Circular RNA: Design Criteria for Optimal Therapeutical Utility

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:41121328">http://nrs.harvard.edu/urn-3:HUL.InstRepos:41121328</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Circular RNA: Design Criteria for Optimal Therapeutic Utility

A dissertation presented

by

Ellese Marie Carmona

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

January 2019
Circular RNA: Design Criteria for Optimal Therapeutic Utility

Abstract

A growing body of work has shown mRNA-based therapeutics are a promising new potential class of drugs, despite the longstanding belief that RNA was too unstable to be efficiently utilized for such applications. Previous studies have established several RNA modifications (e.g., modified nucleotides, cap analogs, polyadenylation, and codon optimization) that enable production of therapeutic levels of protein. Although the established modifications function, in part, by enhancing the stability of RNA, none does so by providing protection against the large family of exonucleases that are primarily responsible for RNA degradation. Thus, we examined whether exonuclease-resistant, circular RNA (circRNA) could be utilized as a superior alternative to conventional, linear RNA. As there are significant differences between linear and circular RNA there was a need to establish appropriate design criteria specific to circRNA. Therefore, we concentrated our work on three areas of focus, the synthesis, immunity, and translation efficiency specifications of circRNA. Regarding synthesis, we have developed a methodology in which large, circRNAs can be efficiently and reliably generated. We showed that the incorporation of a complement-reverse complement (CRC) motif into circRNA and HPLC-mediated purification respectively improve ligation efficiency and drastically reduce immune activation. Additionally, we have identified optimal regulatory motifs that, when incorporated into circRNA, yield enhanced RNA stability, robust protein translation, and minimal immune stimulation both in vitro and in vivo. Lastly, our proof-of-concept in vivo studies using circRNA encoding erythropoietin (EPO) showed significantly longer protein expression and elevated reticulocyte levels compared to standard, modified linear mRNA. Together, this work offers details regarding the design criteria
necessary for circRNA to function well in the therapeutic setting and provides the first *in vivo*-based evidence for the utility of circRNA as a beneficial modification for enhancing the stability and overall efficacy of mRNA therapeutics.
Acknowledgements

I would first like to acknowledge my dissertation advisory committee (Shoba Vasudevan, Lee Gehrke and Paul Anderson) for their help and insight these past 5 years as well as my thesis committee (Lee Gehrke – Chair, Frank Slack, Richard Gregory, and Claire Moore) for allowing me to share my work with them. I would also like to thank my P.I. Michael Goldberg for all his help and advice over my graduate career.

I would not have been in any way qualified or equipped for graduate school without the immense help I received during my time at San Diego State University. The Compact Scholar Program, which has been a part of my life since 7th grade helped me, and countless others stay on track to go to college and continued to help me and other first-generation college students thrive at the university level. As I discovered my interest in science at SDSU, I was helped along this career path by the Minority Access to Research Careers (MARC) Program and the fantastic people that run it (Thelma Chavez and Cathie Atkins). I would not have been a qualified candidate for graduate school if it weren’t for the help of this program. Finally, by far the most impactful relationship I had during my time at State was with Dr. Kelly Doran, my undergraduate PI and mentor. I am forever grateful for the time I got to spend in her lab.

Once I entered graduate school the BBS program administrators (Kate Hodgins, Maria Bollinger, Danny, Anne O’ Shea) made the transition much less daunting and I am very grateful for their frank and informative answers to any of the questions I’ve had throughout graduate school. My BBS class was amazing and down to earth and made grad school an unforgettable experience.

I would like to acknowledge the Cancer Immunology & Virology (CIV) department at DFCI and the incredibly supportive administrative team (Charles Thomas & Kathleen Kennedy).
I would also like to acknowledge the students (Ashley and Abigail) I had the pleasure of teaching and also learning from during my time in the Goldberg lab. Ashley and Abigail both helped tremendously with this project, and I am very grateful for their input and willingness to work hard during our time together. I had no experience with animal work before starting graduate school, but the associates at DFCI’s Animal Resource Facility (ARF) were incredibly helpful in getting my in vivo work up and running (Carmen Da Silva & Catherine Sypher).

