3 µl of *Trans*IT®-LT1 (Mirus) were used per µg of DNA (unless otherwise specified). The transfection agent was diluted to 200 µl using serum-free DMEM, mixed with

DNA for each segment, incubated at room temperature for 20 - 30 min, and then added in a drop wise manner to the monolayer of cells. Viral titer was quantified at reported timepoints via $TCID_{50}$ with MDCK cells.

$TCID_{50}$ Assay

The 50% Tissue culture Infective Dose ($TCID_{50}$) is a measure of infectious viral titer. The endpoint dilution assay quantifies virus by the amount of virus required to kill 50% of the host cells, producing a visual cytopathic effect (CPE). First 1×10^4 MDCK cells are seeded per well in a 96-well flat bottom culture plate in DMEM containing Pen-Strep and 10% FBS at 37 °C and 5% CO_2. Once MDCK cells are ~95% confluent in the plate, growth media is removed from the cells and are washed 3 x with 200 µl of PBS per well and replaced with 100 µl infection media (DMEM containing Pen-Strep and 2% BSA, 0.5 µg/ml TrypZean (Sigma)). In a separate 96-well round bottom culture plate 180ul infection media is added to each well. Then 20 µl of each viral sample is added to the first row. Changing tips, each row is serially diluted 10x by the following row by adding 20 µl of the previous row to the following row until the plate is complete. Then 100 µl of each virus dilution is added to the corresponding well on the flat bottom plate with MDCK cells. Typically, influenza virus will produce complete CPE in a culture in 3-5 days. $TCID_{50}$ results are calculated from wells run in triplicate using the infectivity calculator generously provided by Brett D. Lindenbach [205].

GFP quantification by flow cytometry

Plates were observed on a Nikon Eclipse Ti fluorescence microscope. To prepare samples for flow cytometry fused HEK293 cells were trypsinized, filtered using a 70-micron filter (Corning), washed, and resuspended in PBS. Using an Accuri C6 flow cytometer (BD Bio-sciences. Analysis was based on light-scatter and fluorescence signals produced from 20 mW laser illumination at 488 nm. A 533/530 nm filter was used to detect GFP. The flow cytometer was routinely operated at the Slow Flow Rate setting (14 lL sample/minute). Data acquisition for a single sample took 3–5 minutes.

Results

Method of plasmid constriction

Genes of interest for expression plasmids were constructed from 60 base oligonucleotides ordered from IDT (San Jose, California). Oligonucleotides were assembled using Gibson Assembly [201] so that the first 30 bases were complementary to the previous oligo. Detailed methods for constructing influenza genes are presented in **Appendix A** and an explanation of error correction is presented in **Appendix B**. The same general approach was used to construct E2, TFAM, binding sites, etc. The oligos form an ungaped construct of the gene, assembling dsDNA with 40 base overlaps at the completion of the Gibson Assembly reaction. This construct was PCR amplified, error corrected using Errase (Novici Biotech, Vacaville, California) and then further assembled into an expression vector using 40-base overlaps. Mammalian cells expressing genes of interest were seeded and transfected with the influenza gene segment containing plasmids. Genes of interest were assembled into a pUC based

expression plasmid containing resistance markers for ampicillin and neomycin, allowing for the creation of stable cell lines expressing the nuclear targeting protein. Under the control of a different CMV promoter, mCherry, was used as a screen for positive cells (**figure 2-3**).

A
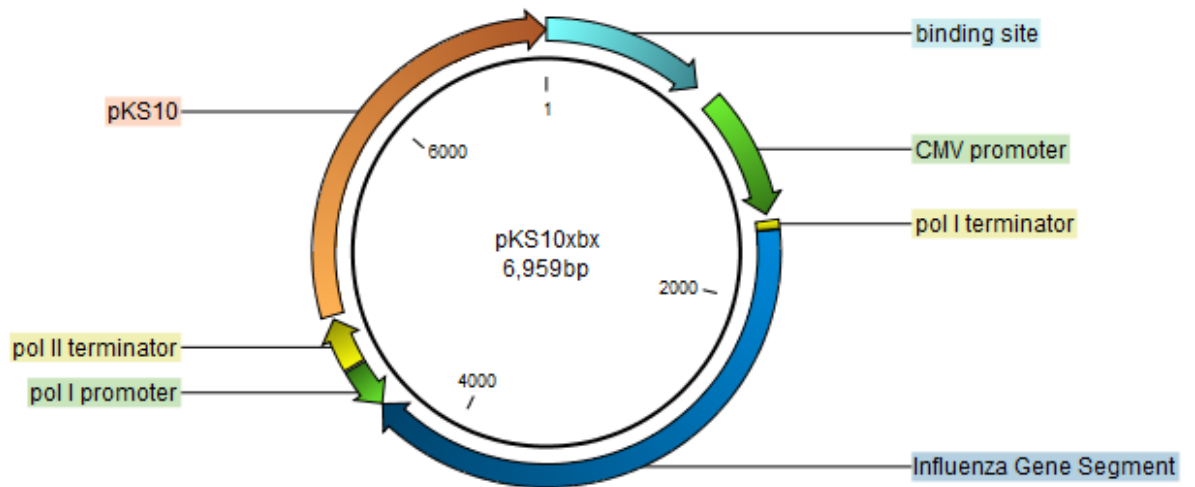
B
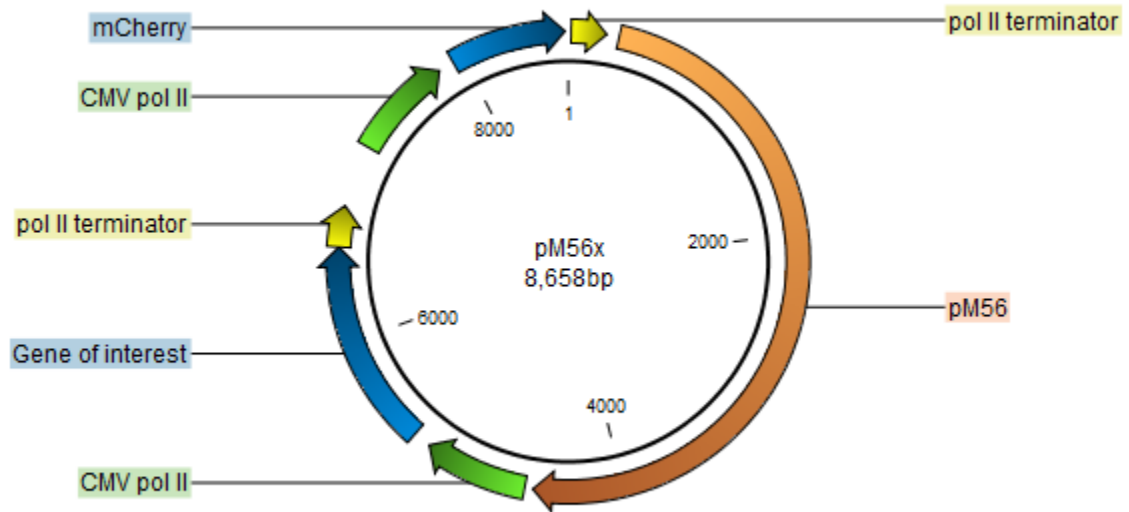
Figure 2-3 Diagram of plasmids used in this study.

(A) Schematic diagram of influenza gene expression plasmid. Influenza coding sequences are blue, pKS10 sequences are orange, promoter sequences are green, terminator sequences are yellow, and binding site is aqua. The binding site and influenza gene segment can be swapped out depending on the expression system. (B) Schematic diagram of the expression plasmids used to generate stable cell lines. Expression genes are blue, pM56 sequences are orange, promoter sequences are green, terminator sequences are yellow, and binding site is aqua. pM56 contains the neomycin resistance marker.

Evaluation of each nuclear delivery system

The E2, TFAM, NF-KB and SV40 nuclear delivery systems were evaluated independently. First, the viral rescue efficiency was determined by evaluating viral titer over time via $TCID_{50}$ assay. If there was an indication that virus generation efficiency was improved than a minigenome assay was performed to determine the transfection efficiency. For the most successful system, E2 system, the rescue consistency, the likelihood of successfully generating virus, was also determined by performing several transfections with a difficult-to-rescue influenza strain (A/Perth/16/2009).

The E2 papillomavirus protein

Testing the E2 system involved the creation of two different E2 expression plasmids (pM56X-E2 and pM56X-E2a). pM56X-E2 was generated by adding the gene for the BPV-E2 protein into the expression plasmid pM56. pM56X-E2a was generated by modifying pM56X-E2 by removing the coding region for the DNA binding domain of BPV-E2 at the C-terminus. The BRD4 binding region should remain intact. Removing the DNA binding domain should disable E2 tethering.

HEK293 cells expressing E2 shows some improvement in influenza rescue efficiency with difficult-to-rescue strains.

The cell line adapted, easy-to-rescue strain, PR8X showed no change in rescue efficiency when E2, E2a, or no BPV protein was expressed (**figure 2-4A**). However, when the HA and NA segments from the difficult-to-rescue strain A/Perth/16/2009

43

were used, virus generation efficiency was significantly enhanced beginning 72 h post

transfection (**figure 2-4B**). However, this difference is in averaged viral titers not the

result of an increase in viral replication rate, or greater virus generation per transfection.

Instead it is due to inconsistencies in whether a transfection was successful at rescuing

virus. Not every transfection of A/Perth/16/2009 successfully rescued virus, therefore

if a greater proportion of transfections were successful, higher averaged titers are

observed. To determine if E2 influenced the virus generation or replication rate the

unsuccessful transfections were removed from the dataset and only the samples that

produced CPE are plotted in (**figure 2-4C**). Once removed, no difference in viral

generation/replication rate is observed.

Non-optimal transfections are more consistent when E2 is expressed

No difference in transfection efficiency was observed via minigenome assay when the

manufacturer's specification for the amounts of transfection reagent and DNA to be

used in a reaction were followed (data not shown). However, when either a sub-optimal

amount of cationic lipid or DNA was used, the transfection efficiency drops

substantially. We tested 100 ng of each plasmid was used with 1 µl of Mirus

transfection agent and obtained <1% transfection efficiency with each of the 5 plasmids

(**figure 2-5**) as measured by flow cytometry. However, when the sub-optimal

transfection is used with E2 expressing cells this sub-optimal transfection still produced

a high efficiency transfection. When E2 is expressed by transfected cells, there may be

a more consistent transfection even if conditions are non-optimal.

Difficult-to-rescue strains are much more likely to be successfully rescued from cells expressing E2.

The rescue consistency of A/Perth/16/2009 was evaluated by performing blocks of 6 independent transfections on the same day with cells seeded and grown at the same time. **Figure 2-6A** shows the 72 h timepoint of 1 set of 6 transfections. The results show that 2 out of 6 transfections generated virus without any E2 expression, 4 out of 6 transfections generated virus when E2 was expressed, and 1 out of 6 when E2a was expressed. Viral titer amounts are not important to measure virus rescue consistency because once virus is generated it will continue to replicate. To measure viral rescue consistency one must look at the proportion of transfections that rescued virus, therefore no conclusions can be drawn from **Figure 2-6A** alone since the sample size is too small. Experiments like this must be combined and are reported in **figure 2-6B**. The experiment reported in **figure 2-6A** was repeated 8 times so that 8 blocks of 6 transfections were performed, for a total of 48 transfections per condition. Each block was performed on a different day. In addition to the conditions represented in panel A, plasmids lacking the E2 binding site were also tested for viral rescue consistency. The percentage of transfections that scored positively was tabulated and summarized in (**figure 2-6C**).

Transfected cells expressing functional E2 have a greater likelihood (25% compared to ~4%) of generating difficult-to-rescue virus ($p < 0.01$); however easy-to-rescue virus is generated with or without E2 expression at an observed 100% rate (only one block

performed). The effect is lost when using plasmids lacking the E2 binding site. This supports our hypothesis that E2 is binding to both the transfected plasmid and the BRD4 protein on the host chromosome scaffolding the plasmids near where the nucleus reforms following mitosis, enhancing the consistency of virus rescue and transfection.

A
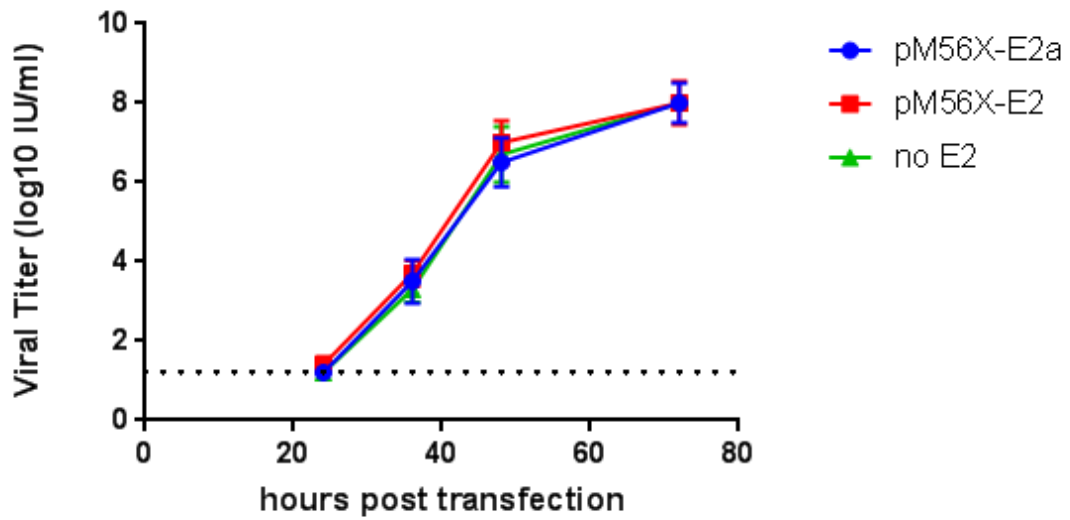
B
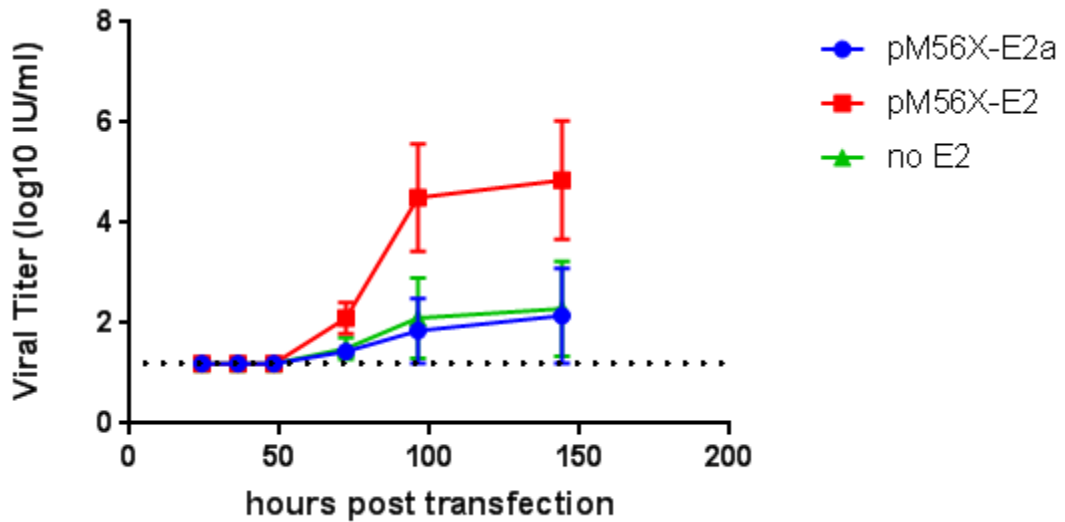
C
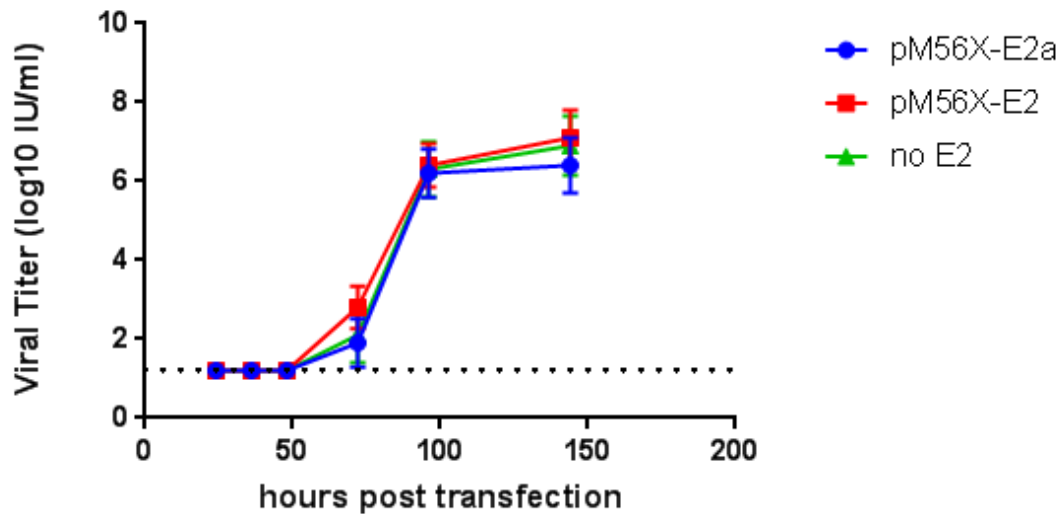
Figure 2-4 Viral Titers in cells expressing E2

The HA and NA segments from PR8X (panel A) and A/Perth/16/2009 (panel B) along with the PR8X backbone containing the E2 binding site were used in a transfection with HEK293 cells already expressing the proteins E2 or E2a. Supernatant samples were collected at select time points and viral titer was measured via a $TCID_{50}$ assay. Shown are the averaged data from 6 independent transfections. Not every transfection in panel B successfully generated virus. Removing transfections from the dataset that resulted in no virus generation is shown in panel C. The dotted line indicates the lower limit of detection.
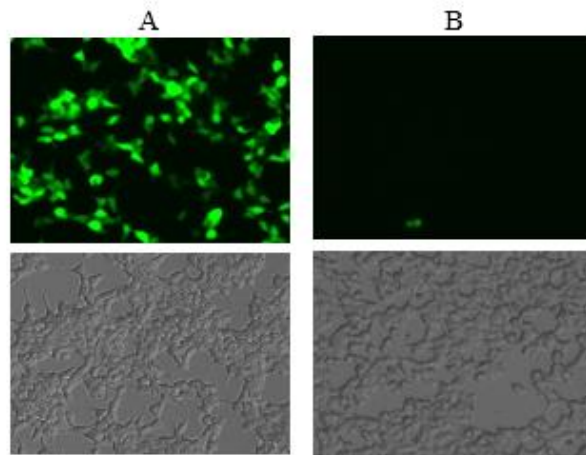
Figure 2-5 Effect of E2 on transfection efficiency using a sub-optimal transfection.
Influenza segments from PR8 (PA, PB1, PB2, NP) and the pPolI-NS-EGFP influenza
expression reporter containing the E2 binding site were transfected into cells expressing
E2, or no E2.   One hundred ng of each plasmid were transfected using 1 µl of Mirus
transfection reagent. A: HEK293 cells expressing E2, B: HEK293 cells not expressing
any E2. Micrographs are representative samples.

A
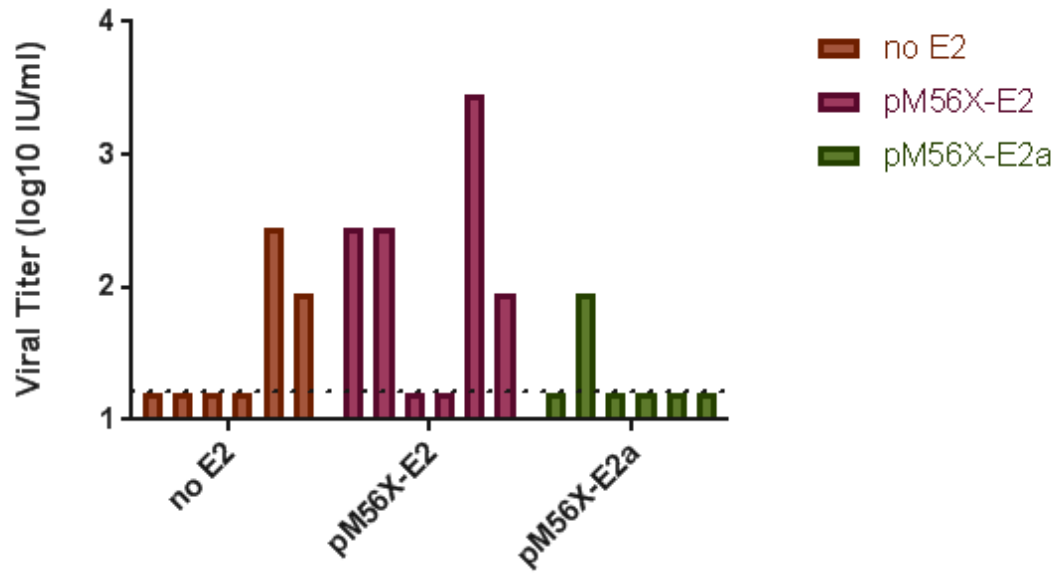
B
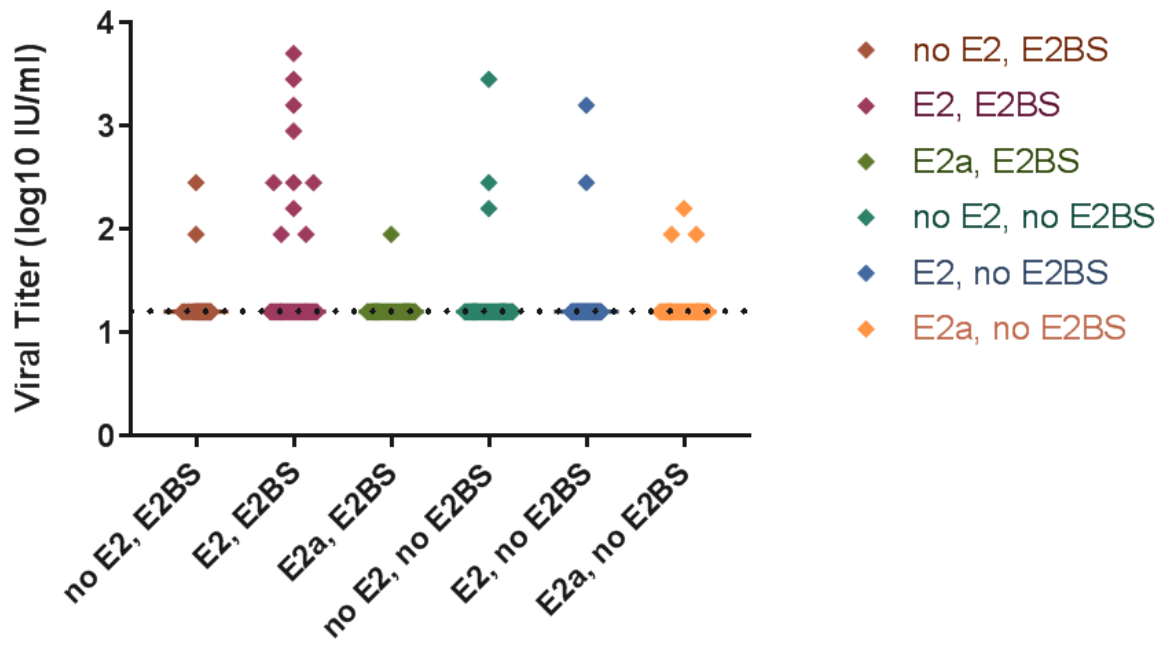
C
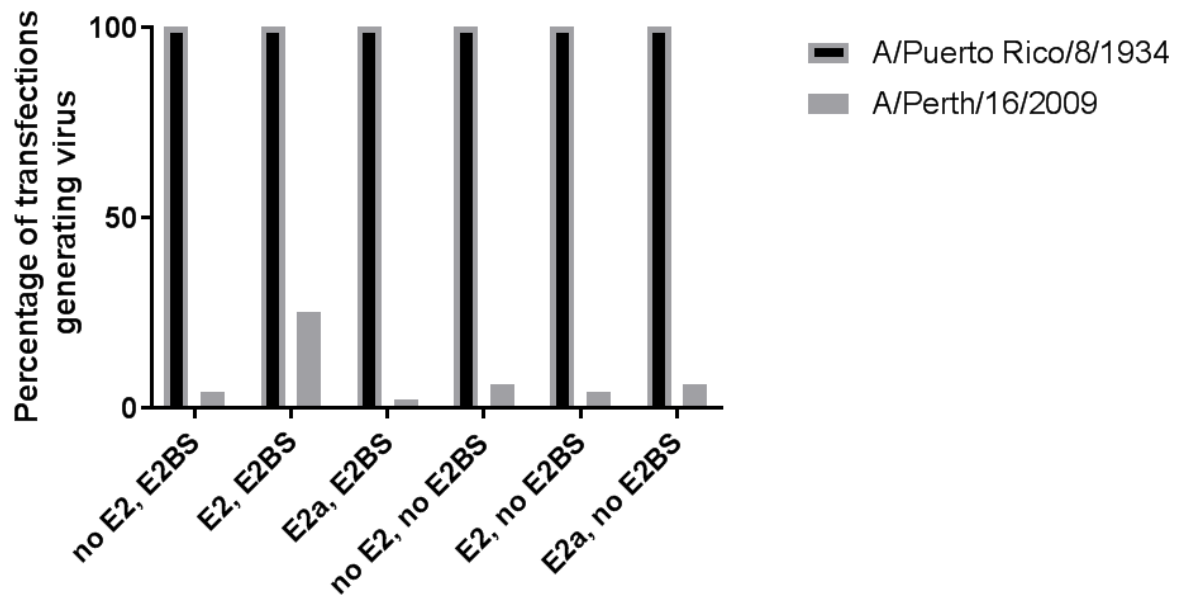
Figure 2-6 Determining rescue consistency.

The HA and NA segments from A/Perth/16/2009 and the 6 backbone segments from PR8X, each containing the E2 binding site, were used in a transfection with HEK293 cells already expressing the E2 protein, E2a, or no E2 protein. Each transfection condition was repeated 6 times. $TCID_{50}$ assay was performed 72 hours post transfection to determine if virus was produced (panel A). The dotted line indicates the lower limit of detection. The experiment in panel A was repeated, 8 blocks of 6 experimental conditions were performed and are summarized in (panel B) for a total of 48 transfections for each condition. If any virus was produced above the limit of detection that transfection was scored as producing virus. The percentage of transfections that scored positively was tabulated and summarized in (panel C).

Modified TFAM protein

We hypothesized that a modified TFAM protein (NLS-TFAM) could bind a 5'-TGTTAGTTGGGGGGGTGACTGTTAAAAGT-3' motif present on an influenza virus reverse genetics plasmids transfected into a mammalian cell, and then transport the plasmids to the nucleus and nucleolus thereby improving the efficiency of the viral reverse genetics process. Preliminary results (not shown) indicated that the full TFAM protein with the NLS had a lower transfection efficiency compared to control. We hypothesized that it is possible that the TFAM protein itself is responsible for a decrease in viral expression efficiency. While the TFAM protein may effectively transport the DNA to the nucleus, it may be interfering with the viral expression once there. TFAM binds strongly to the TFAM binding site designed and incorporated on the plasmid; however, it also binds nonspecifically to other DNA sites. If TFAM is expressed at a high level in these cells, it may be possible that the plasmids are covered in numerous TFAM molecules interfering and silencing successful expression. To eliminate this problem, we modified the TFAM molecule so it no longer contained the nonspecific DNA binding sites. Eliminating the non-specific DNA binding region of TFAM can be accomplished by removing the HMG-2 region of the protein. The protein domain HMG − 1 is responsible for the specific DNA binding and is fully functional without HMG − 2.

A total of three TFAM molecules were produced as described in the methods section. First native TFAM expression was driven by a CMV promoter in a stable cell line of HEK293 cells from pM56X-TFAM. Next, AAA-TFAM was produced from pM56X-

AAA-TFAM by removing the HMG-2 domain from TFAM. Finally, pM56X-NLS-TFAM was produced by conjugating a NLS to AAA-TFAM as described in the introduction. We added the modified human TFAM genes to HEK293 cells and transfected those cells with influenza virus reverse genetics plasmids harboring a TFAM binding site. Unmodified TFAM results in slower virus generation than TFAM lacking the HMG-2 region. Viral titers are not improved when plasmids containing a TFAM binding site were transfected into HEK293 cells expressing NLS-TFAM (**figure 2-7**). NLS-TFAM protein showed no improvement in rescue or transfection efficiency. Results indicated no significant difference in virus generation efficiency. Similar TFAM expression levels were observed via Western blot.

A
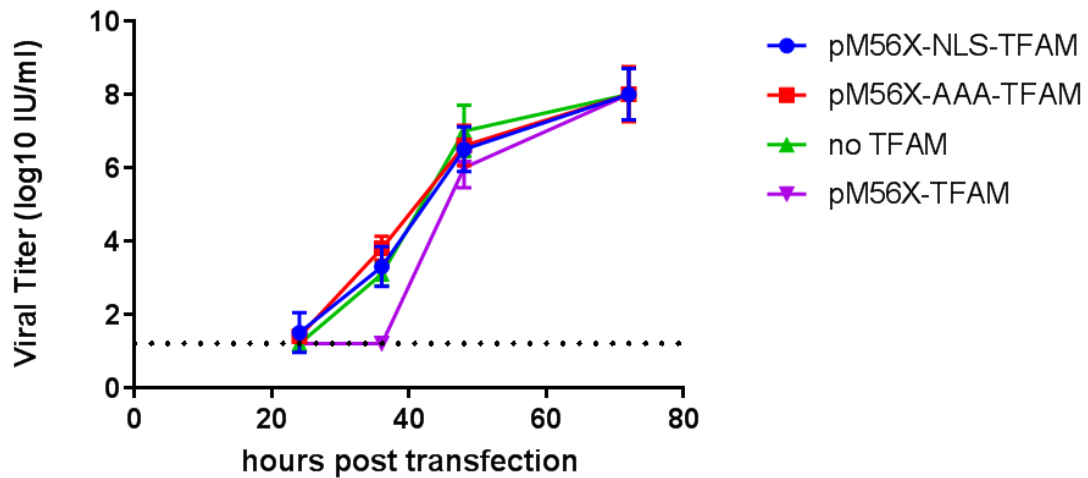
B
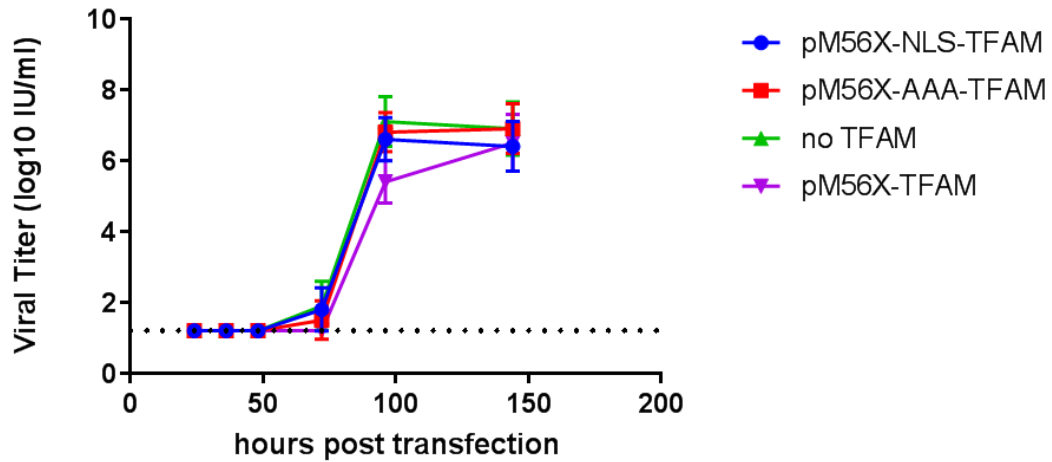
Figure 2-7 Viral Titers in cells expressing TFAM.

The HA and NA segments from PR8X (panel A) and A/Perth/16/2009 (panel B) along

with the PR8X backbone containing the E2 binding site were used in a transfection

with HEK293 cells already expressing the NLS-TFAM or AAA-TFAM. Supernatant

was collected at select time points and viral titer was measured via a $TCID_{50}$ assay.

Shown is the averaged data from 3 independent positive transfections. The dotted line

indicates the lower limit of detection.

Plasmids containing the NF-κB binding sites transfect more efficiently than plasmids lacking in HEK293 cells induced with TNF-alpha.

A dose response curve measuring GFP expression via mini-genome assay was observed

when TNFα (Miltenyi Biotec, Inc.) was added in increasing amounts to HEK293 cells.

Figure 2-8 reveals that TNFα alone increases transfection efficiency in a dose

dependent manner, rapidly increasing transfection efficiency until ~4 ul/ml (4 ng/ml)

when transfection efficiency decreases. Plasmids containing the NF-κB binding site were transfected at a higher rate ($p < 0.05$) along the entire spectrum of concentrations tested and remained high despite a drop off in transfection efficiency with plasmids lacking the NF-κB binding site.

Influenza rescue efficiency decreases after induction with TNF-alpha unless expression plasmids contain NF-κB binding site.

Viral rescue efficiency decreased when inducing HEK293 cells with TNFα. At 2 µl/ml (2 ng/ml) over a 100-fold decrease in rescue efficiency was observed at late timepoints. As TNFα concentration increased to 10 µl/ml (10 ng/ml), viral rescue efficiency improved (**Figure 2-9**). This effect was eliminated by transfecting plasmids containing a NF-κB binding site.

Percentage of GFP expressing
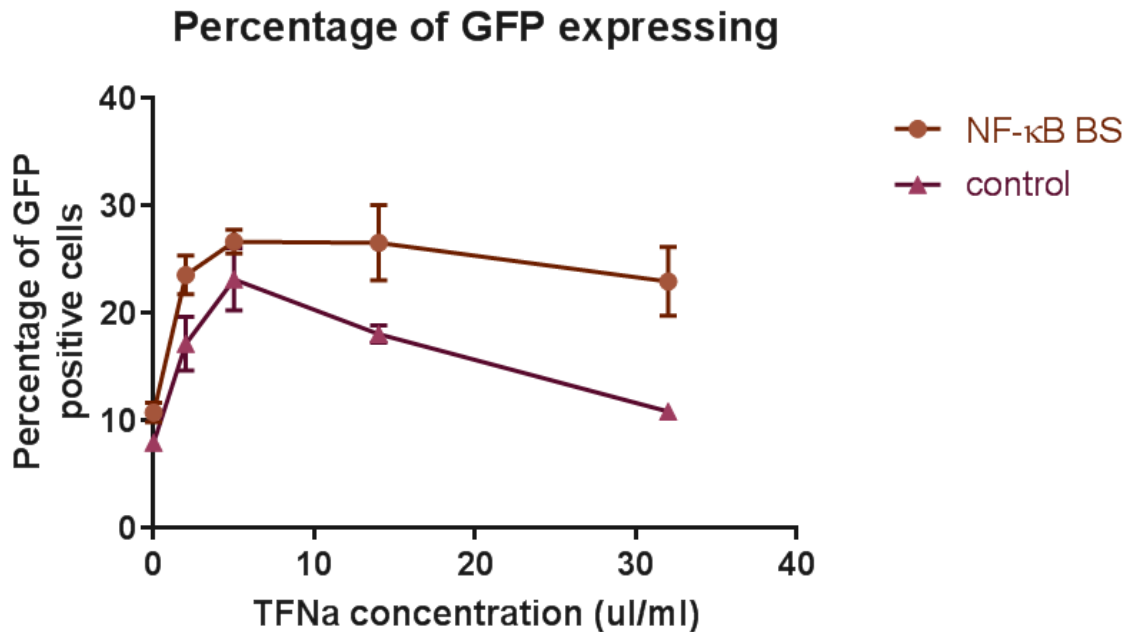
Figure 2-8 Effect of TNFα on transfection efficiency.

Influenza segments from PR8 (PA, PB1, PB2, NP) and the pPolI-NS-EGFP influenza expression reporter containing either the NF-KB binding site or no binding site were transfected into induced HEK293 cells at various TNFα concentrations. Eighteen hours post transfection the number of cells expressing GFP was counted in triplicate via flow cytometry.

A
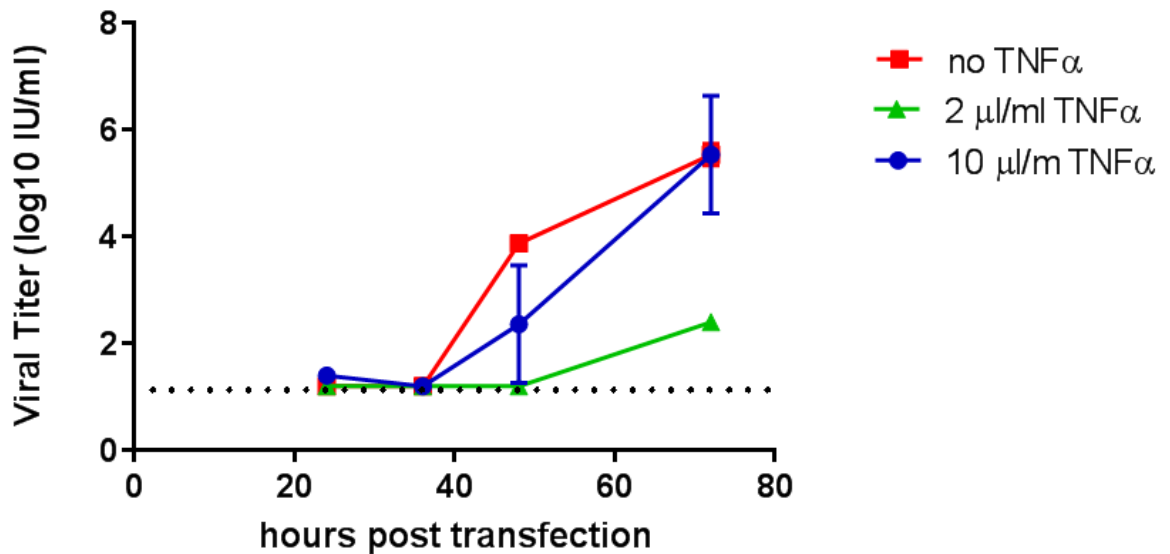
B
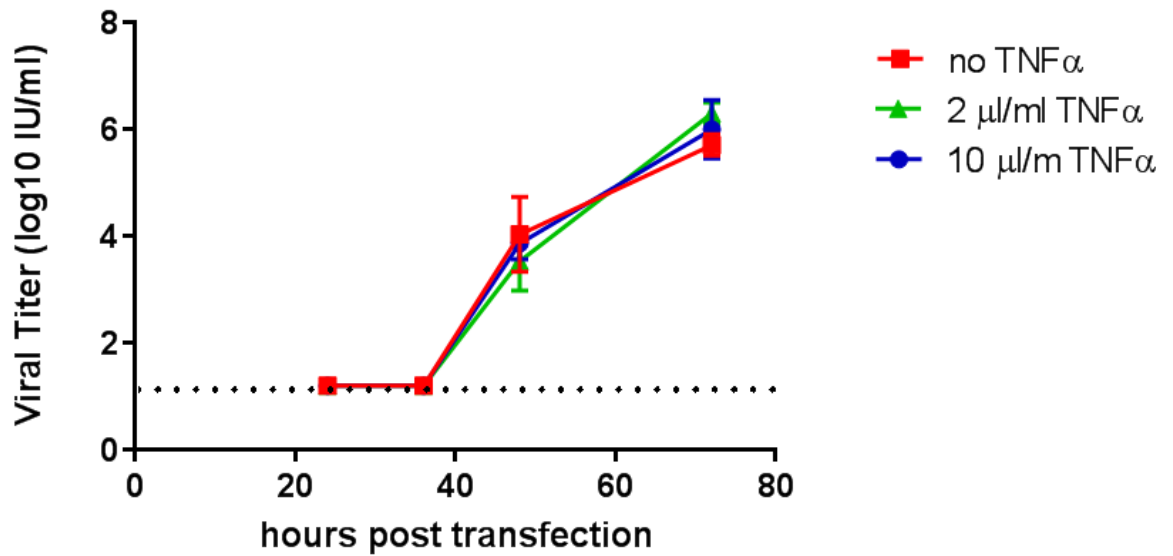
Figure 2-9 Viral rescue efficiency of NF-KB system.

Influenza segments from PR8 each containing either no binding site (panel A) of the NF-κB binding site (panel B) were used in a transfection with HEK293 cells either induced or not induced with TNFα at various concentrations. Supernatant was collected at select time points and its viral titer was measured via a TCID$_{50}$ assay. Shown is the averaged data from 3 independent positive transfections. The dotted line indicates the lower limit of detection.

A
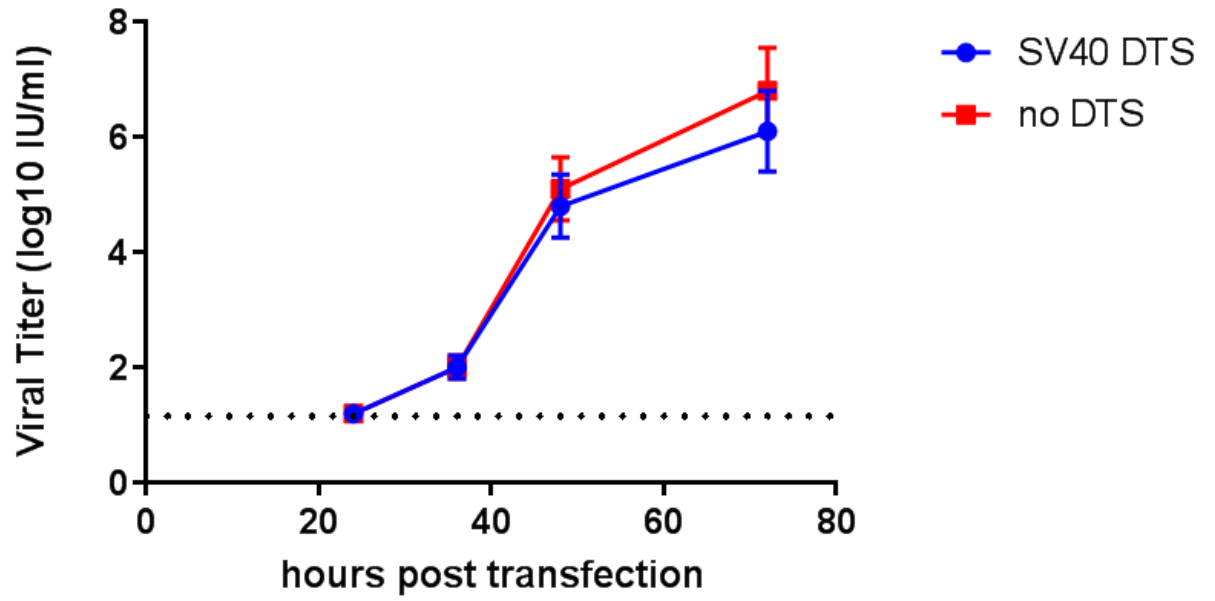
B
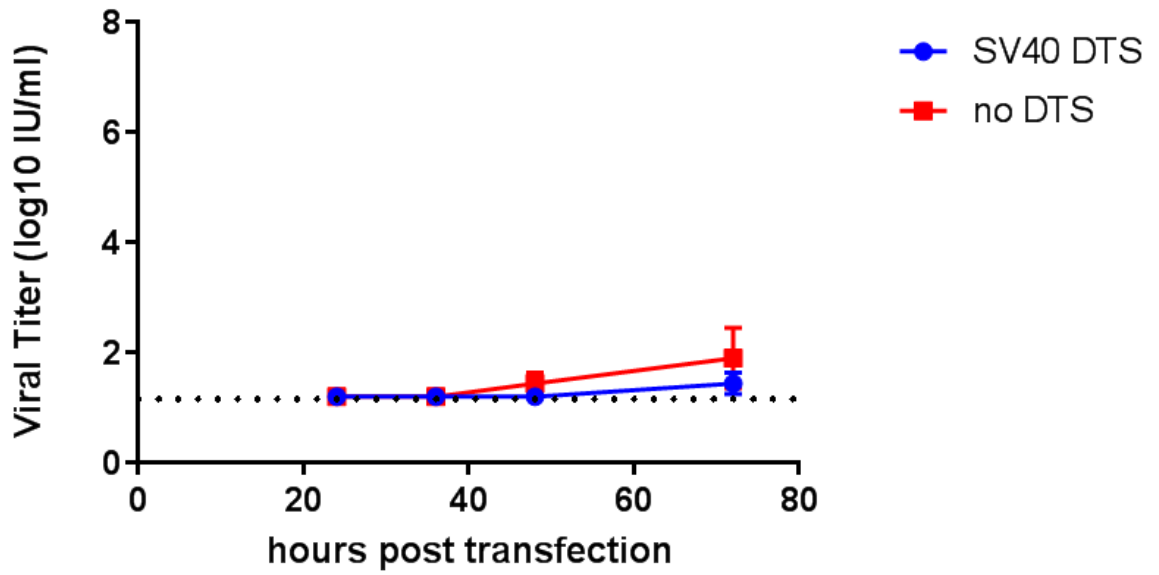
Figure 2-10: Viral generation efficiency of plasmids containing SV40-DTS.
Influenza segments HA and NA from either PR8 (panel A) or A/Perth/16/2009 (panel
B) each containing either the SV40-DTS or no binding site was used in a transfection
with HEK293 cells with the backbone segments from PR8. Supernatant was collected
at select time points and viral titer was measured via a $TCID_{50}$ assay. We show the
averaged data from 3 independent transfections. The dotted line indicates the lower
limit of detection.

No difference in viral generation efficiency was observed between plasmids containing the SV40-DTS binding site and those that didn't.

We anticipated that the SV40 system would enhance influenza rescue efficiency
however we observed that transfected plasmids containing the SV40-DTS rescued at
the same rate as plasmids that did not (**Figure 2-10**) for both the easy-to-rescue strain
and the difficult-to-rescue strain. Our results agree with other groups that have seen no
difference in gene delivery when using DTSs [197, 198].

Discussion

Potential to use E2 system in difficult to transfect systems

The E2 transfection system improved the consistency of transfection. It might be useful
in applications where transfection efficiency is very low, when a high proportion of
cells are required to be successfully transfected, or when only a limited amount of
transfectable material is available. Little difference in transfection efficiency was
observed when using the E2 system when conditions are optimal. The utility of the

system is only when transfections are already inefficient. The Mirus Transit-LT1 lipid is very efficient at delivering DNA in numerous situations. However, when only a limited amount of viral DNA is available a significant increase in transfection efficiency is needed and the E2 system may be justified.

The difficult-to-rescue strain A/Perth/16/2009 may be harder to rescue for a number of reasons. Weak compatibility between the HA and NA fragments and the PR8X backbone, or lower viral expression could result in a shift of the limiting step of the process. Enhancing the nuclear delivery step or ensuring that the plasmids persist in the nucleus longer may be enhancing viral rescue consistency. The E2 system may ensure that numerous copies of each plasmid are delivered to the nucleus ensuring higher expression of viral proteins increasing the likelihood of a successful viral gene expression and rescue of viral replication.

Industrial influenza vaccine production must take place in non-human derived cell lines such as the canine MDCK cell line. To use the E2 system for vaccine production it would need to be tested in a cell line approved for vaccine production. Future work may demonstrate the E2 system is capable of functioning in clinically relevant cell lines.

TNFα concentration dependent phenomena on transfection and rescue efficiency

Tumor necrosis factor alpha (TNFα) is a cell signaling cytokine that activates the NF-κB/ReI protein family and has a central role in the inducible expression of more than 500 genes including those involved in inflammation, host defense, cell survival and proliferation. Tay et. al., 2010 studied the dynamics of TNFα concentrations in mouse fibroblast 3T3 cells. They found that activation of late and intermediate genes (particularly RANTES, Capd4, Ccl2, Trail and Ripk2) are only expressed at the highest doses whereas only early genes are expressed at low doses. We observed TNFα caused two effects on the capacity of the cells to replicate influenza virus via a reverse genetics reaction. First, we saw an increase in transfection efficiency at low TNFα concentrations that was enhanced when we transfected plasmids containing a NF-κB binding site. The other effect was that TNFα suppressed influenza gene expression and virus replication at mid-range concentrations, an effect that could be eliminated by adding NF-κB binding sites to plasmids. Due to the large number of genes involved in the NF-κB signaling pathway mathematical modeling is likely needed to determine the dynamics of the system. Future work may be able to identify which genes are responsible for the phenomena observed in this report.

NLS based nuclear targeting of reverse genetics plasmids using the TFAM and SV40 systems

Each of these systems was designed to eliminate or reduce the nuclear barrier for transfection. Using NLS based nuclear targeting was unsuccessful in this report. The TFAM and SV40 systems were unsuccessful at obtaining increased transfection or

rescue efficiency, suggesting that nuclear entry is not a limiting step when using the cationic lipid tested in this report. The Mirus Transit-LT1 lipid has been used to successfully transfect slow growing human induced pluripotent stem cells iPSCs, as well as primary cell lines. The chemical formula of the reagent is proprietary; however, to obtain successful transfection of these slow growing or non-dividing cell lines it is possible that the cation already contains NLSs and the addition of more NLSs is not effective at increasing transfection rates.

Conclusion

We tested four approaches for improving the gene delivery, and applied those approaches to improve the efficiency of influenza virus reverse genetics. The E2 nuclear targeting system was the most successful. We found that when using this system, the viral rescue consistency (how reliably a virus can be generated in each transfection/reverse genetics experiment) was significantly enhanced when using a difficult-to-rescue strain. Easy-to-rescue viruses did not see an improvement in rescue consistency with this system, likely because the rescue consistency is already very high. Increases in transfection efficiency were observed using E2 when transfection conditions were sub-optimal, such as when there was too much DNA for the amount of transfection reagent. We did not observe any transfection efficiency improvement when transfections were performed under optimal conditions. The increased transfection efficiency when DNA amounts are decreased to a sub-optimal level may be the reason for the increase in viral rescue consistency observed in the A/Perth/16/2009 strain. These results indicate that using the E2 system may be

warranted with difficult-to-rescue viruses or when sub-optimal transfection conditions
are needed.

Table 2-1 List of plasmids

Vector name	*Description*
pKS10	From appendix A
pKS10xbx	Modified pKS10 shown in figure 2-3A
Standard System	
pKS10-HAPR8	pKS10 + HA segment from PR8X
pKS10-NAPR8	pKS10 + NA segment from PR8X
pKS10-HAperth	pKS10 + HA segment from A/Perth/16/2009
pKS10-NAperth	pKS10 + NA segment from A/Perth/16/2009
pKS10-PA	pKS10 + PA segment from PR8X
pKS10-PB1	pKS10 + PB1 segment from PR8X
pKS10-PB2	pKS10 + PB2 segment from PR8X
pKS10-NP	pKS10 + NP segment from PR8X
pKS10-NS	pKS10 + NS segment from PR8X
pKS10-M	pKS10 + M segment from PR8X
pPolI-NS-EGFP	See chapter 2 methods section
E2 System	
pKS10-HAPR8-E2	pKS10-HAPR8 + E2 binding site
pKS10-NAPR8-E2	pKS10-NAPR8 + E2 binding site
pKS10-HAperth-E2	pKS10-HAperth + E2 binding site
pKS10-NAperth-E2	pKS10-NAperth + E2 binding site
pKS10-PA-E2	pKS10-PA + E2 binding site
pKS10-PB1-E2	pKS10-PB1 + E2 binding site
pKS10-PB2-E2	pKS10-PB2 + E2 binding site

pKS10-NP-E2	pKS10-NP + E2 binding site
pKS10-NS-E2	pKS10-NS + E2 binding site
pKS10-M-E2	pKS10-M + E2 binding site
pPolI-NS-EGFP-E2	pPolI-NS-EGFP + E2 binding site
TFAM System	
pKS10-HAPR8-TFAM	pKS10-HAPR8 + TFAM binding site
pKS10-NAPR8-TFAM	pKS10-NAPR8 + TFAM binding site
pKS10-HAperth-TFAM	pKS10-HAperth + TFAM binding site
pKS10-NAperth-TFAM	pKS10-NAperth + TFAM binding site
pKS10-PA-TFAM	pKS10-PA + TFAM binding site
pKS10-PB1-TFAM	pKS10-PB1 + TFAM binding site
pKS10-PB2-TFAM	pKS10-PB2 + TFAM binding site
pKS10-NP-TFAM	pKS10-NP + TFAM binding site
pKS10-NS-TFAM	pKS10-NS + TFAM binding site
pKS10-M-TFAM	pKS10-M + TFAM binding site
pPolI-NS-EGFP-TFAM	pPolI-NS-EGFP + TFAM binding site
NF-κB system	
pKS10-HAPR8-NF-κB	pKS10-HAPR8 + NF-κB binding site
pKS10-NAPR8- NF-κB	pKS10-NAPR8 + NF-κB binding site

pKS10-HAperth- NF-κB	pKS10-HAperth + NF-κB binding site
pKS10-NAperth- NF-κB	pKS10-NAperth + NF-κB binding site
pKS10-PA-NF-κB	pKS10-PA + NF-κB binding site
pKS10-PB1-NF-κB	pKS10-PB1 + NF-κB binding site
pKS10-PB2-NF-κB	pKS10-PB2 + NF-κB binding site
pKS10-NP-NF-κB	pKS10-NP + NF-κB binding site
pKS10-NS-NF-κB	pKS10-NS + NF-κB binding site
pKS10-M-NF-κB	pKS10-M + NF-κB binding site
pPolI-NS-EGFP- NF-κB	pPolI-NS-EGFP + NF-κB binding site
pKS10-HAPR8-SV40	pKS10-HAPR8 + SV40 binding site
pKS10-NAPR8-SV40	pKS10-NAPR8 + SV40 binding site
pKS10-HAperth-SV40	pKS10-HAperth + SV40 binding site
pKS10-NAperth-SV40	pKS10-NAperth + SV40 binding site
pKS10-PA-SV40	pKS10-PA + SV40 binding site
pKS10-PB1-SV40	pKS10-PB1 + SV40 binding site
pKS10-PB2-SV40	pKS10-PB2 + SV40 binding site
pKS10-NP-SV40	pKS10-NP + SV40 binding site
pKS10-NS-SV40	pKS10-NS + SV40 binding site

pKS10-M-SV40	pKS10-M + SV40 binding site
pPolI-NS-EGFP- SV40	pPolI-NS-EGFP + SV40 binding site
Expression plasmids	
pM56	From Genecopedia: EX-mCHER-M56
pM56X	Modified from pM56 shown in figure 2-3B
pM56X-E2	BPV-E2 + pM56X
pM56X-E2a	Modified pM56X-E2 removing 125 bp from the 3' end of BPV-E2 to remove the BRD4 binding site region
pM56X-TFAM	TFAM + pM56X
pM56X-AAA-TFAM	Modified pM56X-TFAM removing 194 bp from the 3' end of TFAM to remove the HMG-2 region
pM56X-NLS-TFAM	pM56X-AAA-TFAM + CCGAAAAAAAAACGCAAAGTG added to the 5' end of the TFAM region, adding a NLS

Chapter 3 Efficient size-independent chromosome delivery from yeast to cultured cell lines

Authors: David M. Brown*[1], Yujia A. Chan[2,3], Prashant J. Desai[4], Peter Grzesik[4], Lauren M. Oldfield[1], Sanjay Vashee[1], Jeffrey C. Way[3], Pamela A. Silver[2,3], and John I. Glass*[1] [206]

Affiliations:

[1] Synthetic Biology and Bioenergy, J. Craig Venter Institute, Rockville, MD, 20850, United States.

[2] Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, United States.

[3] Wyss Institute for Biologically Inspired Engineering, Boston, Massachusetts 02115, United States.

[4] Johns Hopkins University, Sidney Kimmel Comprehensive Cancer Center Johns Hopkins, Viral Oncology Program, Baltimore, Maryland 21231, United States.

ABSTRACT

The delivery of large DNA vectors (>100,000 bp) remains a limiting step in the engineering of mammalian cells and the development of human artificial chromosomes (HACs). Yeast is commonly used to assemble genetic constructs in the megabase size range, and has previously been used to transfer constructs directly into cultured cells.

We improved this method to efficiently deliver large (1.1 Mb) synthetic yeast centromeric plasmids (YCps) to cultured cell lines at rates similar to that of 12 kb YCps. Synchronizing cells in mitosis improved the delivery efficiency by 10-fold and a statistical design of experiments approach was employed to boost the vector delivery rate by nearly 300-fold from 1/250,000 to 1/840 cells, and subsequently optimize the delivery process for multiple mammalian, avian, and insect cell lines. We adapted this method to rapidly deliver a 152 kb herpes simplex virus 1 genome cloned in yeast into mammalian cells to produce infectious virus.

INTRODUCTION

The delivery of large segments of DNA to the mammalian nucleus remains a significant challenge for gene therapy, large DNA virus reverse genetics, and Human Artificial Chromosome (HAC) development. HACs have been in development since the 1990s to address the limitations of viral-based mammalian vectors [141] and allow megabase-scale cloning capacities, copy number control, and long-term gene expression. The current method for transferring large DNA vectors between cells is microcell-mediated chromosome transfer (MMCT), which is a time-consuming, low efficiency, and difficult technique performed by few labs. MMCT works only for select donor rodent cell lines and a limited range of recipient cell lines [117, 118]. One alternative, polyethylene glycol (PEG)-mediated cell fusion is used to transfer yeast centromeric plasmids (YCps) [207] into cultured mammalian cells, where encoded genes are then expressed [143, 208]. However, this is conventionally a low efficiency delivery technique [110, 112, 114, 209]. Other delivery methods such as lipofection[30] and

67

microinjection expose large DNA molecules to shear forces and breakage, decreasing delivery efficiency as the DNA molecule becomes larger requiring the use of agarose plugs to avoid DNA damage due to shear [73, 74, 78]. In contrast, PEG-mediated fusion does not require isolation and exposure of the YCp to shear damage.

In this report, we sought to enhance the YCp delivery rate for large DNA constructs. There are numerous cellular barriers that prevent the successful delivery of a DNA construct to a mammalian nucleus. We hypothesized that synchronizing cells in M-phase, when the nuclear membrane and cytoskeleton is remodeled, could eliminate a rate-limiting step to achieving successful delivery. Targeting of the nuclear membrane as a barrier for effective DNA delivery has been reported previously through the use of nuclear localization signals [11] and fusogenic proteins [123]. In addition, we employed a design of experiments (DoE) methodology to systematically screen and evaluate numerous factors thought to play a role in YCp delivery. Here we describe an improved YCp delivery protocol using PEG-mediated fusion of donor yeast cells with recipient mammalian cells. Our technique increased conventional delivery rates by ~300-fold for HEK293 cells.

Another necessary step for cell line engineering and HAC development is the synthesis and cloning of large DNA molecules. Efficient genetic tools such as yeast recombination-based assembly methods and capacity of yeast to replicate YCps over 1 Mb [150, 210] make *Saccharomyces cerevisiae* a good choice for manipulating large DNA vectors. YCps have diverse utility and have been used to study human genetic elements in transgenic mice [105], assemble exogenous biosynthetic clusters to

produce various compounds in yeast [151], and construct entire bacterial genomes, including the 1.1 Mb *Mycoplasma mycoides* bacterial genome [3]. In addition, large circular DNA molecules can be assembled from >20 DNA fragments in a single transformation step in yeast [158]. Transformation-associated recombination (TAR)-cloning based technology [211, 212] was used to assemble the YCps in this study. Using the same organism to both construct and deliver DNA speed up the process and reduces costs.

We further demonstrated the benefits of our improved delivery technique in the field of reverse virus genetics. Generating virus from an engineered viral genome is essential for the study of viral genes, vaccine development, and clinical trials. Viruses with small genomes can readily be obtained by reverse genetics protocols: transfecting cloned viral genes or genomes as plasmids into a susceptible cell culture [203, 213]. However, larger viruses can be problematic to clone in *Escherichia coli.* Herpes simplex virus type 1 (HSV-1) is a 152 kb double-stranded DNA virus. Using a HSV-1 genome cloned as a YCp (unpublished data), we validated YCp delivery and expression using our improved technique. The infectious HSV-1 genome resulted in generation of infectious virus. We also demonstrated that functional proteins such as the anti-STING VP35 protein [214, 215] could be expressed in yeast and delivered simultaneously to mammalian cells alongside the HSV-1 genome to enable infectious virus production in Vero and HeLa cells, which normally do not support HSV-1 replication.

MATERIALS AND METHODS

Yeast strains, cell culture, and materials

The *S. cerevisiae* strain VL6-48 (ATCC MYA-3666: *MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 cir0*)[216] was used for fusions. HEK293(ATCC CRL-3216) Vero (ATCC CRL-1586), HeLa (ATCC CCL-2) , CHO (ATCC CCL-61), A549 (ATCC CCL-185), DF-1 (ATCC CRL-12203) and C6/36 cells (ATCC CRL-1660) were cultured according to manufacturer's instructions supplemented with, 100 units/mL of penicillin, 100 µg/mL of streptomycin, and 0.25 µg/mL Amphotericin B at 37°C and 5% CO_2.

The codon optimized for yeast Ebola-derived VP35 expression vector was synthesized by GenScript (USA) and cloned into pAG413-GPD in yeast (YYC001). Constitutive VP35 expression was confirmed by western blot (data not shown). The test strain (YDB012) was constructed via TAR cloning using pRS313. Test strain YDB123 and YDB1100 by inserting approximately 100 kb and 1.1 Mb respectively of DNA from *Mycoplasma mycoides* [4] via TAR cloning [211, 212, 217]. A list of plasmids, YCps and yeast strains can be found in supplementary information **(Supplementary Table 3-1; Supplementary Table 3-2).**

PEG MW 2000 solution was prepared by adding 12.5 g of PEG 2000 and up to 25 mL of selected HEPES buffer in a 50 mL tube. This was gently shaken until dissolved. PEG solution was kept at room temperature for up to 1 week.

Construction of Reporter YCp

We constructed reporter YCps and yeast strains (YDB012, YDB123, YDB1100) that monitor movement of cytoplasmic elements from yeast into mammalian cells. These

were circular YCps carrying an mCherry fluorescent protein-encoding gene that is only expressed in yeast using the constitutive GPD (TDH3) promoter, a green fluorescent protein (GFP)-encoding gene under the control of a cytomegalovirus (CMV) promoter only expressed in mammalian cells, and/or stuffer DNA to modulate the vector size. The fluorescent markers were used to quantify the two steps of the delivery process: the fusion rate of yeast spheroplasts and mammalian cells was measured by the post-fusion presence of mCherry in mammalian cells as the yeast cell contents are delivered into the mammalian cytoplasm (**Fig. 3-1**); the vector delivery rate, which is the rate at which the YCp is delivered to the mammalian nucleus and expressed, was measured by GFP expression in mammalian cells (**Fig. 3-1**).

Preparation of yeast spheroplasts

Yeast synthetic complete medium lacking histidine, tryptophan, and/or uracil, supplemented with adenine (120 mg/L) (Teknova, Inc.) was inoculated with VL6-48 and incubated overnight at 30°C with 200 revolutions per minute (rpm) agitation. The culture was diluted 10-fold with YPD medium [211] and grown to an OD_{600} of 2.0–3.0. Cells were centrifuged at 2,500 relative centrifugal force (RCF) for 5 minutes at 10°C, and the supernatant was decanted. The pellet was resuspended in 20 mL of filter-sterilized 1 M sorbitol and kept at 4°C for 2 hours. If needed cells can be kept at 4°C for up to 24 h with no effect on vector delivery rate. This was pelleted as before and resuspended in 20 mL of SPEM solution (1 M sorbitol, 10 mM EDTA, pH 8.0, 2.08 g/L $Na_2HPO_4 \cdot 7H_2O$, and 0.32 g/L $NaH_2PO_4 \cdot 1H_2O$) with 30 μL of β-mercaptoethanol and 40 μL of Zymolyase-20T solution (200 mg Zymolyase (USB), 9 mL of dH_2O, 1

mL of 1M Tris-HCl pH 7.5, and 10 mL of 50% glycerol, stored at −20°C). After digestion for 20 minutes at 30°C with 75 rpm agitation, the OD_{600} was determined for 0.2 mL of SPEM solution containing the yeast cells combined with 0.8 mL of (i) 1 M sorbitol or (ii) 2% SDS solution. The digest continued until the ratio (i/ii) was ~3–4. 30 mL of 1 M sorbitol was added to the remaining SPEM solution containing the yeast cells and mixed by inversion. Spheroplasts were collected by centrifugation at 1,000 RCF for 5 minutes at 10°C. The pellet was gently resuspended in 50 mL of 1 M sorbitol using a 10-mL pipette and centrifuged as before. Finally, the pellet was resuspended in 2 mL of STC solution (1 M sorbitol, 10 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$, and 2.5 mM $MgCl_2$) and kept at room temperature for 10–60 minutes.

YCp delivery protocol for HEK293 cells

Preparation of yeast spheroplasts was completed on the same day as the yeast-HEK293 cell fusion reaction. HEK293 cells were grown in a 75 mL culture flask to 70–80% confluence. Two hours prior to performing the fusion reaction, 6 µL/mL of KaryoMAX® Colcemid™ Solution in HBSS (Thermo Scientific) was added to the HEK293 culture. For the fusion reaction, HEK293 cells were trypsinized using 0.25% Trypsin-EDTA (Thermo Fisher). HEK293 cells and yeast spheroplasts were counted on a hemocytometer. One million HEK293 cells were centrifuged in a 1.5 mL tube at 1957 RCF for 1 minute and washed with phosphate buffered saline pH 7.4 (PBS) (Gibco). Yeast spheroplasts were added to PBS and mixed with the HEK293 cells so that the ratio of yeast to HEK293 cells was 100:1. The number of yeast spheroplasts was determined using a spectrophotometer at OD_{600} after normalizing by counting spheroplasts on a hemocytometer. Typically, this involved using 50 - 200ul of prepared

spheroplasts in STC. This was centrifuged as before and the supernatant was decanted.

50 µL of DMSO (Sigma) and 450 µL of PEG MW 2000 solution (12.5 g of PEG 2000

(Aldrich) in 25 mL of HEPES buffer kept at room temperature up to 1 week) was added

to the cell mixture and mixed by pipetting. No more than 1 minute was allowed to pass

before adding 1 mL of serum free HEK293 media. This was centrifuged as before, the

supernatant was decanted, and the cell pellet was resuspended in 1 mL of HEK293

media for plating into a 6-well tissue culture plate. The media was replaced after 12

hours. Evidence of YCp delivery can be seen 48 hours post-fusion. The vector delivery

rate was determined by counting the number of GFP fluorescent cells and comparing

that number to the number of recipient cells used in the fusion.

Optimized YCp delivery protocol for Vero cells

The protocol for Vero cells was modified from the HEK293 protocol as follows: 24

hours prior to fusion, 50 µL of KaryoMAX® Colcemid™ Solution was added to the

Vero cell culture. Then, Vero media was removed and reserved in a 50 mL tube. Vero

cells were trypsinized according to the manufacturer's instructions using Trypsin-

EDTA (0.25%) and phenol red (Thermo Fisher), and resuspended in the reserved

media. After the step where yeast was added, 5 minutes were allowed to pass before

centrifugation. PEG-cell mixture was placed at 30°C for 5.5 minutes before adding 1

mL of serum-free Vero media.

Optimized YAC delivery protocol for C6/36 cells

The protocol for C6/36 cells was modified from the Vero protocol as follows: The ratio

of yeast cells to C6/36 cells was 50:1. After the step where yeast was added, 5 minutes

were allowed to pass before centrifugation. After the step where PEG solution was added, 5 minutes were allowed to pass before adding 1 mL serum free C6/36 media.

Fractional factorial design

Fractional factorial designs or more specifically Plackett-Burman designs of resolution III were used for screening studies these are powerful design of experiment (DOE) tools to estimate the effects of several factors. Resolution III designs do not confound with main effects however there may be confounding with two-factor interactions [218, 219]. Resolution III was used due to the large number of factors being tested. Designs were created in JMP 10.0 (SAS institute Inc., Cary, NC, USA) and used to screen for important factors prior to optimization and OFAT analysis. The independent variables were studied at 2 or 3 different levels; low (-1), medium (0), and high ($+1$) The designs can be seen for HEK-293 cells (**Supplementary Table 3-3**) Vero cells (**Supplementary Table 3-4**) and C6/36 cells (**Supplementary Table 3-5**).

Response surface methodology (RSM)

RSM based on a custom central composite design with 6 coded levels was employed to determine the optimal conditions of the screened factors in the Plackett-Burman and/or fractional factorial designs. All models and analysis were performed with JMP statistical software version 10.0.0 (SAS Institute Inc., Cary, NC, USA). The final experimental design consisted of 23 runs varying two of the most significant factors identified (**Supplementary table 3-6**). Statistical significance was checked by *F-test*, ANOVA, and goodness of fit.

Detection of HSV-1 virus replication

The HSV-1 genome was constructed in yeast as a YCp and delivered to recipient cells. The recipient cells were harvested after 4 days: freeze-thawed once and sonicated in a cup horn sonicator. Virus in the lysate were enumerated using serial dilutions in PBS that were plated on Vero cell monolayers in tissue culture trays. Following absorption, the culture was overlayed with carboxymethylcellulose in growth medium and incubated for 4 days. Plaques were visualized by crystal violet stain. Viral titers were determined by $TCID_{50}$ assay using Vero cells.

Determining fusion rate of HEK293 using flow cytometry or fluorescence microscopy.

Upon fusing HEK293 cells with the optimized YCp delivery protocol, samples were plated into a 6-well tissue culture plate. After HEK293 cells adhered to the plate (3-4 h post-fusion), excess yeast was removed by washing HEK293 cells with medium, and then adding fresh medium. Plates were observed on a Nikon eclipse Ti fluorescence microscope. To prepare samples for flow cytometry fused HEK293 cells were trypsinized, filtered using a 70 micron filter (Corning), washed, and resuspended in PBS. Using an Accuri C6 flow cytometer (BD Bio-sciences), we analyzed post-fusion cell cultures. Analysis was based on light-scatter and fluorescence signals produced from 20 mW laser illumination at 488 nm. A 670 nm long-pass filter was used to detect mCherry. Threshold levels were empirically set to eliminate detection of yeast cells and debris. The flow cytometer was routinely operated at the Slow Flow Rate setting (14 lL sample/minute). Data acquisition for a single sample took 3–5 minutes.

Mitotic Index Experiments

Several concentrations of colcemid, nocodazole, STLC, or okadaic acid were added directly to the media of a HEK293 cells growing in a 96-well plate. 2 h after treatment with the drug, cells were fixed in 4% (w/v) PFA (paraformaldehyde, Merck) solution followed by a second fixation with ice cold methanol-acetone 1:1 mix for 5 min at -20ºC. After PBS washing, cells permeabilized with PBS-0.1% (v/v) Triton X 100 for 5 min, and were blocked with PBS-3% (w/v) BSA for 30 min at room temperature. After PBS washing, cells were stained with DAPI (4′,6-diamino-2-phenylindole dihydrochloride; 100 ng/ml; Sigma) for 5 min at room temperature protected from light. The mitotic index was measured by counting the number of cells with condensed chromosomes, and percentages calculated. At least 4 fields comprising a total of >100 cells were analyzed for each measurement. Plates were observed on a Nikon eclipse Ti fluorescence microscope.

RESULTS

Synchronizing mammalian cells in mitosis improves vector delivery

Using drugs to synchronize mammalian cells in mitosis prior to fusion with yeast spheroplasts resulted in a 10–25-fold increase in vector delivery rates (**Fig. 3-2**); Colcemid (Gibco), which inhibits microtubule polymerization [220], S-trityl-L-cysteine (STLC) (Sigma), a kinesin 5 inhibitor [221], and nocodazole (Sigma), another microtubule polymerization inhibitor [222] all improved vector delivery rates at different optimum concentrations. Okadaic acid (Sigma), which inhibits

76

serine/threonine phosphatases and synchronizes cells in G1/S-phase at high concentrations [223, 224], did not improve vector delivery at low concentrations and eliminated vector delivery at high concentrations, indicating that mitotic arrest and nuclear envelope breakdown are required for efficient vector delivery. Calculation of the mitotic index for each mitotic drug by analysis of condensed chromosomes (**Supplementary Fig. 3**) showed that two hours of treatment with nocodazole, STLC or colcemid presented an index of approximately 20% at different minimum concentrations. As the concentration of each drug increased the mitotic index increased until leveling off for each drug at approximately 20%, untreated cells presented an index of 5%. The optimum concentration for each drug for YCp delivery was approximately the lowest concentration necessary to enhance the mitotic index to 20%. Subsequent experiments were carried out with Colcemid as it was observed to produce consistently higher rates of delivery.

Design of experiments optimizing delivery in multiple cell lines

The conventional "one-factor-at-a-time" (OFAT) approach, in which one independent variable is studied while maintaining all the other factors at a constant level, is frequently used in biotechnology to obtain higher product yields or to optimize a process. However, OFAT is time consuming, and disregards potential complex interactions among factors. Design of experiments (DoE) is a collection of statistical techniques and methods that can be used to overcome these problems reducing time and cost in process optimization [218, 219, 225]. DoE is performed in two main stages. First, a fractional factorial design is used to screen for the main effects and to eliminate

factors that have no or little effect on the process. Second, a higher resolution experiment using response surface methodology (RSM) is performed with fewer factors to obtain an optimized protocol. Using this methodology, fewer experimental trials are needed compared with OFAT and potential interactions among factors can be identified. Factors tested were primarily selected from differences reported from various protocols of YCp or YAC delivery [103, 109-112, 226-228].

HEK293 cells were the first cell line optimized for vector delivery. Factors were first screened using a fractional factorial design (**Supplementary Table 3-3**) and the regression results (**Supplementary Table 3-7**) predicted the major effects on vector delivery. Using either an OFAT approach or a fractional factorial design eighteen factors were systematically tested for their effect on the post-fusion fraction of mammalian cells expressing GFP (**Table 3-1**). The most critical continuous factor identified was the ratio of yeast spheroplasts to recipient cells. Using an OFAT experiment, the vector delivery rate peaked when the yeast to mammalian cell ratio was 100:1 (**Fig. 3-3**). A higher cell ratio adversely affects vector delivery and post-fusion cell viability likely due to nutrient competition from the excess yeast. We observed the same result when measuring fusion efficiency, which we define as the fraction of mCherry-positive HEK293 cells in the total population. Increasing the ratio of yeast to recipient cells 3-fold decreased the fusion efficiency from 20% to 2% (**Fig. 3-4, Supplementary Fig. 3-1**). When observing GFP expression as a measure of vector delivery rate, mCherry can still be observed dimly in cells 48 h post fusion; however, the magnitude of fluorescence is significantly reduced from 24 h post fusion.

Fusion time was also identified as a highly significant factor (*p-value* = 0.0039). Several categorical factors were also identified as highly significant; diluting the media after adding PEG, osmotically stabilizing the yeast cells in sorbitol and the use of the fungicide Amphotericin B in media. Major factors were optimized by OFAT or RSM (data not shown), which predicted the optimal yeast to mammalian cell ratio, PEG incubation time, and cell line-specific conditions (**Table 3-1).** The vector was delivered and expressed in $1,210 \pm 60$ yeast cells per million HEK293 cells using the optimized protocol.

Vero cells were optimized following a similar methodology. The initial vector delivery rate (2.0 ± 0.7 per 300,000 cells) using the HEK293 optimized protocol was too low to muster sufficient statistical power in the fractional factorial design to identify major factors. Therefore, initially we optimized the treatment with mitotic arrest drugs and the ratio of yeast to Vero cells by OFAT (data not shown). Following initial optimization via OFAT, eight factors were evaluated by a Plackett-Burman fractional factorial resolution III design (**Supplementary Table 3-4**). Significant categorical factors that were identified when optimizing HEK293 cells were not optimized for Vero cells since it was highly unlikely the condition that resulted in no YCp delivery would be advantageous in delivering YCps to Vero cells. Regression results from the Plackett-Burman design identified and eliminated non-major factors (**Supplementary Table 3-8**). A resolution V RSM design (**Supplementary Table 3-3-6**) was used with the major factors identified optimized the protocol (**Supplementary Fig. 3-2**). The model obtained had a very low *P*-value (<0.001) and the lack of fit *P*-value of 0.063

implied that the lack of fit was not significant relative to the pure error (**Supplementary Table 3-9**). A ~100x improvement of the vector delivery rate for Vero cells was obtained when using the Vero-specific protocol compared to the HEK293-specific protocol (**Table 3-2**).

Optimizing Vero cells created a pipeline for the optimization of new cell lines. The insect cell line C6/36 followed this pipeline to create an optimized protocol. The initial vector delivery rate was 1.5 ± 1.0 per 300,000 cells using the HEK293 protocol. First, OFAT optimize the treatment with mitotic arrest drugs. The differences in the efficiency of synchronizing cells in mitosis could explain differences in YCp delivery rates between different cell lines (Rieder & Cole, n.d.). Second, evaluate major factors (**Supplementary Table 3-10**) by a Plackett-Burman fractional factorial design (**Supplementary Table 3-5**). Third, use the central composite RSM design (**Supplementary Table 3-6**) to optimize PEG incubation time and cell ratio. ANOVA results (*P*-value = 0.0021) for C6/36 are given in (**Supplementary Table 3-11**). The optimized C6/36 protocol improved the vector delivery rate to 112 ± 14 per 300,000 cells. The central composite RSM design was also evaluated using HEK293 cells (**Supplementary Table 3-12**) and gave a *P*-value = 0.0021.

Vector delivery rate was independent of vector size

We found that the vector delivery rate was independent of vector size. *M. mycoides* genomic DNA was used to modulate the size of the YCp between 0.1–1.1 Mb. *Mycoplasma* DNA should be transcriptionally and translationally inert in eukaryotic

cells due to its bacterial origin and non-standard genetic code [229]. The vector delivery rate in HEK293 cells remained constant at all tested YCp sizes (12 kb; YDB012, 100 kb; YDB123, or 1.1 Mb; YDB1100) (**Table 3-2**). The YCp was confirmed to be delivered intact using multiplex PCR amplification on eight sites distributed along the 1.1 Mb YCp [106]. Seventy-two hours post fusion, lysed HEK293 cells were used as a template. Each primer set (**Supplementary table 3-13**) amplifies a segment at least 100 kb from any other set, spanning the entire 1.1 Mb YCp (**Supplementary Fig. 3-4**).

Delivery of the HSV-1 genome via fusion produced infectious virus

GFP synthesis from the delivered YCps demonstrates that mammalian polymerases can transcribe the YCp delivered to the mammalian cells as yeast chromatin. This indicates that transit and packaging in yeast does not interfere with mammalian expression as has been previously suggested [143, 208]. As a next step, we tested whether YCps delivered to mammalian cells as yeast chromatin could be replicated in the mammalian cell. Naked herpes virus DNA is infectious when transfected into mammalian cells [230-232]. We delivered a HSV-1 genome assembled as a YCp from yeast to mammalian cells. HSV-1 plaques appeared 4 days post-passage as indicative of infectious virus production (**Supplementary Fig. 3-5A**). This demonstrated that the delivered YCps could be recognized and replicated by viral DNA polymerases and used to generate virus from cloned large viral genomes.

Exogenous DNA that enters mammalian cells is commonly silenced via the STING (Stimulator of Interferon Genes) pathway [233, 234], which facilitates interferon production in response to cytosolic naked DNA or RNA from viruses. Supplying DNA as yeast chromatin may avoid gene silencing to a certain extent. In HEK293 cells, we were able to produce HSV-1 via fusion with yeast carrying the HSV-1 YCp. However, HeLa cells did not produce virus after fusion. This is likely because the STING pathway is not highly activated in HEK-293 cells [235].

To counter STING-mediated silencing, fusion experiments were performed with yeast expressing the Ebola-derived VP35 protein, which inhibits the activation of IRF-3 kinases and the IFN-beta promoter downstream of STING [214, 215]. Fusion using a yeast strain that expressed VP35 protein and harbored the HSV-1 genome enabled HSV-1 production in HeLa cells and enhanced the frequency of virus generation in Vero cells. In Vero cells, virus was observed in only 1 of 4 trials without VP35, versus in all 4 trials with VP35; viral titers were also higher **(Fig. 3-5).** We demonstrated that the delivery of proteins and genomes could be modularized: the VP35 could be expressed in yeast separate from the yeast harboring the HSV-1 genome. Fusion experiments using both strains of yeast resulted in similar HSV-1 genome expression compared to fusion using a strain of yeast that expressed VP35 and carried the HSV-1 genome (**Supplementary Fig. 3-5B**).

DISCUSSION

In this work, we developed an efficient method for transferring large YCps from yeast to different cell lines. PEG-mediated fusion eliminates the need for the vector to cross the cytoplasmic membranes of the donor and recipient cells. It also avoids hydrodynamic shear forces damaging DNA molecules, which could happen when genomic DNA is isolated from cells prior to transfer. We improved vector delivery by arresting the recipient mammalian cell in the M-phase of the cell cycle when the mammalian nuclear envelope is broken down, thus allowing for easier diffusion of the vector to where the mammalian nucleus re-forms. We used three different mitotic arrest drugs: Colcemid and nocodozole both interfere with microtubule polymerization. STLC synchronizes cells by inhibiting kinesin 5. The different drug mechanisms suggest that there is not another confounding variable causing an increase in vector delivery rate. This was confirmed upon analyzing the mitotic index of each drug. Each drug had a different minimum effective concentration, increasing the mitotic index from 5% to 20%. The optimum concentration for each drug for vector delivery correlated to this minimum effective concentration. However, as the concentration of microtubule interfering drugs increased, the mitotic index remained unchanged as the vector delivery rate decreased possibly due to cellular toxicity. The differences in vector delivery rate between each drug could not be completely explained by mitotic index.

Subtle differences in protocol can have a significant effect on vector delivery. We employed DoE along with OFAT to improve fusion efficiencies for 3 different cell

lines. The pipeline and optimization protocol presented here provides a versatile protocol that could be utilized to optimize vector delivery to other cell types. In addition, an insect and avian cell line were also amenable to delivery, suggesting that the protocol may be applicable to other animal phyla.

The ability to deliver functional proteins expressed in yeast alongside genetic constructs assembled in yeast will advance and diversify the tools for gene delivery to mammalian cells. We showed that VP35 protein delivery made HSV-1 genome delivery and expression possible in HeLa cells where it was previously undetectable when the HSV-1 genome was delivered alone. VP35 co-delivery also enhanced HSV-1 production in Vero cells. As expected, delivery rates in HEK293 cells were unimproved by VP35 co-delivery because HEK293 cells have low or undetectable cGAS and STING protein levels [235]. We determined that the co-delivery of protein and DNA within the same yeast cell was not required. Because of the high yeast to recipient cell ratio, one yeast strain can deliver the VP35 and the other can deliver the DNA. This is likely due to the high fusion efficiency (20% of recipient cells obtain fluorescent protein in the yeast cytoplasm after fusion) relative to the lower DNA delivery rate. The co-delivery of other proteins that alter cellular processes resulting in protection of exogenous DNA may be equally useful for reverse virus genetics or HACs. There is also the potential to engineer fusion proteins that can bind to the genetic constructs and target their delivery into the mammalian cell nucleus using a nuclear localization sequence. Our method of fusing yeast to recipient cells makes the co-delivery of functional proteins an easy and effective option.

We used the delivery of yeast carrying HSV-1 genomic DNA as a surrogate for a HAC (in development). We showed that yeast chromatin can be replicated by a HSV-1 DNA polymerase that evolved to replicate DNA associated with mammalian histones and other nuclear associated proteins [236]. This suggests that HACs delivered into mammalian cells using yeast vectors will similarly be processed by mammalian DNA replication machinery. In addition to answering our questions about the compatibility of yeast chromatin with mammalian DNA polymerases, these experiments also offer new possibilities for doing reverse genetics of viruses with large DNA genomes such as other herpes viruses, pox viruses, and the African Swine Fever Virus.

This improved protocol can be a powerful tool for the delivery of large DNA molecules into mammalian cells. MMCT, the method most commonly used for the delivery of HACs into mammalian cells has a yield frequency of no greater than 1/1,000,000– 1/100,000 [123, 237] although recent advancements have reported improved delivery efficiencies [238, 239]. We boosted YCp vector delivery rate to 1/840 for HEK293 cells. Importantly, we are able to consistently obtain efficient YCp delivery and showed that vector sizes up to 1.1 Mb do not impede the YCp delivery rate. The largest reported YCp is 2.3 Mb [208], and is likely not an upper limit or obstacle for our technique. We demonstrated that this protocol could be reliably optimized for different cell lines from higher eukaryotes (**Table 3-2**). It is likely that following the described optimization pipeline many more cell lines can be optimized. The YCp delivery method represents an inexpensive and rapid technique that can be replicated in most laboratories without specialized equipment. YCp-based vector delivery will be useful for studies of gene

85

cluster regulation, transgenic animal construction, and *de novo* HAC development. Advances in the synthesis of mammalian chromosomes in yeast cells will be complemented by this approach for chromosome delivery from yeast to animal cells.

TABLE AND FIGURES LEGENDS

Figure 3-1 Genetic construct delivery via yeast to mammalian cells.

YCp construction and delivery: (A) Overlapping genetic constructs are assembled by transformation-associated recombination (TAR) cloning in yeast. (B) The yeast spheroplasts are fused to mammalian cells with PEG. (C) Delivery of mCherry expressed in yeast to mammalian cells measures the fusion efficiency. (D) The construct escapes from the yeast nucleus into the mammalian nucleus. (E) GFP expression from the construct indicates vector delivery efficiency.
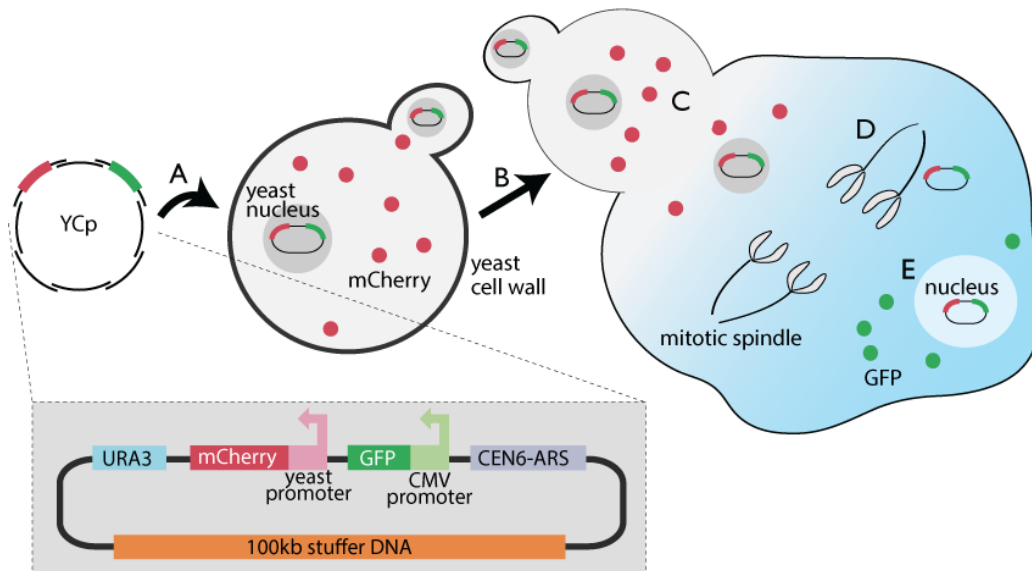
Figure 3-2 Mitotic arrest of mammalian cells significantly increases the vector delivery rate.

Vector delivery rate was determined by counting the number of recipient cells expressing the delivered fluorescent reporter. Colcemid (A), STLC (B) and nocodazole (C) arrest cells in mitosis. Okadaic acid (D) arrests cells in S-phase. Colcemid at 54 nM most strongly enhanced vector delivery efficiency. Error bars represent the standard deviation of 4 replicates.
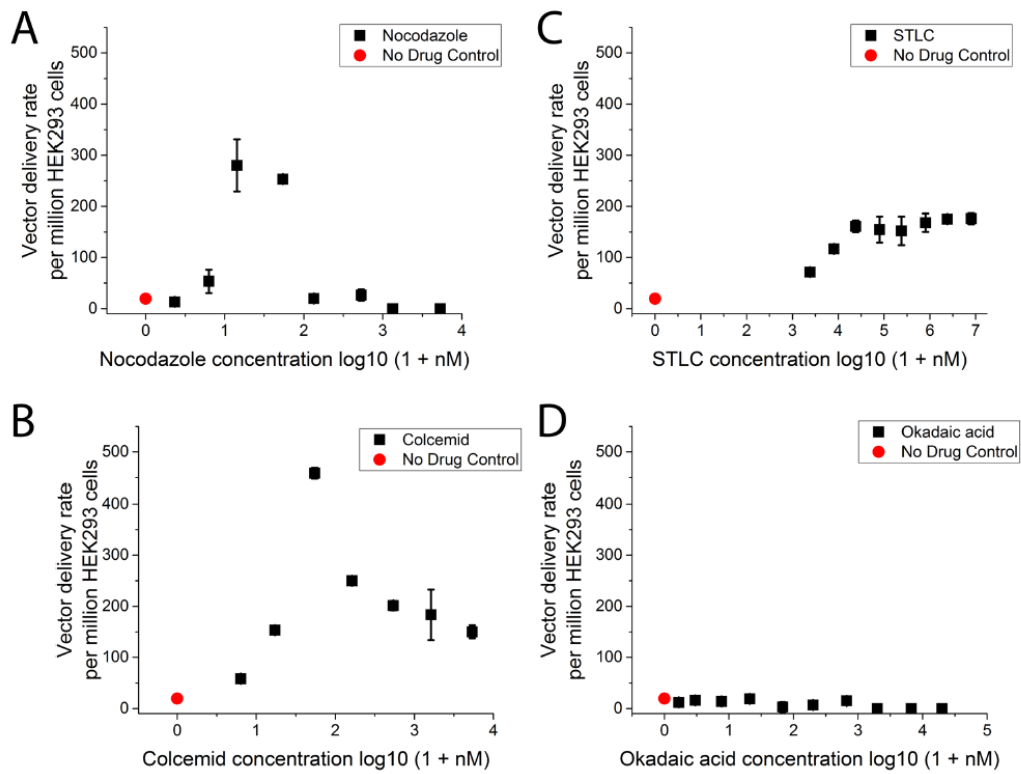
Figure 3-3 The vector delivery rate is significantly affected by the ratio of yeast to recipient cells.

The optimal ratio of yeast to HEK293 cells was 100:1. Error bars represent the

standard deviation of 4 replicates.
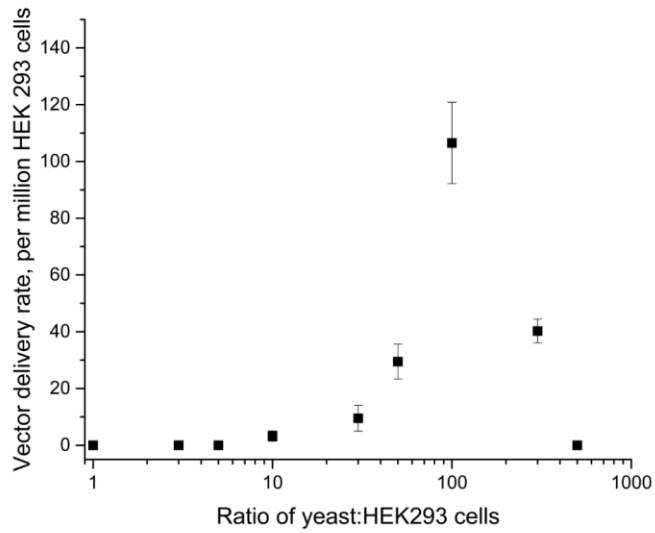
Figure 3-4 Fusion efficiency is significantly affected by the ratio of yeast to recipient cells.

The fusion efficiency was determined by the percentage of mCherry-positive HEK293 cells post-fusion. Microscopy images at 200x magnification show mCherry fluorescence (top panels) and differential interference contrast (DIC; bottom panels). The percentage of HEK293 cells that were mCherry-positive post-fusion is stated on the top left corner and was determined by counting mCherry-positive and negative cells on a hemocytometer. The PEG incubation time, HEK293 cell number, and yeast to HEK293 cell ratio were 1 minute, 1×10^5 cells, and 100:1, respectively.
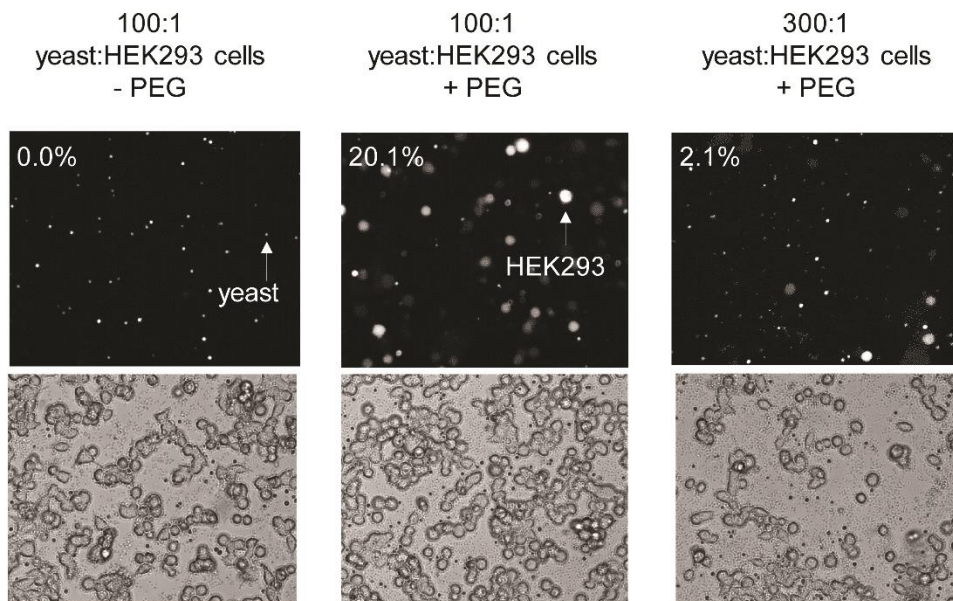
Figure 3-5 VP35 enhances HSV-1 generation from a viral genome cloned in and delivered by yeast.

HEK293, HeLa, or Vero cells were fused with a yeast strain containing the HSV-1 genome and/or a VP35 expression plasmid. Supernatants were collected 72 hours post-fusion and viral titer was determined. The limit of detection is indicated by the solid horizontal line. Samples that produced no virus are indicated at the limit of detection. Each condition had 4 replicates.
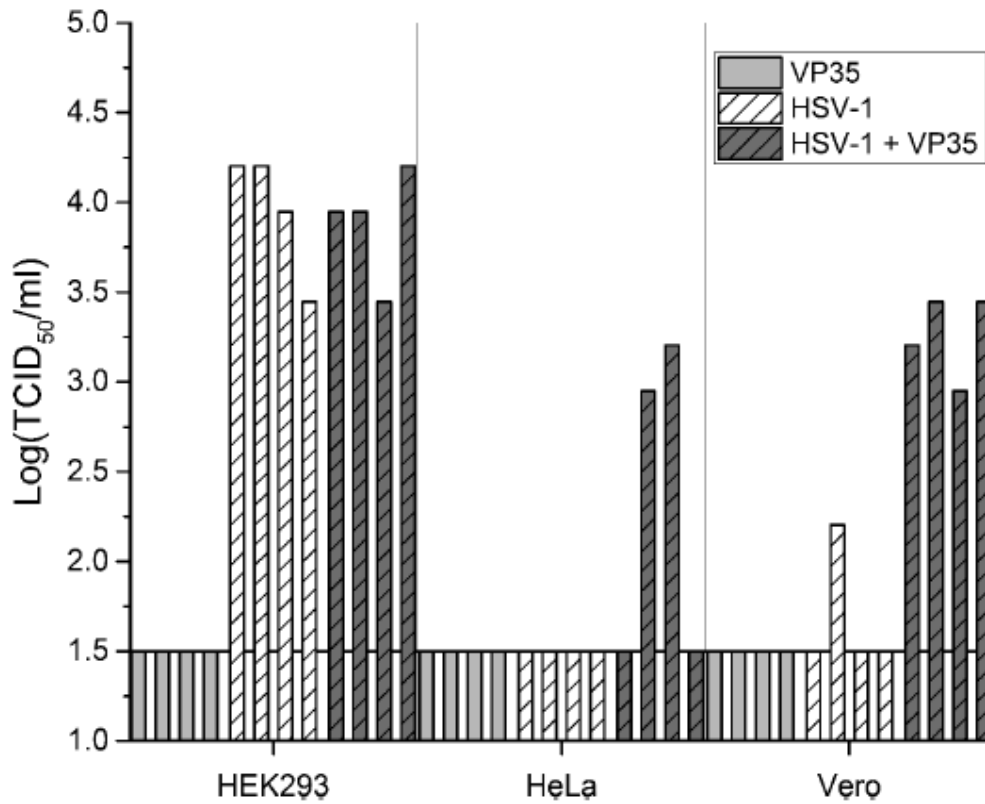
Supplementary Figure 3-1. The yeast to HEK293 cell ratio significantly affects fusion efficiency. Post-fusion flow cytometry profiles are shown for each fusion reaction condition as labeled. The percentage of mCherry-positive HEK293 cells detected is labeled on the top right corner of each flow cytometry profile. The highest fusion efficiency (17.4% of the HEK293 cells are mCherry-positive post-fusion) was attained using a ratio of 100:1 yeast to HEK293 cells.
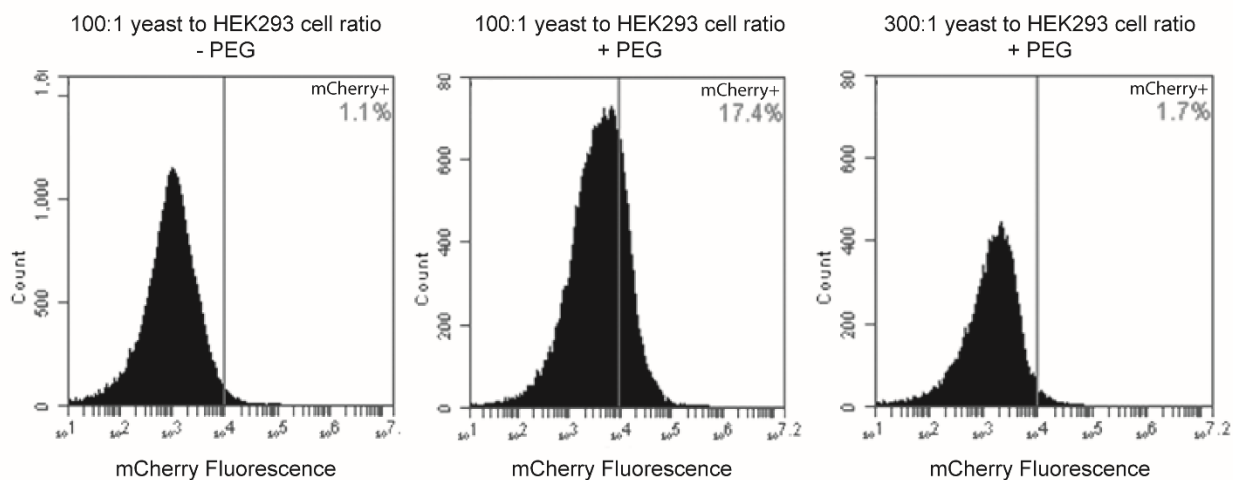
Supplementary Figure 3-2. Changing PEG incubation time and the yeast to mammalian cell ratio significantly increases fusion efficiency for different cell lines. A design of experiments (DOE) approach was adopted to identify the optimal combination of PEG incubation time and number of yeast cells used for fusion with Vero cells. The response surface method (RSM) plot indicated a maximum vector delivery rate with a PEG incubation time and ratio of yeast to Vero cells at 5.7 minutes and 100:1 respectively.
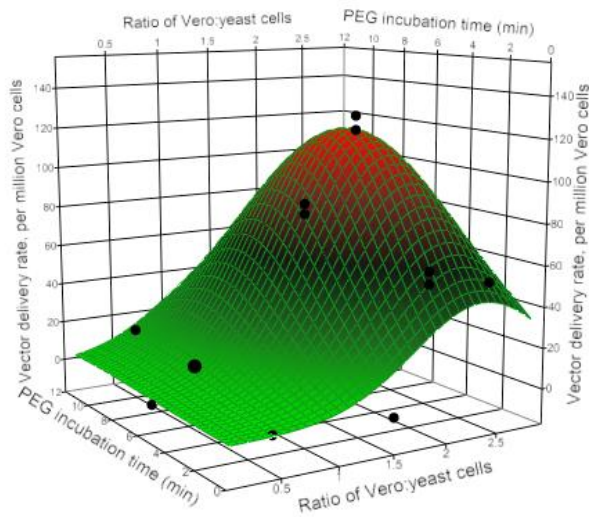
Supplementary Figure 3-3. Cellular response to mitotic agents. Mitotic accumulation during exposure to different concentrations of Nocodazole (A), Colcemid (B) and STLC (C) and Okadaic acid (D) for 2 h. Error bars represent the standard error of 4 replicates.
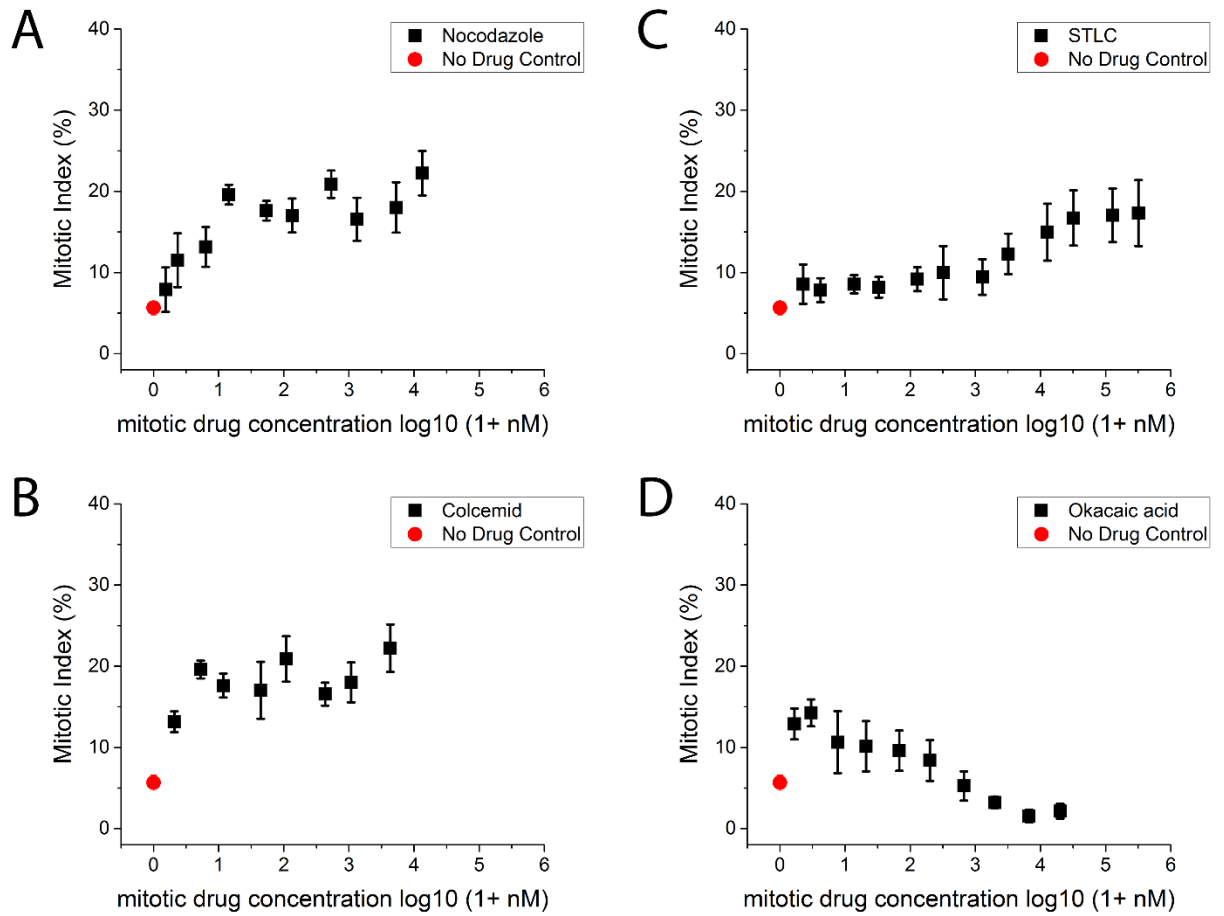
Supplementary Figure 3-4. 1.1Mb YCp can be delivered to HEK293 cells intact.

72h post fusion lysed HEK293 cells (lane C) were used as a template in multiplex PCR amplification using eight individual primers shown in supplementary table 3-11. The resulting amplicons were analyzed by gel electrophoresis. Lane A is a positive control (the *M. mycoides* genome). Lane B is a negative control (unfused HEK293 cells). Molecular ladder is shown in lane L.
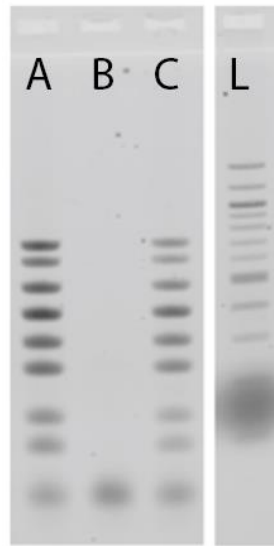
Supplementary Figure 3-5. HSV-1 production from viral genome cloned in and delivered by yeast. (A) HEK293 cells were fused with a yeast strain containing the HSV-1 genome (YLO141). Virus titer present in supernatant was evaluated using a plaque assay, where Vero cells are fixed and stained with crystal violet. Plaques in the monolayer indicate infectious HSV-1 virus production. (B) Vero cells were fused with a yeast strain containing the HSV-1 genome (YLO141), a yeast strain expressing VP35 and containing the HSV-1 genome (YLODB1), a yeast strain expressing VP35 only (YYC001), or with a mixed culture of two yeast strains one expressing VP35 (YYC001) and the other containing the HSV-1 genome (YLO141). Viral titers in the supernatant were evaluated using plaque assays at an appropriate dilution. These plaque assays were compared to HSV-1 produced from a transfection with the DNA vector LOKOS^{YA}

A

No HSV-1
genome   HSV-1 genome
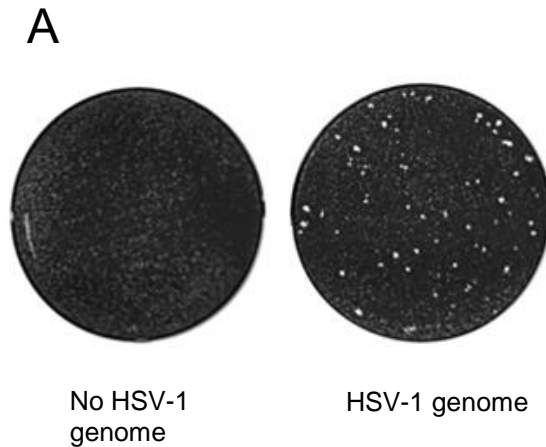
B
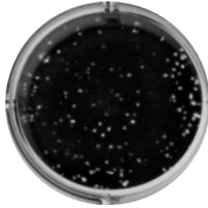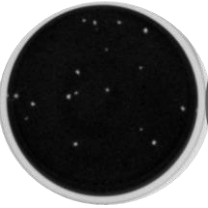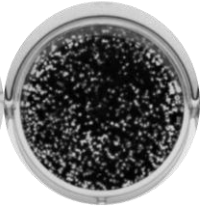
Sample description	VP35 + HSV-1 genome, (mixed culture) YYC001+ YLO141	VP35 + HSV-1 genome, (single strain) YLODB1	HSV-1 genome only, YLO141	VP35 expression only, YYC001	HSV-1 produced from LOKOS[YA]
Plaque formation	+	+	-	-	+

96

Table 3-1 Significance of factors impacting vector delivery and recipient cell viability for HEK293 cells.

Factor	*P-value*	*P-value source*	*Optimized condition for HEK293 cells*	*Range of conditions tested*
Concentration of drug	<0.001*	OFAT	(**Fig. 3-2**)	(**Fig. 3-2**)
Ratio of yeast to recipient cells	<0.001*	OFAT	100:1	1:1000–1:1
Fusion time	<0.001*	Fractional factorial	1 minute	1–20 minutes
PEG concentration	<0.001*	OFAT	50%	20–50%
PEG molecular weight	<0.001*	OFAT	2000 g/mol	1000–8000 g/mol
PEG buffer	<0.001*	OFAT	25 mM HEPES	10–100mM HEPES and Tris
Use of Amphotericin B in media	<0.001*	OFAT	2.5 µg/mL	0-2.5 µg/mL
Dilute media after adding PEG	0.0469*	Fractional factorial	Dilute with 1 mL	Dilution versus no dilution
Osmotically stabilize in sorbitol	0.0808	Fractional factorial	Stabilize at 4°C for 16 hours	Stabilize at 4°C for 0–24h
Pre-fusion incubation time	0.2789	Fractional factorial	1 minute	1–20 minutes
Calcium concentration	0.3643	Fractional factorial	0 M	0–0.1 M
DMSO amount	0.4032	Fractional factorial	10%	0–10%
Replace media after 12 hours	0.4051	Fractional factorial	Replace with DMEM + 10% FBS	Replacement versus no replacement
Stirring during PEG incubation	0.4431	Fractional factorial	Stirring not necessary	Stirring versus no stirring
Recovery time	0.4892	Fractional factorial	All conditions tested	0–5 minutes
FBS amount	0.6071	OFAT	All conditions tested	5–10%
Recovery temperature	0.837	Fractional factorial	25°C	4–42°C

PEG incubation temperature	0.8798	Fractional factorial	25°C	25–42°C
Yeast growth phase	0.928	Fractional factorial	All conditions tested	Log to stationary phase

(OFAT – One Factor At a Time)

Table 3-2 Vector delivery efficiency according to cell line and vector size.

Cell line	Fusion protocol optimized for	Number of reporter expressing cells observed (per 300,000)	Vector size
HEK293	HEK293	326 ± 26*	1.1 Mb
HEK293	HEK293	364 ± 19*	100 kb
HEK293	HEK293	341 ± 21*	12 kb
HEK293	HEK293	0	No vector
HEK293	Vero	23 ± 6	100 kb
HEK293	C6/36	32 ± 8	100 kb
HeLa	HEK293	3.5 ± 2	100 kb
CHO	HEK293	3.0 ± 0.7	100 kb
A549	HEK293	1.5 ± 0.5	100 kb
Vero	HEK293	2.0 ± 0.7	100 kb
Vero	Vero	102 ± 6	100 kb
C6/36 (insect)	HEK293	1.5 ± 1.0	100 kb
C6/36 (insect)	C6/36	112 ± 14	100 kb
DF-1 (chicken)	C6/36	1.0 ± 0.5	100 kb
DF-1 (chicken)	HEK293	0	100 kb
DF-1 (chicken)	Vero	0	100 kb

*No significant difference (*F ratio* = 2.15, *F critical* = 5.14, *P-value* = 0.1975).

Supplementary Table 3-1. List of plasmids and YCps.

Vector name	Description
pRS313	ATCC 77142
pRS313-mCHerGFP	pRS313 + GFP expression with mammalian CMV promoter + mCherry expression
pRS313-100mCher	pRS313 + 100kb *Mycoplasma mycoides* genome (segment 1)
pRS30x-VP35	Yeast codon optimized Ebola VP35 synthesized by GenScript inserted into pRS30X
pRS313-1100mCher	pRS313 + 1.1Mb *Mycoplasma mycoides* genome
LOKOS[YA]	Yeast strain with HSV-1 genome (21)

Supplementary Table 3-2. List of yeast strains.

Strain name	Relevant Genotype	Source
VL6-48	MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 cir0	ATCC MYA-3666
YLO141	MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 LOKOS[YA]	(21)
YYC001	MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 pRS30x-VP35	this study
YDB012	MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 pRS313-mCHerGFP	this study
YDB123	MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 pRS313-100mCHerGFP	this study
YDB1100	MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 pRS313-1100mCHerGFP	this study
YLODB1	MATα his3-Δ200 trp1-Δ1 ura3-52 lys2 ade2-1 met14 LOKOS[YA] pRS30x-VP35	this study

Supplementary Table 3-3. Fractional factorial design for the screening of eleven components for HEK293 cells and the design response.

Pattern	DMSO amount (mL)	Calcium concentration (mM)	PEG incubation temperature (°C)	Stabilize in sorbitol	Growth phase of yeast	Fusion time (minutes)	Stirring during PEG incubation	Dilute media after adding PEG	Recovery time (minutes)	Recovery temperature (°C)	Replace media after 12 hours	Cells with delivered vector
——————++——+——+	0	1	25	no	stationary	1	yes	no	1	4	no	8
+—+————++——+	10	10	25	yes	log	1	no	no	1	4	no	9
+——++++—+—+—	10	1	37	no	stationary	10	yes	no	1	25	yes	2
+—+—+————++—	10	10	25	no	log	1	yes	yes	10	25	yes	35
——++——+——+++	0	10	37	yes	log	10	yes	yes	10	25	no	1
——++++—+—+——	0	10	37	no	stationary	1	no	yes	10	4	yes	12
—++——+——++++	0	10	25	yes	stationary	1	yes	no	10	25	no	5
++——+——+++++—	10	1	25	no	log	1	no	no	10	25	yes	0
+——+——++++—+	10	1	37	yes	log	10	no	no	10	4	no	1
—++++—+—+———	0	10	37	no	log	10	yes	no	1	4	yes	0
++—+—+————++	10	1	37	yes	stationary	1	yes	yes	1	25	no	0
————++——+——++	0	1	37	no	log	1	no	yes	1	25	no	6
—+————++——+—	0	1	25	yes	log	10	no	yes	1	25	yes	0

100

—–+––++++–+–	0	10	25	yes	stationary	10	no	no	1	25	yes	1
–+-+–––++––	0	1	37	yes	log	1	yes	no	10	4	yes	0
–+––++++–+–+	0	1	25	no	stationary	10	no	yes	10	4	no	0
++++–+–+––––	10	10	37	yes	stationary	1	no	yes	1	4	yes	0
+–––++––+––	10	1	25	yes	stationary	10	yes	yes	10	4	yes	2
++++++++++++	10	10	37	no	stationary	10	no	no	10	25	no	0
+++–+-+––––+	10	10	25	no	log	10	yes	yes	1	4	no	0

Supplementary Table 3-4. Plackett-Burman design for the screening of seven components for Vero cells and the design response.

Pattern	Fusion time (minutes)	PEG incubation temperature (°C)	Pre-fusion incubation time (minutes)	Dilute media after adding PEG	Recovery time (minutes)	Recovery temperature (°C)	Stirring during PEG incubation	Cells with delivered vector
+−−+−++	10	25	1	no	1	25	no	15
+++++++	10	37	10	no	10	25	no	32
−−+−+++	1	25	10	yes	10	25	no	15
−+−−+−+	1	37	1	yes	10	4	no	11
−−+−−+−	1	25	10	yes	1	25	yes	14
++−−−+−	10	37	1	yes	1	25	yes	40
+−+++−−	10	25	10	no	10	4	yes	22
−−−+−−+	1	25	1	no	1	4	no	2
+++−−−+	10	37	10	yes	1	4	no	39
−+−+++−	1	37	1	no	10	25	yes	4
−+++−−−	1	37	10	no	1	4	yes	13
+−−−+−−	10	25	1	yes	10	4	yes	26

102

Supplementary Table 3-5. Plackett-Burman design for the screening of five components for C6/36 cells and the design response.

Pattern	Pre-fusion incubation time (minutes)	PEG incubation temperature (°C)	Fusion time (minutes)	Recovery time (minutes)	Recovery temperature (°C)	Cells with delivered vector
−+−++	1	37	1	10	25	13
−+++−	1	37	10	10	4	43
+−−+−	10	25	1	10	4	23
−−+−−	1	25	10	1	4	44
−+−−+	1	37	1	1	25	22
−−+−+	1	25	10	1	25	46
++−−−	10	37	1	1	4	22
+−−−+	10	25	1	1	25	25
+++++	10	37	10	10	25	48
+−+++	10	25	10	10	25	52
−−−+−	1	25	1	10	4	17
+++−−	10	37	10	1	4	44

Supplementary Table 3-6. Central-Composite (RSM) Design for the optimization of a new cell line.

Run	Log (ratio yeast: recipient cell)	PEG incubation time (min)
1	1.5	0
2	0.5	0.5
3	1.5	5.25
4	2.5	0.5
5	2	5.25
6	2.5	10
7	2	5.25
8	0.1	5.25
9	2	5.25
10	1.5	5.25
11	0.5	10
12	2.9	5.25
13	2	12
14	1.5	10
15	2.5	10
16	2	1
17	2	1
18	0.5	10
19	0.5	5.5
20	2.5	1
21	2.5	5.5
22	2	10
23	2	10

Supplementary Table 3-7. Regression results for the Fractional factorial design for HEK293 cells.

Term	Contrast	Lenth t-Ratio	Individual p-value	Simultaneous p-value
Fusion time	-9.95	-6.34	0.0006	0.0061
Dilute media after adding PEG	-3.35	-2.13	0.0477	0.4568
Osmotically stabilize in sorbitol	-2.85	-1.81	0.0822	0.6669
Pre-fusion time	1.65	1.05	0.2767	0.9963
DMSO amount	1.25	0.8	0.4032	1
Stirring during PEG incubation	-1.15	-0.73	0.4412	1
Recovery time	-1.05	-0.67	0.4863	1
Recovery temperature	-0.35	-0.22	0.8302	1
PEG incubation temperature	-0.25	-0.16	0.8798	1
Growth phase of yeast	0.15	0.1	0.929	1
Fusion time x dilute media after adding PEG	3.16	2.02	0.0562	0.5287
Fusion time x osmotically stabilize in sorbitol	1.63	1.04	0.2801	0.9965
Dilute media after adding PEG x osmotically stabilize in sorbitol	1.26	0.81	0.3957	1
Fusion time x pre-fusion time	-0.09	-0.06	0.9526	1
Dilute media after adding PEG x pre-fusion time	-0.39	-0.25	0.8098	1
Osmotically stabilize in sorbitol x pre-fusion time	-0.66	-0.42	0.6886	1
Fusion time x DMSO amount	-0.57	-0.37	0.7267	1
Dilute media after adding PEG x DMSO amount	1.04	0.66	0.5281	1
Osmotically stabilize in sorbitol x DMSO amount	-0.67	-0.43	0.6829	1

Supplementary Table 3-8. Regression results for the Plackett-Burman design for Vero cells.

Term	Contrast	Lenth t-Ratio	Individual p-value	Simultaneous p-value
Fusion time (minutes)	9.58	5.37	0.0039	0.0264
Dilute media after adding PEG	-4.75	-2.66	0.0292	0.207
PEG incubation temperature (°C)	3.75	2.1	0.0548	0.3575
Pre-fusion incubation time (minutes)	3.08	1.73	0.0953	0.5615
Recovery time (minutes)	-1.08	-0.61	0.5881	1
Recovery temperature (°C)	0.58	0.33	0.7674	1
Stirring during PEG incubation	-0.42	-0.23	0.832	1
Fusion time (minutes) x Dilute media after adding PEG	-1.08	-0.61	0.5881	1
Fusion time (minutes) x PEG incubation temperature (°C)	1.49	0.84	0.3653	0.9975
Dilute media after adding PEG x PEG incubation temperature (°C)	0.07	0.04	0.9732	1
Fusion time (minutes) x Pre-fusion incubation time (minutes)	-1.30	-0.73	0.4312	1

Supplementary Table 3-9. ANOVA of regression model optimizing vector delivery or Vero cells. The model P-value was very low (<0.001). The lack of fit P-value of 0.0629 implied that the lack of fit was not significant relative to the pure error. ANOVA table for the model shown in supplementary figure 3-1.

Source	Degrees of freedom	Sum of Squares	Mean Square	F ratio	Prob > F
Model	5	27515.54	5503.11	18.032	<0.001*
Log (Ratio of Yeast: Vero cells), A	1	5126.07	5126.07	16.79	0.0007*
PEG incubation time, B	1	20.356	20.356	0.067	0.7993
A x A	1	6774.44	6774.44	22.19	0.0002*
B x B	1	11294.69	11294.69	37.01	<0.0001*
A x B	1	131.45	131.45	0.43	0.5204
Residual	17	5188.19	305.19		
Lack of fit	10	4282.19	428.22	3.308	0.0629
Pure error	7	906.01	129.42		
C. Total	22	32703.73			

Supplementary Table 3-10. Regression results for the Plackett-Burman design for C6/36 cells.

Term	Contrast	Lenth t-Ratio	Individual p-value	Simultaneous p-value
Fusion time	12.9167	9.62	0.0003	0.0034
Pre-fusion incubation time	2.4167	1.8	0.0891	0.5177
PEG incubation temperature	-1.25	-0.93	0.3182	0.9907
Recovery temperature	1.0833	0.81	0.3844	0.9995
Recovery time	-0.5833	-0.43	0.7	1
Fusion time x Pre-fusion incubation time	-0.4082	-0.3	0.7859	1
Fusion time x PEG incubation temperature	-0.2041	-0.15	0.8917	1
Pre-fusion incubation time x PEG incubation temperature	-0.5774	-0.43	0.7036	1
Fusion time x Recovery temperature	1.3093	0.98	0.2983	0.9818
Pre-fusion incubation time x Recovery temperature	1.1581	0.86	0.3515	0.9968
PEG incubation temperature x Recovery temperature	0.7071	0.53	0.6478	1

Supplementary Table 3-11. ANOVA of regression model optimizing vector delivery or C6/36 cells. The model *P*-value was very low (<0.001). The lack of fit *P*-value of 0.0508 implied that the lack of fit was not significant relative to the pure error.

Source	Degrees of freedom	Sum of Squares	Mean Square	F ratio	Prob > F
Model	5	18380.805	3676.16	18.5796	<0.001
Log (Ratio of Yeast: C6/36 cells), A	1	9107.217	9107.217	46.0285	<0.001*
PEG incubation time, B	1	7.166	7.166	0.0362	0.8513
A x A	1	4.616	4.616	0.0233	0.8804
B x B	1	10589.643	10589.643	53.5207	<0.001*
A x B	1	3188.129	3188.129	16.1130	<0.001*
Residual	17	3363.630	197.86		
Lack of fit	10	2817.6298	281.763	3.6123	0.0508
Pure error	7	546.0000	78.000		
C. Total	22	3363.6298			

Supplementary Table 3-12. ANOVA of regression model optimizing vector delivery or HEK293 cells. The model *P*-value was very low (0.0021). The lack of fit *P*-value of <0.001 implied that the lack of fit was significant relative to the pure error.

Source	*Degrees of freedom*	*Sum of Squares*	*Mean Square*	*F ratio*	*Prob > F*
Model	5	114936.85	22987.4	6.0930	0.0021*
Log (Ratio of Yeast: HEK293 cells), A	1	18798.428	18798.428	4.9827	0.0393*
PEG incubation time, B	1	19007.606	19007.606	5.0381	0.0384*
A x A	1	10146.952	10146.952	2.6895	0.1194
B x B	1	24238.213	24238.213	6.4245	0.0214*
A x B	1	1021.736	1021.736	0.2708	0.6095
Residual	17	5188.199	305.19		
Lack of fit	10	62824.807	6282.48	33.5193	<0.001*
Pure error	7	1312.000	187.43		
C. Total	22	179073.65			

Supplementary Table 3-13. List of multiplex primers

Primer name	sequence
Mmyc mp 100 F	CAA CTG ATA CAC CAA CCA TC
Mmyc mp 100 R	TTA TGG TAG TGG TTT TCA CAT
Mmyc mp 150 F	GCT TTG GTT ATC ATA TGT GAA C
Mmyc mp 150 R	CAA ATC CTT GAT CTT TAA TTA CTT G
Mmyc mp 300 F	TAT TGG TGA ACC AGT GGG
Mmyc mp 300 R	CCT TGT TCA ACA CGT AAT ACT G
Mmyc mp 400 F	GTG AGC AAC AAT GTT TTG AG
Mmyc mp 400 R	CAA CTC CAC CAA GTA CTC C
Mmyc mp 600 F	CTA AAC CAT CAG AAT TAG GTT C
Mmyc mp 600 R	GCA AAG TCA CAG ATC AAC AA
Mmyc mp 700 F	CAA CTC CAG AAG GTG CTC
Mmyc mp 700 R	CTA AAC TAA TTC TAA TAG CAC CC
Mmyc mp 800 F	TCG ATC ATT ATT TTA TAT GTT GTG
Mmyc mp 800 R	TAT AAT TCT TAC TCC AGC ATT TC
Mmyc mp 900 F	GCT CAT CAG CTT GAC TAA TTT G
Mmyc mp 900 R	CTA ATC TCA GAT ATT CAA GCA G

Chapter 4 Discussion and Future Directions

YCp delivery use in Gene Therapy

For years, the holy grail of DNA delivery has been the prospect of applying effective gene therapy to permanently alter genetically deficient cells using designed synthetic genes for therapeutic purposes. To date, gene therapy has remained largely confined to theoretical applications or phase I clinical trials [240, 241] with limited gene therapy treatments being commercially available. Notably, Glybera (UniQure), which delivers an intact copy of the human lipoprotein lipase gene to muscle cells using an adeno-associated virus to treat lipoprotein lipase deficiency [242] and Strimvelis (GSK), an *ex vivo* retroviral gene therapy vector to treat Severe Combined Immunodeficiency due to Adenosine Deaminase deficiency (ADA-SCID) [243, 244].

To break the barrier for more mainstream clinically relevant gene therapy there are several requirements that need to be addressed, including 1) ability to synthesize full length genes, or genetic clusters 2) ensuring proper safety 3) maintenance of stable expression over the long-term and 4) the ability to deliver genes effectively to the target cell. These requirements have not been met so far, however continued progress has put us on the precipice of developing successful gene therapy techniques.

The first requirement, reliable and cost effective gene synthesis, has been largely addressed with advances in synthetic biology assembly techniques. Synthesis of large genes is now relatively common [3, 97, 106, 154, 201, 245]. Functional genetic devices such as light induced optogenetic switches [246, 247], Boolean logic gates [248], tunable oscillators [249], and toggle switches [250] have been developed in mammalian

cells. Inexpensive DNA assembly techniques have enabled the construction of megabase sized DNA structures, opening up the potential for engineering genetic circuits resembling the intricate complexities of native mammalian gene regulation. Although improvements in DNA assembly techniques are desirable, current technology is sufficient to manufacture DNA encoding large mammalian gene clusters.

The second requirement for successful gene therapy, eliminating the risk of adverse events arising from altering patient genomes, is not sufficiently addressed by current gene therapy techniques. Viral transduction as a gene therapy vector presents a high risk of safety concerns. Viral vectors function by insertion into the chromosome. This raises the possibility of deleterious mutations being introduced into the target cell. In one pilot clinical trial, an *ex vivo* retroviral vector was used to treat a rare disorder, X-linked severe combined immunodeficiency (SCID). While effective, 5 of 20 patients three years post-treatment presented T-cell leukemia, two of whom are now deceased. A linkage between the retroviral vector and the presented leukemia was identified due to the common region of insertion upstream of a resident LMO2 oncogene [251]. Similarly, in another clinical trial the retrovirus used in a clinical trial to treat granulomatous disease has led to leukemia and one death [252-254]. Viral vectors typically insert into similar regions of the target cell chromosome, so it is possible to predict the frequency of a deleterious insertion. Adverse mutagenesis has not yet been associated with adenovirus vectors [240, 242-244]; however, mutations caused by adenovirus insertions cannot be avoided and insertion is still a somewhat random event. Additionally, host immune responses to viral vectors have been reported [255-257].

The third concern, maintenance of stable expression over the long-term, remains a problem for gene therapy. Current gene delivery methods deliver DNA only transiently or involve integration into the target cell chromosome. As a consequence, the expression level of the same delivered gene can vary depending on the position of insertion. Gene expression of a delivered DNA sequence is regulated by chromatin accessibility as well as the presence of promoters and enhancers [258, 259]. Further confounding predictability of delivered genes is the possibility of delivering multiple copies. HACs have the potential to overcome the limitations of integration-based vectors. They can be stably maintained if a functional kinetochore is present and likely have near unlimited cloning size capacity, including all the required regulatory elements is possible. Additionally, only one copy of the target gene is present, which minimizes the risk of confounding the expression level of introduced genes through gene duplication. It is worth noting that prior to the late 1980s, before the first construction of YACs [260] was possible, progress in identifying yeast gene function was slow and complex [261-263]. YACs introduced copy number control and mitotic stability, allowing the extensive genetic tools in yeast to be developed. Developing HAC technology could bring that same level of control to human cells as YACs have done for yeast.

Precise delivery of genes to target cells so that they are expressed as needed is the fourth concern of gene therapy. This needs to be addressed in future work. One main limitation of HACs is their reliance on microcell mediated chromosome transfer (MMCT) as a delivery technique. As discussed earlier, MMCT is a low efficiency technique that is only possible in a few cell lines. The YCp delivery protocol in

113

Chapter 3 could be used to enhance the delivery of HACs. We report that the delivery efficiency of transporting intact DNA chromosomes to be 10^3-10^5 more efficient than conventional MMCT. Even compared to newer more efficient MMCT delivery methods involving fusion with viral proteins, YCp delivery is at least 10^1-10^2 more efficient. Additionally, we have not encountered a cell type we have not been able to deliver a YCp to and optimize using our optimization protocol. Cell type specificity is likely not a concern for our YCp delivery protocol. Two separate phyla of *Animalia* can be transfected, *Chordata* and *Anthropada*. We attempted and achieved YCp delivery to cultured cells through PEG-mediated fusion of yeast spheroplasts with the cultured cells from six species from six separate families: *Mus musculus, Chlorocebus sabaeus, Cricetulus griseus, Gallus gallus domesticus, Aedes albopictus, and Homo sapiens*.

Potential challenges from delivering yeast chromatin

Mammalian chromatin is a highly complex network of proteins involved in DNA replication, differentiation, silencing, and gene expression. The possibility of a failure to deliver HACs to mammalian cells as yeast chromatin, to both replicate and segregate in mammalian cells needs to be addressed. One could speculate that delivered yeast chromatin may not properly express, segregate, or replicate in the mammalian cell.

HSV-1 DNA replication involves seven HSV-1 essential proteins along three origins of replication [264] including single-strand binding proteins, a DNA-helicase-primase complex, and a DNA polymerase. Inside a virion, HSV-1 is not associated with histones [265] nor are histones present on the viral genome [266, 267]. However, once

an HSV-1 genome enters a nucleus, the host mammalian cell will attempt to assemble chromatin onto the naked viral genome to silence incoming genes. Transfected DNA, generally, will not remain free in the nucleus. Histones and basic nuclear matrix proteins will quickly assemble transfected DNA into mammalian chromatin-like structures condensing DNA, which may lead to silencing of the transfected DNA [268-270]. Further, chromatin immunoprecipitation (ChIP) studies have demonstrated that histones are associated with lytic genomes during lytic infection [271, 272]. In addition, in a latent infection HSV-1 is assembled into nucleosomes, but is not methylated [273-275]. HSV-1 replication progresses once HSV-1 has been present as euchromatin, suggesting that HSV-1 replication evolved to process euchromatin [271, 272]. We observe efficient replication of delivered methylated [276] yeast chromatin by enzymes evolved to process unmethylated mammalian euchromatin suggesting that yeast chromatin may also be recognized by mammalian DNA polymerases. We demonstrated that the mammalian RNA polymerase II can express yeast chromatin, as evidenced by experiments using our mCherry reporter plasmid and a HSV-1 genome cloned as a YCp. We also demonstrated successful HSV-1 replication from HSV DNA polymerase. Considering HSV-1 polymerase can recognize and replicate yeast chromatin we hypothesize that the delivered YCps may be treated like naked DNA becoming similar to mammalian chromatin by the addition of mammalian histones, thus capable of replicating.

Before recent advances in HAC technology, yeast spheroplast fusion to deliver YACs was used to try to determine the essential elements of a human chromosome. At the

115

time, the results were difficult to interpret and the approach was set aside in lieu of more consistent methods of generating HACs, which will be discussed in the following section. In retrospect, interpreting YAC delivery to mammalian cells is easier. In 1993, YACs transferred to mouse A-9 cells by spheroplast fusion were found to be successfully integrated into the target cell chromosome and maintained at least 27 cell divisions after fusion [277]. This mirrors results for microinjected DNA into mouse cells [84]. Even DNA constructs in excess of 500 kb can be microinjected into mouse cell nuclei or be fused with mouse embryonic stem cells to become successfully integrated and can be used to generate transgenic mice [71, 101, 278]. These results suggest that yeast chromatin delivered by spheroplast fusion or microinjected into mammalian cells can be replicated by mammalian DNA polymerases and properly segregate once integrated into the genome.

Replication of delivered YACs existing as extrachromosomal elements in mammalian cells has been identified in three instances [277, 279, 280]; however, the elements were only partially stable. McGuigan and Huxley tracked an introduced a YAC into mouse A-9 cells by spheroplast fusion. Out of 17 clones tested, two were found to be relatively stable extrachromosomal elements as determined by FISH. The presence of both integrated and extreachromosomal elements after 26 cell divisions supports the conclusion that efficient replication of yeast DNA is occurring. However, after 20 days without selection, 98% of cells lost the extrachromosomal element. Transient extrachromosomal clones are lost rapidly in the absence of selection (~16% loss per cell division) [279] whereas the relatively stable yeast extrachromosomal elements

116

have a reported loss rate of ~3% per cell division in the absence of selection. This suggests that the centromeres randomly introduced into the delivered YACs were partially functional. After careful examination of the introduced YACs, they were found to contain mostly YAC DNA. However mouse centromeric satellite DNA was also identified. This suggests that a recombination occurred and that the mouse centromeric DNA was driving centromere formation. The following year Willard and colleges developed the "bottom up" HAC [142].

Potential for use in human artificial chromosomes

There are two general types of HACs, "top down" and "bottom up". "Top down" HACs focus on engineering HACs from existing chromosome structures. Small chromosomes such as the human Y chromosome are whittled down to a more manageable size for experimentation. This is typically performed based on telomere-associated chromosome fragmentation in the homologous recombination proficient chicken DT40 cell line. This approach has been extensively reviewed [115, 154, 281] [141]; however, it is not amenable to yeast fusion. These HACs are larger than the largest reported YACs [157]. *De novo* HACs or "bottom up" engineered HACs use alpha DNA cloned in BACs or YACs and are transfected into mammalian cells, which multimerize and form stable HACs 1-10 Mb from the amplification of alpha DNA. The Willard HAC [142] was produced from arrays of a 2.7 kb repeats from the chromosome D17Z1 to produce a final linear 173 kb product. Upon transfection into HT1080 cells, the HAC further multimerized and formed HACs 6-10 Mb. Once expanded to that size the HAC stably replicated and segregated for over 6 months as determined by FISH.

The principle of using repeated segments of alphoid satellite DNA first employed by the Willard HAC, largely remains today. One notable advancement in HAC technology is the development of the aplhoid^{tetO}-HAC (**Figure 4-1**). In this HAC, a ~50 kb synthetic alphoid DNA array with a tetracycline operator (tetO) were combined to form a conditional centromere [147, 148, 282] such that when a fusion protein of the Tet Repressor (TetR) with a KRAB silencing domain (tTS) [283] is expressed the kinetochore disassociates and the HAC is lost. This aplhoid^{tetO}-HAC was used in subsequent studies to determine how kinetochore formation occurs and other proteins involved in the process [156]. Another notable advancement is the insertion of a loxP landing pad site into a HAC present in CHO cells allowing the HAC to be manipulated using Cre-LoxP recombination [143, 147].

Until recently, *de novo* HAC formation has been limited to very few human cell lines, notably HT1080. Forming a stable kinetochore is required for accurate chromosome segregation and *de novo* HAC formation. The mechanism by which a kinetochore forms is still somewhat unclear; however recent advances have improved our understanding. Kinetochores assemble around chromatin containing the histone H3 variant CENP-A which binds to the 17 bp CENP-B box motif [155] in repeated sections of alpha satellite DNA but this highly complex process involves the recruitment of over 100 different proteins [284]. Alpha satellite repeats are sufficient for CENP-A assembly in many cell types, however the stability of the assembly, until recently, was only available in HT1080 cells. Recently Ohzeki *et al.* [124] discovered that the H3K9 acetyl/methyl balance explains the cell specificity for *de novo* HAC formation.

118

HT1080 cells appear to be naturally high in histone acetyltransferases relative to histone methyltransferases. Using the aplhoid^{tetO}-HAC, a fusion protein of TetR and histone acetyltransferases tethered to the TetO domains allowed the HAC to be stably maintained in HeLa cells, a cell line previously inaccessible to stable HAC formation.

Considering these all factors, it seems highly likely that YACs delivered to mammalian cells using spheroplast fusion could form stable HACs. Yeast chromatin can be successfully expressed, replicate, and segregate in mammalian cells. Cloning alpha satellite DNA in a yeast vector while including binding sites for histone acetyltransferases and Cenp-A will likely form stable HACs once delivered into mammalian cells. Using yeast to manipulate HACs would greatly extend the applicability of HACs and may enable HACs able to be easily constructed, modified, and delivered through yeast.
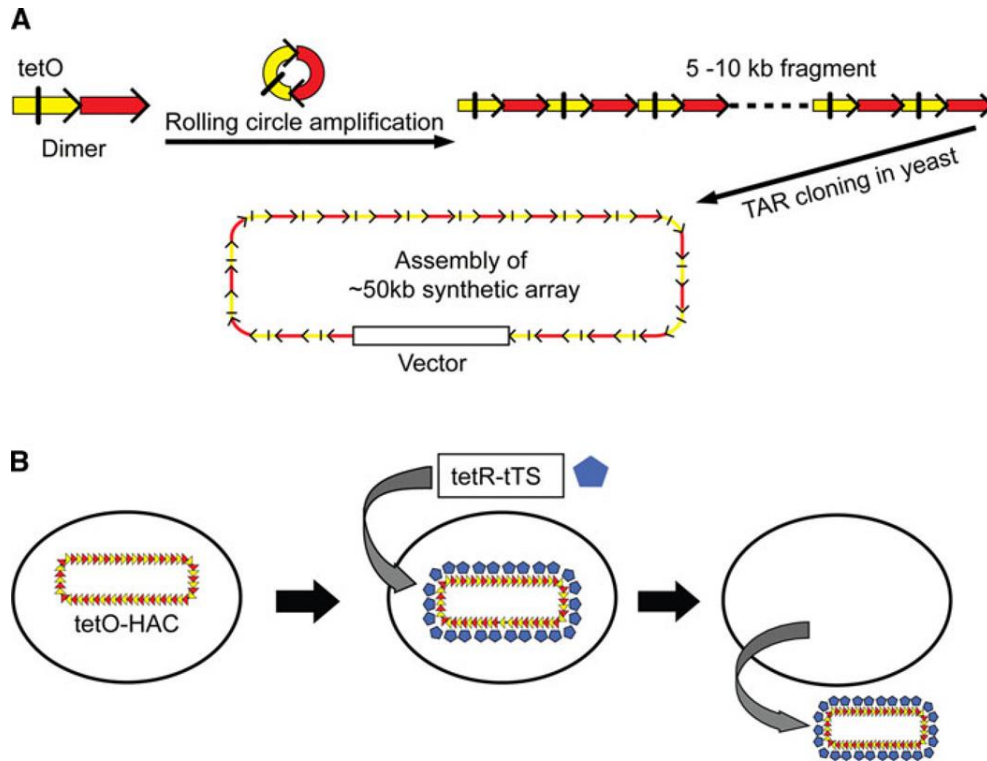
Figure 4-1 Schematic demonstrating how the aplhoidtetO-HAC is constructed, delivered and lost. Adapted from [141]

An example of construction of the de novo generated human artificial chromosome via bottom-up approach using a synthetic alphoid DNA array. **a** Schematic representation of construction of the synthetic tetO-containing DNA tandem repeats array by rolling circle amplification (RCA) in vitro and transformation-associated recombination (TAR) cloning in yeast cells. The first step included amplification of the dimer by RCA up to 5–10 kb in size fragments. One monomer of the dimer is derived from a chromosome 17 alphoid type I 16-mer unit and contains a CENP-B box. The second monomer is a wholly synthetic sequence derived from alohoid DNA consensus, with sequences corresponding to the CENP-B box replaced by a 42-bp tetO motif. The second step included co-transformation of the RCA-amplified fragments into yeast cells along with a vector containing alphoid-specific hooks. End-to-end recombination

120

of alphoid DNA fragments followed by interaction of the recombined fragments with the vector resulted in the rescue of an approximately 50-kb array as a circular molecule in yeast. The targeting vector contains a yeast cassette, *HIS3/CEN/ARS* (a selectable marker *HIS3*, a centromere sequence *CEN6* from yeast chromosome VI and yeast origin of replication *ARSH4*) and a mammalian selectable marker (the Blasticidin resistance gene) and a BAC replicon that allows a YAC clone to be transferred into *E. coli* cells. **b** The alphoid^tetO^-HAC loss induced by targeting of the transcriptional silencer (tTS) fused with the tet-repressor (tetR) into the HAC kinetochore. After expression of a chromatin modifier gene (tTS), the HAC is maintained stably when cells are growing in the presence of doxycyclin that prevents binding of tTS to tetO sequences or the HAC is destabilized when cells are growing in the absence of doxycycline. Image and description adapted from [141]

Other applications of delivering DNA with yeast synthesized functional protein

Applications that involve the delivery of DNA along with functional protein could be developed using the YCp delivery method. The YCp delivery technology described in this thesis has already been used to demonstrate the delivery of protein and DNA to the same mammalian cell twice. First, the reporter YCp expressed GFP while the yeast spheroplast also delivered mCherry to the mammalian cytoplasm. Second, the HSV-1 genome along with Ebola VP35 was delivered to the same mammalian cell and demonstrated a differential in infectivity between different cell lines for HSV-1

generation. In **Chapter 3**, we showed that a yeast strain engineered to contain the entire HSV-1 genome and to also express the Ebola derived VP35, which blocks the host cell antiviral response [214, 215], could shut down host cell interferon response and obtain viral replication in cells which HSV-1 had not replicated. Yeast fused with Vero cells, HeLa cells, and HEK293 cells showed differentials in HSV-1 generation levels depending on whether VP35 was included or not. HEK293 cells, which do not express interferon, and are normally permissive for HSV-1 replication were unaffected by the VP35, whereas in Vero and HeLa cells, virus generation was enhanced if VP35 was also delivered during cellular fusion. This differential in virus generation efficiency and viral rescue consistency could be explored further, and expanded into other applications. Indeed, we have developed an assay that could be used in any application needing to rapidly test different sets of proteins with different sets of DNA. One potential application is to determine the effects of different proteins on the reverse genetics of HSV-1. We have evidence that the difference in HSV-1 rescue consistency from fusion is retained if a mixture of yeast strains is used in a spheroplast fusion, one strain containing the protein of interest, and another strain containing the HSV-1 genome of interest. Different HSV-1 genomes could be engineered in yeast, and different proteins involved in host defense inhibition could be expressed in yeast. Using this methodology, we could quickly screen permutations of different viral genomes with different proteins and observe the effect on viral replication and/or generation. This may have applications in the development of HSV-1 viral transduction vectors.

Investigating mechanism of nuclei escape

We successfully optimized the process of YCp delivery from a yeast cell to the nucleus of a recipient cell by eliminating one cellular barrier. We arrested the cell cycle in M-phase so that the yeast fusion took place while the nuclear membrane was absent, thus, eliminating the cell's nuclear barrier and allowing the donor DNA to be encapsulated by the new nuclear membrane as cell division proceeds. However, there are other barriers to consider. The cell membranes of the yeast spheroplast and the recipient cell are not significant barriers. We observe ~25% of recipient cells received significant levels of cytoplasmic contents from yeast spheroplasts indicating that yeast spheroplasts and recipient cells fuse readily. Since the fusion rate is very high and the YCp delivery rate is relatively low, the rate limiting step must proceed after fusion of the spheroplasts with recipient cell (the cell membrane barrier). Eliminating the recipient cell nuclear barrier significantly increased YCp delivery rates; however, YCp escape from the yeast nuclei could potentially remain a significant barrier.

YCp escape from the yeast nuclei must be occurring because YCps are successfully delivered to the recipient cell nucleus, although the mechanism by which YCps escape the yeast nuclei remains unclear. It might be that once a yeast nuclei enters a mammalian cell it is not degraded and just floats in the cytoplasm. Instead, YCp entry into the mammalian cell may occur when the yeast spheroplasts lyse and free yeast nuclei fuse with recipient mammalian cells, releasing the inner nuclear vesicle into the recipient cell cytoplasm. However, we also know that at least most yeast spheroplasts must remain intact prior to fusion because we observe the mCherry reporter expressed

123

by the yeast cells in a large percentage of mammalian cells (the mCherry protein in the yeast cytoplasm was dispersed into the mammalian cytoplasm upon cell fusion). This explains why so many mammalian cells contain the mCherry reporter and so few mammalian cells express the GFP reporter encoded on the YCp.

We first speculated that YCp escape from the yeast nuclei might be the result of an analogous effect to the proton-sponge effect used by DNA to escape from the endosome discussed in **Chapter 1**. The donor yeast nuclei likely have a different osmotic environment than the mammalian cell cytoplasm. If the difference in osmotic pressure is hypotonic, nuclear lysis might release the YCp into the cytoplasm. However, this is extremely unlikely for two reasons. First, the yeast nucleus will be released to the cytoplasm as a double membrane organelle making lysis of two membranes far less likely. Second, prior to yeast fusion, the yeast is incubated in the osmolyte sorbitol. After incubation in sorbitol yeast cells will contain high concentrations of the osmolyte through the transport of sorbitol through the sorbitol/mannitol ABC transporter [285]. The yeast nuclear pore is large enough for small molecules to diffuse, including sorbitol [35, 286-288], therefore a high sorbitol concentration will exist in delivered yeast nuclei. Sorbitol functions to remove excess water from the yeast nucleus by carrying bound water down the osmolyte concentration gradient out of the yeast nucleus and would likely create a hypertonic environment inside the yeast nucleus making nuclear lysis highly unlikely. Indeed, resistance of yeast spheroplasts to lysis in hypotonic environments is observed when incubated with sorbitol [108, 289].

We hypothesize that a far more likely mechanism for chromosome escape from the yeast nuclei is what we term membrane system disassociation (MSD). This is when the entire yeast nuclear and endoplasmic system breaks down inside the mammalian cytoplasm or possibly inside the mammalian nucleus. The yeast nucleus is a double membrane bound organelle contiguous with the endoplasmic reticulum. The structure of the mammalian nucleus is ensured by structural proteins along the ER and nucleus. Analogous yeast structural proteins are unknown but likely. Breakdown of the nucleus occurs naturally in nearly all eukaryotes at the initiation of the cell cycle when CDC2 phosphorylates lamin proteins, the proteins responsible for maintaining the structure of the nuclear envelope. The phosphorylated lamins then depolymerize and nuclear vesicles form upon fusion of the inner and outer membranes of the nucleus. Budding yeast lacks lamins [290, 291]; although a similar lamin-like protein is likely present, but has not yet been identified. MSD might be occurring during YCp delivery after the lamin-like structural proteins breakdown or are exposed to a mammalian cytoplasmic environment. MDS occurs when the inner and outer membranes of the nucleus fuse forming nuclear vesicles and, like in mitosis, would result in the exposure of the yeast chromosomes and yeast nuclear DNA to the cytoplasm.

Alternatively, MSD may occur when yeast proteins that function to normally produce spindle pole bodies or nuclear pore complexs by inducing fusion of the inner and outer nuclear membranes act unrestricted without the regulatory environment of the yeast cell. Reticulons and the related protein Yop1/DP1 are thought to induce curvature in the yeast nuclear membrane, the first step of membrane fusion, and are required for

normal spindle pole body or nuclear pore complex formation in yeast [292, 293]. Nbp1 is a stabilizing protein that binds to curved membranes that interestingly shows no specificity for curved or planar membranes in vitro [293]. It is possible that the regulation of spindle pole body or nuclear pore complex forming proteins may be altered in the context of the mammalian cell, inducing fusion of the inner and outer nuclear membranes revealing the yeast nuclear contents to the mammalian cytoplasm. MDS would occur if enough of these proteins induced pore formation creating an unstable yeast nucleus. Further understanding of the mechanism behind yeast nuclear structure and/or the regulation of nuclear pore formation may elucidate a mechanism for how YCp yeast nuclear escape occurs.

Viral rescue efficiency of delivered YCps

We wanted to determine if YCp delivery was more efficient at generating HSV-1 than naked transfected BACs. If transfected BACs were more efficient at generating virus than YCps, it could be that yeast chromatin may be interfering with HSV-1 gene expression or replication. Or the escape from yeast nuclei inhibit reverse genetics more than shear forces from isolating BACs. If YCp delivery is more efficient than BAC transfection, YCp delivery could have applications with other large DNA viruses reverse genetics such as cytomegalovirus, poxviruses, or African swine fever virus leading to vaccine development or in the initial construction of viral delivery vectors. Generating virus from spheroplast fusion has the advantage of not subjecting the genome to shear forces. The efficiency of virus generation from both YCp delivery and a transfection of a BAC using cationic lipids can be estimated from the single

plaque assay performed, shown in **Supplementary Figure 3-5**. Although this is just a

preliminary comparison, it could indicate which method is more efficient.

To compare each delivery method, the efficiency of viral generation per genome was

calculated. First, 2.5 µg of the HSV genome was used for a transfection. Calculating

for a 152 kb genome, 1.5×10^{10} molecules were used in the transfection. Five days

post transfection 15×10^5 PFU/ml was observed, 2×10^{-4} PFU/genome. The number

of yeast cells used in a fusion was 1×10^8 and on average 1 genome per yeast cell was

expected. Three days post fusion 60×10^3 PFU/ml was observed, 1.2×10^{-3}

PFU/genome. We see that the efficiency of viral generation per genome is 10x higher

using YCp delivery; however more HSV genomes can be used in a transfection.

Therefore, even though the efficiency per genome decreased, the viral titers are higher.

However, this experiment needs to be repeated and titers taken at the same day post

fusion/transfection. It is likely that HSV continued to replicate following transfection

leading to higher titers. If a straight comparison was performed, it is likely YCp

delivery would show more improvement relative to BAC transfection. Additionally,

152 kb is not exceedingly large, a 1 Mb construct would require using 16 µg of DNA

to be transfected to keep the same number of copies. Therefore, the comparative

advantage of YCp delivery efficiency, per genome, could improve the larger the desired

DNA vector. YCp delivery can likely enhance the reverse genetics of other large DNA

viruses because of the enhancement of the per genome delivery efficiency, increased

genome stability in yeast relative to *E. coli*, and the ability to clone larger constructs in

yeast.

Use of design of experiments (DOE) in biology and YCp delivery

In **Chapter 3** we discuss an application of DOE to optimize a biological process (YCp delivery). DOE is significantly underused in biology but there have been successful applications [294-298]. In **Appendix D** I discuss the basics of DOE using an example from Anderson and Witcomb [299] and discuss how it can be superior to optimize a process rather than optimizing one factor at a time. Although not ideal, we demonstrated how DOE had the statistical power to optimize a very low efficiency process. The most significant gain in utility was from the fractional factorial screening designs. Testing numerous factors can be burdensome and time consuming. DOE quickly screens out trivial factors and allows the optimization of the select few. Particularly for biological processes, there are frequently numerous and sometimes unknown factors. Using fractional factorial screening designs can greatly reduce the number of experimental runs and potentially focus a researcher to the important variables that affect a biological process.

To optimize YCp delivery we first screened 11 factors at once for delivery to HEK293 cells. We were able to reduce the number of factors screened following each cell line. By screening out trivial factors we derived a central composite RSM design (**supplementary table 3-6**) that likely could be applied to many more cell lines. Through the pipeline described in **Chapter 3** we could optimize C6/36 cells in just 2 weeks increasing YCp delivery rates nearly 100-fold. This was possible because the fractional factorial design revealed the same two significant factors as the previous cell lines. If we assume those are the only two critical factors in DOE optimization it is

possible that simply using the RSM design given in supplementary table 3-6 could yield optimization for a new cell line following optimization with a mitotic drug.

Significance of using the E2 system to enhance gene delivery

The significance of this project is in using the E2 plasmid delivery system when conditions are not ideal. Frequently a cell line can be hard to transfect [300-305]. Other viral strains have shown to be difficult to rescue, [306] sometimes due to tandem repeats[307] or cytotoxic genes [308]. When time is a factor and consistent transfection is needed the E2 system may be warranted.

Mycoplasma mycoides contamination concern

Mycoplasma contamination is a common problem across multiple research labs [309-311]. *Mycoplasma* biology is studied at JCVI and in particular *Mycoplasma mycoides* was used as a base template organism for the creation of the first synthetic cell [97]. There was a concern that contamination could occur across labs potentially confounding the results shown in this dissertation. To address this concern, in **Appendix C** I developed an infectivity assay based on a modified color changing units (CCU) assay to detect if *Mycoplasma* could be contaminating mammalian cell cultures. In the process discovered that the deletion of 8 genes in *Mycoplasma mycoides* resulted in the failure to contaminate mammalian cell cultures.

Conclusion

The aim of this thesis was to enhance the delivery of synthetic DNA through the nuclear membrane barrier. First, we developed a plasmid delivery system utilizing the bovine papillomavirus E2 scaffolding protein that when expressed by the transfected cell line, more consistently delivered plasmids bearing the E2 binding site to the nucleus of mammalian cells. This technique enabled us to more consistently deliver DNA when conditions were not ideal or a strain was difficult-to-rescue. Second, we improved YCp delivery by synchronizing recipient cells in mitosis when the nuclear envelope is broken down. This technique allowed us to deliver large megabase sized YCps to the nucleus at the same rate as smaller YCps. Improvements in DNA delivery of both large megabased sized YCps and small plasmids resulted from the targeting of the nuclear membrane barrier. We conclude that the nuclear barrier is frequently the limiting factor affecting the delivery of synthetic DNA.

Appendix A: Synthetic generation of influenza vaccine viruses for rapid response to pandemics

Authors: Philip R. Dormitzer[1*], Pirada Suphaphiphat[1], Daniel G. Gibson[2,3], David E. Wentworth[2], Timothy B. Stockwell[2], Mikkel A. Algire[2], Nina Alperovich[2], Mario Barro[4], David M. Brown[2], Stewart Craig[1], Brian M. Dattilo[4], Evgeniya A. Denisova[2], Ivna De Souza[1], Markus Eickmann[5], Vivien G. Dugan[2], Annette Ferrari[1], Raul C.

Gomila[1], Liqun Han[1], Casey Judge[1], Sarthak Mane[1], Mikhail Matrosovich[5], Chuck Merryman[2], Giuseppe Palladino[1], Gene A. Palmer[1], Terika Spencer[1], Thomas Strecker[5], Heidi Trusheim[1], Jennifer Uhlendorff[5], Yingxia Wen[1], Anthony C. Yee[2], Jayshree Zaveri[2], Bin Zhou[2], Stephan Becker[5], Armen Donabedian[4], Peter W. Mason[1], John I. Glass[2], Rino Rappuoli[1], J. Craig Venter[2,3] [8]

Affiliations:

[1]Novartis Vaccines and Diagnostics, Cambridge, MA 02139, USA; Holly Springs, NC 25740, USA; and 53100 Siena, Italy.

[2]The J. Craig Venter Institute, Rockville, MD 20850, and San Diego, CA 92121, USA.

[3]Synthetic Genomics, Inc., La Jolla, CA 92037, USA.

[4]The United States Biomedical Advanced Research and Development Authority, US Department of Health and Human Services, Washington, DC 20201, USA.

[5]Institut für Virologie, Philipps-Universität Marburg, D-35032 Marburg, Germany

Abstract

During the 2009 H1N1 influenza pandemic, vaccines became available in large quantities only after human infections peaked. To accelerate vaccine availability for future pandemics, we developed a synthetic approach that very rapidly generates vaccine viruses from sequence data. Beginning with HA and NA gene sequences, we combined an enzymatic, cell-free gene assembly technique with enzymatic error correction to allow rapid, accurate gene synthesis. Viruses for use in vaccines were

131

rescued from an MDCK cell line qualified for vaccine manufacture that had been transfected with synthetic viral RNA expression constructs encoding HA and NA and plasmid DNAs encoding viral backbone genes. Performing this rescue with improved vaccine virus backbones increased the yield of the essential vaccine antigen, hemagglutinin. Generation of synthetic vaccine seeds, together with more efficient vaccine release assays, would accelerate responses to influenza pandemics through a system of instantaneous electronic data exchange followed by real time, geographically dispersed vaccine production.

Introduction

The response to the 2009 H1N1 influenza pandemic was the fastest global vaccine development effort in history. Within 6 months of the pandemic declaration, vaccine companies had developed, produced, and distributed hundreds of millions of doses of licensed pandemic vaccines. Unfortunately, the response was not fast enough. Substantial vaccine quantities were available only after the second pandemic wave had peaked [312]. Manufacture of influenza virus subunit vaccines requires a vaccine virus that grows well enough in eggs or cultured mammalian cells to produce sufficient amounts of the essential vaccine antigen, hemagglutinin (HA), to meet vaccine needs. Late availability of a high-yielding vaccine virus contributed to the delay in vaccine supply.

In parallel with other efforts [313, 314] (Novartis Vaccine and Diagnostics (NV&D) used recombinant DNA methods to generate a potential vaccine virus on a standard

132

(A/Puerto Rico/8/1934 (H1N1) [PR8]) vaccine backbone on May 11, 2009, 11 days after receiving influenza strain A/California/04/2009 (H1N1) viral RNA from the US Centers for Disease Control (CDC) (**Figure A-1a and A-2a-c**). NV&D first received a conventional reassortant vaccine virus (X179A) from a WHO Collaborating Center at one of its vaccine manufacturing facilities on May 29, 18 days after we had generated our potential vaccine virus (**Figure A-1a**). Adaptation of the research-based process of recombinant vaccine virus generation to one that used a manufacturing-qualified cell line and good manufacturing practice (GMP) standards was completed with the Philipps-Universität Marburg on June 1, 2009, three days after receipt of X179A [315] [316] (**Figure A-3a**). Our recombinant vaccine virus was not used to produce the vaccines that were distributed to the public in the 2009 pandemic response because the regulatory hurdles for using a new process to produce an urgently needed vaccine were too great. Instead, although the yield of HA from X179A was only 30-50% of the yield from H1N1 vaccine viruses typically used for seasonal vaccine manufacture [313], the pandemic vaccine manufacturing campaign was initiated with an X179A-derived vaccine seed virus.

A. Conventional and reverse genetics systems during the 2009 pandemic

1 April
Nasal swab obtained from first US case

20 April
First sequence posted on GISAID

27 April
Egg-adapted A/Cal/07/09 selected as reference strain

11 May
Rescue on PR8 backbone in NV&D research lab and wild type virus received in GMP seed lab

27 May
WHO announces availability of reassortant vaccine viruses

1 June
Rescue under GMP on PR8 backbone at Uni. Marburg

6 August
WHO announces availability of extended passage reassortant (NIBRG-121xp) with yield similar to typical seasonal vaccine viruses

April — May — June — August

15 April
CDC identifies novel H1N1v virus in the US

24 April
BARDA announces need for vaccine

30 April
Viral RNA from CDC received at NV&D

21-26 May
Conventional and reverse genetics reassortants made at WHO labs, antigenic testing started

29 May - 4 June
First reassortants from WHO received at NV&D manufacturing sites; working seed production started; all found to have low yield but used for initial vaccine production

21 August
NIBRG-121xp received at NV&D manufacturing site; working seed production started

B. Synthetic system with enzymatic error correction

Day 1
HA and NA sequences downloaded, oligos synthesized, genes assembled

Day 5
Panel of synthetic reassortant viruses rescued

Day 12
Top vaccine virus(es) selected based on preliminary tests of yield, antigenicity, and sequence; working seed production started, with confirmatory testing in parallel

Day 2
MDCKs transfected with synthetic genes and multiple backbones in parallel

Day 10
Panel of synthetic viruses expanded in MDCKs

Figure A-1. Timelines to the availability of influenza vaccine viruses at manufacturing sites [313, 317-319]. (**A**) The 2009 timeline from the detection of the H1N1 pandemic strain in the US to the receipt at a Novartis manufacturing site of a vaccine virus with a yield equivalent to that typical for a seasonal vaccine virus. The timeline includes conventional and reverse genetics efforts from WHO Collaborating Centers, Novartis, and Philipps-Universität Marburg. (**B**) Projected future timeline for generation of a vaccine virus using the synthetic system with enzymatic error correction. This timeline assumes virus generation at a manufacturing site, eliminating

134

shipping delays. The projected timeline in **(B)** is aligned with the actual 2009 timeline in (**A**) based on HA and NA gene synthesis being initiated at the time when BARDA announced the need for a vaccine (April 24, 2009). The projected time savings with the synthetic approach relative to receipt at any Novartis manufacturing site of any reassortant vaccine virus in 2009 is 23 days. The time savings relative to receipt in 2009 of a vaccine virus with a yield typical of seasonal vaccine viruses would be 107 days.

A. NV&D rescue of a reverse genetic reassortant from cloned genes - 2009

Viral RNA received from CDC 5 days after sequence downloaded from GISAID

HA and NA genes cloned into rescue vector

First rescue attempt

Rescue of first known reassortant vaccine virus (non-GMP) for H1N1v globally

1 2 3 4 5 6 7 8 9 10 11 12
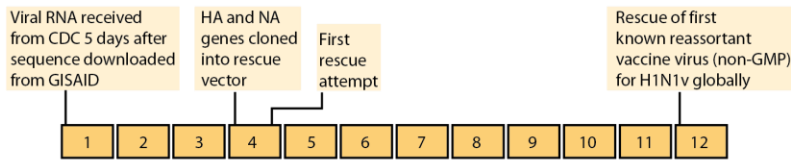
B. NV&D attempt to rescue virus with commercially synthesized genes - 2009

HA and NA sequences downloaded

Gene synthesis order placed with reconstructed ends

Synthetic genes received

Synthetic genes cloned into rescue vector

Clone sequences confirmed

Earliest potential transfection

Failure to rescue virus

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
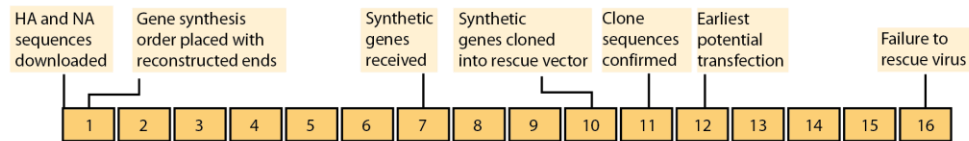
C. CSL attempt at virus rescue with commercially synthesized genes - 2009

HA and NA gene synthesis order placed

Synthetic HA gene received

Synthetic NA gene received

Transfection of 293T/MDCK co-culture with cloned synthetic genes

Failure to rescue virus from eggs inoculated with tranfection culture supernatant

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

D. Rescue with synthesized genes and enzymatic error correction

HA and NA sequences downloaded

Oligos synthesized

MDCKs transfected

Rescue confirmed

1 2 3 4 5

Oligos designed

Genes assembled and error corrected

Virus rescued

E. NIBSC rescue of A/Fujian/411/02 by reverse genetics with a quality system

Virus received for RNA cloning

Transfection of Vero cells

Vero cell harvest (HI positive)

Virus recovered from egg inoculated with Vero cell transfection medium

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

F. Reverse genetic rescue with genes cloned from respiratory specimens

Respiratory specimen received

RNA extraction, RT-PCR, and cloning

Plasmids purified

Clones sequenced and selected

Transfection

Blind passage

Virus rescued

1 2 3 4 5 6 7 8 9 10 11

Figure A-2. Timelines for synthetic and conventional reverse genetic influenza virus rescue. For comparison, the generation of a conventional reassortant vaccine

virus requires approximately 21 days from the time that a wild type virus is received [314]. Each box represents one day.

(**A**) NV&D rescue of a reverse genetic reassortant from cloned genes – 2009. The timeline begins with the receipt of H1N1v RNA from the CDC, 5 days after the HA and NA sequences were downloaded from GISAID, and 10 days after H1N1v sequences were first posted on GISAID. Therefore, for accurate comparison to the synthetic technique, 5 to 10 days must be added to the reverse genetics timeline. This effort resulted in the first PR8 reassortant potential H1N1v vaccine virus globally. The reassortant was not used for vaccine manufacture because the virus was rescued in 293T/MDCK co-cultures under research conditions. The rescue was repeated with manufacturing-suitable MDCK cells alone under highly controlled conditions, but this virus also was not used for vaccine manufacture, due to the regulatory hurdles to manufacturing with a vaccine virus produced by a new process.

 (**B**) NV&D attempt at rescue with commercially synthesized genes – 2009. During the pandemic response, synthetic HA and NA genes were received 6 days after they were ordered from a commercial supplier. Because the supplying and receiving sites were both in Germany, shipping did not contribute significantly to the timeline. Initial attempts at rescue with the synthetic genes were not successful, although subsequent attempts after process optimization did succeed.

(**C**) CSL Limited (CSL) attempt at rescue with commercially synthesized genes – 2009 [314]. The failure of the synthetic virus generation attempt was attributed to lack of egg-adaptive mutations in the sequences available. A subsequent recombinant addition

of egg-adaptive mutations led to rescue in the system, which involves a final egg passage. Information on the timeline of the rescue of the virus with the egg-adaptive mutations was not provided. Because this technique includes rescue in 293T cells, which are not qualified for vaccine manufacture, the technique is not suitable for generating viruses to be used to produce vaccines for human use.

(**D**) Rescue with synthesized genes and enzymatic error correction. This timeline differs from the synthetic rescue timeline in Figure A-9 because shipping between sites is eliminated (by consolidating all activities at a single manufacturing site), and the time for virus rescue is extended (because rescue in MDCK cells alone is less rapid than rescue in a mixed 293T/MDCK cell co-culture). Because the synthetic technique allows the rescue of many viral variants in parallel (including coding sequence, genome segment end, and backbone variants), the probability of obtaining a high yield, antigenically correct strain is increased.

(**E**) National Institute for Biological Standards and Control (NIBSC) rescue of A/Fujian/411/02 by reverse genetics with a quality system [138]. The timeline starts with receipt of a wild type virus. For comparison to the synthetic timeline, the interval from the public availability of sequence data to the receipt by a manufacturer of a sample of an emerging virus must be added to the reverse genetics timeline.

(**F**) Reverse genetic rescue with genes cloned from respiratory specimens [320]. This technique requires receipt of a shipped respiratory specimen rather than an electronically transmitted gene sequence.

138
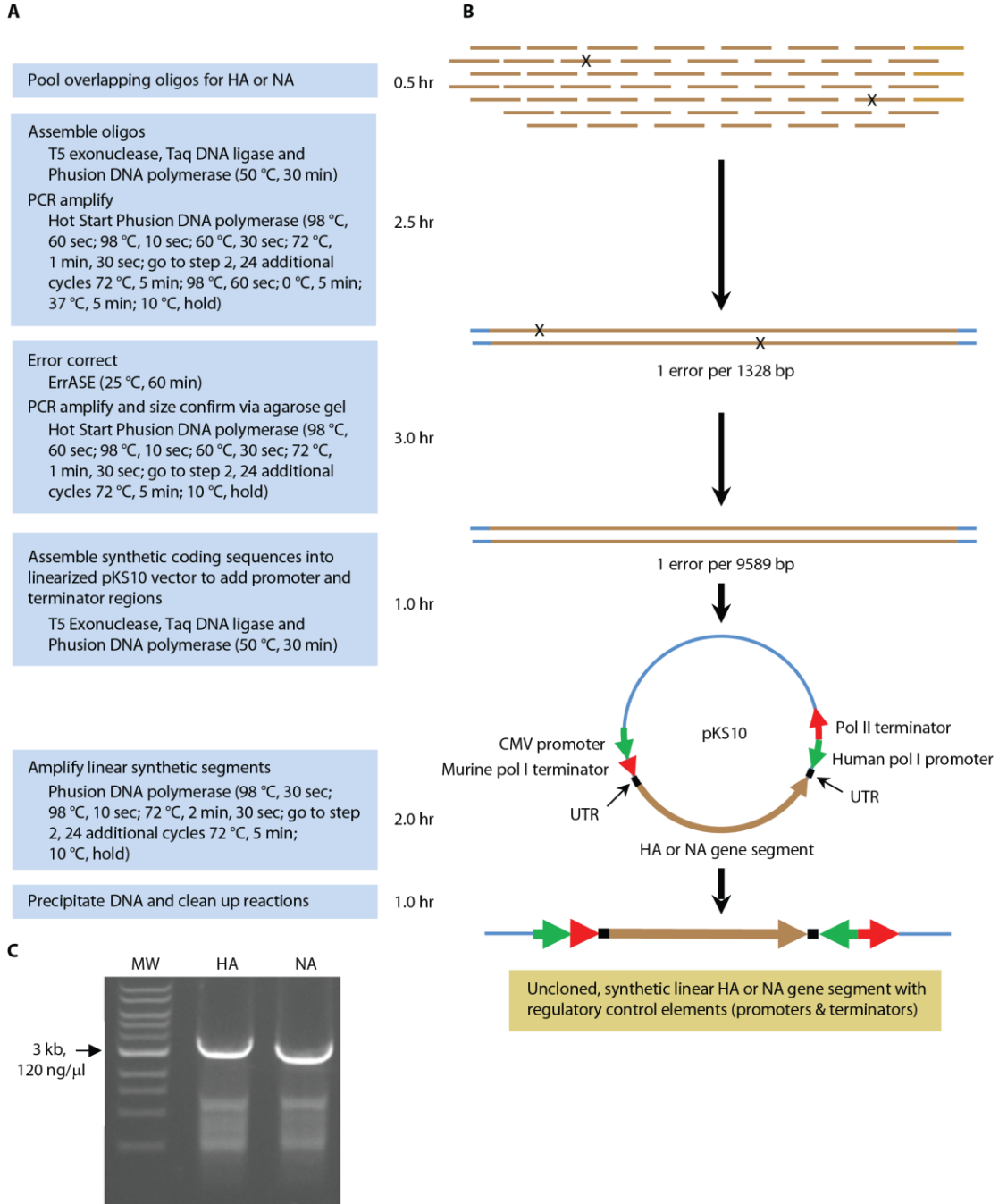
Figure A-3. Synthetic gene segment assembly and error correction.

(**A**) Process for synthesis of HA and NA gene segments. (**B**) Schematic diagram of process. "X" indicates sites of oligonucleotide synthesis errors. Influenza coding

sequences are brown, pKS10 sequences are blue, promoter sequences are green, terminator sequences are red, and UTRs are black. (**C**) Ethidium bromide-stained agarose electrophoresis gel of 2 µg each of linear synthesized HA and NA genes, including regulatory elements used for virus rescue. MW - molecular weight marker.

This experience with the 2009 H1N1 virus provided lessons for the next pandemic vaccine response. (i) Synthetic genomics techniques to produce influenza genome segments rapidly, accurately, and reliably are needed, so that instantaneous exchange of electronic sequence data followed by local gene synthesis can replace the isolation of viruses, preparation of high-growth reassortant vaccine viruses, and shipment of viruses and nucleic acids between geographically dispersed sites where vaccines are manufactured. (ii) A reverse genetics system that uses these synthesized genes to generate viruses that are suitable for use in GMP-compliant vaccine manufacturing should be in place. (iii) To increase the reliability of generating high-yielding, antigenically correct vaccine viruses, a greater variety of influenza HA and neuraminidase (NA) variants should be rescued in the context of multiple combinations of other influenza genome segments. (iv) The use of synthetic and reverse genetic technologies for pandemic responses in seasonal influenza vaccine manufacture could establish regulatory and public acceptance and familiarity to allow reliable application of these approaches during a public health emergency.

On the basis of these lessons and the US government's interest in improving the influenza vaccine manufacturing enterprise, NV&D, the J. Craig Venter Institute

140

(JCVI), Synthetic Genomics Vaccines, Inc. (SGVI), and the US Biomedical Advanced Research and Development Authority (BARDA, HHS) initiated a collaboration to develop a rapid process for synthetic vaccine virus generation. We addressed three major technical barriers to more rapid and reliable pandemic responses: the speed of synthesizing DNA cassettes to drive production of influenza RNA genome segments, the accuracy of rapid gene synthesis, and the yield of HA from vaccine viruses.

Materials and Methods

Cells.

293T cells were maintained in DMEM (Lonza) containing 10% fetal bovine serum (FBS; Omega Scientific) and penicillin-streptomycin (Pen-Strep; Lonza) at 37 °C with 5% CO_2. Suspension MDCK 33016PF cells were maintained in a shake flask in CDM (Lonza) at 37 °C and 5% CO_2 with shaking at 145 rpm. For reverse genetics experiments, six-well plates were seeded with 293T or MDCK 33016PF cells in DMEM plus 10% fetal bovine serum and Pen-Strep and were used for transfection when approximately 70-80% confluent.

Design of oligonucleotides.

The oligonucleotide sequences used to program chemical synthesis were based on HA and NA gene sequences transmitted via email from BARDA and were designed with a precursor to the HANABOD software program developed at JCVI (Supplementary Material). The program initially added set constant regions, "AGTACTGGTCGACCTCCGAAGTTGGGGGGG" and "AATAACCCGGCGGCCCAAAATGCCGACTCG", to the 5´ and 3´ ends,

respectively, of a matched pair of HA and NA genes to generate overlaps homologous to the linearized plasmid (pKS10) that provides the genetic control elements to the DNA cassettes used for transfection. The program then divided the HA and NA gene sequences flanking constant regions into oligonucleotide sequences according to a set of rules that minimized the time for chemical synthesis and maximized the reliability of synthesis and subsequent assembly and error correction. These rules were [312] all bases of both strands are represented, with no unfilled gaps; [313] the length of the oligonucleotides allows the entire oligonucleotide set for a HA, NA pair to be contained in a single 96-well plate of oligonucleotides; [314]within an assembly, differences in length between oligonucleotides are limited to 2 bases (e.g. 66 to 68 bases) to minimize differences in annealing temperatures; one oligonucleotide of an intermediate length (e.g. 67 bases) must bridge the transition between shorter and longer oligonucleotides when the oligonucleotide set is annealed. The length and endpoints of the oligonucleotides are not adjusted based on their nucleotide composition.

Assembly of oligonucleotides into HA and NA gene segments

Overlapping oligonucleotides were pooled for the HA and NA gene segments. For the HA gene segment, 10 µl of each of fifty-two 10 µM oligonucleotides were pooled. For the NA gene segment, 10 µl of each of forty-four 10 µM oligonucleotides were pooled. The pooled oligonucleotides were assembled using a one-step isothermal assembly protocol. Twenty microliter reactions were prepared by combining 10 µl of either the pooled HA or NA oligonucleotides with 10 µl of 2x Assembly Mastermix. The 2x

Assembly Mastermix was prepared by combining 320 µl of 5x ISO Reaction Buffer (25% PEG-8000, 500 mM Tris-HCl pH 8.0, 50 mM $MgCl_2$, 50 mM DTT, 1.0 mM each of the 4 dNTPs, 5 mM NAD), 6.4 µl of 1 U/µl T5 Exonuclease (Epicentre), 20 µl of 2 U/µl Phusion polymerase (New England Biolabs [NEB]), 80 µl of 40 U/µl Taq ligase (NEB) and 374 µl dH_2O. The assembly reaction was incubated at 50 °C for 30 min. The assembled HA and NA gene segments were PCR amplified. The PCR reaction mixture contained 5 µl of HA or NA assembly reaction, 71 µl of dH_2O, 20 µl of HF Buffer (NEB), 2 µl of 10 mM dNTPs, 0.5 µl of 100 µM primer BMP_3 (CGAGTCGGCATTTTGGGCCGCCGGGTTATT), 0.5 µl of 100 µM primer BMP_4 (AGTACTGGTCGACCTCCGAAGTTGGGGGGG), and 1 µl of Hot Start Phusion polymerase (NEB). The cycling conditions were 98 °C for 60 seconds; 98 °C for 10 seconds; 60 °C for 30 seconds; 72 °C, for 1 minute and 30 seconds; go to Step 2 for 24 additional cycles; 72°C for 5 minutes; 98 °C for 60 seconds; 0 °C for 5 minutes; 37 °C for 5 minutes; and 10 °C hold.

The HA and NA PCR products were error correcting using ErrASE (Novici Biotech). Ten µl of each gene segment were added to each of six wells of an ErrASE kit strip. One strip of 6 ErrASE wells was required for each gene segment. The error correction reaction was incubated for 60 minutes at 25 °C. To PCR amplify each of the 12 ErrASE wells, 2 µl of each ErrASE reaction was added to 48 µl of PCR mastermix (36 µl of dH_2O, 10 µl of 5x HF buffer, 1.0 µl of 10 mM dNTPs, 0.25 µl of 100 µM primer BMP_3, 0.25 µl of 100 µM primer BMP_4, and 0.5 µl of Hot Start Phusion polymerase). The amplification was carried out under the following conditions: 98 °C for 60 seconds; 98 °C for 10 seconds; 60 °C for 30 seconds; 72 °C for 1 minute and 30

143

seconds; go to step 2 for 24 additional cycles; 72 °C for 5 minutes; 10 °C hold. The sizes of the gene segments were confirmed on 1.2% E-gels (Invitrogen). DNA bands were visualized and quantitated on an Amersham Typhoon 9410 Fluorescence Imager.

To assemble error corrected, PCR amplified HA and NA gene segments into a linear pKS10 plasmid, 20 ng of the PCR amplicons were added to 10 µl of 2x Assembly Mastermix and 0.2 µl of 70 ng/µl linear pKS10 plasmid. Distilled H_2O was added to a total reaction volume of 20 µl. The assembly reaction was incubated at 50 °C for 30 minutes.

Linear HA and NA gene segments were amplified from the circularized plasmids. Thirty replicate PCR reactions were set up for the HA and the NA gene segments. Each reaction consisted of 5 µl of HA or NA assembled into pKS10 plasmid, 71 µl of dH_2O, 20 µl of 5x HF Buffer, 2.0 µl of 10 mM DNTPs, 0.5 µl of 100 µM primer BMP_13 (CGAAAGGGGGGATGTGCTGCAAGGCGA), 0.5 µl of 100 µM primer BMP_14 (CTTCCGGCTCGTATGTTGTGTGGAATTG), and 1 µl of Phusion polymerase. The reactions were amplified under the following conditions: 98 °C for 30 seconds; 98 °C for 10 seconds; 72 °C for 2 minutes and 30 seconds; go to step 2 for 24 additional cycles; 72 °C for 5 minutes; 10 °C hold. The 30 reactions for each of the gene segments were pooled following PCR amplification. The sizes of the gene segments were confirmed on 1.2% E-gels. DNA bands were visualized and quantified on an Amersham Typhoon 9410 Fluorescence Imager.

Final HA and NA amplicons were then phenol-chloroform extracted and ethanol precipitated. DNA pellets were dissolved in 50 µl of Tris-EDTA (TE) buffer, pH 8.0.

The final products are uncloned, synthetic linear HA or NA gene segments with regulatory control elements (promoters and terminators).

Reverse genetics.

For transfection of 293T cells, 27 µl of FuGENE (Promega) was added to 73 µl of serum-free and antibiotic-free Dulbecco's Modified Eagle's medium (DMEM, Lonza) and incubated at room temperature for 5 minutes. Plasmid DNA (consisting of 1 µg each of plasmids encoding the six influenza backbone genes PA, PB1, PB2, NP, NS, and M and a helper plasmid encoding the serine protease TMPRSS2) was combined with 1 µg each of synthetic HA and NA linear cDNA before addition of the mixture to the diluted FuGENE. The mixture was then incubated for 20 minutes at room temperature. Transfection complexes were added directly to a well of a six-well plate containing 1.5×10^6 293T cells in 2 ml of cell seeding medium (DMEM containing Pen-Strep and 10% FBS) and incubated at 37 °C and 5% CO_2. At 18 hours after transfection, the medium was aspirated from each well and the cells were washed twice with culture medium (serum-free DMEM containing Pen-Strep) and replenished with 2 ml culture medium containing 0.5 µg/ml TrypZean (Sigma) and 3×10^5 feeder MDCK 33016PF cells. Cells were incubated for another 2 days before the harvest of transfection supernatants.

To amplify recovered viruses, a blind passage was initiated in which 50-500 µl of transfection supernatant were added to a well of a six-well plate containing a confluent monolayer of MDCK 33016PF cells in 2 ml culture medium containing 0.5 µg/ml TrypZean. For transfection of MDCK 33016PF cells, 1 µg of each DNA was combined

145

as described above followed by the addition of 500 µl of serum-free and antibiotic-free DMEM and 10 µl of Plus reagent (Invitrogen) and incubated at room temperature for 5 minutes. Twenty-five µl of Lipofectamine LTX (Invitrogen) was then added to the diluted DNA mixture and incubated at room temperature for 30 minutes. Transfection complexes were directed added to a well of a six-well plate containing 0.6×10^6 MDCK 33016PF cells in 2 ml of cell seeding medium. The rest of the protocol is identical to that described above.

Propagation of virus in MDCK cells.

Ten ml of suspension MDCK 33016PF cultures (1×10^6 cells/ml) in propagation medium [70% PFM (Invitrogen), 30% spent CDM (Lonza), 0.5 µg/ml TrypZean (Sigma)] were inoculated with virus at a multiplicity of infection of 0.001 and incubated in a flask at 34 °C and 5% CO_2 with shaking at 145 rpm. Samples were taken at 0, 12, 24, 36, 48, and 60 hours after infection and frozen at -80 °C until used to titer viruses and to test for red blood cell agglutination activity.

Propagation of virus in embryonated chicken eggs.

 Nine-day old, specific pathogen-free, embryonated chicken eggs (Charles River) were incubated at 35 °C in 70% humidity for 2 days. Eleven-day old eggs were inoculated with 200 µl of virus diluted in Dulbecco's phosphate buffered saline (PBS, Invitrogen), typically at a virus dose ranging from 28-2800 infectious units per egg. Inoculated eggs were incubated at 35 °C, 70% humidity for 72 hours. Eggs were placed at 4 °C for 14-24 hours before the allantoic fluid was harvested. Allantoic fluid was clarified

by centrifugation at 3000 rpm for 5 minutes, and samples were assayed for viral infectious titer and hemagglutination activity.

Focus formation assay.

In a slight modification of a previously described focus formation assay [321], culture medium samples were cleared at 10,000 rpm for 1.5 minutes in a microcentrifuge. A 10-fold serial dilution was performed on cleared samples with DMEM containing 1% FBS and 1x Pen-Strep as diluent. Diluted samples were added to confluent MDCK monolayers in a 96-well plate and incubated at 37 °C and 5% CO_2. After 16-18 hours, the supernatant was aspirated, and the cells were washed once with PBS and fixed with a 1:1 mixture of acetone-methanol for 30 minutes at -20 °C. After fixation, the supernatant was aspirated, the cells were washed once with PBST (PBS plus 0.1% Tween 20), and blocked with PBS plus 2% BSA for 30 minutes at room temperature. After removal of the blocking buffer, 50 µl of a mixture of mouse monoclonal antibodies recognizing influenza NP (Millipore) diluted 1:6000 in blocking buffer was added to each well and incubated for 1 hour at room temperature. Wells were washed three times with PBST, and 50 µl of goat anti-mouse IgG (Jackson Immunoresearch) diluted 1:2000 in blocking buffer were added to each well and incubated for 1 hour at room temperature. Cells were washed three times with PBST, and 50 µl of True Blue peroxidase substrate (Kirkegaard and Perry Laboratories, Inc.) were added to each well and incubated for 5 minutes at room temperature. Wells were washed one time with distilled water, dried, and the number of blue cells was counted under a microscope.

147

Sucrose density gradient separation.

Sixty ml of suspension MDCK 33016PF cultures (1 x 10^6 cells/ml) in propagation medium were inoculated with virus at a multiplicity of infection of 0.001 and incubated in a flask at 34 °C and 5% CO_2 with shaking at 145 rpm. Medium samples were harvested at 60 hours after infection by centrifugation at 1000 x g for 5 minutes. Forty ml of the harvested media were concentrated 20-fold by ultrafiltration (Vivaspin 20 with 300 kD MWCO, Sartorius-Stedim Biotech). A continuous sucrose density gradient, consisting of 20%-60% sucrose in PBS pH 7.5, was prepared in an open-top polyclear centrifuge tube (Seton Scientific) with a gradient maker (BioComp). A 1.8 ml sample of the concentrated virus was overlaid on top of the sucrose gradient and centrifuged in the Beckman Coulter Optima L-100 XP-Ultra at 100,000 x g for 4 hours. After centrifugation, sucrose gradients were fractionated (roughly 1 ml per fraction) into 96-well deep well plates. A hemagglutination assay with 0.5% guinea pig red blood cells (Cleveland Scientific) was performed on all fractions to determine fractions with the most concentrated virus. Virus-containing fractions were verified by SDS-PAGE and pooled.

HA ELISA.

Ninety six-well plates were coated overnight with *Galanthus nivalis* (GNA) lectin (Sigma). Plates were washed four times with wash buffer (PBS+ 0.05% Tween20) and blocked with blocking buffer (10 mM Tris-HCl, 150 mM NaCl, 3% sucrose, 1% BSA, pH 7.68) for 1 hour at room temperature. Three-fold serial dilutions of the samples containing a final concentration of 1% Zwittergent 3-14 (Sigma) were prepared, added in duplicate to the plates, and incubated at 37 °C for 30 minutes in a shaker. After

incubation, biotinylated-IgG purified from pooled sheep antisera (National Institute for Biological Standards and Control [NIBSC]) raised against specific virus strains were added and further incubated at 37 °C for 30 minutes in a shaker. Plates were then washed four times with wash buffer and incubated with streptavidin-alkaline phosphatase (KPL) in wash buffer at 37 °C for 30 minutes in a shaker. Plates were washed four times with wash buffer and developed using 1 mg/ml p-nitrophenyl phosphate (pNPP, Sigma) in DEA buffer phosphatase substrate (KPL). Plates were read after 40-50 minutes incubation in the dark at 405 nm using an Infinite^{TM} 200 PRO plate reader (Tecan).

Reversed-phase high pressure liquid chromatography (RP-HPLC).

Each 225 µl virus sample was mixed with 25 µl of 10% Zwittergent 3-14 (Sigma) for a final detergent concentration of 1% by gently pipetting, followed by incubation at room temperature for 30 minutes. Next, 7.5 µl of 1 M DTT was added (for a final concentration of 25 µM) and mixed by gentle pipetting. Samples were heated at 90 °C for 10 minutes, cooled at room temperature for 5 minutes, and then spun down briefly to remove condensation. For each run, a 100 µl sample was injected onto a POROS R1/10 column (2.1 mm x 100 mm, Applied Biosystems), and HA1 was eluted with a 30-35% acetonitrile gradient containing 0.1% TFA. The HA1 concentration was quantified by using purified HA as reference.

Statistics.

All statistical evaluations were performed using GraphPad Prism version 5.00 for Windows (GraphPad Software). One-way ANOVA followed by uncorrected Fisher's

least significant difference test was used to analyze data sets with 3 or more groups. A paired t-test was used to analyze data sets with 2 groups.

Results

Enzymatic assembly and *in vitro* error correction.

We adapted an enzymatic one-step, isothermal assembly method for gene assembly, previously used to synthesize the 16,299 base pair mouse mitochondrial genome from 600 overlapping oligonucleotides [322], to generate synthetic DNA copies of influenza virus genome segments. 5´ T5 exonuclease, Phusion DNA polymerase and Taq DNA ligase join multiple DNA fragments during a brief 50 °C reaction[201]. We selected this method to assemble genes for our synthetic vaccine seeds because it is rapid and readily automated.

When phosphoramidite chemistry is used to build the precursor oligonucleotides of synthetic DNA constructs, one error per 500-750 bp is generally observed (with some variation between oligonucleotide manufacturers) [322]. During the mouse mitochondrial genome synthesis, subassemblies of less than 500 bp were cloned and sequenced, and sets of error-free sequences were selected for subsequent rounds of assembly[322]. Each cycle of synthesis, assembly, cloning, and sequencing that was required to collect a complete set of correct sequence fragments lasted two to three days, depending on the time needed for sufficient *E. coli* growth for colony formation and plasmid preparation (**Figure A-4a**). For rapid influenza vaccine seed virus generation, we felt that this method of error correction would introduce unnecessary delays. However, when DNA copies of the 1.7 kb HA and 1.5 kb NA viral RNA

150

genome segments were assembled with oligonucleotides approximately 60 bases in length with 30 bases of overlap between oligonucleotides on opposite strands without cloning and sequencing steps to select subassemblies without errors, an average of 3% of the assembled products had the correct sequence (1 of 48 HA clones and 2 of 48 NA clones).

We solved the problem of synthesizing DNA copies of HA and NA genome segments with both accuracy and speed by (i) increasing the overlap between oligonucleotides, (ii) introducing an enzymatic error correction step, and (iii) increasing the number of oligonucleotides assembled at once, eliminating the need for stepwise assembly via sub-assemblies (**Figure A-3a and b**). Specifically, the length of oligonucleotides was increased to 60-74 bases, and full-length genes (including 5´ and 3´ un-translated regions) were assembled from staggered sets of oligonucleotides that contained all residues of a double-stranded DNA molecule so that, prior to ligation, the full double-stranded gene could be annealed. In practice, the HA and NA BARDA Oligo Designer (HANABOD) software program generates a set of sequences for oligonucleotides (a maximum of 96 oligonucleotides per HA-NA pair to fit in a 96-well plate) that meet these criteria. After chemical synthesis of the oligonucleotides, enzymatic isothermal assembly, and PCR amplification, error-containing DNA was removed enzymatically by treating melted and re-annealed DNA with a commercially available error correction kit that excises areas of base mismatch in double-stranded DNA molecules before another round of PCR amplification.

After the products' sizes were verified on agarose gels (**Figure A-3c),** we added the control sequences (including Pol I and Pol II promoters and their terminator and polyadenylation signals) needed to generate RNA genome segments and mRNA for virus rescue by isothermally coupling the synthetic DNA with a linearized plasmid (pKS10) that contained these regulatory sequences [323]. Nucleotide identity between the ends of the linearized plasmid and the 5′ and 3′ primers used for gene synthesis guided this assembly. The assembled molecule was then used as the substrate for a round of high-fidelity PCR amplification with primers from outside the transcription control regions.

After purification and concentration of the amplicons, we obtained approximately 10 µg of assembled linear DNA cassettes that contained the influenza HA or NA gene flanked by control sequences and which were ready for transfection into the MDCK 33016PF cell line for influenza virus rescue. The time from receipt of oligonucleotides to a purified HA- or NA-encoding DNA cassette ready for transfection was approximately 10 hours (**Figure A-3 1a and b and A-4b**). While virus rescue was underway with the enzymatically assembled, error corrected, and amplified DNA, we verified the sequence of the assembled genes by cloning and sequencing. When we applied one rounds of enzymatic error correction, the rate of sequence errors in synthesized genes was < 1 in 9,000 base pairs, and ≥ 80% of the full-length HA and NA gene sequences obtained were correct (**Table A-1**). This gene synthesis and assembly method enabled the rapid and accurate conversion of digitally transmitted sequence information to biologically active DNA genes.

152

A.

Download HA and NA sequences | Synthesize oligos

Colony formation, E. coli growth in broth, DNA preparation | Pool correct clones, assemble into a vector, transform E. coli

Sequence DNA

| Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |

Design oligos | Pool oligos, assemble, transform *E. coli*

Sequence DNA

Colony formation, *E. coli* growth in broth, DNA preparation

B.

Download HA and NA sequences | Synthesize and pool oligos | Error correct and amplify | Amplify linear gene segments

| Day 1 |

Design oligos | Assemble oligos and amplify | Assemble coding sequence into vector to add control elements | Precipitate and clean up DNA

Figure A-4. Timelines for HA and NA gene synthesis.

The state-of-the-art process at the start of the project (A) produced plasmids containing HA and NA genes approximately 4.5 to 5.5 days after sequences were obtained. The optimized assembly and error correction process (B) produces gene cassettes, ready for transfection, approximately 21.5 hours after sequences are obtained.

Table A-1. Error rate of synthesized genes.

The improvement in accuracy without error correction from 3% of clones correct when the program started to 25%-44% of the control clones correct reflects changes other than enzymatic error correction in the synthesis method that also improve accuracy. These other changes include the use of oligonucleotide sets without gaps, the elimination of one PCR reaction, and the elimination of one round of sub-assembly. The synthesized HA and NA genes are from A/Uruguay/716/2007 (H3N2).

Error correction description	Gene	# clones correct	# clones analyzed	% clones correct	# bases synthe-sized	# errors	Error rate (1 error per X bp)	Grand total		
								# bases synthe-sized per method	# errors	Error rate (1 error per X bp)
One round correction	HA	15	16	93.75	28192	1	28192	57532	6	9589
	NA	16	20	80.00	29340	5	5868			
One round control[a]	HA	4	15	26.67	26430	22	1201	55770	42	1328
	NA	5	20	25.00	29340	20	1467			
Two rounds correction	HA	22	24	91.67	42288	2	21144	77496	3	25832
	NA	23	24	95.83	35208	1	35208			
Two rounds control[a]	HA	4	9	44.44	15858	8	1982	43731	28	1562
	NA	8	19	42.11	27873	20	1394			

[a]In the controls for one and two rounds of error correction, the error correction enzymes were omitted from the reactions.

154

Rescue of synthetic influenza viruses in a manufacturing cell line.

The rescue protocol for synthetic seed virus generation is adapted from a previously described eight-plasmid ambisense system in which each expression plasmid has a cDNA copy of a viral gene segment bounded at the 5´ end by a Pol II promoter to drive transcription of messenger RNA and at the 3´ end by a human Pol I promoter to drive transcription of negative-stranded influenza RNA genome segments [324]. The manufacturing-qualified MDCK 33016PF cell line is a less efficient substrate for transfection and influenza virus rescue by reverse genetics than 293T cells (which are not qualified for vaccine production). Influenza virus reverse genetic rescue has been described in Vero cells (some banks of which are qualified for vaccine production) [138, 139]. However, using one cell line for vaccine virus rescue and a different cell line for antigen production would add adventitious agent risk and regulatory and manufacturing complexity. We therefore elected to increase the efficiency of reverse genetic DNA rescue in MDCK 33016PF cells so that a single cell line could be used for seed generation and vaccine antigen production. Although Pol I promoters are generally species specific [325, 326], human Pol I efficiently drives transcription in MDCK 33016PF cells, which are of canine origin [327].

One microgram of each linear synthetic cassette that encodes HA or NA was co-transfected into MDCK 33016PF cells together with 1 μg of each ambisense plasmid that encodes PA, PB1, PB2, NP, NS, or M and a helper plasmid that encodes the protease TMPRSS2 [328]. To increase rescue efficiency, we added cultures of fresh (un-transfected) MDCK 33016PF cells and additional TrypZean after transfection [327]. Adding fresh cells after transfection increases the efficiency of virus recovery

from DNA-transfected cells, presumably by providing a healthier population of cells in which rescued viruses can further amplify (**Figure A-5**). Viruses were detected in cell culture medium within 72 hours after transfection (approximately 24 hours later than after transfection of 293T cells), with a focus-formation assay in which the medium from the transfected culture was added to a fresh MDCK cell monolayer, and infectious virus was detected by immuno-staining for expressed influenza virus nucleoprotein (NP). After virus amplification in MDCK cells, which generally does not select for HA-adaptive mutations [329], we verified HA and NA identity by sequencing HA and NA gene-specific reverse transcriptase-polymerase chain reaction (RT-PCR) amplicons from the viral cultures and comparing the sequences to those used to program the oligonucleotide synthesis.
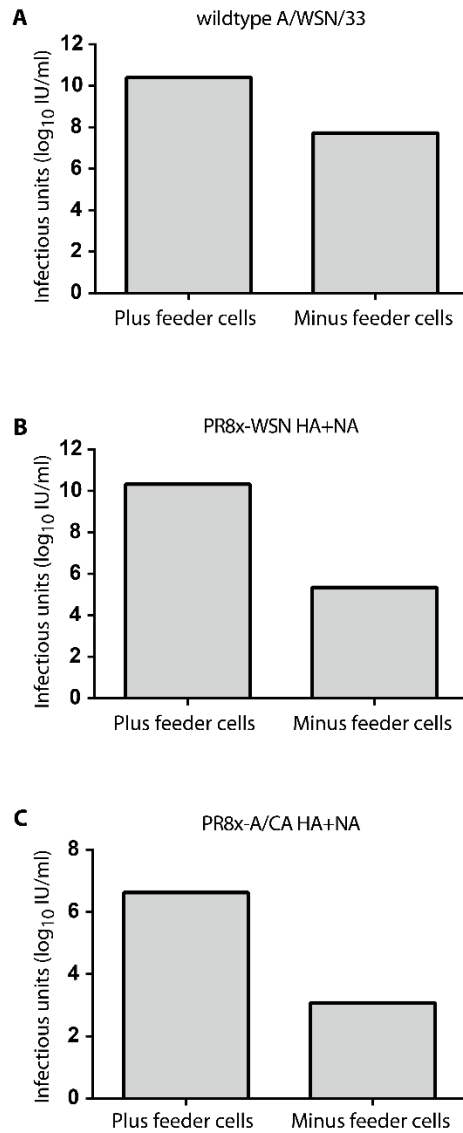
Figure A-5. Effect of MDCK feeder cell addition on titers of viruses recovered from DNA-transfected cells.

Fresh MDCK 33016PF cells were added after DNA transfection of MDCK 33016PF cells in serum-free medium. Titers of (**A**) wild type virus A/WSN/1933, or recombinant viruses containing the PR8x backbone with HA and NA segments from (**B**) A/WSN/1933 or (**C**) A/California/04/2009 were measured 72 hours after transfection

157

by a focus formation assay. Data in each panel are from one experiment with each measurement made in duplicate.

Backbones for synthetic virus rescue.

We further increased rescue efficiency by using improved influenza backbones (sets of genome segments encoding influenza virus proteins other than HA and NA) (**Figure A-6 and Table A-2**). The initial improvement in the backbone resulted from the use of genes from a PR8 variant (designated PR8x) that had been adapted over five passages to growth in MDCK 33016PF cells. Additional improvements resulted from combining PR8x backbone genome segments with backbone genome segments from biosafety level 2 human influenza isolates that were not highly pathogenic. During pilot manufacturing of influenza vaccines in MDCK 33016PF cells, several human influenza viruses, including strain A/Hessen/105/2007 (an A/New Caledonia/20/1999 (H1N1)-like strain that was isolated in 2007 from a patient without severe disease and then passaged 30 times in MDCK 33016PF cells), were adapted to grow efficiently in MDCK 33016PF cells, although not as efficiently as strain PR8x. Synthesized viruses with HA and NA genes from historical H3N2 strains and all possible backbone combinations of 5 PR8x genome segments and one A/Hessen/105/2007 genome segment or 3 PR8x genome segments and 3 A/Hessen/105/2007 genome segments were screened for virus growth by a focus formation assay and for HA production by an ELISA. This screen identified a backbone (designated #19), composed of NP, PB1, and PB2 genome segments from strain A/Hessen/105/2007 and M, NS, and PA genome segments from strain PR8x, that often outperformed equivalent viruses with entirely PR8x backbones in reverse genetic rescue efficiency and yield of HA (Tables A-2 and

158

A-3 and Figure A-6 and A-7 a-d). Similarly, synthesized viruses with HA and NA genes from H1N1 strains and a backbone (designated #21) with the PB1 genome segment of the 2009 H1N1 pandemic strain, A/California/7/2009, and the other genome segments from strain PR8x often had greater rescue efficiencies and HA yields than equivalent viruses with entirely PR8x backbones (Tables A-2 and A-3 and Figure A-6 and A-7 a-d). This finding is consistent with a report that the A/California PB1 genome segment is preferentially found in the reassortant progeny of co-infections of chicken eggs with A/California/7/2009 and a donor strain that has a PR8 backbone[330].

Historically, most influenza type B vaccine seeds have been wild type viruses, not reassortants, because wild type influenza B viruses generally provide adequate yields. To use our synthetic procedures for influenza B viruses more readily, we developed two optimized type B backbones that provided consistent rescue of synthetic influenza B viruses (**Tables A-2 and A-3 and Figure A-6, and A-7 e and f**). In the first (designated Brisbane), all backbone genome segments originated from B/Brisbane/60/2008 (B/Victoria lineage); in the second (designated B34), the genome segments encoding PA, PB1, PB2, and NP originated from B/Brisbane/60/2008, and those encoding M and NS originated from B/Panama/45/1990 (B/Yamagata lineage).

Overall, the use of optimized backbones for A strains increased rescue efficiencies up to 1000 times (as measured by infectious titers obtained after transfection) (Figure A-7 a-d) and increased HA yields in research scale infections of MDCK 33016PF cells by 0.3 to 9.9 times relative to reference strains, depending on the strain and assay used for HA detection (**Table A-3**). In general, yields of HA from these viruses were also

159

increased relative to those from viruses with PR8 backbones when the viruses were propagated in embryonated chicken eggs (**Table A-4**). To make use of such strain-specific differences, an optimal synthetic seed generation strategy would combine the HAs and NAs from circulating strains of interest with a panel of alternative backbones to maximize the chances of isolating a high-yielding vaccine virus.
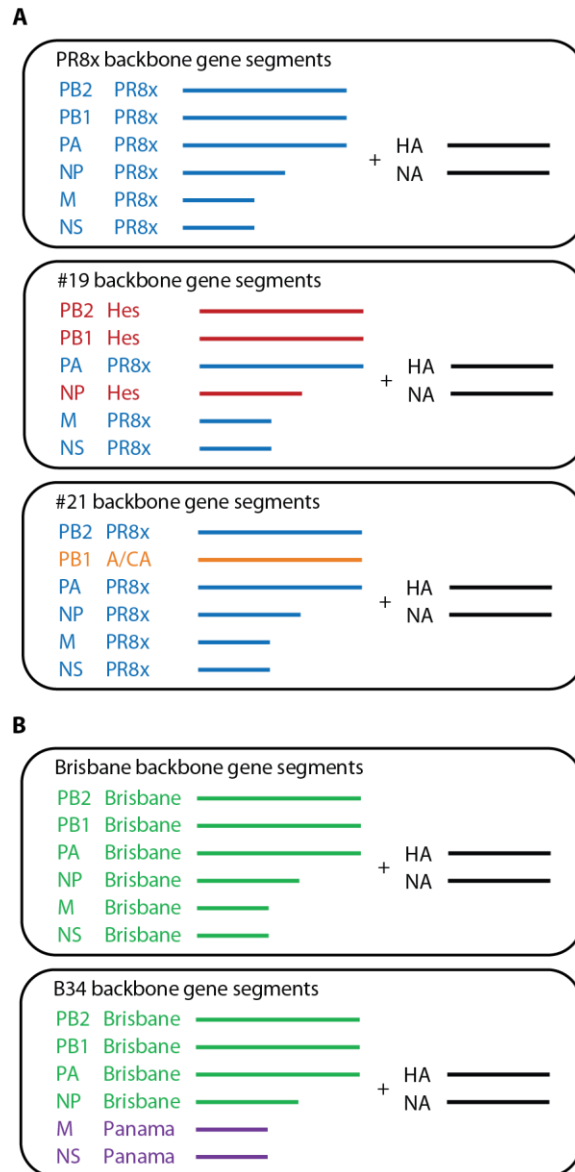
Figure A-6. Panel of backbones used to rescue synthetic influenza viruses. Backbones for type A (**A**) and type B (**B**) influenza strains are shown. Synthesized HA and NA genes from circulating strains are matched with backbone genes from plasmids. GenBank accession codes for backbone gene sequences are listed in table A-2. Hes - A/Hessen/105/2007 (H1N1); A/CA - A/California/7/2009 (H1N1); Brisbane - B/Brisbane/60/2008 (Victoria lineage); Panama - B/Panama/45/1990 (Yamagata lineage).
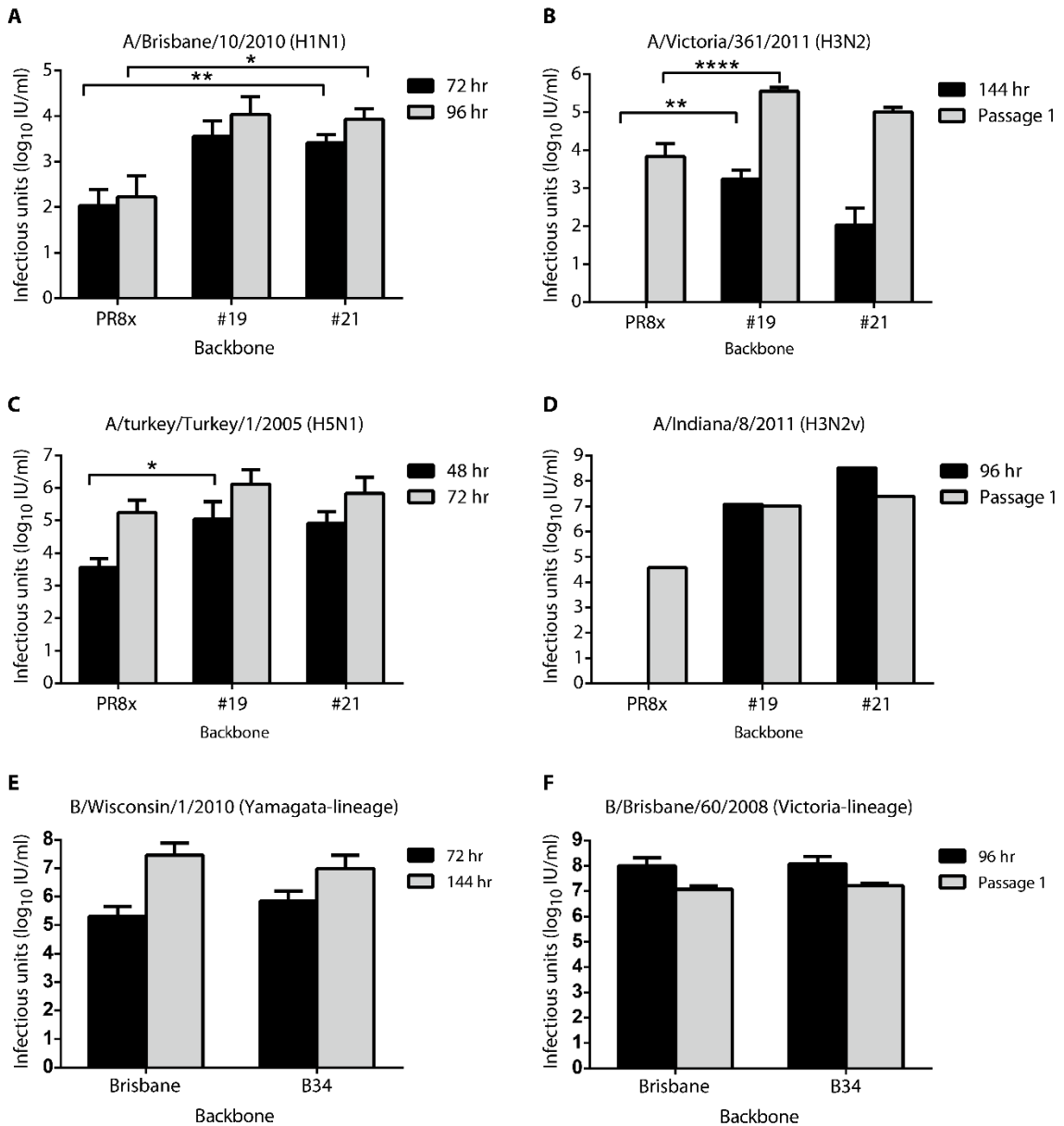
Figure A-7. Effect of optimized backbones on synthetic influenza virus rescue efficiencies.

Influenza viruses with the indicated backbones and synthetic HAs and NAs were detected in MDCK 33016PF cell culture fluid harvested at the indicated times after transfection or harvested 24-48 h after a blind passage of 500 µl of the primary culture fluid on fresh MDCK 33016PF cell monolayers (Passage 1). Viral titers were

determined with a focus formation assay for **(A)** an H1N1 strain (3 independent experiments for each time point), **(B)** an H3N2 strain (2 independent experiments for each time point), **(C)** an attenuated H5N1 strain (4 or 5 independent experiments, depending on the backbone, for the 48 hr time point and 5 or 6 independent experiments, depending on the backbone, for the 72 hr time point), **(D)** a swine origin H3N2v strain (one experiment), **(E)** a B/Yamagata lineage strain (2 independent experiments), and **(F)** a B/Victoria lineage strain (3 independent experiments for Brisbane; 4 independent experiments for B34). All measurements were made in duplicate (with the exception of one experiment in panel C at the 48 hr time point, for which single measurements were made). Error bars indicate the standard error of the mean. For experiments with sufficient replicates to allow reliable statistical evaluation, the significance of the difference between backbones and time points that produced the highest and lowest infectious titers was assessed by a one-way ANOVA, Fisher's LSD. * - p-value < 0.05; ** - p-value ≤ 0.01; **** - p-value ≤ 0.0001.

Table A-2. **GenBank accession codes of backbone gene sequences.**

Strain	Gene	GenBank accession code
A/Puerto Rico/8/1934	NS	KC866599
	M	KC866600
	NP	KC866598
	PA	KC866595
	PB1	KC866596
	PB2	KC866597
A/Hessen/105/2007	NP	KC866612
	PB1	KC866610
	PB2	KC866611
A/California/7/2009	PB1	KC866601
B/Brisbane/60/2008	NS	KC866606
	M	KC866606
	NP	KC866605
	PA	KC866602
	PB1	KC866603
	PB2	KC866604
B/Panama/45/1990	NS	KC866608
	N	KC866609

Table A-4. Virus titers and HA yields from influenza viruses with optimized backbones relative to conventional vaccine viruses when propagated in embryonated chicken eggs.

HA yields in mass per volume of allantoic fluid are normalized and shown as fold-improvement relative to yields from reference strains, which are set to 1.0. Guinea pig red blood cell (GP-RBC) agglutination, RP-HPLC, or lectin-capture ELISA was used to detect HA. Reference strains were obtained from the US CDC or the UK NIBSC.

Strain with alternative backbone	Reference strain	FFA titer	HA titer by GP-RBC agglutination	HA yield by RP-HPLC	HA yield by ELISA	Backbone
A/Christchurch/16/2010 (H1N1)[a,b]	NIB74	3.0	3.5	18	8.4	#21
A/Victoria/210/2009 (H3N2)[a,b]	X187	0.9	1.3	n/t	1.2	PR8x
A/Victoria/361/2011 (H3N2)[c,d]	IVR-165	6.4	2.6	n/t	3.4	#21
A/Indiana/8/2011 (H3N2v)[b,c]	X213	n/t	3.0	1.6	n/a	PR8x
B/Wisconsin/1/2010 (Yam)[c,d]	wild-type	4.7	3.4	n/t	3.5	Brisbane
B/Brisbane/60/2008 (Vic)[c,d]	wild-type	1.1	0.8	n/t	0.8	Brisbane

[a]virus contains HA and NA genome segments from plasmids

[b]HA yields determined from virus purified from egg allantoic fluid by a sucrose density gradient

[c]virus contains synthetic HA and NA genome segments

[d]HA yields determined directly from egg allantoic fluid

n/t, not tested

n/a, data not available because strain-specific antisera were not available for ELISA

Yam, B/Yamagata lineage

Vic, B/Victoria lineage

Table A-3. Virus titers and HA yields from influenza viruses with optimized backbones in MDCK 33016PF cells.

Test strain yields are presented as fold-improvement over reference strain yields, which are normalized to 1.0. Reference strains were obtained from the US CDC or the UK NIBSC. RP-HPLC or lectin-capture ELISA was used to detect HA from the culture medium of virus-infected cells (m.o.i = 0.001 or 0.0001).

Test strain	Reference strain	HA yield by ELISA		HA yield by RP-HPLC		Best backbones
		Fold increase	p-value[a]	Fold increase	p-value[a]	
H1N1						
A/Christchurch/16/2010	NIB74	4.3 **	0.0025 ($N = 8$)	2.8	0.0544 ($N=5$)	#21
A/Brisbane/10/2010	wild-type	6.9 ***	0.0005 ($N=13$)	4.2	0.1207 ($N=5$)	#21
A/Brisbane/59/2007	IVR-148	2.9	-- ($N=1$)	1.9	-- ($N=1$)	#21
A/Solomon/3/2006	IVR-145	6.4	-- ($N=2$)	1.8	-- ($N=1$)	#21
H3N2						
A/Victoria/361/2011	IVR-165	1.8	0.1548 ($N=4$)	1.8	0.0740 ($N=4$)	PR8x
A/Victoria/210/2009	X187	1.8 *	0.0299 ($N=7$)	3.4	0.2741 ($N=3$)	PR8x
A/Wisconsin/15/2009	X183	9.9	-- ($N=2$)	Ref. undetect.	--	#19
A/Uruguay/716/2007	X175C	1.4 *	0.0191 ($N=7$)	1.4	-- ($N=1$)	#19
H5N1						
A/turkey/Turkey/1/2005	NIBRG23	2.1 *	0.0479 ($N=8$)	3.8	0.1324 ($N=4$)	#19, #21[b]
H3N2v						
A/Indiana/8/2011	X213	n/a	n/a	2.4 ***	0.0005 ($N=8$)	#21
B Yamagata						
B/Wisconsin/1/2010	wild-type	2.1	0.2502 ($N=4$)	1.1	0.6266 ($N=3$)	Brisbane
B/Brisbane/3/2007	wild-type	2.9	-- ($N=2$)	2.0	-- ($N=2$)	#B34
B/Florida/3/2006	wild-type	1.6	0.3892 ($N=3$)	not tested	not tested	#B34
B Victoria						
B/Brisbane/60/2008	wild-type	1.0	0.9555 ($N=15$)	1.5 **	0.0088 ($N=7$)	Brisbane
B/Brisbane/32/2002	wild-type	2.3	0.4060 ($N=3$)	1.8	0.1813 ($N=3$)	Brisbane

[a] Two-tailed p-values were calculated using a paired t-test. (--), not enough replicates for statistical evaluation; *, p-value ≤ 0.05; **, p-value ≤ 0.01; ***,p-value ≤ 0.001; *N*, number of replicates.

[b] For A/turkey/Turkey/1/2005, statistical analyses for HA yields were based on results from viruses with two different backbones and equivalent yields relative to the reference strain.

n/a, data not available because strain-specific anti-sera were not available for ELISA.

Ref. undetect., data not available because the reference strain had undetectable HA levels by RP-HPLC.

Synthetic vaccine virus generation in a simulated pandemic response.

In a proof-of-concept test of the synthetic system's first iteration, BARDA collaborators not directly involved in the synthesis provided our virus synthesis group with unidentified, partial HA and NA genome segment sequences at 10:09 am EDT on Wednesday, August 24th, 2011 (**Figure A-7**). The sequences included complete coding regions but incomplete untranslated regions (UTRs), mimicking the information likely to be available in the early days of a pandemic. Sequence analysis of the HA genome segment showed that it was very closely related (96% nucleotide sequence identity by Blast to GenBank) to a low pathogenicity North American avian H7N3 virus (A/Canada goose/BC/3752/2007, GenBank accession code EU500844), and that the NA genome segment was very closely related (96% nucleotide sequence identity by Blast to GenBank) to a low pathogenicity North American avian H10N9 virus (A/king eider/Alaska/44397-858/2008, GenBank accession code HM060036). Although the HANABOD software generates the sequences of the oligonucleotides for assembly of

166

the synthetic genes, user intervention is needed when there are ambiguities in the available sequence data, such as heterogeneity in the UTRs of fully sequenced genome segments that are similar to a partially sequenced, new HA or NA variant of interest. Heterogeneity in UTRs is found in multiple influenza subtypes. In the test case, the unknown terminal UTR sequences were reconstructed by sequence alignment of the test coding sequences with a limited number of related full-length, high quality H7 HA genome segment and N9 NA genome segment sequences in GenBank, followed by comparison of the identified UTRs. This analysis revealed heterogeneity in the non-coding regions of NA genes of H7N9 strains (U/C at 1434 in the positive-sense orientation). So, alternative sets of 5´ NA oligonucleotides were used to construct two variants of the NA cassettes, one containing T at 1434 and one containing C. UTR design was completed at 12.40 pm EDT on Wednesday, August 24th, 2011 (**Figure A-7**).

By agreement with BARDA, the timed portion of the test began at a pre-specified hour, 8:00 am EDT on Monday, August 29, 2011, when the full length HA and NA genome segment nucleotide sequences, including reconstructed UTRs, were used as input to a precursor of the HANABOD software program, which generated the oligonucleotide sequences for synthesis (**Figure A-9**). By noon on Friday, September 4, immunostaining of a secondary culture confirmed that the virus had been rescued. The 4 days and 4 hours from start of oligonucleotide design to detection of rescued virus (4 days and 6.5 hours if the time to reconstruct the UTRs is added) included time spent shipping DNA from the oligonucleotide synthesis and gene assembly laboratories in California to the virus rescue laboratory in Massachusetts. If all functions were

167

consolidated in one location, the potential for delays and mishaps from shipping would be reduced.

These original proof-of-concept rescues were conducted with 293T cells; rescue of the strains with MDCK cells, as in an actual pandemic response, slows detection of rescued virus by approximately 24 hours (**Figure A-8**). The sequences of the HA and NA genome segments of the synthetic H7N9 reassortant viruses from the proof-of-concept exercise were determined after two rounds of virus amplification in MDCK 33016PF cells and were identical to those used to program oligonucleotide synthesis. Two-way hemagglutination inhibition (HI) testing (reciprocal HI assays using antigen from the synthetic and natural strains and ferret sera drawn after synthetic and natural virus infection) [331], conducted by the US CDC, demonstrated antigenic identity of the synthetic virus to the source wild type virus and confirmed that the synthesized virus was immunologically active (**Table A-5**). A/goose/Nebraska/17097-4/2011 (H7N9) (Genbank accession codes JX899805 and JX899806) was subsequently revealed as the source wild type virus for the sequences that were electronically transmitted to the virus synthesis group.

The A/goose/Nebraska/17097-4/2011 HA and NA genes were rescued with PR8x, #19, and #21 backbones. Virus rescue was more efficient with the #19 and #21 backbones than the PR8x backbone, based on the titers of viruses harvested 48 and 72 hours after transfection (Figure A-10a). To test growth characteristics, the synthetic viruses were amplified once in MDCK 33016PF monolayers and then used to infect suspension MDCK 33016PF cultures at a multiplicity-of-infection of 0.001. Despite differences

in the efficiency of virus recovery, viruses exhibited similar growth characteristics, regardless of backbone (**Figure A-10b**). The H7N9a set of viruses (C1434 positive sense NA) achieved infectious titers approximately 10 times higher than their H7N9b counterparts (U1434 positive sense NA) (figure A-12). The viruses with the highest infectious yields also produced the most HA per volume of infected MDCK suspension culture (**Figure A-10c and d**). Thus, the single nucleotide substitution in the 5´ NA non-coding region of the genomic RNA strongly influenced both infectious titer and HA yield. The H7N9a virus with the #19 backbone produced approximately 1.5 times more HA than a virus with the same HA and NA in the context of the standard PR8x backbone (**Figure A-10c and d**). This result confirmed the importance of rescuing multiple HA or NA variants with multiple backbones to increase the probability of identifying high yielding vaccine virus strains early in the vaccine seed generation process. Simultaneous rescue of multiple variants is faster and more easily accomplished with the synthetic approach than standard plasmid mutagenesis approaches. Our results also illustrate the importance, for effective response to pandemics, of including the most complete genome segment sequences possible in genetic databases and of clearly delineating terminal sequences originating from viral genome segments from those originating from sequencing primers.

Table A-6. Diversity of synthetic influenza virus strains rescued.

All H1N1 strains listed are of the post-2009 H1N1 pandemic lineage.

Seasonal serotype A virus		Backbone		
Source of synthetic HA, NA genes	Subtype	PR8X	#19	#21
A/Brisbane/10/2010	H1N1	+	+	+
A/Christchurch/16/2010 (NIB74)	H1N1	+	+	+
A/Christchurch/16/2010 NIB74-K170E	H1N1	n/a	n/a	+
A/Christchurch/16/2010 NIB74-K171E	H1N1	n/a	n/a	+
A/Christchurch/16/2010 NIB74-G172E	H1N1	n/a	n/a	+
A/Christchurch/16/2010 NIB74-G173D	H1N1	n/a	n/a	+
A/Uruguay/716/2007	H3N2	+	+	+
A/Victoria/210/2009 (X187)	H3N2	+	+	+
A/Victoria/361/2011 (CDC E3)	H3N2	+	+	+
A/Victoria/361/2011 (WHO E3)	H3N2	+	+	+
A/Victoria/361/2011 (MDCK)	H3N2	+	+	+
A/Berlin/93/2011 (egg-derived)	H3N2	+	+	+
A/Berlin/93/2011 (cell-derived)	H3N2	+	+	+
A/Brisbane/402/2011	H3N2	+	+	+
A/Victoria/304/2011 NVD p2/E3	H3N2	-	-	+
A/Brisbane/256/2011 MDCK P2	H3N2	+	+	+
A/Brisbane/256/2011 P2/E3	H3N2	-	+	+
A/South Australia/34/2011	H3N2	-	+	+
A/Brisbane/299/2011 (IVR164)	H3N2	+	+	+
A/Brisbane/299/2011 (E5)	H3N2	+	+	+
A//South Australia/3/2011	H3N2	+	+	+
A/Wisconsin/1/2011	H3N2	+	+	+

Seasonal serotype B virus		Backbone	
Source of synthetic HA, NA genes	Lineage	Brisbane	B34
B/Hubei-Wujiangang/158/2009	Yam	+	+
B/Wisconsin/1/2010	Yam	+	+
B/Brisbane/3/2007	Yam	+	+
B/Jiangsu/10/2003	Yam	+	+
B/Johannesburg/05/1999	Yam	+	+
B/Yamanashi/166/1998	Yam	+	+
B/Yamagata/16/1988	Yam	+	+
B/Texas/6/2011	Yam	+	-
B/Brisbane/36/2012	Yam	-	+
B/New Hampshire/1/2012	Vic	+	+
B/Malaysia/2506/2004	Vic	+	+
B/Brisbane/32/2002	Vic	+	+
B/Brisbane/60/2008 (cell)	Vic	+	+
B/Brisbane/60/2008 (egg)	Vic	+	n/a
B/Nevada/3/2011	Vic	+	+

Pre-pandemic viruses		Backbone		
Source of synthetic HA, NA genes	Subtype	PR8X	#19	#21
A/Hubei/1/2010	H5N1	+	+	+
A/Egypt/N03072/2010	H5N1	+	+	+
A/Turkey/Turkey/1/2005	H5N1	+	+	+
A/goose/Nebraska/11-017097-4/2011	H7N9	+	+	+
A/Indiana/8/2011	H3N2v	+	+	+

n/a, not attempted; +, virus recovered in ≤6 days post-transfection; -, virus not recovered by 6 days post-transfection; Yam, Yamagata; Vic, Victoria.
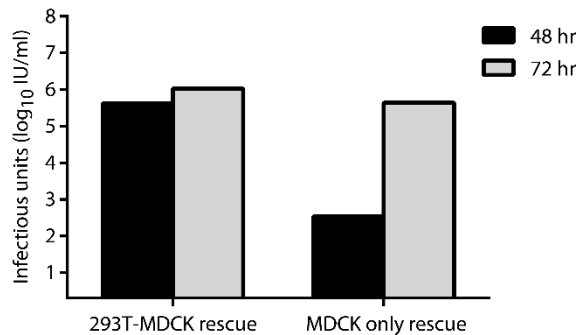
Figure A-8. Synthetic H7N9a virus rescue efficiency from MDCK-supplemented 293T cells or from MDCK cells only.

Influenza viruses were detected in culture fluid harvested 48 and 72 h after transfection with the #19 backbone plasmids and synthetic H7 HA and N9 NA gene constructs. Viral titers were determined on MDCK cell monolayers with a focus formation assay. Data are from a single experiment with 2 replicate measurements.

Table A-5. Confirmation of the antigenic authenticity of synthetic H7N9 viruses by two-way HI testing.

HI was performed with turkey RBCs as described [331]. Strains with abbreviations listed under both the influenza strains and antiserum columns were used to inoculate ferrets to generate antisera and as antigen targets for HI, with self-HI reactions in red. The other strains were used only as antigen targets. Strain origins are synthetic (S) or natural (N). The backbones of the synthetic viruses are listed after their abbreviations and full names. The codes for passage indicate the passage substrate in the first position and the passage number in the second. Passages before arrival at the CDC are left of the backslash; passages at the CDC are on the right. C, MDCK cell-passaged; E, egg-passaged; X, unknown passage number. "Norm sheep" indicates normal sheep serum, used as a negative control, from uninfected, influenza-negative sheep. A PBS control was used to ensure that there was not non-specific RBC agglutination by the serum. "OK" indicates no RBC agglutination.

Influenza strain				Antiserum				
Abbrev.	Full name	Subtype (origin)	Passage	GS/NE PR8x	GS/NE	TK/MN	TK/VA/ 4529	Norm sheep
GS/NE PR8x	A/goose/Nebraska/17097-4/2011 PR8x	H7N9 (S)	C2/E1	**320**	640	160	40	<10
GS/NE	A/goose/Nebraska/17097-4/2011	H7N9 (N)	EX/E1	320	**640**	160	40	<10
TK/MN	A/turkey/Minnesota/0141354/2009	H7N9 (N)	EX/E1	80	160	**160**	10	<10
TK/VA/4529	A/turkey/Virginia/4529/2002	H7N2 (N)	E3/E1	160	640	8	**320**	<10
GS/NE #19	A/goose/Nebraska/17097-4/2011 #19	H7N9 (S)	C2/E1	640	1280	640	40	<10
GS/NE #21	A/goose/Nebraska/17097-4/2011 #21	H7N9 (S)	C2/E1	160	640	160	40	<10

H7N9	A/Guinea fowl/Nebraska/17096-1/2011	N7N9 (N)	EX/E1	320	640	640	40	<10
PBS control				OK	OK	OK	OK	OK
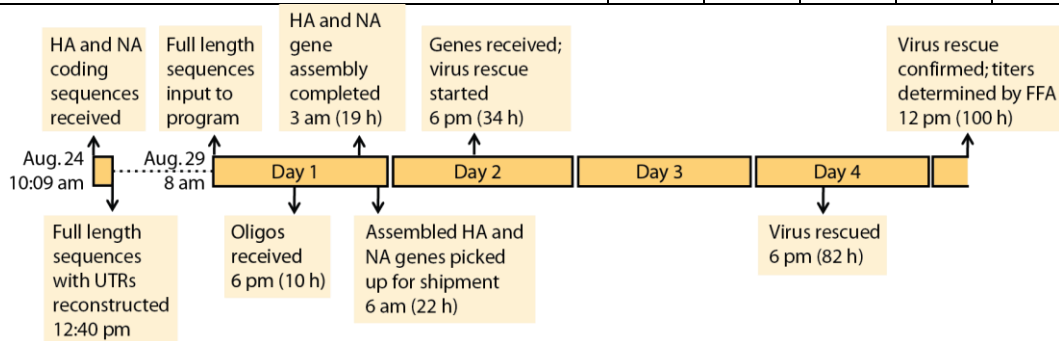
Figure A-9. Timeline of the proof-of-concept synthesis of H7N9 influenza viruses.

HA and NA genome segment UTR sequence reconstruction took 2.5 hours and was performed in advance of the timed test. The dashed line indicates the waiting period from the end of UTR sequence reconstruction to the time set by BARDA for the start of oligonucleotide design and synthesis. The listed times in hours are from the start of processing reconstructed full length gene segment sequences by a precursor to the HANABOD software program to generate the oligonucleotide sequences for synthesis.
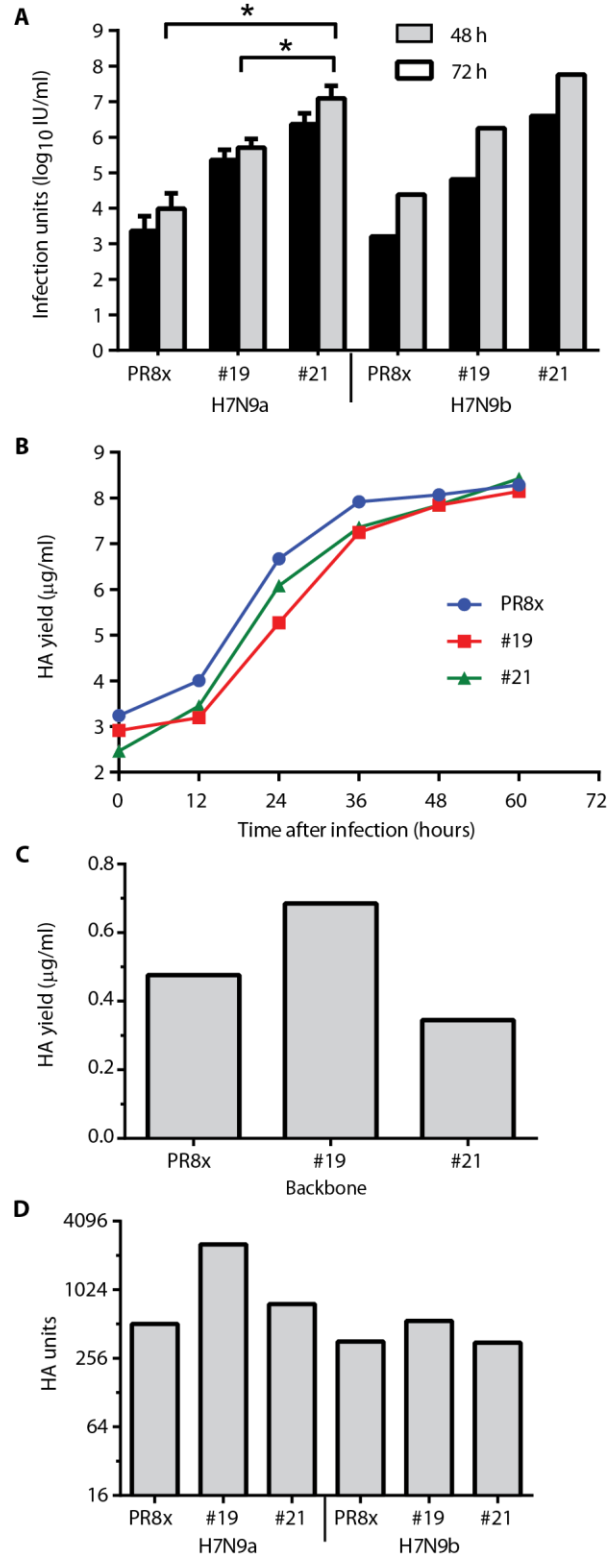
Figure A-10. Performance of synthetic H7N9 reassortant viruses from the proof-of-concept test.

(**A**) Titers of influenza viruses synthesized from 3 alternative backbones and two alternative NA UTRs. The NA UTR in the H7N9a set of viruses has C1434 in the positive sense; the NA UTR in the H7N9b set of viruses has U1434. Culture fluid was harvested from MDCK 33016PF-supplemented 293T cells 48 and 72 h after co-transfection with the indicated backbone plasmids and synthetic HA and NA gene constructs. Viral titers were determined by a focus formation assay with MDCK cell monolayers. H7N9a data are from 4 independent experiments; the error bars represent the standard error of the mean of two replicate measurements. H7N9b data are from 2 independent experiments. Analysis of bracketed differences in titers from replicate experiments and measurements is by one-way ANOVA, uncorrected Fisher's LSD. * - $p < 0.05$, **** - $p < 0.0001$. (**B**) Replication kinetics of synthetic H7N9a reassortant viruses in MDCK 33016PF suspension cultures. Data are from two independent experiments. (**C**) HA yields from synthetic H7N9 viruses in MDCK suspension cultures, determined by RP-HPLC after purification of viruses on sucrose density gradients. Data are from a single experiment. (**D**) HA yield from synthetic H7N9 viruses in MDCK suspension cultures, as measured by turkey red blood cell agglutination. Data are from two independent experiments. All measurements in A-D were made in duplicate.
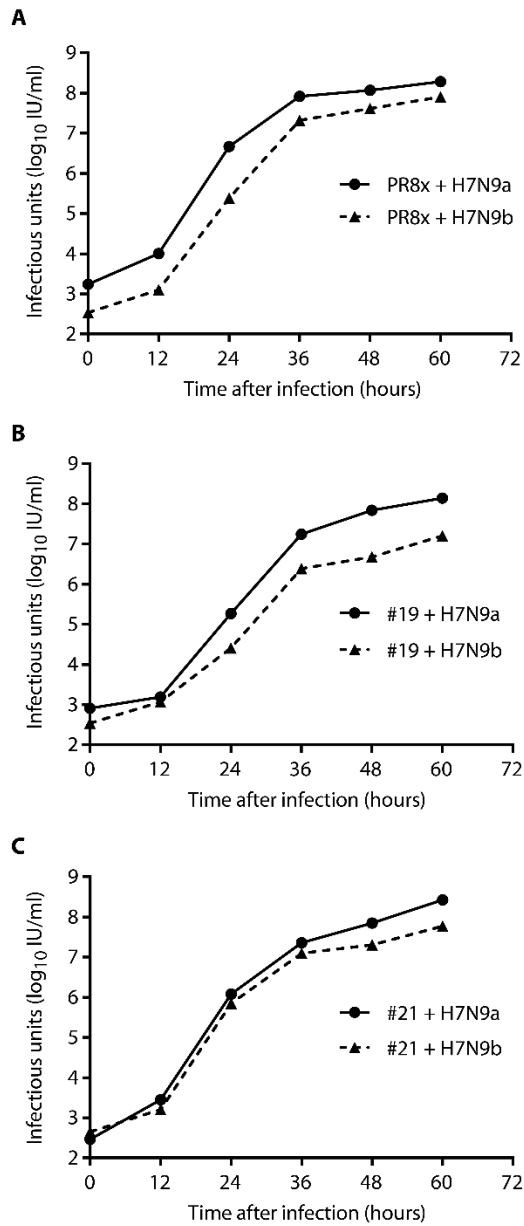
Figure A-11. Replication kinetics of synthetic H7N9 viruses with alternative NA UTRs and different backbones.

The NA UTR in the H7N9a set of viruses had C1434 in the positive sense; the NA UTR in the H7N9b set of viruses had U1434. The tested viruses had PR8x (**A**), #19

175

(**B**), or #21 (**C**) backbones and were propagated in MDCK 33016PF suspension cultures. The starting multiplicity of infection was 0.001. Each line depicts the results of 2 independent experiments. All measurements were made in duplicate.

Robustness of synthetic vaccine virus generation.

The synthetic process described here has been used to generate various influenza strains, including H1N1, seasonal H3N2, swine origin H3N2v, B (Yamagata and Victoria lineages), attenuated H5N1, and H7N9 strains (**Table A-6**). To date, we have not encountered any influenza virus strain that cannot be rescued synthetically. The robustness of synthetic influenza virus recovery from MDCK cells is in contrast to the unreliability of conventional vaccine virus isolation using eggs, particularly for recent H3N2 strains.

Discussion

We have demonstrated that an enzymatic system of gene assembly with error correction lowers error rates sufficiently to allow HA and NA gene cassettes, assembled from chemically synthesized oligonucleotides, to be used to rescue influenza vaccine viruses with pre-specified HA and NA sequences and backbones without the need to pause for cloning in *E. coli* and sequencing before transfection (**Table A-1**). The increased fidelity decreases the time from obtaining a sequence to transfecting cells with synthetic HA and NA genes from approximately 4.5-5.5 days to less than 1 day (**Figure A-4**). The reliability of influenza virus rescue on a vaccine manufacturing cell line and yields of HA are increased through the use of improved vaccine backbones (**Table A-**

3 and Figure A-7). A timed test of vaccine virus synthesis in response to a simulated H7N9 influenza outbreak demonstrates that the system can generate vaccine viruses very rapidly, and the use of the system to rescue a diversity of synthetic influenza viruses demonstrates the system's robustness.

A fast, straightforward and accurate method of synthesizing higher yielding influenza vaccine viruses, as described here, could enable more rapid pandemic responses and a more reliable supply of better matched seasonal and pandemic vaccines than available at present. In addition to increasing the speed of gene synthesis and virus rescue, this method also alleviates the need to ship viruses and clinical specimens between laboratories and eliminates the need for a separate set of viral manipulations, classic reassortment (the "mating" of influenza viruses with stochastic exchange of genome segments during co-infection), to generate high-yielding vaccine strains.

Today, more than 120 National Influenza Centers (NICs) conduct influenza surveillance and periodically ship clinical specimens to WHO Collaborating Centers, which propagate wild type viruses in MDCK cells, assess their antigenicity and sequence, re-isolate selected viruses in chicken eggs, and reassort them with high-growth backbones. High growth reassortants recommended by the WHO are then shipped to vaccine manufacturers for further adaptation to make vaccine seed viruses. A system that employs synthetic gene generation could be more efficient. National Influenza Centers could sequence HA and NA genomic RNAs in clinical specimens and post the data on publicly accessible websites for immediate download by manufacturers, public health agencies, and researchers worldwide.   Continuous

177

comparison of newly posted sequences to databases of sequence and HI data by algorithms now under development could identify the emerging viruses most likely to differ antigenically from current vaccine strains. Efficient synthetic virus rescue with a panel of high growth backbones (or alternatively with native backbones, if additional influenza genome segment sequences were provided) could simultaneously generate viruses for antigenic testing and the best vaccine virus candidates for strains found to be antigenically distinct. The speed and accuracy of gene synthesis by the cell-free, enzymatic techniques could also enable more rapid production of the strain-specific genetic constructs used to manufacture alternative, recombinant influenza vaccines [332]. Today, vaccine viruses are only shipped to manufacturers after a regimen of testing at WHO-associated laboratories, which often takes longer than generating the vaccine strains. Decentralized generation of synthetic viruses could allow manufacturers to undertake scale up and process development with strains they generate immediately after the NICs post sequences. Carrying out manufacturing activities simultaneously with seed testing could cut additional weeks from pandemic response times. Maintenance of libraries of synthetic influenza genes could further accelerate pandemic responses, if they contain pre-synthesized genes that match future pandemic strains.

Technical and non-technical hurdles must be overcome to realize fully the potential benefits of synthetic vaccine viruses for pandemic responses. Research protocols must be translated into performance-tested, at-scale manufacturing processes that have been approved by regulatory authorities. More prompt and open sequence and antigenic data sharing and biosafety standards for virus synthesis that ensure safety without

unduly slowing vaccine responses are needed. The efficiency of testing synthetic vaccine viruses must increase. In particular, panels of post-infection ferret sera against previous reference viral strains are needed for initial antigenic screening of vaccine viruses, and procedures must be agreed for decentralized generation of ferret sera against selected synthetic strains for definitive antigenic testing. The current potency assay for vaccine release, which requires a lengthy immunization protocol to generate strain-specific sheep antisera, must be replaced by a more rapid, non-serum-requiring release assay, potentially one based on the physical characteristics of properly folded, immunogenic HA. The regulatory pathway for strain changes in licensed influenza vaccines must be adapted for synthetic vaccine viruses. Automation could enable access to the technologies needed for a global pandemic response system based on sequencing HA and NA genes in clinical specimens and rapidly synthesizing vaccine viruses.

On April 17, 2009, a new influenza virus strain that had caused outbreaks in Mexico was detected in southern California [317, 318]. An initial qualified vaccine virus (X179A) was not received at one of our manufacturing facilities until May 29, and a vaccine virus (NIBRG-121xp) with ordinary HA yield (10-15 mg of HA per 100 eggs) was not received until August 21, too late to manufacture vaccines for the primary pandemic response (**Figure A-1**) [313]. If the laboratory demonstrations of synthetic vaccine technology prove themselves in manufacturing and field implementation before the next pandemic, a high-yielding vaccine virus could be available to manufacturers for testing, scale-up, and process optimization days, not months, after the new virus is first detected.

179

Appendix B: Development of cell-free *in vitro* cloning for influenza virus synthetic vaccine seeds

Authors: Kyung Moon[1], David M. Brown[1, 2], Mikkel A. Algire[1], John I. Glass[1]

Affiliations:

[1] Synthetic Biology and Bioenergy, J. Craig Venter Institute, Rockville, MD, 20850, United States.

[2] Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA.

Abstract

The rapid production of accurate synthetic DNA has great potential for numerous applications in therapeutics, diagnostics and vaccines. Herein we present one such technique, *in vitro* cloning: Multiple Displacement Amplification (MDA)-High Fidelity Polymerase Chain Reaction (HFPCR), capable of accurately producing influenza vaccine seeds directly from sequence data. *In vitro* cloning was used to select for correct DNA assemblies from a pool of several hundred thousand molecules containing sequencing errors emerging from oligonucleotide synthesis. Amplification of a single molecule, without generating by-products, was possible using Phi29 DNA polymerase and 20 sequence specific heptamers. We demonstrate 1-3 µg of *in vitro* cloned influenza virus DNA can be produced in an MDA-HFPCR reaction 8 hours after obtaining the synthetic oligonucleotides used to assemble the viral gene segments. This amount of DNA was sufficient to use in reverse genetics protocols and reliably yield

infectious influenza virus. Selecting correct assemblies using this *in vitro* cloning method is possible more quickly than the more commonly used *Escherichia coli in vivo* cloning, and is a suitable alternative to enzymatic error correction.

Introduction

In the event of an influenza virus pandemic, the more rapidly vaccines can be developed the fewer people will be infected with the virus. Recently, we reported in Dormitzer *et al.* one such technique that was used to produce influenza vaccine seeds directly from sequence data. Synthetic DNA was assembled from oligonucleotides constructed for use for in vaccine manufacturing [8]. This represents a major departure from the protocols used by influenza virus vaccine makers for the last 60 years. That process relied on mating two viral strains and then screening for the desired viral strain containing an exact combination of genes derived from each strain. The resulting vaccine seed virus expresses the needed antigens and also grows in fertilized chicken eggs, a process that in extreme cases can take as long 6 months [333]. Starting from digital information, i.e. the genome sequence of a virus isolated from a patient, we used a novel DNA synthesis approach to make synthetic copies of the immunogenic viral genes in only 21.5 hours [8]. Then, we used that DNA to program a reverse genetics procedure to generate the desired chimeric viral seed in a little more than 4 days from the receipt of the viral sequences. In the event of an influenza virus pandemic, this synthetic biology-reverse genetics approach could result in vaccine being available to protect the public as much as one month sooner than it could be using the old viral mating approach.

For some years, the production of synthetic DNA has become easier and easier. Currently, most DNA foundries provide customers with synthetic DNA that is cloned in *Escherichia coli* (*E. coli*) plasmids or as yeast artificial chromosomes. In virology and vaccine manufacture, these DNAs can be used in the reverse genetics process to produce infectious influenza virus. *E. coli* cloning is necessary because a high number of errors are produced in synthetic oligonucleotide construction, usually single base deletions. By cloning the synthetic DNA, it is possible to select clones with the correct sequence. However, for the slightly less than 2 kb molecules like the influenza virus hemagglutinin (HA) or neuraminidase (NA) genome segments used in vaccine seed virus production, standard commercial DNA synthesis requires about 10 days [334]. In the event of an influenza pandemic, that may be too long. Dormitzer *et al.* reduced the time necessary to produce accurate DNA assemblies by relying on a critical enzymatic error correction step (ErrASE®) to reduce the errors resulting from oligonucleotide construction instead of selecting for correct assemblies. However, one of the potential problems with this approach is that by programming reverse genetics reactions using a pool of PCR amplicons, generated from a pool of synthetic DNA oligonucleotides, is that it generates a more heterogeneous population of molecules than the *in vivo* approach. With *in vivo* cloning of a synthetic DNA molecule in *E. coli*, which begins with a single synthetic molecule that is amplified during bacterial DNA replication with an error rate of as low as low as 10^{-9} to 10^{-11} errors per base pair [335]. With Dormitzer *et al.,* at the outset of the 25 cycles of PCR that generates approximately 10 µg of DNA, there are ~700,000 molecules of template synthetic DNA. When Dormitzer *et al.* applied one round of enzymatic error correction, the rate

of sequence errors in the final PCR amplified synthesized DNA was <1 per 9000 bp. Thus, ~20% of the full length HA and NA gene sequences obtained had at least 1 error [8]. The PCR using Phusion DNA polymerase is a high fidelity process (4.4 x 10^{-7} errors per bp [336]), and is an insignificant source of the errors in this pool of DNA molecules. Even with enzymatic error correction as part of the DNA assembly process, the principal source of DNA sequence errors is from incorporation of error containing synthetic oligonucleotides into the DNA assembly reaction. The phosphorimidite chemistry used to make these precursor oligonucleotides of synthetic DNA constructs generally results in one error per 500 to 750 bp (with some variation between oligonucleotide manufacturers) [4].

In this paper, we describe a different process to generate the ~1 µg amount of DNA needed to program an influenza virus reverse genetics reaction. Rather than using PCR to amplify the entire ~700,000 molecule pool of assembled DNA, we diluted that pool so that we could amplify a single template molecule using a combination of rolling circle multiple displacement amplification (MDA) using phi29 DNA polymerase and high fidelity PCR (HFPCR). We expected this process to produce a population of DNA molecules with fewer errors than the process we described in Dormitzer *et al.* [8]. We call this entirely enzymatic process *in vitro* cloning. While this method is susceptible to amplification of assembled molecules that contain errors, the key is to perform many replicates of the single molecule amplification and to synthesize enough DNA so that while part of that DNA is used to program reverse genetics reactions to make virus, another fraction of the DNA pool derived from a single template molecule can be

183

sequence verified. During the several days required to transfect the DNA into mammalian cells and produce usable amounts of virus, the DNA can be sequenced. Only viral production reactions containing *in vitro* cloned DNAs with the correct sequence would be used for vaccine production. To produce influenza virus vaccine seeds, the only viral genes that must be made specifically for each new viral strain are the genes that encode the HA and NA antigens. Sets of the other viral genes, which are referred to as the viral "backbone", are usually held constant, and so need not be re-synthesized for each new vaccine [171, 203, 324, 337]. Thus, in our viral rescue reactions, only the HA and NA genes were provided as *in vitro* synthesized molecules. We were able to produce ~ 1 µg of each of these two gene segments needed for reverse genetics from sequence data in 19.5 hours.

One of the challenges in developing *in vitro* cloning was to find conditions that yielded efficient MDA of single molecules with a low error rate. Hutchison *et al.* [338] reported cell-free *in vitro* cloning using phi29 DNA polymerase and exonuclease resistant DNA random hexamers [338]. Unfortunately, undesirable by-products generated by phi29 DNA polymerase during the single molecule amplification became problematic unless the reactions were performed in nanoliter volumes. In other approaches to mitigate unwanted MDA by-products, Takahashi *et al.* (2009) employed of RNA primers and Woyke *et al.* (2011) reduced nonspecific product accumulation by eliminating exogenous DNA from whole genome amplification reagents including enzymes, using UV irradiation [339, 340]. In this study, we developed an *in vitro* cloning technique to produce DNA assemblies for vaccine production by combining MDA and HFPCR,

184

utilizing sequence specific DNA heptamer primers and phi29 DNA polymerase. Moreover, we demonstrated that by using MDA-HFPCR *in vitro* cloning, we were able to sort out and rescue a clone with the right sequence without an extra error correction step (i.e. ErrASE®). MDA-HFPCR could be used as a more rapid alternative to our previously reported method employing enzymatic error correction steps. Furthermore, Variations of *in vitro* cloning could be used in a number of applications where extremely rapid synthesis of DNA could lessen morbidity and mortality resulting from other new and emerging pathogens in instances when speed is vital.

MATERIALS and METHODS

DNA, primers, strains and materials

pUC19 DNA was obtained from NEB (New England Biolabs, MA). Synthetic influenza DNAs for use in reverse genetics reactions, DW 26 (A/Moscow/10/99(H3N2) HA) and DW29 (A/Sydney/5/97(H3N2) NA), were made at the J. Craig Venter Institute (JCVI) [8]. The details of the cell strains and plasmids for this study are in Table B-1. All DNA and RNA primers were synthesized by Integrated DNA Technologies, except for DNA random hexamers (6 mers; Fidelity Systems, Rockville, MD) [338, 341]. DNA heptamers (7 mers) were synthesized with a triphosphate linkage between two nucleotides at 3′ end. Each RNA base in RNA 7 mers was modified with 3′ triphosphate linkage based on Takahashi *et al.* (2009) [339]. The sequence of each primer can be found in Table B-2. phi29 DNA

polymerase, DNA standard and all restriction enzymes were purchased from NEB.

Inorganic pyrophophatase (5 U/µl) was obtained from Affymetrix.
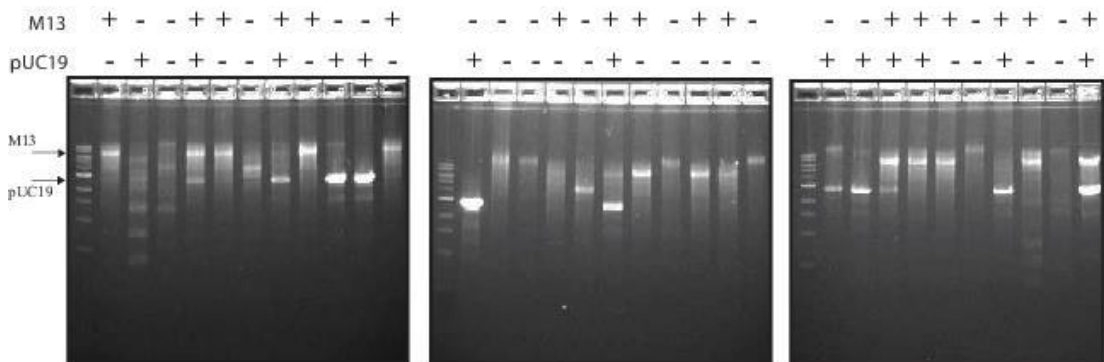
Figure B-1. Limiting dilution provides single molecule of the templates.

A DNA template for the MDA reaction was prepared based on Hutchison *et al.* [338]. Two different DNA preps were combined: 19 molecules of M13 single stranded DNA (New England Biolabs) and 13 molecules of pUC19 in a 30 µl solution (15 µl TE and 15 µl 100mM NaOH). A 1 µl aliquot of the prepared template was distributed in each reaction of 30 replicates. The reaction products after 16 h of MDA were digested with PstI and visualized on 1.2% agarose gel. The + and − symbols above the gel indicate the presence or absence of bands with molecular weights of linear M13 (7.2 kb) or pUC19 (2.8 kb) DNAs. Sixteen positive clones for M13 and 12 positive ones for pUC19 were observed; overall, 8 samples were empty and 6 samples had both products. This phenomenon is expected due to the variability in the DNA template arising from reactions caused by the adherence of molecules to tubes and pipettes when the DNA concentration is very low, as reported in the previous study [338]. The expected number of positive clones for M13 was close to the expected number (16 vs. 19), which suggests that the limiting dilution provides a single molecule

template. However, the number of positive clones for pUC19 was one half of the expected number. This may be due to the fact that pUC19 is a double stranded DNA and that the Phi29 DNA polymerase prefers a single stranded DNA for amplification.

Table B-1. Strains and plasmids in this study

strains or plasmids	Description	Reference or Source
E. coli Strains		
EPI300™	Phage T1-Resistant; For cloning	Epicentre
TOP10F'	an MC1061 derivative with F plasmid containing lacI^q and Tn10; for cloning	Invitrogen
Mammalian cell lines		
MDCK cells	Madin-Darby canine kidney	
293T cells	Human Embryonic Kidney cells	
DNA and Plasmids		
pUC19	Amp^R; a Cloning vector; pBR322 origin	NEB
M13mp 18	Single stranded M13 DNA	NEB
pGFK2001	Amp^R; pKS10-DW26 cloned in and purified from *E. coli*	In this study
pGFK2002	Amp^R; pKS10-DW26 cloned in and purified from *E. coli*	In this study
pKS10	Amp^R; pUC19 origin	Dormitzer et al. (2013)
pKS10-DW26	Amp^R; the HA segment from A/Moscow/10/99(H3N2)HA cloned into pKS10 by GIBSON ASSEMBLY	N. Alperovich & M. Algire
pKS10-DW29	Amp^R; the HA segment from A/Sydney/5/97(H3N2)NA cloned into pKS10 by GIBSON ASSEMBLY	N. Alperovich & M. Algire
PR8NA	Amp^R; the NA segment from A/Puerto Rico/8/1934 (PR8X) cloned in pKS10	E. Denisova
PR8HA	Amp^R; the HA segment from A/Puerto Rico/8/1934 (PR8X) cloned in pKS10	E. Denisova

pCC1BAC™	Cm^R; low copy cloning vector		Epicentre
pDB6S	Cm^R; The 6 segments(PB2, PB2, PA, NS, M, and NP) from A/Puerto Rico/8/1934 cloned into pCC1BAC		D. Brown
pCR®-4Blunt-TOPO®	Amp^R; a Cloning vector; pUC origin		Invitrogen

Table B-2. Oligonucleotide and TaqMan® Probes in this study

Oligo names	Sequences (5′→3′)	Oligo names	Sequences (5′→3′)
Heptamers for pUC19			
pUC19-6F-3'PT	CGTTT*C*G	pUC19-473F-3'PT	TGTTT*C*C
pUC19-600F-3'PT	CGCTT*T*C	pUC19-904F-3'PT	AATCG*A*C
pUC19-1194F-3'PT	ACACG*A*C	pUC19-1673F-3'PT	TTCGT*T*C
pUC19-1804F-3'PT	AACCA*G*C	pUC19-2278F-3'PT	AAAGT*G*C
pUC19-2405F-3'PT	GCGTT*T*C	pUC19-2605F-3'PT	AAAGT*G*C
pUC19-2684R-3'PT	CGAAA*G*G	pUC19-2516R-3'PT	ATGCT*T*C
pUC19-2110R-3'PT	CCATG*A*G	pUC19-1941R-3'PT	AGCAA*T*G
pUC19-1500R-3'PT	GTTCC*A*C	pUC19-1390R-3'PT	TGGTT*T*G
pUC19-900R-3'PT	GTGAT*G*C	pUC19-783R-3'PT	TTCTG*T*G
pUC19-298R-3'PT	AGCGA*A*G	pUC19-40R-3'PT	ATGTG*T*C
Heptamers for pKS10			
pKS10-26F	ACTGT*T*G	pKS10-5088R	AAGAT*G*C
pKS10-409F	TCAAG*T*G	pKS10-4642R	CAACA*A*C
pKS10-778F	TGCTT*A*C	pKS10-4212R	TTCGT*T*C

188

pKS10-2944F	TGTTT*G*C	pKS10-3768R	AGAAA*G*C
pKS10-3224F	ACATA*C*G	pKS10-3211R	TTGTG*T*G
pKS10-4081F	AAACC*A*C	pKS10-3158R	TTGAG*T*G
pKS10-5045F	TTCGA*T*G	pKS10-2691R	TCTTT*C*G
pKS10-4422F	AACTA*C*G	pKS10-152R	TTCAC*T*G
pKS10-4872F	TTCTG*T*G	pKS10-253R	TGTAA*C*G
pKS10-3613F	AATCG*A*C	pKS10-637R	ATTGA*C*G
Oligos for TaqMan® analysis			
pKS10-DW26 qF	CATATCATGTTTTTTGCTTTGTGTTG		
pKS10-DW26 qR	GATGGTGGCGTTTTTGGG		
DW26 probe [a]	/TAMRA/CATTAGGCGACGCGCGACACG/FAM/		
DW29-q805F	GACTCACTATAGGGAGACCCAAGC		
DW29-q993R	ACAGTAGTTACCAGGATGGCAATTT		
DW29-918probe [a]	/TAMRA/TAACGATTGGCTCTGTTTCTCTCACTATT GCC/FAM/		
Oligos for High Fidelity PCR			
BMP_13	CGAAAGGGGGATGTGCTGCAAGGCGA		
BMP_14	CTTCCGGCTCGTATGTTGTGTGGAATTG		
BMP_21	GCATTTAGGTGACCGCCGGGA		
BMP_22	GGCTAGCAGTTAACCCGAGTACTGGTC		
Oligos for sequencing			
Uni-pKS10F	CGAGTCGGCATTTTGGGCCGCCGGGTTATT		
Uni-pKS10R	AGTACTGGTCGACCTCCGAAGTTGGGGGGG		

189

DW26-654F	AGCGTATATGTCCAAGCATCAGGGAG
DW26-1208R	GATTTGGTTGATTGCTGCTTGAGTG
DW26-1402F	TTGAAAGAACAAGGAAGCAACTGAG
DW26-257R	ATCAAGGATTTGGTGAGGACTGTCG
DW29-745F	TGATGGGAGTGCTTCAGGAAGAG
DW29-1319F	ACTGAAGTCTGGTGGACCTCAAAC
DW29-858R	TCTGACACCAGGATATCGAGGATAAC
DW29-246R	TCCTTCTCTATGGTGGTGTTGGTC
H3-1F	AGCAAAAGCAGGGGATAATTCT
H5-1770R	AGTAGAAACAAGGGTGTTTTTAA
huN2-1	AGCAAAAGCAGGAGTAAAGATG
hun2-1476	AGTAGAAACAAGGAGTTTTTTTCTAAAATTG

* stands for 3′ thio modification

[a]:All TaqMan® probes are dual labeled with a 5′ reporter dye 6-carboxyfluorescein (FAM) and a 3′ quencher dye tetramethylrhodamine (TAMRA).

Preparation of DNA templates

All DNA templates were diluted to a single molecule level by limiting dilution. Synthetic influenza virus HA) and NA genome segments prepared by the Gibson Assembly reaction (with or without the ErrASE® error correction step) were diluted from 10^{-1} to 10^{-8} in Tris-EDTA (TE) buffer (pH 8.0) by limiting dilution in order to find the dilutions that had only one molecule per sample. Real time PCR analyses were performed to quantify the copy number of the template in the various dilutions ranging from 10^{-4} to 10^{-8} using TaqMan® probes and primer sets (Supplementary Table S2).

Plasmids purified from *E. coli* (pGFK2001 and pGFK2002; Supplementary Table S1) were quantified by both the Typhoon® 9410 scanner (Amersham Bioscience) analysis of electrophoresis gels loaded with the primers and direct analysis (NanoDrop®, ThermoScientific), and then were used to generate standard curves. Three replicates of each diluted samples were analyzed.

MDA in vitro cloning

All primers were designed for both the sense and anti-sense strands with 3´ terminal PTO (phosphorothioate) modifications to the last two nucleotides of their 3´ ends; this modification protects the primers from the 3´→ 5´ exonuclease activity of phi29 DNA polymerase [341]. DNA templates used in this study were prepared in TE buffer (pH 8) by the limiting dilution. For DNA primers, the template was first denatured with an equal volume of 100 mM sodium hydroxide (NaOH) for 10 min and added to the phi29 reaction mixture containing DNA primers to carry out multiple displacement amplification (MDA) [338]. To anneal RNA primers to a template, we denatured the mixture of template and primers at 95 °C, and then cooled it to 25 °C over 35 min based on Takahashi's study [339]. This mixture was used for further MDA reactions. At least 20 replicates of each reaction were set up for MDA. The conditions of the phi29 reaction have been modified from the manufacturer's protocol, and the details are as follow: 1 X phi29 Buffer (NEB), 20 µM primers, 2 % Tween 20, 0.01 U pyrophosphotase, 5.5 U phi29 DNA polymerase in 10 µl, and either 1 mM or 200 µM dNTPs. The reaction mixtures were incubated at 30 °C for a designated time and then inactivated at 65 °C for 10 min [338]. The final products were digested with restriction

191

enzymes (from NEB); EcoRI for pUC19; PstI for pUC19; NcoI for pKS10-HA/ NA. The digested products were visualized on 1.2 % agarose gels.

MDA-HFPCR in vitro cloning

The MDA reaction for MDA-HF PCR was modified based on the standard MDA reaction mentioned above; either 1 mM or 200 µM dNTPs were used for each reaction as indicated in the text. Twenty different pKS10 sequence specific primers were used for each reaction and the incubation times at 30 °C were varied; 2, 2.5, 3, and 4 h. After incubation at 30 °C, each reaction was inactivated at 65 °C for 10 min and then used for HFPCR directly without restriction enzyme digestion. Each reaction contains 1 X Phusion® HF Buffer (NEB), 0.5 µM of each primer (sense and antisense), 200 µM dNTPs, 2 U Phusion® polymerase in 100 µl volume. For HFPCR, two different primer sets were used: BMP 13 - BMP 14, which generated an approximately 3 kb amplicons, and BMP 21 - BMP 22, which generated an approximately 2 kb final product. Two different HFPCR conditions were used, depending on the primer set. For BMP 13 - BMP 14 primer set, the cycle condition was (i) 30 s at 98 °C; (ii) 25 cycles of 30 s at 98 °C; 4 min at 72 °C; and (iii) final extension of 5 min at 72 °C. The cycle condition of BMP 21 - BMP 22 primer set was as follows: (i) 6 cycles of 15 s at 98 °C; 15 s at 65 °C, and 1.5 min at 72 °C; (ii) 25 cycles of 15 s at 98 °C, 15 s at 62 °C, and 1.5 min at 72 °C; and (iii) 5 min at 72 °C. After HFPCR, 6 µl of each reaction was loaded onto a 1.2% E gel® (Invitrogen) to select positive clones.

TaqMan analysis

For TaqMan® quantitative PCR analysis, primers and probes were designed using Primer Express Software (version 2.0.0; Applied Biosystems, Foster City, CA). The primers amplified the junction between the HA and NA inserts and a plasmid vector in which they were ligated, pKS10. The designed amplicons were either 259 bp for HA (DW 26: A/Moscow/10/99(H3N2) HA) or 189 bp for NA (DW 29: A/Sydney/5/97(H3N2) NA). TaqMan® probes were designed to detect the junction of the pKS10 vector and the insert of either the HA or NA segment in order to count the perfect circular molecules (HA or NA ligated in pKS10) that were generated during the gene assembly (Sequence information in **Table S2**). The probes were labeled on the 5′ end with 6-carboxyfluorescein (FAM) and the 3′ end with tetramethylrhodamine (TAMRA). Real-time PCR was conducted in a 25 µl reaction mixture containing 12.5 µl of 2 × TaqMan Universal Master Mix (Applied Biosystems, CA), 500 nM each forward and reverse primer (pKS10-DW26F+ pKS10-DW26R; DW29-q805F+DW29-q993R), 200 nM TaqMan probe (DW26 probe, DW29-918 probe), and 0.45 µl water. Finally, 2.5 µl of DNA template was added to each reaction mixture. Standard curves were generated using either pGFK2001 or pGFK2002 with various concentrations in order to quantify the actual number of DNA molecules in reactions. Real-time PCR was performed as follows: 1 cycle at 50 °C for 2 min, 1 cycle at 95 °C for 10 min, and 50 cycles (for detecting a single molecule) / 45 cycles (for quantifying final products from MDA-HFPCR) of denaturation at 95 °C for 15 s, followed by extension and annealing at 60 °C for 1 min. All reactions were done in an Applied Biosystems 7900HT® Fast real time PCR instrument (Life Technologies). We did each reaction set

in triplicate and averaged the results. Only the experiment sets with R > 0.90 were used for quantification.

DNA sequencing

For the sequencing analysis, the products from MDA-HFPCR were cloned into Zero-blunt® vector (Invitrogen) according to the manufacture's protocol and then transformed in *E. coli* Top10F´ cells (Invitrogen). Colonies were cultured in LB with 100 µg/ml ampicillin at 37 °C for overnight, and DNA plasmids were purified using QIAGEN Mini Prep Kits. Each plasmid was sequenced with 6 different primers to obtain complete coverage on both strands of the HA or NA DNAs. Sanger DNA sequencing was done by the JCVI Joint Technology Center. Sequence data were analyzed using CLC Genomic Workbench.

Virus generation

Influenza virus was generated via reverse genetics. Co-cultured HEK 293T cells (1.0 x 10^6) and MDCK cells (0.3 x 10^6) were transfected with the eight DNA fragments carrying the cDNA of each influenza gene segment using a protocol adapted from Hoffmann and Webster [171]. Briefly, according to the manufacturer's instructions, either 5 µl of X-tremeGENE 9 (Roche Diagnostics) or 3 µl of *Trans*IT®-LT1 (Mirus) were used per µg of DNA, the transfection agent was diluted to 200 µl using serum-free DMEM, mixed with DNA for each segment, incubated at room temperature for 20 - 30 min, and then added in a drop wise manner to the monolayer of cells.

Viral Sequencing

Virus sequencing was performed on the final positive well of a TCID$_{50}$ assay. A multisegment reverse transcription-PCR (M-RTPCR) approach [204] was utilized to amplify the eight genomic RNA segments from collected virus. Phusion® High-Fidelity DNA Polymerase (NEB) was used according to manufacturer's instructions to PCR amplify the HA and NA segments. The temperature cycle parameters were 98 °C for 1 min, and then 25 cycles (98 °C for 10 s, 61 °C for 30 s, and 72 °C for 2.5 min), followed by 72 °C for 5 min. Sanger sequencing was conducted using a standard capillary platform (ABI 3730).

B.4 Results

The development of MDA-HFPCR involved three main stages. Firstly, the properties of amplification primers were investigated to increase rolling circle amplification. Second, the amplification kinetics of a single template molecule was examined. Third, MDA was paired with HFPCR in accelerating speed and accuracy. Finally, MDA-HFPCR was evaluated via sequencing and reverse genetics.

The primer properties affect the MDA efficiency at low template concentrations

To reduce background by-products, we examined whether various types of primers, could affect the MDA efficiency in reactions containing very low concentration of DNA templates. We tested the effect of a number of different primer factors on the MDA reaction: specific verses random primers, the length of primers, different numbers of specific primers used in the reaction, and DNA verses RNA primers. The different primers were compared in reactions using single copy limited template

dilutions, similar to the approach used by Hutchison *et al.* [338](**Figure B-1**). The single template copy dilution level for each MDA reaction was confirmed by real time PCR. Details of MDA reactions are in Materials and Methods. We first investigated the effect of random versus sequence specific primers. Sequence-specific 6 mers generated fewer by-products amplifying the low amount of DNA template (pUC19) than the random hexamers (**Figure B-2A & 1B**). We tested various lengths of sequence-specific primers (6 mers, 7 mers, 8 mers, and 15 mers). All shorter primers (6 mers, 7 mers, and 8 mers) generated amplicons in reactions loaded with low numbers of the template molecules. We observed amplification using the 6 mers from at least 350 molecules of template, and from 7 mers and 8 mers at least 35 molecules of template (**Figure B-2 A-D**). The 7 mers generated less background than the 8 mers in the reaction with 35 molecules of pUC19. We found that 15 mers generated no measurable amplicons in a sample with less than 1 ng (3.5×10^7 molecules) of template (**Figure B-3**). We tested whether the numbers of different sequence-specific primers (ten, twenty, or thirty respectively) that annealed to various locations on the DNA template were critical to increasing the MDA efficiency. No major difference was detected; ten primers were enough for the amplification (**Figure B-2E & 1F**). RNA primers (either random hexamers or sequence-specific 7 mers) were compared with DNA primers based on the previous study [339]. Unlike the observation made by Takahashi et al. (2009), RNA primers did not significantly change the MDA efficiency in our conditions (**Figure B-2G & 1H; Figure B-4**). Thereafter we used 20 different sequence-specific DNA 7 mers in all reactions.
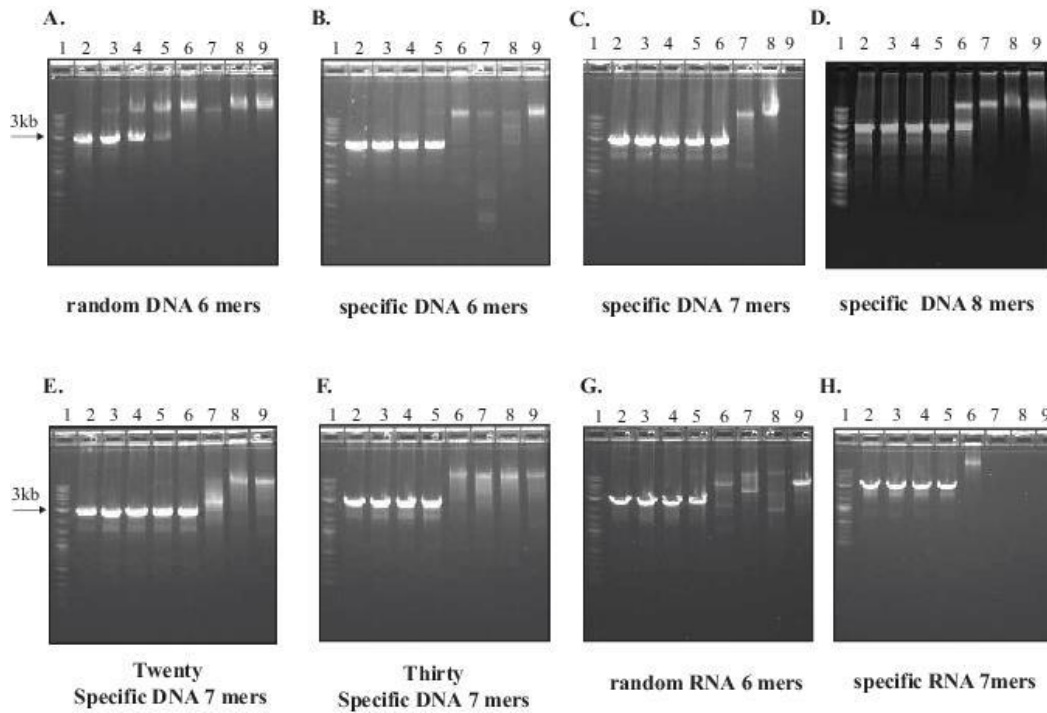
Figure B-2. MDA efficiency using various primer sets.

The same amount of template DNA was used in each lane of each panel. The number of molecules for each lane is:1) 3.5×10^6; 2) 3.5×10^5; 3) 3.5×10^4 ; 4) 3.5×10^3; 5) 350; 6) 35; 7) 3.5; 8) 1; 9) no DNA (as a negative control). 10 sequence specific DNA primers were used for panel B-D and 20 and 30 DNA specific primers were used for panel E and F, respectively. Random DNA hexamer was used in panel A and Random RNA primers (6mers) were used in panel G. Sequence specific RNA primers were used in panel H. lane 1 in each panel is DNA standard (Quick load 2-log DNA ladder; NEB).
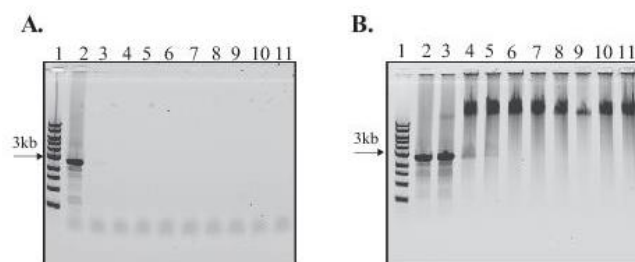
Figure B-3. Comparison of DNA primers and RNA primers on efficiency of a single molecule MDA.

The DNA template was diluted to a single molecule per reaction by the limiting dilution and then denatured with 100mM NaOH before the MDA reaction. Thirty-two replicates of the MDA reaction were set up with the sequence specific primers, either DNA heptamers or RNA heptamers (details in Materials and Methods). After the restriction enzyme digestion of the products, visualization was carried on agarose gel. S indicates 1 kb DNA ladder; N as a negative control without any DNA template; P as a positive control containing 0.1 ng DNA template.
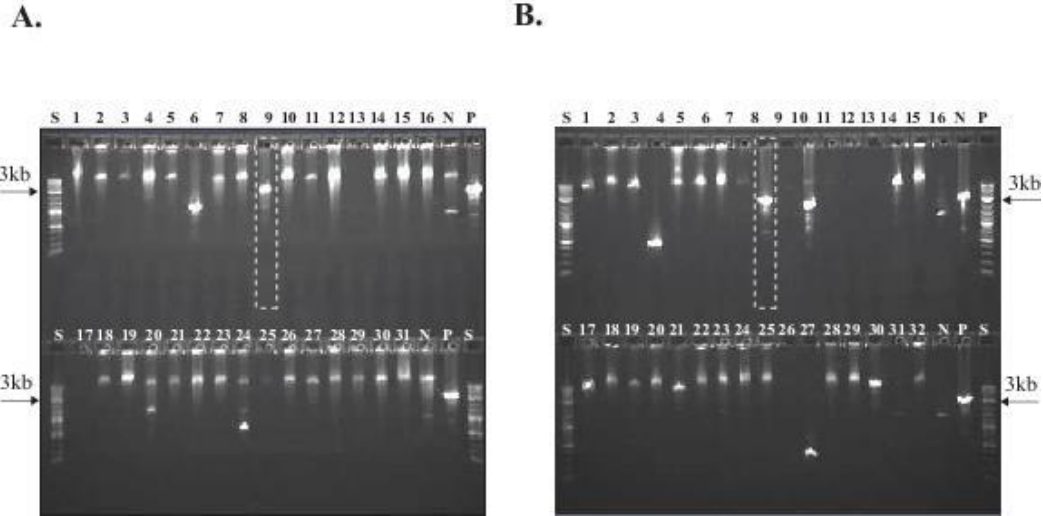
Figure B-4. MDA with 15 mers of sequence specific DNA primers.

Lanes 2-6 contains various number of pUC19 molecules that prepared by limiting dilution: 2. 3.5×10^7; 3. 3.5×10^6; 4. 3.5×10^5; 5. 3.5×10^4; 6. 3.5×10^3; 7. 350; 8. 35; 9. 3.5; 10. 1; 11. No DNA. Ten DNA sequence specific 15mers were used as primers (A) for MDA and compared with DNA random hexamers. After 16 hours of MDA reaction and the EcoRI digestion, the reaction products were confirmed on agarose gels. Lane 1 indicates 1 kb DNA ladder.

In vitro cloning amplification kinetics of a single template molecule

Establishing a procedure for capturing a single molecule from a Gibson Assembly reaction by limiting dilution is an essential task for the *in vitro* cloning of individual synthetic HA and NA molecules. For these experiments our analytic method was real time quantitative PCR analyses with TaqMan® probes and primer sets. The analysis

199

indicated that the 10^{-7} dilutions of the original DNA assembly samples for both DW 26 (A/Moscow/10/99(H3N2) HA) and DW 29 (A/Sydney/5/97(H3N2) NA) in the pKS10 plasmid vector sequence was sufficient to dilute the sample to a single molecule level (**Table B-3**). We performed an independent MDA with various dilutions of the HA and NA templates to determine whether serial dilution can be executed so that on average each MDA reaction has a single molecule of template. If the 10-fold dilution factor for preparing DNA templates was accurately executed, we expected the number of replicates that contained templates (referred to as positives) would decline suddenly as the serial dilutions produced reactions containing on average a single copy of template. Based on our TaqMan® results, we analyzed three different dilutions (10^{-6}, 10^{-7} and 10^{-8}) of DW 29 NA. Twenty replicates of the MDA reaction with each template in three different dilutions were set up. Analyses showed that 2 out of 20 were positive DW 29 NA clones with the template of the 10^{-7} dilution, which had a single molecule copy; 19 out of 20 were positive when the 10^{-6} dilution (10 molecules per reaction) template was used, while no positive clone was observed with the DNA template of the 10^{-8} dilution (presumably with no DNA copy in each reaction) (**Figure B-5**).

We next performed the kinetic study to determine the minimal reaction time for obtaining >1 µg DNA from MDA with a single molecule. As in the previous experiments, the templates were dilutions of a DW 26 HA and DW 29 NA Gibson Assembly reactions diluted to approximately one template molecule per sample. Twenty 7 mers, sequence- specific to the pKS10 vector but not for the inserts (HA or NA), were designed and used for the study (**Supplementary Table S2**). Twenty replicates of the MDA reaction each containing approximately a single molecule of

DW26 template were set up in 40 µl volumes. The reactions were incubated at 30°C, and 7 µl of each reaction was collected at various designated times: 6, 8, 12, and 16 h (**Figure B-6A**). Three replicates of individual samples were analyzed. At least 12 h of incubation were required to obtain 1 µg of DNA products using MDA alone (**Figure B-6B**).
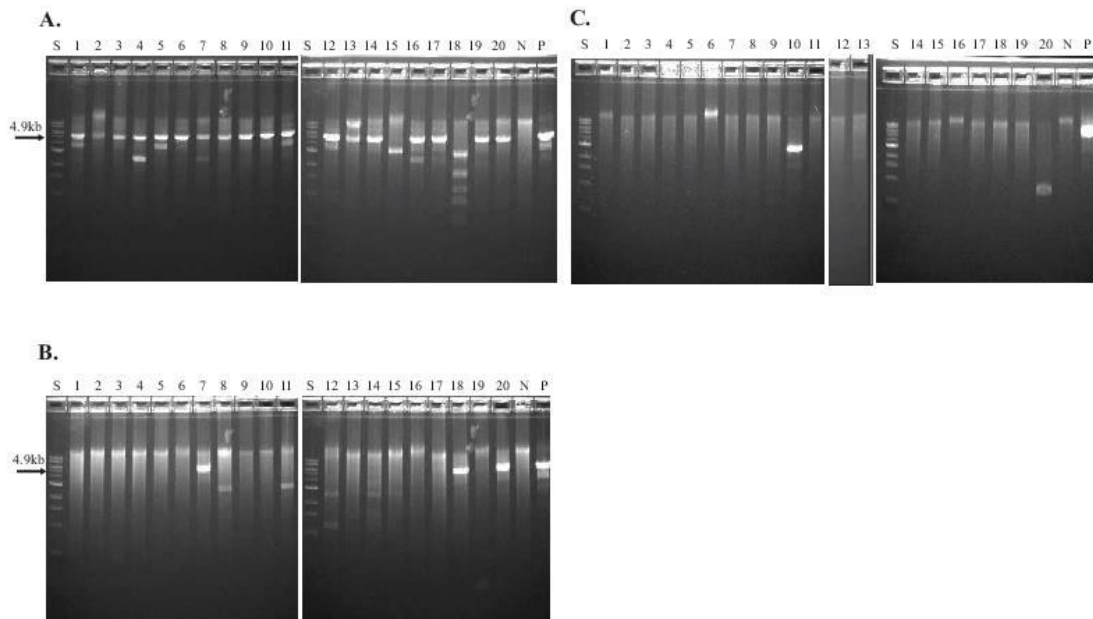
Figure B-5. Various dilutions of a DNA template followed by MDA to determine limiting dilution.

DNA template containing DW 29 (NA segment) was prepared by limiting dilution: containing DW29 (NA segment) was prepared by limiting dilution: 10^{-6} (A), 10^{-7} (B) and 10^{-8} (C). 10^{-7} dilution of the DNA template provided a single molecule level based

on the real time PCR quantification. MDA reactions in twenty replicates for each dilution were set up (see Methods). After incubating for 16hrs, each reaction was visualized on 1.2 % agarose gel after the restriction enzyme digestion with NcoI. Lane "S" indicated 1kb DNA ladder (NEB).

A.

B.
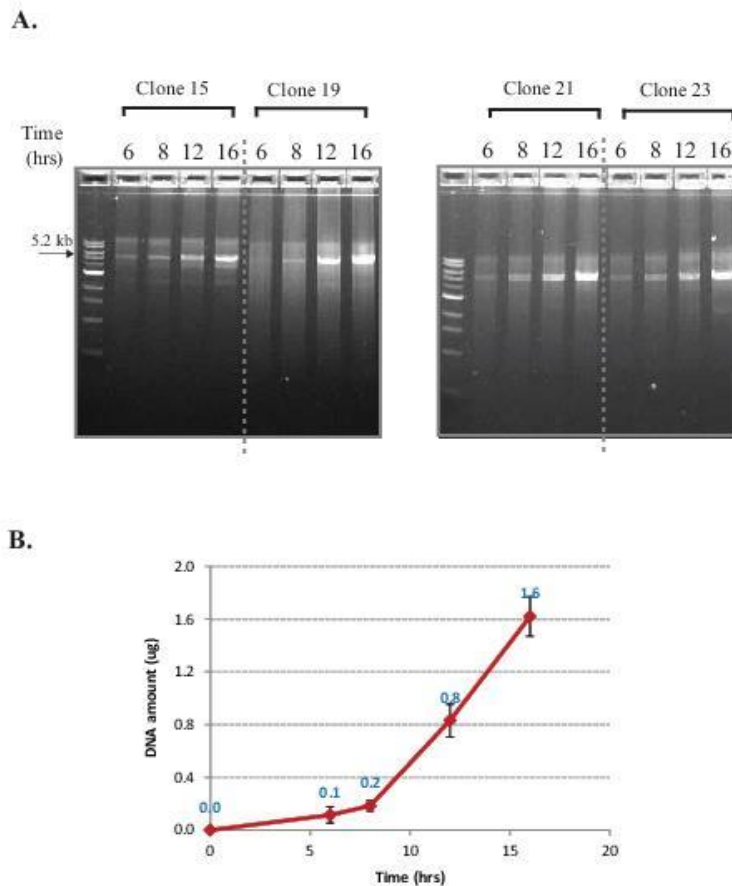
Figure B-6. The kinetics of a single molecule MDA.

(A) The amount of MDA products increases with longer incubation of time. Twenty replicates of MDA reactions of a single molecule dilution were taken at 6, 8, 12, and 16 h following initiation of the MDA incubation. The products collected were

visualized on a 1.2 % agarose gel after the restriction enzyme digestion. Only 4 clones

that were positive were shown. (B) Amount of DNA products from each positive clone

was quantified by real time PCR. Error bars indicated standard deviation of four

replicates.

Table B-3. The quantification of the copy numbers of molecules in serial

dilutions of vaccine seeds.

Vaccine Seed	Serial dilution	CT value	# of molecule (+/- SD)[a]
pKS10-DW26 (5.27kb; 1.76kb)[b]	10^{-4}	27	2209 (+/- 391)
	10^{-5}	31	183 (+/- 93)
	10^{-6}	34	14 (+/- 6)
	10^{-7}	38	1 (+/- 0.5)
	10^{-8}	UD	0 (+/- 0)
pKS10-DW29 (4.9kb; 1.4kb)[c]	10^{-4}	26	4303 (+/- 1040)
	10^{-5}	29	434 (+/- 192)
	10^{-6}	33	18 (+/- 2)
	10^{-7}	37	1 (+/- 0.3)
	10^{-8}	UD	0 (+/- 0)

a.   # of molecules represents an average value of three replicates.

b.   The HA insert is 1.76 kb; the size of total plasmid including the HA insert is 5.27 kb.

c.   The NA insert is 1.4 kb; the size of total plasmid including the NA insert is 4.9 kb.

UD: undetermined

MDA reaction followed by HFPCR shortens the in vitro cloning of DNA vaccine seeds.

The long reaction time (12-16 h) during the MDA reaction are drawbacks of the protocol described above. Our kinetic study showed that 6 h of MDA was sufficient to generate approximately 0.2 µg of DNA products from a single molecule; however, the number of sequence errors introduced in the amplicons made this method unacceptable. This observation raised the question whether a PCR using a high fidelity DNA polymerase following an earlier amplification stage of MDA (e.g. < 6 h of MDA) would yield the 1 µg of DNA needed for virus rescue without resulting in too many sequence errors. A 2.5 hr MDA reaction generated approximately 1 ng of pKS10-DW26 from a single molecule, and the following HFPCR produced approximately 10 µg of final products (**Figure B-7**). To reduce the number of errors introduced during the amplification of single molecules to a level that would support viral rescue, we tested shorter MDA reaction periods. We observed positive clones after 2 h of MDA, but not after 1.5 h. Both reactions were followed by High Fidelity PCR (HFPCR) prior to analysis. The real time PCR analysis showed the 2 hr MDA generated approximately 0.1 ng DNA from a single molecule.
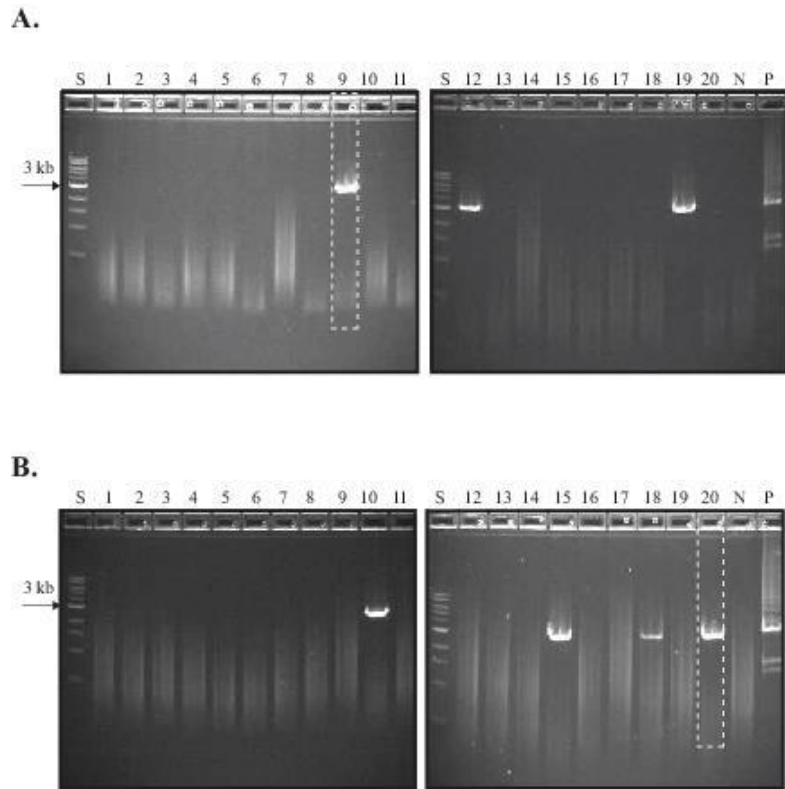
Figure B-7. Modified *in vitro* cloning, MDA-HFPCR, for both HA (A) and NA (B) segments.

20 replicates of MDA-HFPCR were used for each segment. The concentration of dNTP for MDA was 1mM. The primer set used for the PCR part was BMP 13 & 14, which produces longer products (~3 kb). After PCR, the products were directly applied on agarose gel for visualization. The details of modified *in vitro* cloning can be found in text. S: 1kb DNA ladder; N: negative control; P: positive control (4.7×10^6 DNA molecules).

Error rates of in vitro cloning

In order to see whether the positive clones after the single copy amplification have a clonal population of DNA, we did a further sequence analysis of each clone: if the MDA product were from a single DNA copy, only one sequence would be expected to be obtained. We transformed the MDA product into *E. coli* and then the sequence of a set of clones was analyzed by Sanger sequencing. Clone #27, which yielded a product of the correct size in the MDA reaction (**Figure B-8**) from the DW26 HA *in vitro* cloning was selected for sequence analysis. We linearized the MDA products by cutting with NcoI and then self –ligated the resulting molecules to reform the original plasmids. Those were transformed into *E. coli* and we analyzed a population of 76 colonies by Sanger sequencing (Details in Materials and Methods). The entire sequence of the 1,762 bp HA region was read. Twelve among 76 colonies had a single base pair substitution at various places in the HA region. Sixty-four colonies had the expected sequence. Thus the error rate for a 12 hr MDA reaction was ~1 error per 11,000 bp. In our view, the most likely cause of these errors was low phi29 DNA polymerase fidelity, and we will address this issue in the discussion.

After the HFPCR, approximately 0.8 - 2 µg of total DNA products were obtained (**Figure B-9B**). This result demonstrated that *in vitro* cloning of the synthetic DNA from oligonucleotides, capable of a virus rescue, could be completed in 8 h, which is slightly faster than the 10 h needed for the method our team reported on earlier for generating synthetic HA and NA DNAs that used enzymatic error correction [8]. The sequences of positive clones from the 2 hr MDA-HFPCR were analyzed in order to see whether the sequence errors due to phi29 DNA polymerase were reduced relative to

the MDA only protocol. The PCR fragments were cloned into the Zero Blunt® TOPO

PCR cloning kit (Invitrogen) followed by blue/white screening on LB/X-gal plates. We

picked 100 white colonies initially and obtained sequence results from 82 colonies.

Among them, 16 colonies (20%) showed a single mutation (either a base change or

base pair addition) in various positions of the HA segment. This equated to ~1 error

every 9,000 bases synthesized by the MDA-HFPCR.

The components of the phi29 DNA polymerase reaction were re-examined in order to

further reduce the sequence errors due to the MDA reaction. We had elected to use 1

mM dNTPs in the MDA reactions. Various studies reported that more errors occurred

when 1 mM of dNTPs were used for synthesis for different DNA polymerases

[342];Reviewed in [343]. For instance, the fidelity of Pfu DNA polymerase increases

at less than 400 µM dNTPs, and the fidelity of *Thermococcus litoralis* polymerase was

constant in the range of 10 - 200 µM dNTPs [344, 345]. Therefore, we reduced the

dNTPs concentration to 200 µM for the 2 hr MDA reaction and continued HFPCR to

obtain amplicons. The products were analyzed to examine whether this would affect

the DNA production and error incorporation. The real time PCR analysis showed that

changing the dNTPs concentration to 200 µM in the MDA reaction dramatically

reduced the final DNA product (0.01 ng) (**Figure B-9A**). We performed sequencing

analysis of this product to measure the impact on the error occurrence. Two among 54

sequenced colonies had a single mutation in different locations along the sequence of

the HA segment (~4%). That equated to ~1 error per 47,500 bp synthesized. This

indicated that decreasing the concentration of dNTPs to 200 µM reduced the errors

occurring during *in vitro* cloning. To obtain the more than 1 µg of final DNA products

necessary to pursue virus rescues, optimization of the HFPCR, such as using different primers and adding more cycles, was attempted. A new primer set, BMP21 and BMP22, were utilized and 5 additional cycles were added (total 30 cycles) to the HFPCR. This new primer set generated a product approximately 1 kb shorter than using the previously used BMP 13 & BMP14 primer set (**compare Figure B-9 & Figure B-7**). The 2 hr MDA reaction, followed by the HFPCR using the new primer set BMP 21 & BMP 22, produced amounts DNA sufficient for virus rescues: 3 - 4 µg (200 µM dNTPs) and 22 - 25 µg (1 mM dNTPs) (Figure B-9C &D).
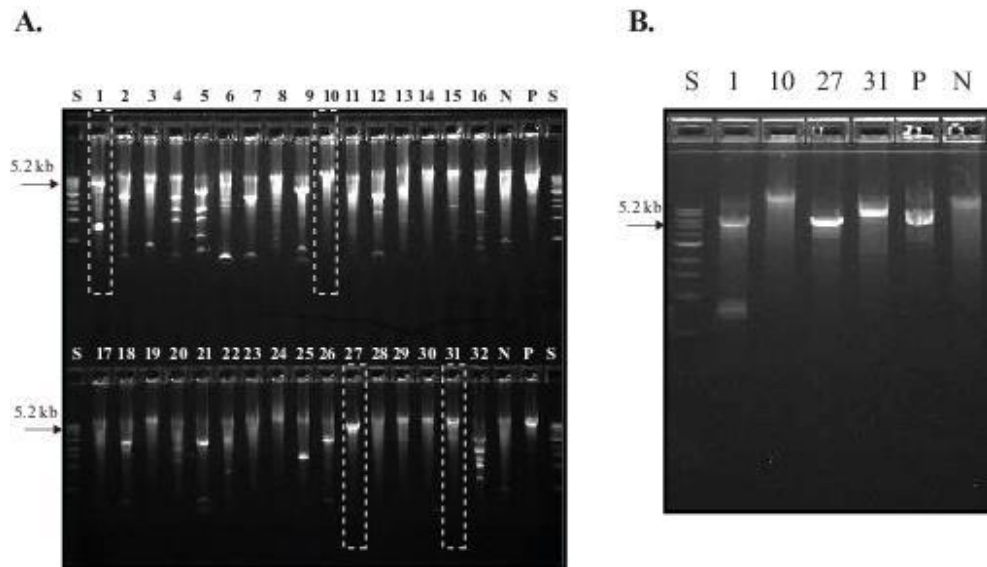
Figure B-8. *In vitro* **Cloning of DW26 (HA) segment by 16 hours MDA.**

DNA template was prepared by limiting dilution mentioned in text. 32 replicates of a single molecule MDA reaction were set up (described in Materials & Methods) and the products were visualized on a 48-well agarose gel (A). 4 possible clones were picked up and rerun on another agarose gel (B). S: 1 kb DNA ladder; N: negative control; P: positive control (4.7X10^6 DNA molecules)
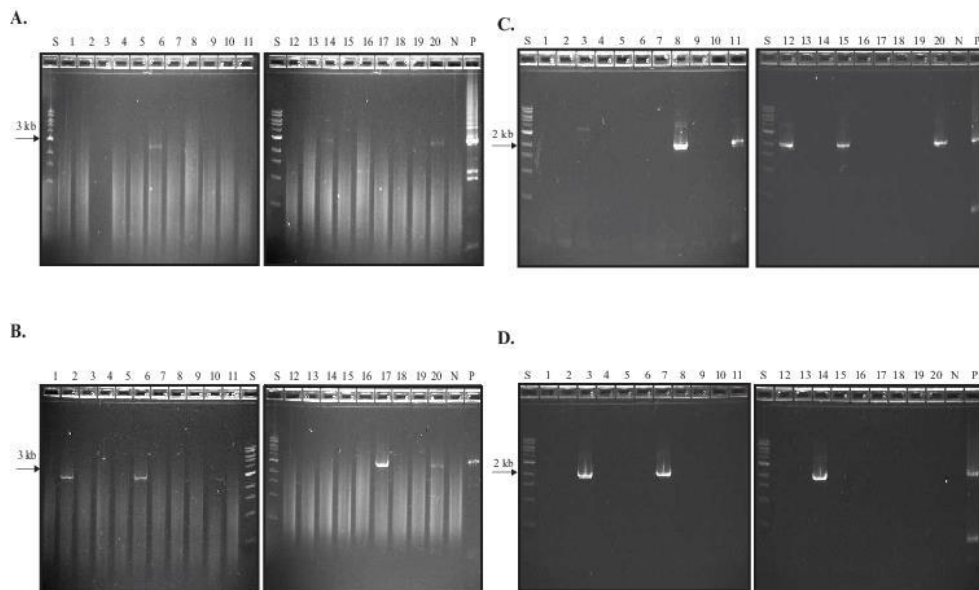
Figure B-9. Modified *in vitro* cloning, MDA-HFPCR, supplemented dNTP concentrations.

pKS10-DW26 was used as a DNA template for *in vitro* cloning. The DNA was diluted in a single molecule level and used for 2 h MDA followed by HFPCR. Two different sets of primers, BMP13 & 14(A & B) and BMP 21 & 22 (C & D), were examined with two different dNTPs concentrations; 200 µM (A & C) and 1 mM (B & D). 20 replicates of each condition were examined. The primer set, BMP 13 & 14 produced the final product about 3kb in size, while BMP 21 & 22 generated shorter products, 2kb. After PCR, final products were visualized on agarose gels. Lane S indicates 1kb DNA ladder. N indicates a negative control containing no DNA template; P shows the positive control starting with about 4.7×10^6 molecules of DNA templates in a reaction. Each gel is representative of three independent experiments.

Table B-4. Error rates of replication processes and assembly techniques.

Process	Error rate (1 error per x bp)
In vivo DNA replication	10^9-10^{11}
HFPCR (Phusion® polymerase)	~10^7
oligonucleotide construction	500 - 750
Assembly reaction with no error correction	1328
Assembly reaction with enzymatic error correction	9589
12 h MDA, not optimized	~11000
4.5 h MDA-HFPCR, not optimized	~9000
4.5 h MDA-HFPCR, optimized	47500

Synthetic DNA vaccine seeds prepared by in vitro cloning were rescued.

To confirm DNA produced by *in vitro* cloning can be used to generate influenza virus via reverse genetics, DNA from a 16 hr MDA reaction was used to generate clones of pKS10-DW26 (HA). The total amount of DNA was measured by real time PCR. 0.5 µg of DNA of each clone was used for virus rescue. These DNA clones were used in a transfection for virus rescue with externally provided plasmids of the influenza A virus backbone segments, as well as NA. In addition, to determine if circular DNA is necessary to produce virus, the generated linear MDA amplicons were ligated with T4 DNA ligase to form circular molecules (the reaction likely produced a small amount of concatamers; although this was not confirmed). pGFK2001, purified from *in vivo* cloning from *E. coli* and containing the HA (DW 26) segment, was used as a positive control, and a transfection performed without the HA segment was used as a negative control. Viral titers at three different time points post transfection using a TCID$_{50}$ assay determined influenza was produced from *in vitro clones* (**Figure B-10A**).

Screening of MDA products in agarose gels identified two clones (DW26-1 & DW26-27) that had the correct size DNA bands. A subsequent transfection showed that Clone DW26-27 was successfully rescued, while clone DW26-1 was not (data not shown). Upon viral sequencing analyses, we found that Clone DW26-27 had the wild type DW 26 HA sequence, but Clone DW26-1 contained a single base pair mutation encoding a base pair substitution at position 526, changing an arginine to an isoleucine. This indicates that if virus is not produced via transfection the most likely cause is a sequence error, implying that *in vitro* cloning requires screening to identify correctly sequenced viruses.

Virus was generated from both HA (DW26) and NA (DW29) fragments generated from the MDA-HFPCR *in vitro* cloning reaction as well as from the combination of the two clones; although the viral titer was lower than that of the positive controls (**Figure B-11A**). In addition, virus was rescued from *in vitro* clones regardless of dNTPs concentrations and primer set used. The primer set BMP 13 & 14 and the primer set BMP21 & BMP22 showed no significant difference in transfection efficiency (data not shown). Viruses were sequence verified (details in Materials & Methods) to confirm whether they were generated from the original *in vitro* DNA.

A.
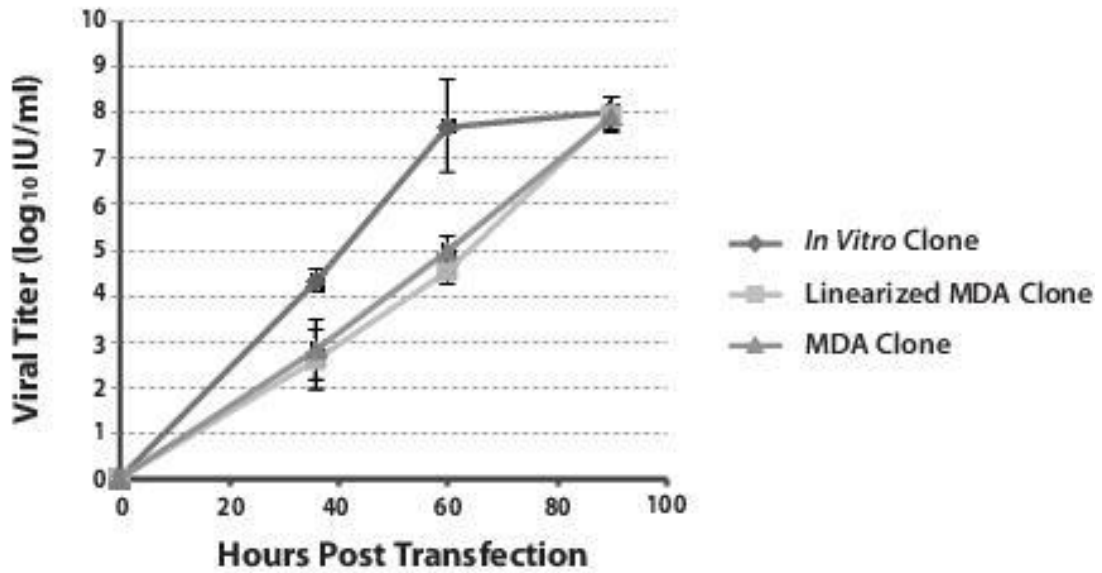
B.
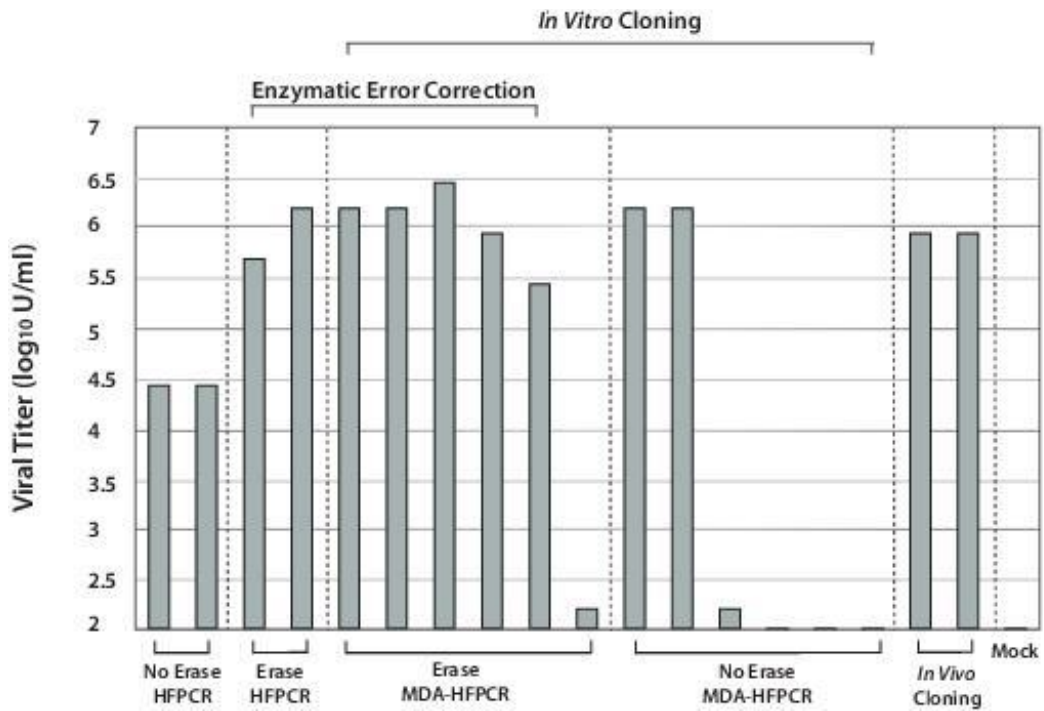
Figure B-10. Virus rescue of *in vitro* cloning products.

(A) The 16 hr MDA reaction viral rescue. Two positive DW 26 HA segment containing clones, were obtained from a MDA reaction (Figure B-8). One of which was ligated to form a circular DNA product. 0.5 µg of each product was used for virus rescue along with PR8 NA and the other DNA components needed for virus production. pGKF2001 was used as a positive control, *in vivo* clone.   The efficiency of virus rescue was determined by TCID$_{50}$ assay. Error bars were obtained from the standard deviation of three replicates. (B) Comparing Influenza rescue from DNA prepared with or without error correction or *in vitro* cloning. Two - six replicates of each condition were compared to a positive control *in vivo* clone and a negative control mock.

A.
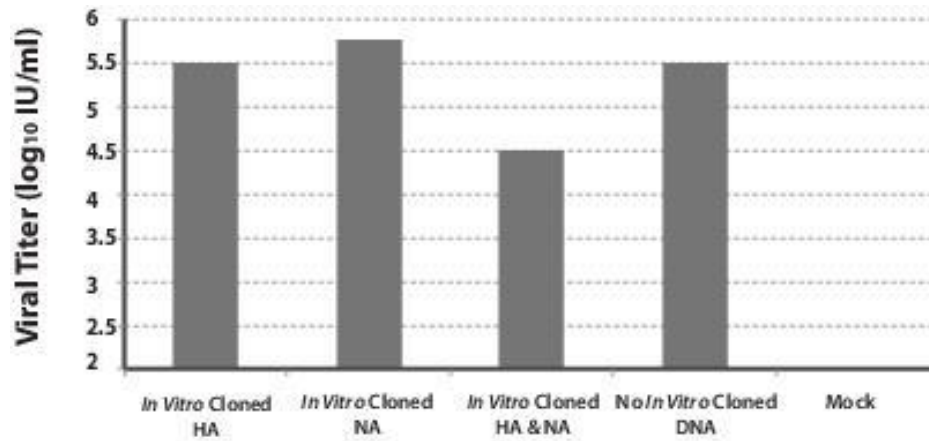
B.
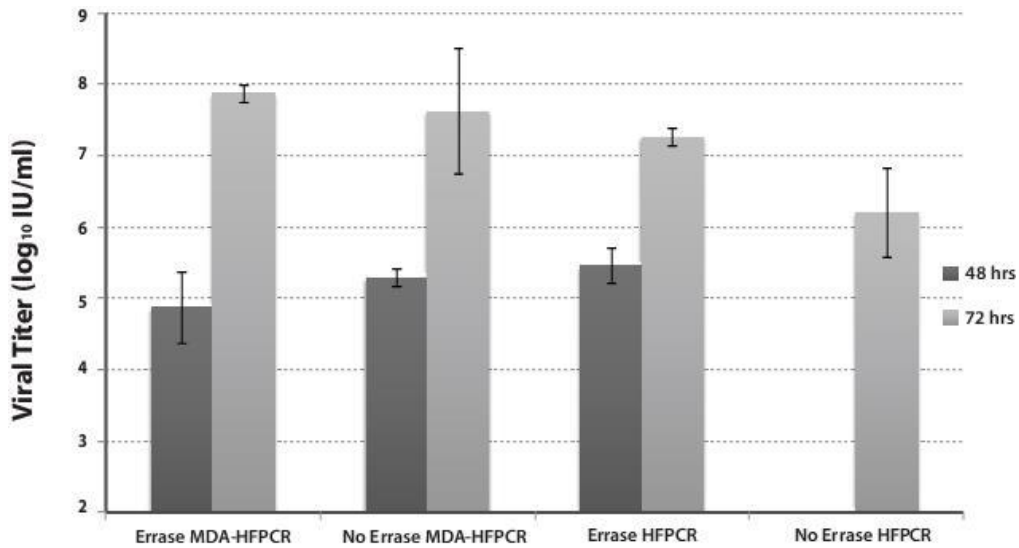
Figure B-11: Virus rescue of vaccine seeds prepared by MDA-HFPCR with or without the error correction step 48 h and 72 h post-transfection.

(A) The combination of genes prepared by MDA-HFPCR or HFPCR. 0.5 μg of each product was used for virus rescue along with an *in vivo* clone of the other DNA components needed for virus production. Efficiency of virus rescue was determined by $TCID_{50}$ assay. **(B)** Vaccine seeds prepared by MDA-HFPCR with or without the error correction step (using ErrASE®) and were observed in two different time points post-transfection: 48 h and 90 h. The error bar of each sample represents the standard deviation of two independent transfections. Efficiency of virus rescue was determined by $TCID_{50}$ assay.

MDA-HFPCR as a substitute for ErrASE® error correction.

To compare enzymatic error correction with *in vitro* cloning, DNA from DW29 (NA) was produced four different ways following the Gibson Assembly step: by MDA-HFPCR without an ErrASE® correction, by HFPCR alone without an ErrASE®

correction, by MDA-HFPCR with an ErrASE® correction, and HFPCR alone with an ErrASE® correction. Meaning, each DNA preparation tested either: *in vitro* cloning as an error screening method, ErrASE® error correction, the combination of the two or neither technology (**Figure B-10B**). *In vitro* cloning was performed via MDA-HFPCR with primer set BMP21 & BMP22. Enzymatic error correction was performed using ErrASE®. We then used equal amounts of DNA in a rescue experiment in conjunction with the other genes necessary for successful influenza virus rescue. Seventy-two h post transfection viral titer was measured by TCID50 assay. Two to six replicates were performed using each method in addition to a positive control produced from an in vivo cloning from E. coli plasmid prep. Without using either *in vitro* cloning or enzymatic error correction a pool of molecules is generated with a relatively high portion of them incorrect. This resulted in lower rescue efficiency due to the mixture of the DNA with correct sequences along with incorrect ones. Clones with the ErrASE® treatment alone reduced the portion of incorrect clones in the pool and were therefore rescued with efficiency similar to a positive control. When only *in vitro* cloning was used single molecules were amplified from a pool of correct and incorrect clones so screening was necessary to identify a correct clone; we observed 2/6 clones tested generated a high titer of virus. ErrASE® in conjunction with *in vitro* cloning revealed that 5/6 clones were rescued at the same level as the positive control. This data further indicated that the pool of molecules that were generated after ErrASE® correction had a higher portion of correct clones than a pool generated without error correction. However, without using enzymatic error correction clones that generated a high titer of virus could still be identified after screening only 6 clones. These results were consistent

215

with another independent experiment. Positive clones which had been treated with either ErrASE® or *in vitro* cloning successfully rescued more efficiently than a pool without any treatment both at 48 h and 90 h post-transfection (**Figure B-11**).

B. 5 Discussion

The success of synthesizing genes and the dramatic reduction in price of gene assembly technology provide great potential to produce viral vaccine seeds from sequence information without requiring virus or viral genomes being shipped to vaccine makers [8, 346]. Herein, we demonstrate *in vitro* cloning can rapidly produce influenza virus. DNA assembly technology can enable us to prepare vaccine seeds before the peak of a pandemic and at a lower cost. Additionally, our *in vitro* cloning should deliver a more homogenous population of DNA molecules to reverse genetics reactions than our previously described approach because it clones from a single isolated template molecule. This is valuable because incorrect viral genome segment DNA molecules can cause several kinds of problems. An infection or transfection containing a fraction of incorrect HA or NA DNAs at a high multiplicity (meaning multiple copies of each DNA go into each cell) may produce viral particles that are noninfectious if the error is an indel or results in a lethal amino acid substitution. These viruses may act as a sort of defective interfering particles that can lower production of the intended virus.

Conditions for single molecule amplification using MDA.

The protocol, we developed for rolling circle MDA from a single template molecule, is more amenable to DNA production than other reported methods. Previous studies addressed the issue of non-specific by-products forming during the MDA using random

216

hexamers. Various ways have been reported to diminish the by-products, such as using a nanoliter scale of reaction volume, cleaning materials with UV-irradiation, and doing MDA with RNA primers [338-340]. In this study, we showed that using sequence-specific DNA 7 mers rather than random hexamers enhanced amplification of a single copy of the MDA template in a 10 µl reaction volume. It was not necessary to concentrate the template by using nanoliter scale reactions as was done by Hutchison *et al.* using random primers [338]. Therefore, it was easy to handle samples. Another advantage of our method over typical MDA using phi29 DNA polymerase was we removed bovine serum albumin (BSA) from the reaction and omitted an additional step of cleaning up starting materials. Moreover, using DNA primers is more cost effective and time saving than using RNA primers, which have been advocated for use in MDA [339]. Our specific DNA heptamer cost $10 each and could be received from oligonucleotide foundries in one day. Equivalent RNA oligos cost $80 each and took a week for synthesis and delivery.

MDA-HFPCR is time efficient.

The initial 16 h *in vitro* cloning step relying solely on MDA was reduced to 4.5 hours by pairing the reaction with HFPCR. By using two DNA polymerases we were able to take advantage of the high processivity of phi29 DNA polymerase and high fidelity of Phusion® polymerase [347-349]. High processivity of phi29 DNA polymerase enables one to amplify from a single copy of DNA in order to get the minimal DNA products for HFPCR, which can produce a higher amount of DNA more efficiently with high fidelity (generating fewer errors). This compares favorably to what we reported in Dormitzer *et al.* [8]. The oligonucleotides to influenza gene time in Dormitzer *et al.*

217

was 10 hours. Whereas the oligonucleotides to influenza gene time for *in vitro* cloning is 8 hours. This would imply a digital to influenza gene time of 21.5 hours reported in Dormitzer *et al.* would be 19.5 hours for *in vitro* cloning. The digital to virus time for *in vitro* cloning we would expect to be 80 hours.

Error rate of phi29 DNA polymerase and in vitro cloning

We used phi29 DNA polymerase because the enzyme can catalyze MDA from a single template molecule with high processivity. The unexpected drawback we found was that phi29 DNA polymerase MDA was a much more error prone process than *in vivo* DNA replication or high fidelity PCR using Phusion DNA polymerase (as we have used previously [8]). Various studies have been measured the elongation fidelity of phi29 DNA polymerase. Esteban *et al.* (1993) first reported it between 10^{-5} and 10^{-6}, which is 10-100X higher than for Phusion DNA polymerase [350, 351].   In our initial experiments using 1 mM dNTPs and long MDA incubation times, we found as much as 36% of the resulting NA molecules had one or more sequence errors. We confirmed that the errors were caused by the phi29 DNA polymerase itself. We used a plasmid containing the correct DW 26 sequence (pGFK2001) and diluted to the single molecule level to do MDA. After collecting the positive clones, one of them was transformed into *E. coli* for the sequencing analysis. We sequenced the DNA samples of 95 colonies and found 30 colonies having independent errors (32% and 1 error per ~5000 bp synthesized) in various locations inside the DW 26 HA segments. This result indicated that the errors did come from phi29 DNA polymerase activity. Importantly, once we adopted 200 µM dNTPs and a 2 h MDA reaction followed by a HFPCR, we were able to produce sufficient DNA for both reverse genetics reactions and sequence

confirmation using Sanger chemistry. Analysis of *in vivo* cloned *in vitro* cloning products showed only 2 among 54 sequenced colonies had a single mutation in different locations along the sequence of the HA segment. That equated to ~1 error per 47,500 bp synthesized. This is approximately as low as the error rate reported in the Dormitzer *et al.* publication.

Among the errors we observed that were introduced by phi29 DNA polymerase, many were insertions of extra base pairs of A/G in the A or G stretch regions, while other errors seemed to be random base pair changes. Also, no deletion mutation was observed. None of the previous studies has reported errors during MDA and it would be useful to characterize the errors and calculate the error rate of phi29 DNA polymerase more thoroughly for future study. Recently, Salas *et al.* reported a manipulated phi29 DNA polymerase that can bind to DNA tighter than wild type [352]. Such an enzyme might have improved the processivity and fidelity over standard phi29 DNA polymerase, which would make MDA even more useful. Our study is the first report to address the errors by phi29 DNA polymerase and address this error problem for successful, *in vitro* cloning.

DNA vaccine seeds prepared by in vitro cloning are an effective alternative to enzymatic error correction for viral rescue.

Our results indicated that MDA alone was capable of producing DNA vaccine seed that was capable of rescuing influenza virus; however, the process was too slow and introduced too many errors. Since linear products from DNA are sufficient for a virus rescue [8], we attempted to shorten the process and reduce error accumulation by utilizing a HFPCR in combination with MDA. This successfully reduced the *in vitro*

cloning time. Rescue experiments confirmed the HA, the NA, and the combination of the two genes produced by MDA-HFPCR were capable of generating virus.

After error rate optimization, we determined that *in vitro* cloning was effective and practical at screening out assembly errors. Rescue experiments comparing enzymatic error correction and *in vitro* cloning indicated that *in vitro* cloning was as effective as enzymatic error correction at producing DNA assemblies capable of virus generation. It is important to note that *in vitro* cloning amplifies from a pool of assembled DNA that contains both correct DNA sequences along with incorrect ones. The number of clones that generate virus is proportional to the number of correct clones in that pool. Our results are consistent with the previous observation by Dormitzer *et al.* (2013) that only ~ 25 % of a DNA pool has the correct DNA sequences without the error correction step [8]. We observed 2/6 clones generate a high titer of virus. Furthermore, when the proportion of incorrect assemblies is reduced, such as when enzymatic error correction is used, the number of clones which generate virus when *in vitro* cloning is used increased. We observed 5/6 clones generate a high titer of virus when *in vitro* cloning is used to select for correct assemblies from a more homogenous pool of assemblies is used for amplification (Figure B-10B).

Significance

One of the most effective responses to influenza outbreaks is how fast the production and distribution of vaccines can be executed. Synthesizing DNA using synthetic biology tools has shown great promise to improve public health. In this study, we report on the "*in vitro* cloning" of synthetic influenza vaccine seeds, as part of the synthetic seed production, to produce and screen the vaccine seeds with correct genetic

information after the gene assembly without relying on living organisms (**Figure B-12**). Our method using MDA-HFPCR enables one to clone DNA with the correct genetic information 8 h upon the arrival of oligonucleotides, and accelerates the entire production procedure compared to using conventional *in vivo* cloning. Additionally, the enzymatic error correction step of the DNA assembly can be skipped, which reduces the time needed for construction and may increase consistency. Further, this technique can provide higher quality vaccine seeds in a shorter period of time. This study, ultimately, can be used to guide the automation of the entire vaccine manufacturing procedure for faster and more equitable distributions.
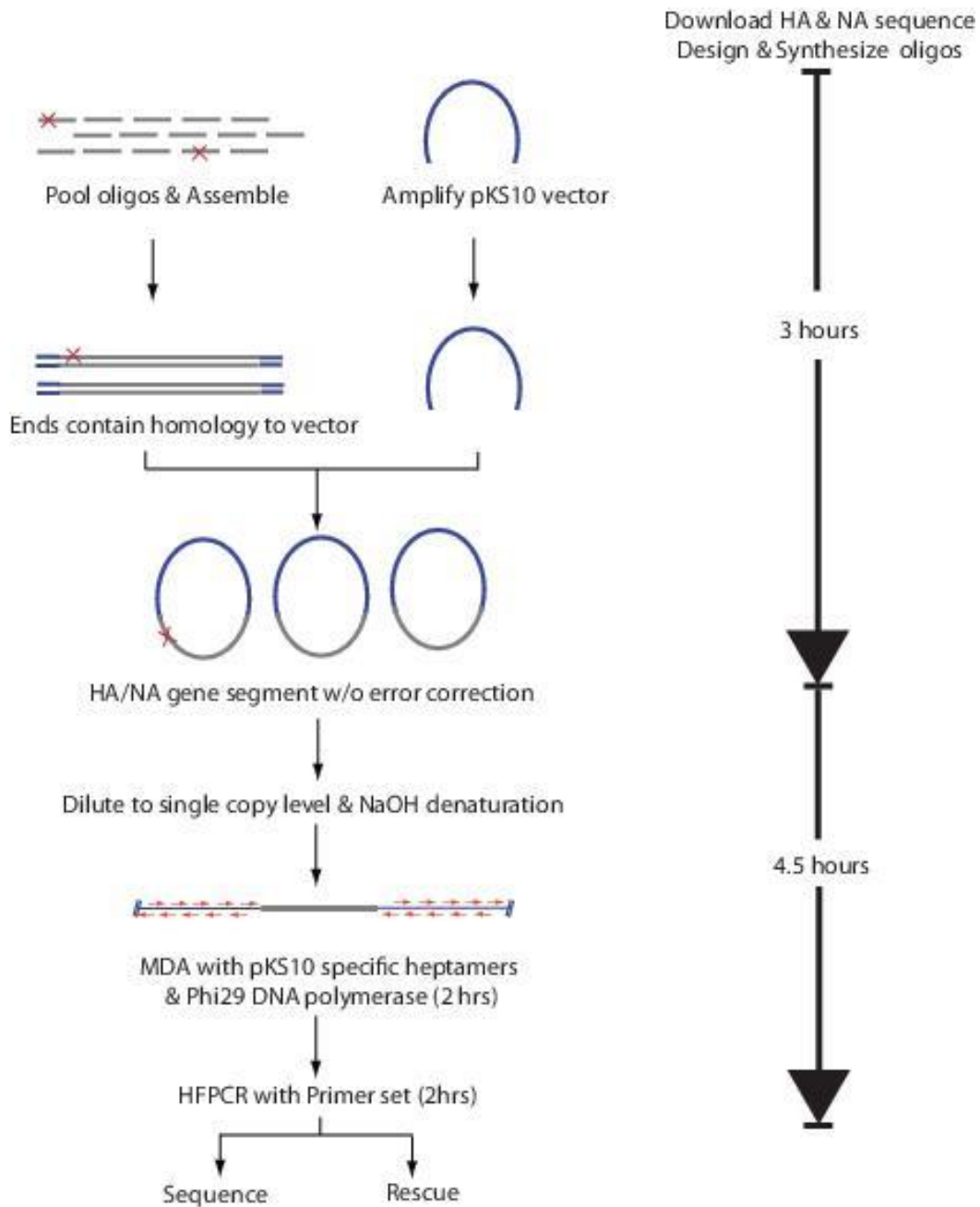
Figure B-12. The scheme of *in vitro* cloning technique, MDA-HFPCR.

(A) Schematic diagram of procedure of MDA-HFPCR. Either HA or NA gene segments was assembled based on the previous study (**Appendix A**). The X symbol designates sites of oligonucleotide synthesis errors. Gray lines represent influenza coding sequences; Blue is the pKS10 vector. Red arrows represent DNA sequence

specific heptamers that can align on pKS10. **(B)** Timelines for MDA-HFPCR. DNA was ready for transfection, about 4.5 h after the gene assembly without the error correction step.

Appendix C: A cluster of eight genes necessary for *Mycoplasma mycoides* Contamination of mammalian cell culture

Authors: David M. Brown[1, 2], Nacyra Assad-Garcia[1], John I Glass[1].

Affiliations:

[1] Synthetic Biology and Bioenergy, J. Craig Venter Institute, Rockville, MD, 20850, United States.

[2] Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA.

Abstract

As the smallest free growing organism, it has been suggested that *Mycoplasmas* should be used as a blueprint to evaluate the genes necessary for life. However, one problem in using *Mycoplasmas* is the potential for contamination of mammalian cell cultures. The need to ensure *Mycoplasma* free cultures is essential. While it is likely that significantly minimized strains lack the capacity to infect mammalian cultures this study demonstrates that the minimal cell JCVI-syn3.0 cannot persist in mammalian cell cultures. We have developed an infectivity assay to evaluate the persistence of *Mycoplasma mycoides* strains JCVI-syn1.0, JCVI-syn3.0 along with other synthetic *M. mycoides* strains. We identified a cluster of eight genes necessary for the persistence of *M. mycoides* in mammalian cell cultures.

INTRODUCTION

Mycoplasma contamination poses a frequent threat to cell cultures and biological materials. Experimental results can become unreliable and biologic products defective if exposed to a viable strain of Mycoplasma [353]. The use of contaminated cells endangers almost all aspects of cell physiology, and often leads to erroneous results or causes the loss of unique cell lines [309-311].

Mycoplasma infection of cell cultures is difficult to detect. Adherence to the cell surface results in no turbidity in the media as a result of bacterial growth. Mycoplasma contamination cannot be detected by the naked eye or under an inverted microscope. Furthermore, *Mycoplasma* contamination is difficult to remove. The complete lack of a bacterial cell wall confers a natural resistance against penicillin [354]. Small cellular volume allows filterability through 200-nm pore diameter membrane filters [355]. The small size of *Mycoplasma* (0.15-0.3nm) is the main reason for their escape through filtering systems and also their growth in high concentration in mammalian cell cultures without any turbidity or other obvious symptoms [310]. Detection via PCR has been used to determine if Mycoplasma contamination is present and which species are persistent [356].

Intense interest has been placed on Mycoplasma species from the synthetic biology community. As the smallest self-replicating bacteria known, their small genome size has been proposed as a possible blueprint for the design of synthetic organisms [357]. The smallest known natural genome of any free-living organism capable of being grown in axenic culture pgenome is 580 kb, for *Mycoplasma genitalium* having only 482 protein-coding genes [358]., In 2010 a 1.1 Mb genome based on the genome of

Mycoplasma mycoides used to create the strain JCVI-syn1.0 [4]. The strain contains all the genes wild type mycoides. Recently our group at JCVI has reported a minimal cell (JCVI-syn 3.0)[7]. Using several cycles of design, strains of intermediate genome size were obtained until the genome was minimized to 531 kb and 473 genes. To obtain this minimal cell several intermediates were obtained. These intermediates were used to evaluate which genes are responsible for *Mycoplasma* persistence in cell culture. In this report we evaluated the infectivity of several synthetic strains produced by JCVI and determined 8 genes necessary for the persistence of *Mycoplasma mycoides* in cell culture.

MATERIALS AND METHODS

Mycoplasma strains and cell culture

All synthetic *Mycoplasma* mycoides strains were obtained from theteam at JCVI. Each strain is described in detail in previous reports [7]. A brief description of each strain can be found in **supplementary table 1**. HEK293 and HeLa cells (ATCC) were maintained in Dulbecco's modified Eagle's medium (DMEM) (Gibco) with 10% fetal bovine serum, 100 units/mL of penicillin, 100 µg/mL of streptomycin, and 0.25 µg/mL Amphotericin B at 37°C and 5% CO_2.

Determination of Color Changing Unit (CCU)

The CCU assay was performed in SP4 media containing a phenol red preparation supplemented with 3 µg/ml tetracycline. Bacterial metabolism leads to a pH shift that in turn causes a color change from red to orange and then to yellow [359]. Briefly, 100

µl of the growing culture or sample was added to 900 µl SP4 media followed by serial 10-fold dilutions as described previously [106, 359-363] Unless otherwise specified CCU assay was allowed to reach saturation after incubating at 37 degrees C for 1 week.

Infectivity assay

Synthetic *M. mycoides* strains were grown in SP4 medium supplemented with tetracycline. Once cultures were orange (about 24 h) 100 µl of sample was added to 5 ml of supernatant of a growing mammalian cell culture (approximately 30-50% confluent) growing in a T25 25 cm^2 flask. At select timepoints a CCU assay was performed. Once the mammalian cell culture reached confluence the medium was removed and 1 ml of trypsin (Gibco) was used to wash the cells. Trypsin was removed and 1 ml fresh trypsin was added to the flask. Mammalian cells were allowed to incubate until they detached from the surface. Once trypsinization was complete cells were passaged into a fresh 25T flask by adding 500 µl of cell suspension into 4.5 ml fresh medium. If *Mycoplasma* contamination was present after 7 d post infection and after to passages strains were designated as persistent.

RESULTS

Detection of *Mycoplasma* infection in HEK 293 cells

To evaluate if *Mycoplasma* is persistent in a culture of HEK293 cells an infectivity assay was performed. **(Table 1)** The synthetic *M. mycoides* JCVI-syn1.0 genome is a non-minimized strain based on the genome sequences of two laboratory strains of *M. mycoides* subspecies *capri* GM12 both strains capable of contaminating cell culture. JCVI-syn2.0 is a partially minimized synthetic strain approximately 576 kb with 479

genes [7]. It is highly likely that the genes responsible for infectivity and pathogenesis have been removed from JCVI-syn2.0. First we had to ensure that JCVI-syn1.0 and JCVI-syn2.0 could not grow in the mammalian medium (DMEM + 10% FBS). Fresh media was infected with each strain then incubated at 37 degrees. 7 d post infection phenol red present in the spent media did not indicate bacterial metabolism and a negative CCU assay determined that not only had bacterial growth arrested, no surviving *Mycoplasma* was present in spent media 7 d post infection. (**Table C-2 , supplementary figure C-1A**)

Once we confirmed *Mycoplasma* cannot grow in mammalian medium, contamination of HEK293 cells was confirmed by infecting an actively growing culture of HEK293 cells with JCVI-syn1.0. Contamination was evaluated at select time points and after several passages. We determined that not only is JCVI-syn1.0 persistent in HEK293 cells, it is growing. Comparing the CCU of a the 5th passage 14 d post infection with the CCU 15 d post infection, a 10-100-fold increase in CCUs were present after 24 h of growth. (**Supplementary Figure C-1B**)

JCVI-syn2.0 was evaluated alongside JCVI-syn1.0. JCVI-syn2.0 could be detected up to 7 d post infection and before the 2nd passage of HEK 293 cells. Results from this infectivity assay formed the basis for determining the threshold of detecting whether each synthetic *Mycoplasma* strain is able to persist in mammalian culture. *Mycoplasma* contamination could not be determined by eye or with an inverted microscope (**supplementary figure C-2**)

Using minimized *Mycoplasma* strains identified genes necessary for infection

JCVI-syn2.0 is significantly minimized from JCVI-syn1.0. To evaluate which genes or clusters of genes are important for infectivity of cell culture the RGD 1.0 strain series [7] was used to screen large section of the JCVI-syn1.0 genome for the genes necessary for persistence. The series is based on the JCVI-syn1.0 genome; each RGD 1.0 strain has $1/8^{th}$ of the genome minimized. For example, RGD 1.0-3 has $7/8^{th}$ of the genome from JCVI-syn1.0 and the 3rd segment of the genome from JCVI syn 2.0. This 8 strain series was used to systematically test each $1/8^{th}$ segment for the persistence phenotype. Testing each of the RGD 1.0 series showed only RGD 1.0-2 lost the persistence phenotype, meaning the deleted genes from segment 2 result in a loss of the persistence phenotype. We used our library of intermediate minimized strains to identify and test with an infectivity assay (**table 2**), a partially minimized strain that also lost the persistence phenotype, designated as syn1.0F. This strain's only difference between JCVI-syn1.0 is eight genes given in table 3. Every strains evaluated via the infectivity assay (five additional strains not reported) lacking these 8 genes resulted in a loss of the persistence phenotype.

Mycoplasma requires mammalian cells when grown in mammalian media

When mammalian cells are grown in culture they consume nutrients and secrete a number of different natural products and proteins. To determine if mammalian cells or

if their secreted products are required for *Mycoplasma* infection, a culture of either HeLa cells or HEK293 cells were grown for 2 d in fresh DMEM media supplemented with 10% FBS. Media was removed from the culture and filtere sterilized to produce spent media. Spent media was infected with synthetic *Mycoplasma* strains then incubated at 37 degrees. *Mycoplasma* contamination was evaluated by CCU assay. 7 d post infection phenol red present in the spent media did not indicate bacterial metabolism and a negative CCU assay determined that not only had bacterial growth arrested no surviving *Mycoplasma* was present in spent media 7 d post infection.

Validation of CCU assay with PCR

Although the CCU assay is the current gold standard for detecting Mycoplasmas in cell culture and requires only one living cell in order to obtain a positive result, it was possible that contamination from another source led to positive CCU assays. To validate that the CCU assay was detecting synthetic strains of *Mycoplasma*, a multiplex PCR assay was used to detect complete genomes of JCVI-syn1.0. Multiplex PCR amplification on eight sites distributed along the 1.1 Mb genome was based off a previously reported multiplex PCR protocol and primers for detecting only JCVI-syn1.0 and its derivatives [106]. Each primer set (**Supplementary table 1**) amplifies a segment at least 100 kb from any other set, spanning the entire 1.1 Mb genome. Syn1.0F is derived from JCVI-syn1.0 so, if present, every segment will be amplified. *Mycoplasma* was evaluated using multiplex primers just prior to passage on the 3rd day post infection; and just after passaging twice on the 7th day post infection. From previous results, we know that on the 7th day post infection cells that cannot persist in

230

culture will not result in a positive CCU assay but there are still enough cells present in the inoculum to result in a positive CCU assay on the 3rd day post infection. We observed JCVI-syn1.0 is detected 7 d post infection while Syn1.0F is not detected; identical results to the infectivity assay evaluated via CCU. **(figure C-1)**

DISCUSSION

The infectivity assay reported here can effectively evaluate if *Mycoplasma* contamination is present in a cell culture. Although the assay is nonspecific for both *Mycoplasma* species or strain, if used with known contaminants the assay can determine if Mycoplasma is present in mammalian cell culture. Both HEK293 cells and HeLa cells were identical when evaluated for *Mycoplasma* persistence, indicating that the genes necessary for the persistence phenotype is not specific for cell line. Further, the secreted products from either cell line are not sufficient for Mycoplasma persistence in spent medium.

We identified a cluster of eight genes which are necessary for *Mycoplasma mycoides* infection of mammalian cell lines. Although all of them might be necessary for *Mycoplasma* persistence there are three excellent candidates for the persistence phenotype. The putative lipoprotein MMSYN1_0180 and the unannotated transport and binding protein MMSYN1_0184 are the best candidates for infectivity.

Mycoplasma infection is likely the result of a number of actions involving several surface membrane components, allowing the pathogen to adhere tightly to the host cell surface and persist. Lipoproteins have been linked to interactions between Mycoplasmas and eukaryotic cells, particularly with respect to adhesion [364]. The putative maltose ABC transporter MMSYN1_0183 might also be involved with

infectivity [365]. A predicted maltose ABC transporter malF was found to be necessary for the persistence of *Mycoplasma galliseptium* strain in infected birds. Further evaluation is necessary to determine if the gene/s necessary persistence phenotype can be identified further.

This report confirms that the minimized synthetic Mycoplasma strains JCVI-syn2.0 and JCVI-syn3.0 lack the cellular machinery necessary for infecting and contaminating mammalian cell cultures. If these strains are to be used as a potential model to blueprint the genes necessary for life the extra precautions for limiting their use for labs concerned about potential contamination.

TABLE AND FIGURE LEGENDS

Table C-1 CCU assay of Synthetic *M. mycoides* persistence in HEK-293T cells.

An actively growing culture of JCVI-syn 1.0 or JCVI-syn 2.0 were used to contaminate a culture of HEK293T cells. At select points over 15 days a sample of supernatant was removed and a CCU assay was performed. Mycoplasma was detected (+) when a color change was identified. Passage number is given in parentheses. Pictures of the CCU assay can be seen in supplementary figure C-1.

Strain	1 d post infection (P0)	3 d post infection (P0)	5 d post infection (P1)	7 d post infection (P2)	11 d post infection (P3)	14 d post infection (P5)	15 d post infection (P5)
JCVI-syn 1.0	+	+	+	+	+	+	+
JCVI-syn 2.0	+	+	+	-	-	-	-
No infection	-	-	-	-	-	-	-

Table 2 Minimized synthetic Mycoplasma strains identified genes necessary for infection.

Actively growing cultures of synthetic Mycoplasma *mycoides* were used to contaminate either a culture of HEK293T cells; a culture of HeLa cells; fresh DMEM media supplemented with 10% FBS; filtere sterilized spent DMEM media supplemented with 10% FBS after being used as growth medium for 2 d for HEK293T cells; or HeLa cells. 7 d post infection a sample of supernatant or medium was removed

and a CCU assay was performed. Mycoplasma was detected (+) when a color change

was identified.

Strain	HEK293T	HeLa	unspent DMEM media	HEK293 spent media	HeLa spent media
JCVI-syn 1.0	+	+	-	-	-
JCVI-syn 2.0	-	-	-	-	-
JCVI-syn 3.0	-	-	-	-	-
JCVI-syn 3.0D	-	-	-	-	-
syn 1.0F	-	-	-	-	-
RDG 1.0-1	+	n/a	n/a	n/a	n/a
RDG 1.0-2	-	n/a	n/a	n/a	n/a
RDG 1.0-3	+	n/a	n/a	n/a	n/a
RDG 1.0-4	+	n/a	n/a	n/a	n/a
RDG 1.0-5	+	n/a	n/a	n/a	n/a
RDG 1.0-6	+	n/a	n/a	n/a	n/a
RDG 1.0-7	+	n/a	n/a	n/a	n/a
RDG 1.0-8	+	n/a	n/a	n/a	n/a

Figure C-1 Synthetic *M. mycoides* persistence in HEK-293T cells verified by multiplex PCR.

Using eight sets of multiplex primers shown in supplementary table 1 a PCR was

performed on the supernatant from a contaminated culture of HEK-293T cells from

JCVI-syn1.0 (syn1), syn1.0F (synF), or a culture of HEK293T cells with no

contamination (-). An actively growing bacterial cultures of each strain was used as a

positive control. A molecular ladder (L) with markers every 100 bases from 100 bp to

1000 bp is shown in both outside lanes.
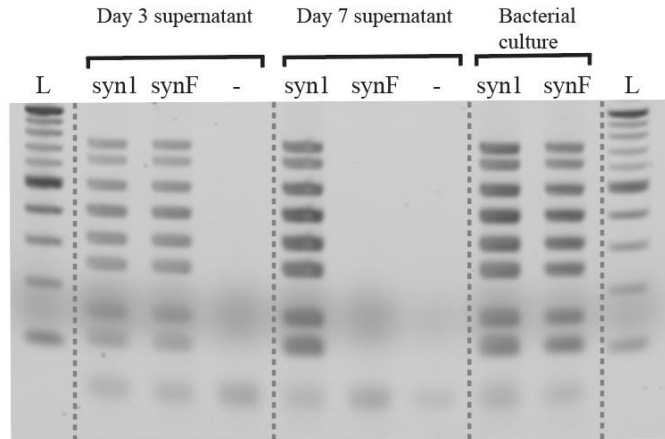
Table 3 Eight required genes necessary for persistence in mammalian cell lines.

A cluster of eight genes removed from JCVI-syn 1.0 to produce Syn1.0F.

Start	End	gene designation	Functional category	SGI annotation
236820	238610	MMSYN1_0180	Cell envelope	putative lipoprotein
238692	239183	MMSYN1_0181	NULL	hypothetical protein
239198	241615	MMSYN1_0182	Transport and binding proteins	maltosaccharide ABC transporter
241630	244170	MMSYN1_0183	Transport and binding proteins	maltose ABC transporter permease protein
244172	245257	MMSYN1_0184	Transport and binding proteins	Multiple
245365	247662	MMSYN1_0185	Central intermediary metabolism	glycosyl hydrolase, family 65
247664	249268	MMSYN1_0186	Energy metabolism	trehalose-6-phosphatehydrolase (Alpha,alpha-phosph

Supplementary Figure C-1 Synthetic *M. mycoides* growth does not affect morphology in HEK293T cells.

Micrograph images were taken using an inverted scope at 30x magnification.
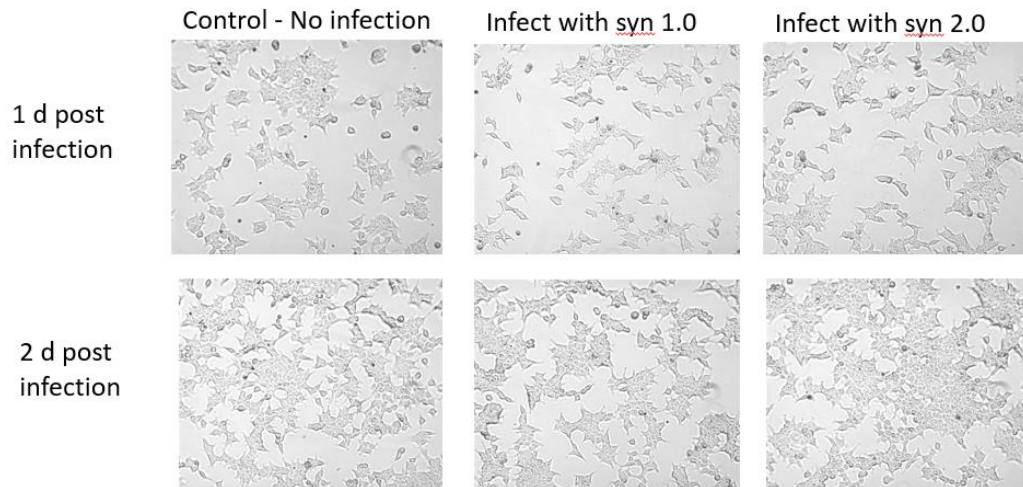
Supplementary figure C-2 CCU assay of synthetic *M. mycoides* persistence in HEK293T cells.

(A) 100 µl of an actively growing culture of JCVI-syn 1.0 (Syn 1.0), JCVI-syn 2.0 (Syn 2.0), or 100 µl of DMEM media (No infection) were used to contaminate a culture of HEK293T cells. At select points over 13 days a sample of supernatant was removed and a CCU assay was performed. Uninfected media is shown as a negative control to show the color without any bacterial growth. Each time point a 4x10x serial dilution was performed. Passage number is given in parentheses. (B) JCVI-Syn1.0 grows with HEK293T cells. A CCU assay was performed on day 14 and day 15 after a new passage (P5). 24 h after removing the sample images were taken of the CCU assay.

A

	Syn1.0	Syn2.0	No infection	Media only
1 d post infection (P0)				
3 d post infection (P0)				
5 d post infection (P1)				
7 d post infection (P2)				
11 d post infection (P3)				
13 d post infection (P4)				

B

day 14 (P5)   -1 -2 -3 -4 -5 -6 -7 -8

day 15 (P5)   -1 -2 -3 -4 -5 -6 -7 -8
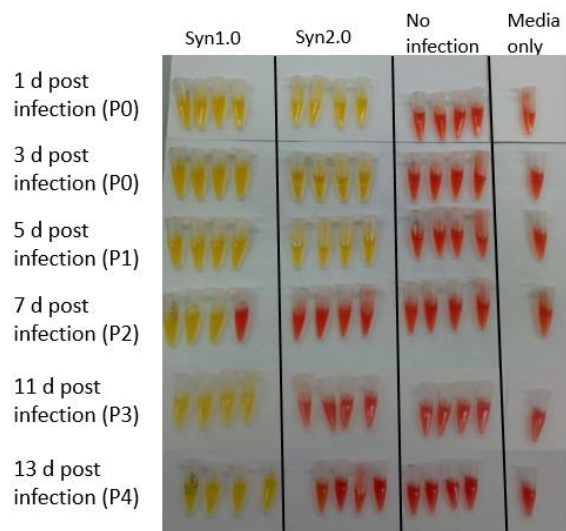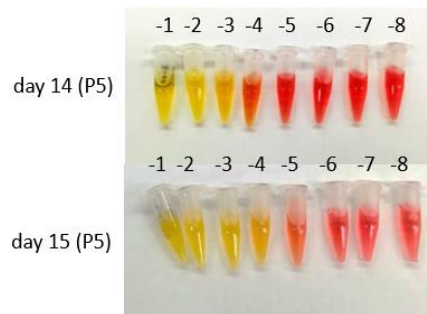
Appendix D: Design of Experiments (DOE)

In this appendix, we review DOE and how it can be applied to optimize a process using an example from Anderson and Witcomb [299] and how it can identify conditions of optimization that may not be able to be identified testing one factor at a time.

Design of experiments (DOE)

Design of experiments is a powerful statistical tool for two main reasons. First, it is a practical approach using the power of statistics to minimize experimentation while optimizing a process. Second, DOE can reveal potential interactions that could not have been identified studying the process one factor at a time (OFAT). Anderson and Witcomb [299] demonstrate the latter principle with example data seen in **Figures D-1, D-2 and D-3**. In the Anderson and Witcomb example, a researcher using a mid-value for factor A and B is given a response of 63. Hoping to increase that response the researcher performs optimization experiments OFAT. In **Figure D-1**, the researcher first optimizes factor A using the initial mid-level value for factor B, keeping factor B constant throughout the experiment. Optimizing for factor A, the response increases to 75 (red arrow). Next factor B is optimized keeping the 'optimum' value of factor A from the first experiment constant. **Figure D-2** shows the results when optimizing for factor B; the response level is increased to 82 (red arrow). Using OFAT this optimization experiment is now complete, with an increase in response from 63 to 82 after 18 runs. However, if every experimental condition possible were performed, a response surface can be generated. **Figure D-3** shows the response surface graph for

238

this example.   The response surface reveals a rising ridge graph with the actual

maximum value of 94 as indicated by the green arrow. The OFAT data points are also

plotted in **Figure D-3** showing how a researcher who is only performing OFAT can

miss the interactions between factors and the absolute maximum.
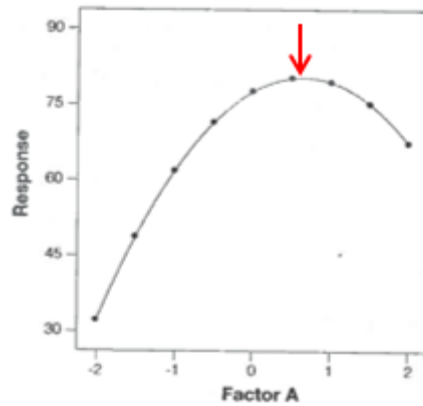
Figure D-1 OFAT plot after experimenting on factor A.

Adapted from [299]
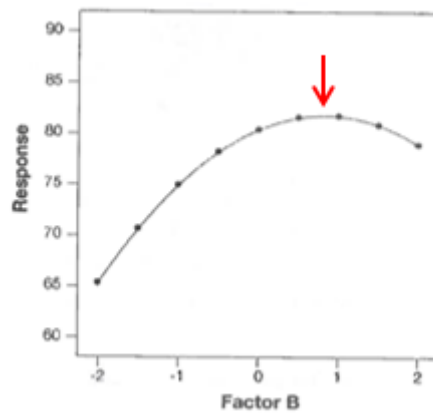
Figure D-2 OFAT plot after experimenting on factor B.

Adapted from [299]
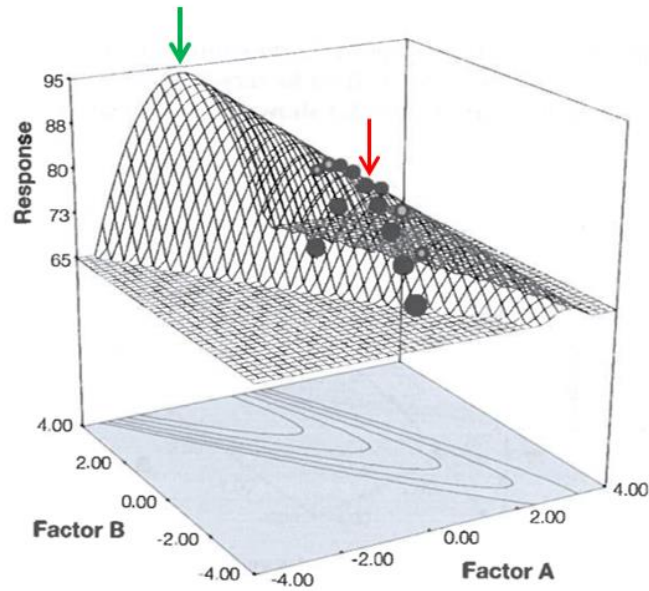
Figure D-3 OFAT points shown on true surface.

Red arrow indicates the maximum response generated from OFAT optimization. Green arrow indicates the absolute maximum across possible ranges. Adapted from [299]

One alternative to OFAT is to perform a full factorial design, where the entire range of possible values in the surface is explored. Commonly, a three-level full factorial design is used, meaning that for each factor three levels are considered; a low, mid-range and high value typically abbreviated as -1, 0, and +1. In the given example, this would result in 9 experimental runs, which would result in identifying a maximum response of 86 (**Table D-1**) when the ranges of the OFAT experiments are used from **Figure D-**

1 and D-2. Using a full factorial design results in a higher response and can be identified with fewer experimental runs. More importantly, the full response surface can be estimated using statistical software potentially revealing an even higher response point, which can be validated with few select runs.

Table D-1 Three level factorial design results.

A	B	Response
0	2	86
0	0	78
2	-2	73
-2	2	71
2	0	67
0	-2	53
2	2	45
-2	0	32
-2	-2	-23

While excellent at identifying the maximum response as well as being used to estimate the response surface, one problem with a three-level full factorial design is that as the number of factors grows the number of runs necessary to complete the design increases exponentially.   The number of experimental runs can be expressed mathematically as 3^k, where k is the number of factors. With just 5 factors being tested, the number of experimental runs quickly increases to 243.   If the process you wish to optimize requires a large number of factors, or if the number of factors affecting the response is unknown, the number of runs can quickly become unmanageable. One technique to reduce the number of runs is to use a fractional factorial design. As the name implies this design simply uses a fraction of runs from the full factorial design. In this design, a carefully chosen subset of experimental conditions are selected ahead of time and a

screening experiment can be performed to eliminate factors that have no or little effect on the process.

In 1946, Plackette and Burman [218] developed a powerful fractional factorial design technique that can be used to estimate the main effects of a process and substantially reduce the number of runs in a DOE experiment down to N experimental runs when N-1 is the number of factors being tested. Although extremely efficient from a practical perspective, a Plackette-Burman design can confound their results with two-factor interactions, meaning the results may not be able to determine if one factor is a main effect or if a two-factor interaction is causing the observed effect. For this reason, Plackette-Burman designs and other low resolution fractional factorial generally are usually used just for screening purposes.

Once trivial factors are eliminated using a fractional factorial design, a higher resolution design closer to a full factorial design can be used for optimization since the number of factors in a process is reduced. The design can then be analyzed using common statistical methods such as ANOVA and a regression analysis and validated using the predicted optimum conditions.

Appendix E: Sequences

E2 expression sequence:

CAGGCTTGAAGGAATTCGGTACCGCCGCCACCATGGCTGAGACAGCATGC
GAACGTTTACATGTAGCGCAAGAAACACAAATGCAGTTGATTGAGAAAA
GTAGTGATAAGTTGCAAGATCATATACTGTACTGGACTGCTGTTAGAACT
GAGAACACACTGCTTTATGCTGCAAGGAAAAAGGGGTGACTGTCCTAGG
ACACTGCAGAGTACCACACTCTGTAGTTTGTCAAGAGAGAGCCAAGCAGG
CCATTGAAATGCAGTTGTCTTTGCAGGAGTTAAGCAAAACTGAGTTTGGG
GATGAACCATGGTCTTTGCTTGACACAAGCTGGGACCGATATATGTCAGA
ACCTAAACGGTGCTTTAAGAAAGGCGCCAGGGTGGTAGAGGTGGAGTTTG
ATGGAAATGCAAGCAATACAAACTGGTACACTGTCTACAGCAATTTGTAC

242

ATGCGCACAGAGGACGGCTGGCAGCTTGCGAAGGCTGGGGCTGACGGAA
CTGGGCTCTACTACTGCACCATGGCCGGTGCTGGACGCATTTACTATTCTC
GCTTTGGTGACGAGGCAGCCAGATTTAGTACAACAGGGCATTACTCTGTA
AGAGATCAGGACAGAGTGTATGCTGGTGTCTCATCCACCTCTTCTGATTTT
AGAGATCGCCCAGACGGAGTCTGGGTCGCATCCGAAGGACCTGAAGGAG
ACCCTGCAGGAAAAGAAGCCGAGCCAGCCCAGCCTGTCTCTTCTTTGCTC
GGCTCCCCCGCCTGCGGTCCCATCAGAGCAGGCCTCGGTTGGGTACGGGA
CGGTCCTCGCTCGCACCCCTACAATTTTCCTGCAGGCTCGGGGGGCTCTAT
TCTCCGCTCTTCCTCCACCCCGGTGCAGGGCACGGTACCGGTGGACTTGG
CATCAAGGCAGGAAGAAGAGGAGCAGTCGCCCGACTCCACAGAGGAAGA
ACCAGTGACTCTCCCAAGGCGCACCACCAATGATGGATTCCACCTGTTAA
AGGCAGGAGGGTCATGCTTTGCTCTAATTTCAGGAACTGCTAACCAGGTA
AAGTGCTATCGCTTTCGGGTGAAAAAGAACCATAGACATCGCTACGAGAA
CTGCACCACCACCTGGTTCACAGTTGCTGACAACGGTGCTGAAAGACAAG
GACAAGCACAAATACTGATCACCTTTGGATCGCCAAGTCAAAGGCAAGA
CTTTCTGAAACATGTACCACTACCTCCTGGAATGAACATTTCCGGCTTTAC
AGCCAGCTTGGACTTCTGATCACTGCCATTGCCTTTTCT

TFAM expression sequence:

GGTGGCAGGGCTCTCTGGCTACTAGAGAACCCACTGCTTACTGGCTTATC
GAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGTTAAGCTT
GATCAAACAAGTTTGTACAAAAAAGCAGGCTTGAAGGAATTCGGTACCAT
GTTCGGTACCATGAGCGAAAGCAGGCAAACCATTTGAATGCGGTACCATG
CGACAGAAACCGTGCACATGAGCAGCGTGCTGGCCAGCTGCCCCAAGAA
GCCCGTGAGCAGCTACCTGAGATTCAGCAAGGAGCAGCTGCCCATCTTCA
AGGCCCAGAACCCCGACGCCAAGACCACCGAGCTGATCAGAAGAATCGC
CCAGAGATGGAGAGAGCTGCCCGACAGCAAGAAGAAGATCTACCAGGAC
GCCTACAGAGCCGAGTGGCAGGTGTACAAGGAGGAGATCAGCAGATTCA
AGGAGCAGCTGACCCCCAGCCAGATCATGAGCCTGGAGAAGGAGATCAT
GGACAAGCACCTGAAGAGAAAGGCCATGACCAAGAAGAAGGAGCTGACC
CTGCTGGGCAAGCCCAAGAGACCCAGAAGCGCCTACAACGTGTACGTGG
CCGAGAGATTCCAGGAGGCCAAGGGCGACAGCCCCCAGGAGAAGCTGAA
GACCGTGAAGGAGAACTGGAAGAACCTGAGCGACAGCGAGAAGGAGCTG
TACATCCAGCACGCCAAGGAGGACGAGACCAGATACCACAACGAGATGA
AGAGCTGGGAGGAGCAGATGATCGAGGTGGGCAGAAAGGACCTGCTGAG
AAGAACCATCAAGAAGCAGAGAAAGTACGGCGCCGAGGAGTGCCCCAAG
AAGAAGAGAAAGGTGTGATAGATAGTCTAGCCGCGCCCGACACTGCAAA
CCTTTACCTCGAGTGCGCGCTACGCTAGCGTGAGCAGGCGAGGAGATTAC
ATGTCATCATCAAGAAGTCATGCGCTCAAGTGCACATGAGGGCTCGTGAC
GTACGAGTTCAGATCGAAGTAAGGCCGAGGGACCGCCTCCTACGAGGGC
AAC

E2 Binding site sequence:

243

ACGGCTGATTACTCTTGTTGTAATAAAAATCACTTAATAGCAATGTGCTGT
GTCAGTTGTTTATTGGAACCACACCCGGTACACATCCTGTCCAGCATTTGC
AGTGCGTGCATTGAATTATTGTGCTGGCTAGACTTCATGGCGCCTGGCACC
GAATCCTGCCTTCTCAGCGAAAATGAATAATTGCTTTGTTGGCAAGAAAC
TAAGCATCAATGGGACGCGTGCAAAGCACCGGCGGCGGTAGATGCGGGG
TAAGTACTGAATTTTAATTCGACCTATCCCGGTAAAGCGAAAGCGACACG
CTTTTTTTTCACACATAGCGGGACCGAACACGTTATAAGTATCGATTAGGT
CTATTTTTGTCTCTCTGTCGGAACCAGAACTGGTAAAAGTTTCCATTGCGT
CTGGGCTTGTCTATCATTGCGTCTCTATGGTTTTTGGAGGATTAGACGGGG
CCACCAGTAATGGTGCATAGCGGATGTCTGTACCGCCATCGGTGCACCGA
TATAGGTTTGGGGCTCCCCAAGGGACTGCTGGGATGACAGCTTCATATTA
TATTGAATGGGCGCATAATCAGCTTAATTGGTGAGGACAAGCTACAAGTT
GTAACCTGATCTCCACAAAGTACGTTGCCGGTCGGGGTCAAACCGTCTTC
GGTGCTCGAAACCGCCTTAAACTACAGACAGGTCCCAGCCAAGTAGGCG
GATCAAAACCTCAAAAAGGCGGGAGCCAATCAAAATGCAGCATTATATTT
TAAGCTCACCGAAACCGGTAAGTAAAGACTATGTATTTTTTCCCACATAG
CACATTCATCGCATC

TFAM Binding site sequence:

TGGAAACCCCGACCTATTTTTCGCCGATGGTGCGATTACCGGGCTCTGCC
ATCTTAACATAGTAATGAACAATAGATAAGTGAAATAGCCTAACTTGAAA
CACCTTCATCTAGTTAAGCGACTAGTTTAATTCATGAGACTATTATCACTA
AGCCTAGATTATGAAAACTTTGCGGGGGGTCGCGAGTACGCTATCTTCTA
CCCATCCCCGACCAGATTTCAAATTTTATCTTTTGGCGGTATGCACTTTTA
ACAGTCACCCCCCAACTAACACATTATTTTCCCCTCCCACTCCCATACTAC
TAATCTCATCAATACAACCCCCGCCCATCCTACCCAGCACACACACACCG
CTGCTAACCCCATACCCCGAACCAACCAAACCCCAAAGACACCCCCCACA
TTTCCCTTGATAAAGCACCCGTTGCACGAGGTGCAATTGTGCTATCCTTCG
ACCGTTTATGTAGTAGTAATGACTCTATAGATAAGTGAGTTTAATTCATGC
TTTCATCTAGTTAAGGAGGTATTATCACTAACTGAATAGCCTAACTTGATG
TTAAGATGGCAGAGCCCGGTAATCGCAGGGCACCCTCCCCATTCTGGTAA
TCATCA

NF-κB binding site sequence

GGGGACTTTCCGGGGACTTTCCGGGGACTTTCCGGGGACTTTCCGGGGAC
TTTCC.

Bibliography

1.   Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age.* Nat
       Rev Genet, 2010. **11**(5): p. 367-79.

2.   Lienert, F., et al., *Synthetic biology in mammalian cells: next generation research tools and therapeutics.* Nature Reviews Molecular Cell Biology, 2014. **15**(2): p. 95-107.

3.   Gibson, D.G., et al., *Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome.* Science, 2008. **319**(5867): p. 1215-20.

4.   Gibson, D.G., et al., *Creation of a bacterial cell controlled by a chemically synthesized genome.* Science, 2010. **329**(5987): p. 52-6.

5.   Auslaender, S., et al., *Programmable single-cell mammalian biocomputers.* Nature, 2012. **487**(7405): p. 123-+.

6.   Benders, G.A., *Cloning whole bacterial genomes in yeast.* Methods Mol Biol, 2012. **852**: p. 165-80.

7.   Hutchison, C.A., 3rd, et al., *Design and synthesis of a minimal bacterial genome.* Science, 2016. **351**(6280): p. aad6253.

8.   Dormitzer, P.R., et al., *Synthetic generation of influenza vaccine viruses for rapid response to pandemics.* Sci Transl Med, 2013. **5**(185): p. 185ra68.

9.   Doherty, A.M.O. and E.M.C. Fisher, *Microcell-mediated chromosome transfer (MMCT): small cells with huge potential.* Mammalian Genome, 2003. **14**(9): p. 583-592.

10.   Glover, D.J., H.J. Lipps, and D.A. Jans, *Towards safe, non-viral therapeutic gene expression in humans.* Nat Rev Genet, 2005. **6**(4): p. 299-310.

11.   Miller, A.M. and D.A. Dean, *Tissue-specific and transcription factor-mediated nuclear entry of DNA.* Adv Drug Deliv Rev, 2009. **61**(7-8): p. 603-13.

12.   Gupta, B., T.S. Levchenko, and V.P. Torchilin, *Intracellular delivery of large molecules and small particles by cell-penetrating proteins and peptides.* Adv Drug Deliv Rev, 2005. **57**(4): p. 637-51.

13.   Yin, H., et al., *Non-viral vectors for gene-based therapy.* Nat Rev Genet, 2014. **15**(8): p. 541-55.

14.   Zhi, D., et al., *The headgroup evolution of cationic lipids for gene delivery.* Bioconjug Chem, 2013. **24**(4): p. 487-519.

15.   Lonez, C., M. Vandenbranden, and J.M. Ruysschaert, *Cationic lipids activate intracellular signaling pathways.* Adv Drug Deliv Rev, 2012. **64**(15): p. 1749-58.

16.   Zuhorn, I.S., R. Kalicharan, and D. Hoekstra, *Lipoplex-mediated transfection of mammalian cells occurs through the cholesterol-dependent clathrin-mediated pathway of endocytosis.* Journal of Biological Chemistry, 2002. **277**(20): p. 18021-18028.

17.   Akita, H., et al., *Quantitative three-dimensional analysis of the intracellular trafficking of plasmid DNA transfected by a nonviral gene delivery system using confocal laser scanning microscopy.* Molecular Therapy, 2004. **9**(3): p. 443-451.

18.   Colin, M., et al., *Cell delivery, intracellular trafficking and expression of an integrin-mediated gene transfer vector in tracheal epithelial cells.* Gene Therapy, 2000. **7**(2): p. 139-152.

19. Khalil, I.A., et al., *Uptake pathways and subsequent intracellular trafficking in nonviral gene delivery.* Pharmacological Reviews, 2006. **58**(1): p. 32-45.

20. Rejman, J., A. Bragonzi, and M. Conese, *Role of clathrin- and caveolae-mediated endocytosis in gene transfer mediated by lipo- and polyplexes.* Molecular Therapy, 2005. **12**(3): p. 468-474.

21. Zuhorn, I.S. and D. Hoekstra, *On the mechanism of cationic amphiphile-mediated transfection. To fuse or not to fuse: Is that the question?* Journal of Membrane Biology, 2002. **189**(3): p. 167-179.

22. Payne, C.K., et al., *Internalization and trafficking of cell surface proteoglycans and proteoglycan-binding ligands.* Traffic, 2007. **8**(4): p. 389-401.

23. Wang, T., J.R. Upponi, and V.P. Torchilin, *Design of multifunctional non-viral gene vectors to overcome physiological barriers: dilemmas and strategies.* Int J Pharm, 2012. **427**(1): p. 3-20.

24. Benjaminsen, R.V., et al., *The possible "proton sponge " effect of polyethylenimine (PEI) does not include change in lysosomal pH.* Mol Ther, 2013. **21**(1): p. 149-57.

25. Akinc, A., et al., *Exploring polyethylenimine-mediated DNA transfection and the proton sponge hypothesis.* Journal of Gene Medicine, 2005. **7**(5): p. 657-663.

26. Sonawane, N.D., F.C. Szoka, and A.S. Verkman, *Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes.* Journal of Biological Chemistry, 2003. **278**(45): p. 44826-44831.

27. Hsu, C.Y. and H. Uludag, *Nucleic-acid based gene therapeutics: delivery challenges and modular design of nonviral gene carriers and expression cassettes to overcome intracellular barriers for sustained targeted expression.* J Drug Target, 2012. **20**(4): p. 301-28.

28. Varkouhi, A.K., et al., *Endosomal escape pathways for delivery of biologicals.* J Control Release, 2011. **151**(3): p. 220-8.

29. Hoekstra, D., et al., *Gene delivery by cationic lipids: in and out of an endosome.* Biochemical Society Transactions, 2007. **35**: p. 68-71.

30. Zabner, J., et al., *Cellular and molecular barriers to gene transfer by a cationic lipid.* J Biol Chem, 1995. **270**(32): p. 18997-9007.

31. Lukacs, G.L., et al., *Size-dependent DNA mobility in cytoplasm and nucleus.* Journal of Biological Chemistry, 2000. **275**(3): p. 1625-1629.

32. Dowty, M.E., et al., *PLASMID DNA ENTRY INTO POSTMITOTIC NUCLEI OF PRIMARY RAT MYOTUBES.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(10): p. 4572-4576.

33. Stoffler, D., B. Fahrenkrog, and U. Aebi, *The nuclear pore complex: from molecular architecture to functional dynamics.* Current Opinion in Cell Biology, 1999. **11**(3): p. 391-401.

34. van der Aa, M.A., et al., *The nuclear pore complex: the gateway to successful nonviral gene delivery.* Pharm Res, 2006. **23**(3): p. 447-59.

35. Feldherr, C.M., D. Akin, and R.J. Cohen, *Regulation of functional nuclear pore size in fibroblasts.* Journal of Cell Science, 2001. **114**(24): p. 4621-4627.

36. Rout, M.P., et al., *The yeast nuclear pore complex: Composition, architecture, and transport mechanism.* Journal of Cell Biology, 2000. **148**(4): p. 635-651.

37. Bryan, A.K., et al., *Measurement of mass, density, and volume during the cell cycle of yeast.* Proc Natl Acad Sci U S A, 2010. **107**(3): p. 999-1004.

38. Wolff, D.A. and H. Pertoft, *SEPARATION OF HELA-CELLS BY COLLOIDAL SILICA DENSITY GRADIENT CENTRIFUGATION .1. SEPARATION AND PARTIAL SYNCHRONY OF MITOTIC CELLS.* Journal of Cell Biology, 1972. **55**(3): p. 579-&.

39. Fasbender, A., et al., *Effect of co-lipids in enhancing cationic lipid-mediated gene transfer in vitro and in vivo.* Gene Therapy, 1997. **4**(7): p. 716-725.

40. Fasbender, A., et al., *A low rate of cell proliferation and reduced DNA uptake limit cationic lipid-mediated gene transfer to primary cultures of ciliated human airway epithelia.* Gene Therapy, 1997. **4**(11): p. 1173-1180.

41. Zanta, M.A., P. Belguise-Valladier, and J.P. Behr, *Gene delivery: A single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(1): p. 91-96.

42. Uherek, C. and W. Wels, *DNA-carrier proteins for targeted gene delivery.* Advanced Drug Delivery Reviews, 2000. **44**(2-3): p. 153-166.

43. Melchior, F. and L. Gerace, *MECHANISMS OF NUCLEAR-PROTEIN IMPORT.* Current Opinion in Cell Biology, 1995. **7**(3): p. 310-318.

44. Akey, C.W., *VISUALIZATION OF TRANSPORT-RELATED CONFIGURATIONS OF THE NUCLEAR-PORE TRANSPORTER.* Biophysical Journal, 1990. **58**(2): p. 341-355.

45. Kiseleva, E., et al., *Active nuclear pore complexes in Chironomus: visualization of transporter configurations related to mRNP export.* Journal of Cell Science, 1998. **111**: p. 223-236.

46. Branden, L.J., B. Christensson, and C.I.E. Smith, *In vivo nuclear delivery of oligonucleotides via hybridizing bifunctional peptides.* Gene Therapy, 2001. **8**(1): p. 84-87.

47. Cartier, R. and R. Reszka, *Utilization of synthetic peptides containing nuclear localization signals for nonviral gene transfer systems.* Gene Therapy, 2002. **9**(3): p. 157-167.

48. Tachibana, R., et al., *Quantitative studies on the nuclear transport of plasmid DNA and gene expression employing nonviral vectors.* Advanced Drug Delivery Reviews, 2001. **52**(3): p. 219-226.

49. Nagasaki, T., et al., *Enhanced nuclear import and transfection efficiency of plasmid DNA using streptavidin-fused importin-beta.* Journal of Controlled Release, 2005. **103**(1): p. 199-207.

50. Akita, H., et al., *Cell cycle dependent transcription, a determinant factor of heterogeneity in cationic tipid-mediated transgene expression.* Journal of Gene Medicine, 2007. **9**(3): p. 197-207.

51. Itaka, K., et al., *In situ single cell observation by fluorescence resonance energy transfer reveals fast intra-cytoplasmic delivery and easy release of*

plasmid DNA complexed with linear polyethylenimine. Journal of Gene Medicine, 2004. **6**(1): p. 76-84.

52.   Bieber, T., et al., *Intracellular route and transcriptional competence of polyethylenimine-DNA complexes.* Journal of Controlled Release, 2002. **82**(2-3): p. 441-454.

53.   Brunner, S., et al., *Overcoming the nuclear barrier: Cell cycle independent nonviral gene transfer with linear polyethylenimine or electroporation.* Molecular Therapy, 2002. **5**(1): p. 80-86.

54.   de Semir, D., et al., *Non-viral vector-mediated uptake, distribution, and stability of chimeraptasts in human airway epithelial cells.* Journal of Gene Medicine, 2002. **4**(3): p. 308-322.

55.   Feng, M., D. Lee, and P. Li, *Intracellular uptake and release of poly (ethyleneimine)-co-poly(methyl methacrylate) nanoparticle/pDNA complexes for gene delivery.* International Journal of Pharmaceutics, 2006. **311**(1-2): p. 209-214.

56.   Wagstaff, K.M. and D.A. Jans, *Importins and Beyond: Non-Conventional Nuclear Transport Mechanisms.* Traffic, 2009. **10**(9): p. 1188-1198.

57.   Grosse, S., et al., *Which mechanism for nuclear import of plasmid DNA complexed with polyethylenimine derivatives?* Journal of Gene Medicine, 2006. **8**(7): p. 845-851.

58.   Mannisto, M., et al., *The role of cell cycle on polyplex-mediated gene transfer into a retinal pigment epithelial cell line.* Journal of Gene Medicine, 2005. **7**(4): p. 466-476.

59.   Chirico, G. and G. Baldini, *Rotational diffusion and internal motions of circular DNA .1. Polarized photon correlation spectroscopy.* Journal of Chemical Physics, 1996. **104**(15): p. 6009-6019.

60.   Fishman, D.M. and G.D. Patterson, *Light scattering studies of supercoiled and nicked DNA.* Biopolymers, 1996. **38**(4): p. 535-552.

61.   Seils, J. and R. Pecora, *DYNAMICS OF A 2311-BASE-PAIR SUPERHELICAL DNA IN DILUTE AND SEMIDILUTE SOLUTIONS.* Macromolecules, 1995. **28**(3): p. 661-673.

62.   Prazeres, D.M., *Prediction of diffusion coefficients of plasmids.* Biotechnol Bioeng, 2008. **99**(4): p. 1040-4.

63.   Robertson, R.M., S. Laib, and D.E. Smith, *Diffusion of isolated DNA molecules: Dependence on length and topology.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(19): p. 7310-7314.

64.   He, L. and B. Niemeyer, *A novel correlation for protein diffusion coefficients based on molecular weight and radius of gyration.* Biotechnology Progress, 2003. **19**(2): p. 544-548.

65.   Tyn, M.T. and T.W. Gusek, *PREDICTION OF DIFFUSION-COEFFICIENTS OF PROTEINS.* Biotechnology and Bioengineering, 1990. **35**(4): p. 327-338.

66.   Young, M.E., P.A. Carroad, and R.L. Bell, *ESTIMATION OF DIFFUSION-COEFFICIENTS OF PROTEINS.* Biotechnology and Bioengineering, 1980. **22**(5): p. 947-955.

67.   Dauty, E. and A.S. Verkman, *Actin cytoskeleton as the principal determinant of size-dependent DNA mobility in cytoplasm: a new barrier for non-viral gene delivery.* J Biol Chem, 2005. **280**(9): p. 7823-8.

68.   Capecchi, M.R., *HIGH-EFFICIENCY TRANSFORMATION BY DIRECT MICRO-INJECTION OF DNA INTO CULTURED MAMMALIAN-CELLS.* Cell, 1980. **22**(2): p. 479-488.

69.   Hagstrom, J.E., et al., *Nuclear import of DNA in digitonin-permeabilized cells.* Journal of Cell Science, 1997. **110**: p. 2323-2331.

70.   Zabner, J., et al., *Comparison of DNA-lipid complexes and DNA alone for gene transfer to cystic fibrosis airway epithelia in vivo.* J Clin Invest, 1997. **100**(6): p. 1529-37.

71.   Peterson, K.R., *Preparation of intact yeast artificial chromosome DNA for transgenesis of mice.* Nat Protoc, 2007. **2**(11): p. 3009-15.

72.   Lengsfeld, C.S. and T.J. Anchordoquy, *Shear-induced degradation of plasmid DNA.* J Pharm Sci, 2002. **91**(7): p. 1581-9.

73.   Bauchwitz, R. and F. Costantini, *YAC transgenesis: a study of conditions to protect YAC DNA from breakage and a protocol for transfection.* Biochim Biophys Acta, 1998. **1401**(1): p. 21-37.

74.   Rocchi, L., et al., *Escherichia coli-cloned CFTR loci relevant for human artificial chromosome therapy.* Hum Gene Ther, 2010. **21**(9): p. 1077-92.

75.   Noskov, V.N., et al., *Assembly of Large, High G plus C Bacterial DNA Fragments in Yeast.* Acs Synthetic Biology, 2012. **1**(7): p. 267-273.

76.   Karas, B.J., et al., *Direct transfer of whole genomes from bacteria to yeast.* Nature Methods, 2013. **10**(5): p. 410-+.

77.   Tagwerker, C., et al., *Sequence analysis of a complete 1.66 Mb Prochlorococcus marinus MED4 genome cloned in yeast.* Nucleic Acids Research, 2012. **40**(20): p. 10375-10383.

78.   Karas, B.J., et al., *Transferring whole genomes from bacteria to yeast spheroplasts using entire bacterial cells to reduce DNA shearing.* Nat Protoc, 2014. **9**(4): p. 743-50.

79.   Lartigue, C., et al., *Genome transplantation in bacteria: changing one species to another.* Science, 2007. **317**(5838): p. 632-8.

80.   Murphy, J.C., et al., *Compaction agent protection of nucleic acids during mechanical lysis.* Biotechnol Prog, 2006. **22**(2): p. 519-22.

81.   Prather, K.J., et al., *Industrial scale production of plasmid DNA for vaccine and gene therapy: plasmid design, production, and purification.* Enzyme and Microbial Technology, 2003. **33**(7): p. 865-883.

82.   Sang, Y., et al., *Salt ions and related parameters affect PEI-DNA particle size and transfection efficiency in Chinese hamster ovary cells.* Cytotechnology, 2015. **67**(1): p. 67-74.

83.   Pedraza, C.E., et al., *The importance of particle size and DNA condensation salt for calcium phosphate nanoparticle transfection.* Biomaterials, 2008. **29**(23): p. 3384-92.

84.   Gnirke, A., et al., *Microinjection of intact 200- to 500-kb fragments of YAC DNA into mammalian cells.* Genomics, 1993. **15**(3): p. 659-67.

85.   Giraldo, P. and L. Montoliu, *Size matters: use of YACs, BACs and PACs in transgenic animals.* Transgenic Res, 2001. **10**(2): p. 83-103.

86.   Alsaggar, M. and D. Liu, *Physical methods for gene transfer.* Adv Genet, 2015. **89**: p. 1-24.

87.   Wang, J., et al., *Gene gun-mediated oral mucosal transfer of interleukin 12 cDNA coupled with an irradiated melanoma vaccine in a hamster model: Successful treatment of oral melanoma and distant skin lesion.* Cancer Gene Therapy, 2001. **8**(10): p. 705-712.

88.   Cevc, G. and H. Richardsen, *Lipid vesicles and membrane fusion.* Advanced Drug Delivery Reviews, 1999. **38**(3): p. 207-232.

89.   Tamm, L.K., J. Crane, and V. Kiessling, *Membrane fusion: a structural perspective on the interplay of lipids and proteins.* Current Opinion in Structural Biology, 2003. **13**(4): p. 453-466.

90.   Lentz, B.R., *PEG as a tool to gain insight into membrane fusion.* Eur Biophys J, 2007. **36**(4-5): p. 315-26.

91.   Saez, R., et al., *DETERGENT-LIKE PROPERTIES OF POLYETHYLENEGLYCOLS IN RELATION TO MODEL MEMBRANES.* Febs Letters, 1982. **137**(2): p. 323-326.

92.   Boni, L.T., T.P. Stewart, and S.W. Hui, *ALTERATIONS IN PHOSPHOLIPID POLYMORPHISM BY POLYETHYLENE-GLYCOL.* Journal of Membrane Biology, 1984. **80**(1): p. 91-104.

93.   Maehara, T., et al., *Effect of Bacillus subtilis BsuM restriction-modification on plasmid transfer by polyethylene glycol-induced protoplast fusion.* Fems Microbiology Letters, 2011. **325**(1): p. 49-55.

94.   Baigori, M., et al., *TRANSFER OF PLASMIDS BETWEEN BACILLUS-SUBTILIS AND STREPTOCOCCUS-LACTIS.* Applied and Environmental Microbiology, 1988. **54**(5): p. 1309-1311.

95.   Dai, M.H., et al., *Visualization of protoplast fusion and quantitation of recombination in fused protoplasts of auxotrophic strains of Escherichia coli.* Metabolic Engineering, 2005. **7**(1): p. 45-52.

96.   Tarshis, M., M. Salman, and S. Rottem, *Fusion of mycoplasmas: the formation of cell hybrids.* FEMS microbiology letters, 1991. **66**(1): p. 67-71.

97.   Gibson, D.G., et al., *Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome.* Science, 2010. **329**(5987): p. 52-56.

98.   Gyuris, J. and E.G. Duda, *HIGH-EFFICIENCY TRANSFORMATION OF SACCHAROMYCES-CEREVISIAE CELLS BY BACTERIAL MINICELL PROTOPLAST FUSION.* Molecular and Cellular Biology, 1986. **6**(9): p. 3295-3297.

99.   Guerratschuschke, I., I. Martin, and M.T. Gonzalez, *POLYETHYLENE GLYCOL-INDUCED INTERNALIZATION OF BACTERIA INTO FUNGAL*

PROTOPLASTS - ELECTRON-MICROSCOPIC STUDY AND OPTIMIZATION OF EXPERIMENTAL CONDITIONS. Applied and Environmental Microbiology, 1991. **57**(5): p. 1516-1522.

100. Schaffner, W., *DIRECT TRANSFER OF CLONED GENES FROM BACTERIA TO MAMMALIAN-CELLS.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1980. **77**(4): p. 2163-2167.

101. Li, L. and T. Blankenstein, *Generation of transgenic mice with megabase-sized human yeast artificial chromosomes by yeast spheroplast-embryonic stem cell fusion.* Nature Protocols, 2013. **8**(8): p. 1567-1582.

102. Vansolingen, P. and J.B. Vanderplaat, *FUSION OF YEAST SPHEROPLASTS.* Journal of Bacteriology, 1977. **130**(2): p. 946-947.

103. Simpson, K. and C. Huxley, *A shuttle system for transfer of YACs between yeast and mammalian cells.* Nucleic Acids Res, 1996. **24**(23): p. 4693-9.

104. Markie, D., *A simple assay for optimizing yeast mammalian cell fusion conditions.* Molecular Biotechnology, 1996. **6**(2): p. 99-104.

105. Li, L. and T. Blankenstein, *Generation of transgenic mice with megabase-sized human yeast artificial chromosomes by yeast spheroplast-embryonic stem cell fusion.* Nat Protoc, 2013. **8**(8): p. 1567-82.

106. Lartigue, C., et al., *Creating bacterial strains from genomes that have been cloned and engineered in yeast.* Science, 2009. **325**(5948): p. 1693-6.

107. Karas, B.J., et al., *Assembly of eukaryotic algal chromosomes in yeast.* J Biol Eng, 2013. **7**(1): p. 30.

108. Kouprina, N. and V. Larionov, *Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology.* Chromosoma, 2016. **125**(4): p. 621-32.

109. Zhang, X.F., et al., *Transfer of an expression YAC into goat fetal fibroblasts by cell fusion for mammary gland bioreactor.* Biochem Biophys Res Commun, 2005. **333**(1): p. 58-63.

110. Huxley, C., et al., *The Human HPRT Gene on a Yeast Artificial Chromosome Is Functional When Transferred to Mouse Cells by Cell Fusion.* Genomics, 1991. **9**(4): p. 742-750.

111. Pachnis, V., et al., *Transfer of a yeast artificial chromosome carrying human DNA from Saccharomyces cerevisiae into mammalian cells.* Proc Natl Acad Sci U S A, 1990. **87**(13): p. 5109-13.

112. Julicher, K., et al., *Yeast artificial chromosome transfer into human renal carcinoma cells by spheroplast fusion.* Genomics, 1997. **43**(1): p. 95-8.

113. Simpson, K., A. McGuigan, and C. Huxley, *Stable episomal maintenance of yeast artificial chromosomes in human cells.* Molecular and Cellular Biology, 1996. **16**(9): p. 5117-5126.

114. Auriche, C., et al., *Functional human CFTR produced by a stable minichromosome.* EMBO Rep, 2002. **3**(9): p. 862-8.

115. Kazuki, Y. and M. Oshimura, *Human Artificial Chromosomes for Gene Delivery and the Development of Animal Models.* Molecular Therapy, 2011. **19**(9): p. 1591-1601.

251

116.   Ege, T. and N.R. Ringertz, *PREPARATION OF MICROCELLS BY ENUCLEATION OF MICRONUCLEATE CELLS.* Experimental Cell Research, 1974. **87**(2): p. 378-382.

117.   Meaburn, K.J., C.N. Parris, and J.M. Bridger, *The manipulation of chromosomes by mankind: the uses of microcell-mediated chromosome transfer.* Chromosoma, 2005. **114**(4): p. 263-274.

118.   Fournier, R.E. and F.H. Ruddle, *Microcell-mediated transfer of murine chromosomes into mouse, Chinese hamster, and human somatic cells.* Proc Natl Acad Sci U S A, 1977. **74**(1): p. 319-23.

119.   Abe, S., et al., *Localization of an hTERT repressor region on human chromosome 3p21.3 using chromosome engineering.* Genome integrity, 2010. **1**(1): p. 6-6.

120.   Qi, D.-L., et al., *Identification of PITX1 as a TERT Suppressor Gene Located on Human Chromosome 5.* Molecular and Cellular Biology, 2011. **31**(8): p. 1624-1636.

121.   Shinohara, T., et al., *Mice containing a human chromosome 21 model behavioral impairment and cardiac anomalies of Down's syndrome.* Human Molecular Genetics, 2001. **10**(11): p. 1163-1175.

122.   Devoy, A., et al., *Genomically humanized mice: technologies and promises.* Nature Reviews Genetics, 2012. **13**(1): p. 14-20.

123.   Hiratsuka, M., et al., *Retargeting of microcell fusion towards recipient cell-oriented transfer of human artificial chromosome.* Bmc Biotechnology, 2015. **15**.

124.   Ohzeki, J., et al., *Breaking the HAC Barrier: histone H3K9 acetyl/methyl balance regulates CENP-A assembly.* EMBO J, 2012. **31**(10): p. 2391-402.

125.   Co, D.O., et al., *Generation of transgenic mice and germline transmission of a mammalian artificial chromosome introduced into embryos by pronuclear microinjection.* Chromosome Research, 2000. **8**(3): p. 183-191.

126.   Kagawa, Y., Y. Inoki, and H. Endo, *Gene therapy by mitochondrial transfer.* Adv Drug Deliv Rev, 2001. **49**(1-2): p. 107-19.

127.   Niidome, T. and L. Huang, *Gene therapy progress and prospects: Nonviral vectors.* Gene Therapy, 2002. **9**(24): p. 1647-1652.

128.   Neumann, E., et al., *Gene transfer into mouse lyoma cells by electroporation in high electric fields.* Embo j, 1982. **1**(7): p. 841-5.

129.   Ibraheem, D., A. Elaissari, and H. Fessi, *Gene therapy and DNA delivery systems.* International Journal of Pharmaceutics, 2014. **459**(1-2): p. 70-83.

130.   Gehl, J., *Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research.* Acta Physiologica Scandinavica, 2003. **177**(4): p. 437-447.

131.   De Smedt, S.C., J. Demeester, and W.E. Hennink, *Cationic polymer based gene delivery systems.* Pharmaceutical Research, 2000. **17**(2): p. 113-126.

132.   Erbacher, P., J.S. Remy, and J.P. Behr, *Gene transfer with synthetic virus-like particles via the integrin-mediated endocytosis pathway.* Gene Therapy, 1999. **6**(1): p. 138-145.

133. Naldini, L., *Gene therapy returns to centre stage.* Nature, 2015. **526**(7573): p. 351-60.
134. Thomas, C.E., A. Ehrhardt, and M.A. Kay, *Progress and problems with the use of viral vectors for gene therapy.* Nature Reviews Genetics, 2003. **4**(5): p. 346-358.
135. Lukashev, A.N. and A.A. Zamyatnin, *Viral vectors for gene therapy: Current state and clinical perspectives.* Biochemistry-Moscow, 2016. **81**(7): p. 700-708.
136. Rappuoli, R. and P.R. Dormitzer, *Influenza: options to improve pandemic preparation.* Science, 2012. **336**(6088): p. 1531-3.
137. Clark, T.W., et al., *Trial of 2009 influenza A (H1N1) monovalent MF59-adjuvanted vaccine.* N Engl J Med, 2009. **361**(25): p. 2424-35.
138. Nicolson, C., et al., *Generation of influenza vaccine viruses on Vero cells by reverse genetics: an H5N1 candidate vaccine strain produced under a quality system.* Vaccine, 2005. **23**(22): p. 2943-2952.
139. Ozaki, H., et al., *Generation of high-yielding influenza A viruses in African green monkey kidney (Vero) cells by reverse genetics.* J Virol, 2004. **78**(4): p. 1851-7.
140. Chu, C., et al., *Conversion of MDCK cell line to suspension culture by transfecting with human siat7e gene and its application for influenza virus production.* Proc Natl Acad Sci U S A, 2009. **106**(35): p. 14802-7.
141. Kouprina, N., et al., *A new generation of human artificial chromosomes for functional genomics and gene therapy.* Cell Mol Life Sci, 2013. **70**(7): p. 1135-48.
142. Harrington, J.J., et al., *Formation of de novo centromeres and construction of first-generation human artificial microchromosomes.* Nature Genetics, 1997. **15**(4): p. 345-355.
143. Kazuki, Y., et al., *Refined human artificial chromosome vectors for gene therapy and animal transgenesis.* Gene Ther, 2011. **18**(4): p. 384-93.
144. Mejia, J.E., et al., *Efficiency of de novo centromere formation in human artificial chromosomes.* Genomics, 2002. **79**(3): p. 297-304.
145. Nakano, M., et al., *Inactivation of a human kinetochore by specific targeting of chromatin modifiers.* Dev Cell, 2008. **14**(4): p. 507-22.
146. Grimes, B.R., A.A. Rhoades, and H.F. Willard, *Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation.* Mol Ther, 2002. **5**(6): p. 798-805.
147. Iida, Y., et al., *Human artificial chromosome with a conditional centromere for gene delivery and gene expression.* DNA Res, 2010. **17**(5): p. 293-301.
148. Kouprina, N., et al., *Organization of synthetic alphoid DNA array in human artificial chromosome (HAC) with a conditional centromere.* ACS Synth Biol, 2012. **1**(12): p. 590-601.
149. Murray, A.W. and J.W. Szostak, *Construction of artificial chromosomes in yeast.* Nature, 1983. **305**(5931): p. 189-93.

150. Burke, D.T., G.F. Carle, and M.V. Olson, *Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors.* Science, 1987. **236**(4803): p. 806-12.

151. Goldman, S., *Genetic chemistry: production of non-native compounds in yeast.* Curr Opin Chem Biol, 2010. **14**(3): p. 390-5.

152. Lee, N.C., et al., *Protecting a transgene expression from the HAC-based vector by different chromatin insulators.* Cell Mol Life Sci, 2013. **70**(19): p. 3723-37.

153. Kononenko, A.V., et al., *Re-engineering an alphoid(tetO)-HAC-based vector to enable high-throughput analyses of gene function.* Nucleic Acids Res, 2013. **41**(10): p. e107.

154. Martella, A., et al., *Mammalian Synthetic Biology: Time for Big MACs.* ACS Synth Biol, 2016. **5**(10): p. 1040-1049.

155. Ohzeki, J.-i., et al., *CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA.* The Journal of Cell Biology, 2002. **159**(5): p. 765-775.

156. Shono, N., et al., *CENP-C and CENP-I are key connecting factors for kinetochore and CENP-A assembly.* J Cell Sci, 2015. **128**(24): p. 4572-87.

157. Marschall, P., N. Malik, and Z. Larin, *Transfer of YACs up to 2.3 Mb intact into human cells with polyethylenimine.* Gene Ther, 1999. **6**(9): p. 1634-7.

158. Gibson, D.G., et al., *One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20404-9.

159. Nair, H., et al., *Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis.* Lancet, 2011. **378**(9807): p. 1917-30.

160. Thompson, W.W., et al., *Estimates of US influenza-associated deaths made using four different methods.* Influenza Other Respir Viruses, 2009. **3**(1): p. 37-49.

161. Neumann, G. and Y. Kawaoka, *Reverse genetics of influenza virus.* Virology, 2001. **287**(2): p. 243-50.

162. de Jong, M.D. and T.T. Hien, *Avian influenza A (H5N1).* Journal of Clinical Virology, 2006. **35**(1): p. 2-13.

163. Olsen, C.W., *DNA vaccination against influenza viruses: a review with emphasis on equine and swine influenza.* Veterinary Microbiology, 2000. **74**(1-2): p. 149-164.

164. Webster, R.G., et al., *EVOLUTION AND ECOLOGY OF INFLUENZA-A VIRUSES.* Microbiological Reviews, 1992. **56**(1): p. 152-179.

165. Kiraly, J. and F. Kostolansky, *Reverse genetics and influenza virus research.* Acta Virologica, 2009. **53**(4): p. 217-224.

166. Ozawa, M., et al., *Contributions of two nuclear localization signals of influenza A virus nucleoprotein to viral replication.* J Virol, 2007. **81**(1): p. 30-41.

167.   Fukuyama, S. and Y. Kawaoka, *The pathogenesis of influenza virus infections: the contributions of virus and host factors.* Curr Opin Immunol, 2011. **23**(4): p. 481-6.

168.   Hoffmann, E., et al., *Eight-plasmid system for rapid generation of influenza virus vaccines.* Vaccine, 2002. **20**(25-26): p. 3165-3170.

169.   Kilbourne, E.D., *Future influenza vaccines and the use of genetic recombinants.* Bull World Health Organ, 1969. **41**(3): p. 643-5.

170.   Pleschka, S., et al., *A plasmid-based reverse genetics system for influenza A virus.* Journal of Virology, 1996. **70**(6): p. 4188-4192.

171.   Hoffmann, E. and R.G. Webster, *Unidirectional RNA polymerase I-polymerase II transcription system for the generation of influenza A virus from eight plasmids.* J Gen Virol, 2000. **81**(Pt 12): p. 2843-7.

172.   Pouton, C.W., et al., *Targeted delivery to the nucleus.* Advanced Drug Delivery Reviews, 2007. **59**(8): p. 698-717.

173.   McBride, A.A., M.G. McPhillips, and J.G. Oliveira, *Brd4: tethering, segregation and beyond.* Trends in Microbiology, 2004. **12**(12): p. 527-529.

174.   McBride, A.A., *The papillomavirus E2 proteins.* Virology, 2013. **445**(1-2): p. 57-79.

175.   Graham, S.V., *Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies.* Future Microbiology, 2010. **5**(10): p. 1493-1506.

176.   Brannon, A.R., et al., *Reconstitution of papillomavirus E2-mediated plasmid maintenance in Saccharomyces cerevisiae by the Brd4 bromodomain protein.* Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2998-3003.

177.   Salman, H., et al., *Nuclear localization signal peptides induce molecular delivery along microtubules.* Biophys J, 2005. **89**(3): p. 2134-45.

178.   Nagasaki, T., et al., *Can nuclear localization signals enhance nuclear localization of plasmid DNA?* Bioconjugate Chemistry, 2003. **14**(2): p. 282-286.

179.   Miyamoto, Y., et al., *Importin alpha can migrate into the nucleus in an importin beta- and Ran-independent manner.* Embo Journal, 2002. **21**(21): p. 5833-5842.

180.   Kakuk, A., et al., *Nuclear and nucleolar localization signals and their targeting function in phosphatidylinositol 4-kinase PI4K230.* Exp Cell Res, 2008. **314**(13): p. 2376-88.

181.   Vaysse, L., et al., *Nuclear-targeted minicircle to enhance gene transfer with non-viral vectors in vitro and in vivo.* Journal of Gene Medicine, 2006. **8**(6): p. 754-763.

182.   Collas, P. and P. Alestrom, *Rapid targeting of plasmid DNA to zebrafish embryo nuclei by the nuclear localization signal of SV40 T antigen.* Molecular Marine Biology and Biotechnology, 1997. **6**(1): p. 48-58.

183.   Arenal, A., et al., *The SV40 T antigen nuclear localization sequence enhances nuclear import of vector DNA in embryos of a crustacean (Litopenaeus schmitti).* Gene, 2004. **337**: p. 71-77.

184. Langle-Rouault, F., et al., *Up to 100-fold increase of apparent gene expression in the presence of Epstein-Barr virus oriP sequences and EBNA1: Implications of the nuclear import of plasmids.* Journal of Virology, 1998. **72**(7): p. 6181-6185.

185. Birbach, A., et al., *Cytosolic, nuclear and nucleolar localization signals determine subcellular distribution and activity of the NF-kappaB inducing kinase NIK.* J Cell Sci, 2004. **117**(Pt 16): p. 3615-24.

186. Emmott, E. and J.A. Hiscox, *Nucleolar targeting: the hub of the matter.* EMBO Rep, 2009. **10**(3): p. 231-8.

187. Behr, J.P., et al., *EFFICIENT GENE-TRANSFER INTO MAMMALIAN PRIMARY ENDOCRINE-CELLS WITH LIPOPOLYAMINE-COATED DNA.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(18): p. 6982-6986.

188. Cros, J.F., A. Garcia-Sastre, and P. Palese, *An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein.* Traffic, 2005. **6**(3): p. 205-13.

189. Kang, D., S.H. Kim, and N. Hamasaki, *Mitochondrial transcription factor A (TFAM): roles in maintenance of mtDNA and cellular functions.* Mitochondrion, 2007. **7**(1-2): p. 39-44.

190. Rubio-Cosials, A., et al., *Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter.* Nat Struct Mol Biol, 2011. **18**(11): p. 1281-9.

191. Campbell, C.T., J.E. Kolesar, and B.A. Kaufman, *Mitochondrial transcription factor A regulates mitochondrial transcription initiation, DNA packaging, and genome copy number.* Biochim Biophys Acta, 2012. **1819**(9-10): p. 921-9.

192. Keeney, P.M., et al., *Mitochondrial Gene Therapy Augments Mitochondrial Physiology in a Parkinson's Disease Cell Model.* Human Gene Therapy, 2009. **20**(8): p. 897-907.

193. Prasad, T.K. and N.M. Rao, *The role of plasmid constructs containing the SV40 DNA nuclear-targeting sequence in cationic lipid-mediated DNA delivery.* Cellular & Molecular Biology Letters, 2005. **10**(2): p. 203-215.

194. Dean, D.A., et al., *Sequence requirements for plasmid nuclear import.* Experimental Cell Research, 1999. **253**(2): p. 713-722.

195. Miller, A.M. and D.A. Dean, *Cell-specific nuclear import of plasmid DNA in smooth muscle requires tissue-specific transcription factors and DNA sequences.* Gene Therapy, 2008. **15**(15): p. 1107-1115.

196. van Gaal, E.V., et al., *DNA nuclear targeting sequences for non-viral gene delivery.* Pharm Res, 2011. **28**(7): p. 1707-22.

197. Zhou, R. and D.A. Dean, *Gene transfer of interieukin 10 to the murine cornea using electroporation.* Experimental Biology and Medicine, 2007. **232**(3): p. 362-369.

198. Tanimoto, M., et al., *No enhancement of nuclear entry by direct conjugation of a nuclear localization signal peptide to linearized DNA.* Bioconjugate Chemistry, 2003. **14**(6): p. 1197-1202.

199. Carlotti, F., et al., *Activation of nuclear factor kappa B in single living cells - Dependence of nuclear translocation and anti-apoptotic function on EGFPRELA concentration.* Journal of Biological Chemistry, 1999. **274**(53): p. 37941-37949.

200. Mesika, A., et al., *A regulated, NFkappaB-assisted import of plasmid DNA into mammalian cell nuclei.* Mol Ther, 2001. **3**(5 Pt 1): p. 653-7.

201. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.

202. Zhou, B., et al., *Engineering temperature sensitive live attenuated influenza vaccines from emerging viruses.* Vaccine, 2012. **30**(24): p. 3691-702.

203. Neumann, G., et al., *Generation of influenza A viruses entirely from cloned cDNAs.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 9345-50.

204. Zhou, B., et al., *Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses.* J Virol, 2009. **83**(19): p. 10309-13.

205. Kohlway, A., et al., *Hepatitis C virus RNA replication and virus particle assembly require specific dimerization of the NS4A protein transmembrane domain.* J Virol, 2014. **88**(1): p. 628-42.

206. Brown, D.M., et al., *Efficient size-independent chromosome delivery from yeast to cultured cell lines.* Nucleic Acids Res, 2016.

207. Christianson, T.W., et al., *Multifunctional yeast high-copy-number shuttle vectors.* Gene, 1992. **110**(1): p. 119-22.

208. Marschall, P., N. Malik, and Z. Larin, *Transfer of YACs up to 2.3 Mb intact into human cells with polyethylenimine.* Gene Ther, 1999. **6**(9): p. 1634-1637.

209. Davies, N.P., I.R. Rosewell, and M. Bruggemann, *Targeted alterations in yeast artificial chromosomes for inter-species gene transfer.* Nucleic Acids Res, 1992. **20**(11): p. 2693-8.

210. Larin, Z., A.P. Monaco, and H. Lehrach, *Yeast artificial chromosome libraries containing large inserts from mouse and human DNA.* Proc Natl Acad Sci U S A, 1991. **88**(10): p. 4123-7.

211. Kouprina, N. and V. Larionov, *Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae.* Nat Protoc, 2008. **3**(3): p. 371-7.

212. Noskov, V.N., et al., *Recombinase-mediated cassette exchange (RMCE) system for functional genomics studies in Mycoplasma mycoides.* Biol Proced Online, 2015. **17**: p. 6.

213. Zhang, X., et al., *A one-plasmid system to generate influenza virus in cultured chicken cells for potential use in influenza vaccine.* J Virol, 2009. **83**(18): p. 9296-303.

214. Basler, C.F., et al., *The Ebola virus VP35 protein inhibits activation of interferon regulatory factor 3.* J Virol, 2003. **77**(14): p. 7945-56.

215. Cardenas, W.B., et al., *Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling.* J Virol, 2006. **80**(11): p. 5168-78.

216. Larionov, V., et al., *Direct isolation of human BRCA2 gene by transformation-associated recombination in yeast.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7384-7.

217. Lee, N.C., V. Larionov, and N. Kouprina, *Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast.* Nucleic Acids Res, 2015. **43**(8): p. e55.

218. Plackett, R.L. and J.P. Burman, *The Design of Optimum Multifactorial Experiments.* Biometrika, 1946. **33**(4): p. 305-325.

219. Cochran, W.G. and G.M. Cox, *Experimental Designs*. Second Edi ed. 1957: Wiley.

220. Rieder, C.L. and R.E. Palazzo, *Colcemid and the mitotic cycle.* J Cell Sci, 1992. **102 (Pt 3)**: p. 387-92.

221. Skoufias, D.A., et al., *S-trityl-L-cysteine is a reversible, tight binding inhibitor of the human kinesin Eg5 that specifically blocks mitotic progression.* J Biol Chem, 2006. **281**(26): p. 17559-69.

222. Zieve, G.W., et al., *Production of large numbers of mitotic mammalian cells by use of the reversible microtubule inhibitor Nocodazole.* Experimental Cell Research, 1980. **126**(2): p. 397-405.

223. Stein, C.A., et al., *PHYSICOCHEMICAL PROPERTIES OF PHOSPHOROTHIOATE OLIGODEOXYNUCLEOTIDES.* Nucleic Acids Research, 1988. **16**(8): p. 3209-3221.

224. Chou, C.F. and M.B. Omary, *Mitotic arrest with anti-microtubule agents or okadaic acid is associated with increased glycoprotein terminal GlcNAc's.* J Cell Sci, 1994. **107 (Pt 7)**: p. 1833-43.

225. Anderson, M.J. and P.J. Whitcomb, *DOE Simplified: Practical Tools for Effective Experimentation*. Second Edi ed. 2007: Productivity Press.

226. Eliceiri, B., et al., *Stable integration and expression in mouse cells of yeast artificial chromosomes harboring human genes.* Proc Natl Acad Sci U S A, 1991. **88**(6): p. 2179-83.

227. Markie, D., et al., *NEW VECTOR FOR TRANSFER OF YEAST ARTIFICIAL CHROMOSOMES TO MAMMALIAN-CELLS.* Somatic Cell and Molecular Genetics, 1993. **19**(2): p. 161-169.

228. Reeves, R.H., W.J. Pavan, and P. Hieter, *MODIFICATION AND MANIPULATION OF MAMMALIAN DNA CLONED AS YACS.* Genetic Analysis-Biomolecular Engineering, 1990. **7**(5): p. 107-113.

229. Noskov, V.N., et al., *Assembly of large, high G+C bacterial DNA fragments in yeast.* ACS Synth Biol, 2012. **1**(7): p. 267-73.

230. Messerle, M., et al., *Cloning and mutagenesis of a herpesvirus genome as an infectious bacterial artificial chromosome.* Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14759-63.

231. Sheldrick, P., et al., *Infectious DNA from herpes simplex virus: infectivity of double-stranded and single-stranded molecules.* Proc Natl Acad Sci U S A, 1973. **70**(12): p. 3621-5.

232.   Desai, P., et al., *Mutations in herpes simplex virus type 1 genes encoding VP5 and VP23 abrogate capsid formation and cleavage of replicated DNA.* J Virol, 1993. **67**(3): p. 1357-64.

233.   Barber, G.N., *Innate immune DNA sensing pathways: STING, AIMII and the regulation of interferon production and inflammatory responses.* Curr Opin Immunol, 2011. **23**(1): p. 10-20.

234.   Ishikawa, H. and G.N. Barber, *STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling.* Nature, 2008. **455**(7213): p. 674-8.

235.   Zhang, Y., et al., *The DNA sensor, cyclic GMP-AMP synthase, is essential for induction of IFN-beta during Chlamydia trachomatis infection.* J Immunol, 2014. **193**(5): p. 2394-404.

236.   Karlin, S., E.S. Mocarski, and G.A. Schachtel, *Molecular evolution of herpesviruses: genomic and protein sequence comparisons.* J Virol, 1994. **68**(3): p. 1886-902.

237.   Tedesco, F.S., et al., *Stem cell-mediated transfer of a human artificial chromosome ameliorates muscular dystrophy.* Sci Transl Med, 2011. **3**(96): p. 96ra78.

238.   Suzuki, T., et al., *Highly Efficient Transfer of Chromosomes to a Broad Range of Target Cells Using Chinese Hamster Ovary Cells Expressing Murine Leukemia Virus-Derived Envelope Proteins.* PLoS One, 2016. **11**(6): p. e0157187.

239.   Liskovykh, M., et al., *Moving toward a higher efficiency of microcell-mediated chromosome transfer.* Mol Ther Methods Clin Dev, 2016. **3**: p. 16043.

240.   Luo, J., et al., *Adeno-associated virus-mediated cancer gene therapy: current status.* Cancer Lett, 2015. **356**(2 Pt B): p. 347-56.

241.   Guo, X. and L. Huang, *Recent Advances in Nonviral Vectors for Gene Delivery.* Accounts of Chemical Research, 2012. **45**(7): p. 971-979.

242.   Gaudet, D., et al., *Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPLS447X) gene therapy for lipoprotein lipase deficiency: an open-label trial.* Gene Therapy, 2013. **20**(4): p. 361-369.

243.   Cicalese, M.P., et al., *Update on the safety and efficacy of retroviral gene therapy for immunodeficiency due to adenosine deaminase deficiency.* Blood, 2016. **128**(1): p. 45-54.

244.   Shorter, S.A., et al., *The potential of toxin-based drug delivery systems for enhanced nucleic acid therapeutic delivery.* Expert Opin Drug Deliv, 2016: p. 1-12.

245.   Algire, M., R. Krishnakumar, and C. Merryman, *Megabases for kilodollars.* Nat Biotechnol, 2010. **28**(12): p. 1272-3.

246.   Mueller, K. and W. Weber, *Optogenetic tools for mammalian systems.* Molecular Biosystems, 2013. **9**(4): p. 596-608.

247.   Pathak, G.P., J.D. Vrana, and C.L. Tucker, *Optogenetic control of cell function using engineered photoreceptors.* Biology of the Cell, 2013. **105**(2): p. 59-72.

248. Kramer, B.P., C. Fischer, and M. Fussenegger, *BioLogic gates enable logical transcription control in mammalian cells.* Biotechnology and Bioengineering, 2004. **87**(4): p. 478-484.

249. Tigges, M., et al., *A tunable synthetic mammalian oscillator.* Nature, 2009. **457**(7227): p. 309-12.

250. Kramer, B.P., et al., *An engineered epigenetic transgene switch in mammalian cells.* Nature Biotechnology, 2004. **22**(7): p. 867-870.

251. Hacein-Bey-Abina, S., et al., *Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy.* N Engl J Med, 2002. **346**(16): p. 1185-93.

252. Hacein-Bey-Abina, S., et al., *LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1.* Science, 2003. **302**(5644): p. 415-419.

253. Hacein-Bey-Abina, S., et al., *A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency.* N Engl J Med, 2003. **348**(3): p. 255-6.

254. Hacein-Bey-Abina, S., et al., *Efficacy of gene therapy for X-linked severe combined immunodeficiency.* N Engl J Med, 2010. **363**(4): p. 355-64.

255. Muruve, D.A., *The innate immune response to adenovirus vectors.* Human Gene Therapy, 2004. **15**(12): p. 1157-1166.

256. Mingozzi, F. and K.A. High, *Immune responses to AAV in clinical trials.* Current Gene Therapy, 2007. **7**(5): p. 316-324.

257. Lundstrom, K. and T. Boulikas, *Viral and non-viral vectors in gene therapy: Technology development and clinical trials.* Technology in Cancer Research & Treatment, 2003. **2**(5): p. 471-485.

258. Lanctot, C., et al., *Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions.* Nature Reviews Genetics, 2007. **8**(2): p. 104-115.

259. Batenchuk, C., et al., *Chromosomal Position Effects Are Linked to Sir2-Mediated Variation in Transcriptional Burst Size.* Biophysical Journal, 2011. **100**(10): p. L56-L58.

260. Clarke, L. and J. Carbon, *ISOLATION OF A YEAST CENTROMERE AND CONSTRUCTION OF FUNCTIONAL SMALL CIRCULAR CHROMOSOMES.* Nature, 1980. **287**(5782): p. 504-509.

261. Ratzkin, B. and J. Carbon, *Functional expression of cloned yeast DNA in Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(2): p. 487-491.

262. Hitzeman, R.A., L. Clarke, and J. Carbon, *Isolation and characterization of the yeast 3-phosphoglycerokinase gene (PGK) by an immunological screening technique.* J Biol Chem, 1980. **255**(24): p. 12073-80.

263. Hinnen, A., J.B. Hicks, and G.R. Fink, *Transformation of yeast.* Proc Natl Acad Sci U S A, 1978. **75**(4): p. 1929-33.

264. Boehmer, P.E. and I.R. Lehman, *Herpes simplex virus DNA replication.* Annu Rev Biochem, 1997. **66**: p. 347-84.

265. Oh, J. and N.W. Fraser, *Temporal association of the herpes simplex virus genome with histone proteins during a lytic infection.* J Virol, 2008. **82**(7): p. 3530-7.

266. Pignatti, P.F. and E. Cassai, *Analysis of herpes simplex virus nucleoprotein complexes extracted from infected cells.* J Virol, 1980. **36**(3): p. 816-28.

267. Cohen, G.H., et al., *Structural analysis of the capsid polypeptides of herpes simplex virus types 1 and 2.* J Virol, 1980. **34**(2): p. 521-31.

268. Cereghini, S. and M. Yaniv, *Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication.* Embo j, 1984. **3**(6): p. 1243-53.

269. Apostolou, E. and K. Hochedlinger, *Chromatin dynamics during cellular reprogramming.* Nature, 2013. **502**(7472): p. 462-471.

270. Jeong, S. and A. Stein, *Micrococcal nuclease digestion of nuclei reveals extended nucleosome ladders having anomalous DNA lengths for chromatin assembled on non-replicating plasmids in transfected cells.* Nucleic Acids Research, 1994. **22**(3): p. 370-375.

271. Herrera, F.J. and S.J. Triezenberg, *VP16-dependent association of chromatin-modifying coactivators and underrepresentation of histones at immediate-early gene promoters during herpes simplex virus infection.* J Virol, 2004. **78**(18): p. 9689-96.

272. Kent, J.R., et al., *During lytic infection herpes simplex virus type 1 is associated with histones bearing modifications that correlate with active transcription.* J Virol, 2004. **78**(18): p. 10178-86.

273. Deshmane, S.L. and N.W. Fraser, *During latency, herpes simplex virus type 1 DNA is associated with nucleosomes in a chromatin structure.* J Virol, 1989. **63**(2): p. 943-7.

274. Kubat, N.J., et al., *The Herpes Simplex Virus Type 1 Latency-Associated Transcript (LAT) Enhancer/rcr Is Hyperacetylated during Latency Independently of LAT Transcription.* Journal of Virology, 2004. **78**(22): p. 12508-12518.

275. Dressler, G.R., D.L. Rock, and N.W. Fraser, *Latent herpes simplex virus type 1 DNA is not extensively methylated in vivo.* J Gen Virol, 1987. **68 (Pt 6)**: p. 1761-5.

276. Tang, Y., et al., *Widespread Existence of Cytosine Methylation in Yeast DNA Measured by Gas Chromatography/Mass Spectrometry.* Analytical chemistry, 2012. **84**(16): p. 7249-7255.

277. Featherstone, T. and C. Huxley, *Extrachromosomal maintenance and amplification of yeast artificial chromosome DNA in mouse cells.* Genomics, 1993. **17**(2): p. 267-78.

278. Utvik, J.K., A. Nja, and K. Gundersen, *DNA injection into single cells of intact mice.* Human Gene Therapy, 1999. **10**(2): p. 291-300.

279. McGuigan, A. and C. Huxley, *Replication of yeast DNA and novel chromosome formation in mouse cells.* Nucleic Acids Research, 1996. **24**(12): p. 2271-2280.

280. Nonet, G.H. and G.M. Wahl, *INTRODUCTION OF YACS CONTAINING A PUTATIVE MAMMALIAN REPLICATION ORIGIN INTO MAMMALIAN-CELLS CAN GENERATE STRUCTURES THAT REPLICATE AUTONOMOUSLY.* Somatic Cell and Molecular Genetics, 1993. **19**(2): p. 171-192.

281. Bergmann, J.H., et al., *HACking the centromere chromatin code: insights from human artificial chromosomes.* Chromosome Research, 2012. **20**(5): p. 505-519.

282. Kim, J.-H., et al., *Human artificial chromosome (HAC) vector with a conditional centromere for correction of genetic deficiencies in human cells.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(50): p. 20048-20053.

283. Zhu, Z., et al., *Use of the tetracycline-controlled transcriptional silencer (tTS) to eliminate transgene leak in inducible overexpression transgenic mice.* J Biol Chem, 2001. **276**(27): p. 25222-9.

284. Nagpal, H. and T. Fukagawa, *Kinetochore assembly and function through the cell cycle.* Chromosoma, 2016. **125**(4): p. 645-59.

285. Garty, H., et al., *Sorbitol permease: an apical membrane transporter in cultured renal papillary epithelial cells.* Am J Physiol, 1991. **260**(5 Pt 2): p. F650-6.

286. Bednenko, J., G. Cingolani, and L. Gerace, *Nucleocytoplasmic transport: Navigating the channel.* Traffic, 2003. **4**(3): p. 127-135.

287. Gorlich, D., *Nuclear protein import.* Current Opinion in Cell Biology, 1997. **9**(3): p. 412-419.

288. Macara, I.G., *Transport into and out of the nucleus.* Microbiology and Molecular Biology Reviews, 2001. **65**(4): p. 570-+.

289. Arnold, W.N. and R.G. Garrison, *Isolation and Characterization of Protoplasts from Saccharomyces rouxii.* Journal of Bacteriology, 1979. **137**(3): p. 1386-1394.

290. Taddei, A., H. Schober, and S.M. Gasser, *The Budding Yeast Nucleus.* Cold Spring Harbor Perspectives in Biology, 2010. **2**(8): p. a000612.

291. Taddei, A. and S.M. Gasser, *Structure and function in the budding yeast nucleus.* Genetics, 2012. **192**(1): p. 107-29.

292. Smoyer, C.J. and S.L. Jaspersen, *Breaking down the wall: the nuclear envelope during mitosis.* Current Opinion in Cell Biology, 2014. **26**: p. 1-9.

293. Kupke, T., et al., *Targeting of Nbp1 to the inner nuclear membrane is essential for spindle pole body duplication.* Embo j, 2011. **30**(16): p. 3337-52.

294. Bie, X., et al., *Screening the Main Factors Affecting Extraction of the Antimicrobial Substance from Bacillus sp. fmbJ using the Plackett–Burman Method.* World Journal of Microbiology and Biotechnology, 2005. **21**(6): p. 925-928.

295. Chen, X.C., et al., *Medium optimization for the production of cyclic adenosine 3',5'-monophosphate by Microbacterium sp. no. 205 using response surface methodology.* Bioresour Technol, 2009. **100**(2): p. 919-24.

296. Borges, E.R. and N. Pereira, Jr., *Succinic acid production from sugarcane bagasse hemicellulose hydrolysate by Actinobacillus succinogenes.* J Ind Microbiol Biotechnol, 2011. **38**(8): p. 1001-11.

297. Salehi, Z., et al., *Statistical medium optimization and biodegradative capacity of Ralstonia eutropha toward p-nitrophenol.* Biodegradation, 2010. **21**(4): p. 645-57.

298. Li, C., et al., *Optimization of a cultural medium for bacteriocin production by Lactococcus lactis using response surface methodology.* J Biotechnol, 2002. **93**(1): p. 27-34.

299. Patrick J. Whitcomb, M.J.A., *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments*. 2004: Productivity Press. 304.

300. Tamm, C., et al., *Fast and Efficient Transfection of Mouse Embryonic Stem Cells Using Non-Viral Reagents.* Stem Cell Reviews and Reports, 2016. **12**(5): p. 584-591.

301. Miller, A.D., *RETROVIRUS PACKAGING CELLS.* Human Gene Therapy, 1990. **1**(1): p. 5-14.

302. Mykhaylyk, O., et al., *Generation of magnetic nonviral gene transfer agents and magnetofection in vitro.* Nature Protocols, 2007. **2**(10): p. 2391-2411.

303. Matsui, H., et al., *Loss of binding and entry of liposome-DNA complexes decreases transfection efficiency in differentiated airway epithelial cells.* Journal of Biological Chemistry, 1997. **272**(2): p. 1117-1126.

304. Campeau, E., et al., *A Versatile Viral System for Expression and Depletion of Proteins in Mammalian Cells.* Plos One, 2009. **4**(8): p. 18.

305. Hofmann, A., G.P. Nolan, and H.M. Blau, *Rapid retroviral delivery of tetracycline-inducible genes in a single autoregulatory cassette.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(11): p. 5185-5190.

306. Tomanin, R., et al., *Development and characterization of a binary gene expression system based on bacteriophage T7 components in adenovirus vectors.* Gene, 1997. **193**(2): p. 129-140.

307. Flowers, J.L. and L.J. Vaillancourt, *Parameters affecting the efficiency of Agrobacterium tumefaciens-mediated transformation of Colletotrichum graminicola.* Current Genetics, 2005. **48**(6): p. 380-388.

308. Zhao, C., et al., *Lac-regulated system for generating adenovirus 5 vaccine vectors expressing cytolytic human immunodeficiency virus 1 genes.* J Virol Methods, 2009. **160**(1-2): p. 101-10.

309. Robinson, L.B. and R.H. Wichelhausen, *Contamination of human cell cultures by pleuropneumonialike organisms.* Science, 1956. **124**(3232): p. 1147-8.

310. Olarerin-George, A.O. and J.B. Hogenesch, *Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI's RNA-seq archive.* Nucleic Acids Res, 2015. **43**(5): p. 2535-42.

311. Miller, C.J., et al., *Mycoplasma infection significantly alters microarray gene expression profiles.* Biotechniques, 2003. **35**(4): p. 812-4.

312. Tizzoni, M., et al., *Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm.* Bmc Medicine, 2012. **10**.

313. Robertson, J.S., et al., *The development of vaccine viruses against pandemic A(H1N1) influenza.* Vaccine, 2011. **29**(9): p. 1836-1843.

314. Verity, E.E., et al., *Rapid generation of pandemic influenza virus vaccine candidate strains using synthetic DNA.* Influenza and Other Respiratory Viruses, 2012. **6**(2): p. 101-109.

315. Strecker, T., et al., *Exploring synergies between academia and vaccine manufacturers: a pilot study on how to rapidly produce vaccines to combat emerging pathogens.* Clinical Chemistry and Laboratory Medicine, 2012. **50**(7): p. 1275-1279.

316. Doroshenko, A. and S.A. Halperin, *Trivalent MDCK cell culture-derived influenza vaccine Optaflu (R) (Novartis Vaccines).* Expert Review of Vaccines, 2009. **8**(6): p. 679-688.

317. Fraser, C., et al., *Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings.* Science, 2009. **324**(5934): p. 1557-1561.

318. (CDC), C.f.D.C.a.P., *Swine influenza A (H1N1) infection in two children – Southern California, March-April 2009.* MMWR Morb. Mortal. Wkly. Rep., 2009. **58**(400-402).

319. http://platform.gisaid.org.

320. Zhou, B., et al., *Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses.* Journal of Virology, 2009. **83**(19): p. 10309-10313.

321. Okuno, Y., et al., *RAPID FOCUS REDUCTION NEUTRALIZATION TEST OF INFLUENZA-A AND INFLUENZA-B VIRUSES IN MICROTITER SYSTEM.* Journal of Clinical Microbiology, 1990. **28**(6): p. 1308-1313.

322. Gibson, D.G., et al., *Chemical synthesis of the mouse mitochondrial genome.* Nat Methods, 2010. **7**(11): p. 901-3.

323. Suphaphiphat, P., et al., *Mutations at positions 186 and 194 in the HA gene of the 2009 H1N1 pandemic influenza virus improve replication in cell culture and eggs.* Virology Journal, 2010. **7**.

324. Hoffmann, E., et al., *A DNA transfection system for generation of influenza A virus from eight plasmids.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 6108-13.

325. Heix, J. and I. Grummt, *SPECIES-SPECIFICITY OF TRANSCRIPTION BY RNA-POLYMERASE-I.* Current Opinion in Genetics & Development, 1995. **5**(5): p. 652-656.

326. Wang, Z. and G.M. Duke, *Cloning of the canine RNA polymerase I promoter and establishment of reverse genetics for influenza A and B in MDCK cells.* Virol J, 2007. **4**: p. 102.

327. Suphaphiphat, P., et al., *Human RNA polymerase I-driven reverse genetics for influenza a virus in canine cells.* J Virol, 2010. **84**(7): p. 3721-5.

328. Bottcher, E., et al., *Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and HAT from human airway epithelium.* Journal of Virology, 2006. **80**(19): p. 9896-9898.

329. Katz, J.M., M.L. Wang, and R.G. Webster, *DIRECT SEQUENCING OF THE HA GENE OF INFLUENZA (H3N2) VIRUS IN ORIGINAL CLINICAL-SAMPLES REVEALS SEQUENCE IDENTITY WITH MAMMALIAN CELL-GROWN VIRUS.* Journal of Virology, 1990. **64**(4): p. 1808-1811.

330. Fulvini, A.A., et al., *Gene Constellation of Influenza A Virus Reassortants with High Growth Phenotype Prepared as Seed Candidates for Vaccine Production.* Plos One, 2011. **6**(6).

331. Klimov, A., et al., *Influenza virus titration, antigenic characterization, and serological methods for antibody detection.* Methods in molecular biology (Clifton, N.J.), 2012. **865**: p. 25-51.

332. Dormitzer, P.R., T.F. Tsai, and G. Del Giudice, *New technologies for influenza vaccines.* Human Vaccines & Immunotherapeutics, 2012. **8**(1): p. 45-58.

333. Matthews, J.T., *Egg-based production of influenza vaccine: 30 years of commercial experience.* The Bridge-Washington-National Academy of Engineering, 2006. **36**(3): p. 17-24.

334. DNA2.0 Inc., *Gene Synthesis. [Online]. Available: https://www.dna20.com/services/gene-synthesis. [2015, January 26].* 2015.

335. Drake, J.W., et al., *Rates of spontaneous mutation.* Genetics, 1998. **148**(4): p. 1667-86.

336. New England BioLabs Inc., *What is the error rate of Phusion® High-Fidelity DNA Polymerase? [Online]. Available: https://www.neb.com/faqs/2012/09/06/what-is-the-error-rate-of-phusion-reg-high-fidelity-dna-polymerase. [2015, January 29].* 2015.

337. Hoffmann, E., et al., *Eight-plasmid system for rapid generation of influenza virus vaccines.* Vaccine, 2002. **20**(25-26): p. 3165-70.

338. Hutchison, C.A., 3rd, et al., *Cell-free cloning using phi29 DNA polymerase.* Proc Natl Acad Sci U S A, 2005. **102**(48): p. 17332-6.

339. Takahashi, H., et al., *Cell-free cloning using multiply-primed rolling circle amplification with modified RNA primers.* Biotechniques, 2009. **47**(1): p. 609-15.

340. Woyke, T., et al., *Decontamination of MDA reagents for single cell whole genome amplification.* PLoS One, 2011. **6**(10): p. e26161.

341. Dean, F.B., et al., *Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification.* Genome Res, 2001. **11**(6): p. 1095-9.

342. Xie, P., *Dynamics of DNA polymerase I (Klenow fragment) under external force.* J Mol Model, 2013. **19**(3): p. 1379-89.

343. Fijalkowska, I.J., R.M. Schaaper, and P. Jonczyk, *DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair.* FEMS Microbiol Rev, 2012. **36**(6): p. 1105-21.

344. Cline, J., J.C. Braman, and H.H. Hogrefe, *PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases.* Nucleic Acids Res, 1996. **24**(18): p. 3546-51.

345. Eckert, K.A. and T.A. Kunkel, *DNA polymerase fidelity and the polymerase chain reaction.* PCR Methods Appl, 1991. **1**(1): p. 17-24.

346. Venter, J.C., *Life at the Speed of Light: From the Double Helix to the Dawn of Digital Life*. 2013, New York, New York: Penguin Group. LLC.

347. Kamtekar, S., et al., *Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29.* Mol Cell, 2004. **16**(4): p. 609-18.

348. Truniger, V., et al., *Phi 29 DNA polymerase requires the N-terminal domain to bind terminal protein and DNA primer substrates.* J Mol Biol, 1998. **278**(4): p. 741-55.

349. Zhang, D.Y., et al., *Detection of rare DNA targets by isothermal ramification amplification.* Gene, 2001. **274**(1-2): p. 209-16.

350. Esteban, J.A., M. Salas, and L. Blanco, *Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization.* J Biol Chem, 1993. **268**(4): p. 2719-26.

351. Paez, J.G., et al., *Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification.* Nucleic Acids Res, 2004. **32**(9): p. e71.

352. de Vega, M., et al., *Improvement of phi29 DNA polymerase amplification performance by fusion of DNA binding motifs.* Proc Natl Acad Sci U S A, 2010. **107**(38): p. 16506-11.

353. Nikfarjam, L. and P. Farzaneh, *Prevention and detection of Mycoplasma contamination in cell culture.* Cell J, 2012. **13**(4): p. 203-12.

354. Davies, G. and W.C. Read, *The use of bacteriocidal agents in the primary isolation of Mycoplasma mycoides.* J Comp Pathol, 1969. **79**(1): p. 121-5.

355. Bruchmuller, I., et al., *Introduction of a validation concept for a PCR-based Mycoplasma detection assay.* Cytotherapy, 2006. **8**(1): p. 62-9.

356. Tola, S., et al., *Rapid and specific detection of Mycoplasma agalactiae by polymerase chain reaction.* Vet Microbiol, 1996. **51**(1-2): p. 77-84.

357. Check, E., *Venter aims for maximum impact with minimal genome.* Nature, 2002. **420**(6914): p. 350.

358. Fraser, C.M., et al., *The minimal gene complement of Mycoplasma genitalium.* Science, 1995. **270**(5235): p. 397-403.

359. Thiaucourt, F. and A. Di Maria, *A new microtitration method for the enumeration of contagious bovine pleuropneumonia (CBPP) vaccines.* Biologicals, 1992. **20**(1): p. 11-3.

360. Dedieu, L. and V. Balcer-Rodrigues, *Viable Mycoplasma mycoides ssp. mycoides small colony-mediated depression of the bovine cell responsiveness to the mitogen concanavalin A.* Scand J Immunol, 2006. **64**(4): p. 376-81.

361. Rweyemamu, M.M., et al., *Contagious bovine pleuropneumonia vaccines: the need for improvements.* Rev Sci Tech, 1995. **14**(3): p. 593-601.

362. Mulongo, M., et al., *Vaccination of cattle with the N terminus of LppQ of Mycoplasma mycoides subsp. mycoides results in type III immune complex disease upon experimental infection.* Infect Immun, 2015. **83**(5): p. 1992-2000.

363. Benders, G.A., et al., *Cloning whole bacterial genomes in yeast.* Nucleic Acids Res, 2010. **38**(8): p. 2558-69.
364. Pilo, P., J. Frey, and E.M. Vilei, *Molecular mechanisms of pathogenicity of Mycoplasma mycoides subsp. mycoides SC.* Vet J, 2007. **174**(3): p. 513-21.
365. Tseng, C.W., et al., *MalF is essential for persistence of Mycoplasma gallisepticum in vivo.* Microbiology, 2013. **159**(Pt 7): p. 1459-70.