I would like to acknowledge my friends and family for their continued support and love. Ashley, Simone, Erin, Alyssa, Vera, my Uncle Jerry, and my brothers (James & Matthew): you guys have helped keep things in perspective for me and kept me calm when things got especially stressful in lab. Marcus – thank you for letting me take over your space as I wrote my thesis and being nothing but supportive since the day we met. Finally, I would like to acknowledge and thank my parents, George and Marie Carmona, for continually encouraging me in my academic pursuits and sacrificing so much to help get me where I am today. Thank you and I love you.
Publications


List of Figures

Figure 1: T4 RNA ligase 1-mediated ligation is more efficient than splint-mediated ligation. ..10
Figure 2: CRC motif enhances single-stranded RNA ligation efficiency/. .......................... 13
Figure 3: CRC motif functionality is enhanced with optimal sequence length & composition. .. 16
Figure 4: Optimized CRC motif does not induce immunity or effect translation efficiency .... 18
Figure 5: RppH is the superior enzyme for obtaining the 5’ monophosphate end .................... 21
Figure 6: HPLC purification is an efficient strategy for increasing RNA functionality. ........... 22
Figure 7: Poly(A) resistance assay can confirm circularization and aids in HPLC isolation..... 24
Figure 8: The optimized outline for RNA ligase I enzymatic ligation & confirmatory assays. .. 25
Figure 9: Comparison of optimized enzymatic ligation to newly developed PIE method....... 28
Figure 10: EMCV IRES produces robust levels of translation across multiple systems ........ 35
Figure 11: 3’ UTR Motifs can significantly affect the translation of circRNA..................... 37
Figure 12: Circularization nor modified nucleotide incorporation affect cellular entry of RNA 38
Figure 13: m6A is not a potent inducer of protein translation in circRNA........................... 41
Figure 14: Incorporation of modified NT into IRES-driven circRNA ablate translation ....... 42
Figure 15: CircRNA extends protein translation in vitro ..................................................... 44
Figure 16: Modified nucleotides have similar ablation effect on CVB IRES.......................... 47
Figure 17: Modified nucleotides prevent immune induction when incorporated into circRNA. 53
Figure 18: Purification method can modulate circRNA immune induction.......................... 55
Figure 19: RIG-I is potently induced by RNA but more so in PIE-generated circRNA....... 57
Figure 20: HPLC-associated reduction in immune induction is observed in primary cells .... 58
Figure 21: HPLC is sufficient to reduce cellular toxicity in circRNA-transfected cells........ 60
Figure 22: Circularization confers protection against serum and cell lysate degradation..... 61
Figure 23: CircRNA persists in vitro and in vivo significantly longer than linear RNA ....... 63
Figure 24: Initial characterization and confirmation of circRNA persistence in vivo.......... 71
Figure 25: CircRNA translates protein robustly in vivo & extends protein expression........ 72
Figure 26: Circularization extends Epo RNA kinetics in vivo........................................... 74
Figure 27: Physiological response derived from Epo-encoded circRNA.......................... 75
Figure 28: CircRNA recapitulates its proposed endogenous function as a molecular sponge ... 79

List of Tables

Table 1: IRES Panel & Source Information............................................................................ 34
Table 2: Mammalian RNA-Sensing Immune Receptors...................................................... 50
List of Supplemental Figures

Supplemental Figure S1: Secondary structure and alignments for 400, 500, 600 NT RNAs....87
Supplemental Figure S2: Secondary structure of 300-1300 NT RNA Panel of Figure 2E......88
Supplemental Figure S3: Alternative HPLC sequences for larger or heavily modified RNAs.89
Supplemental Figure S4: IRES panel controls .................................................................90
Supplemental Figure S5: Celigo cell counts .................................................................91
Supplemental Figure S6: mRNA in vivo optimization .....................................................91
Supplemental Figure S7: Protein expression kinetics in vivo........................................92
Supplemental Figure S8: Confirmation of Epo expression in vitro.................................92

List of Supplemental Tables

Supplemental Table S1: Vector Maps..................................................................................93
Supplemental Table S2: qPCR Primer List .........................................................................106
Supplemental Table S3: General Primer List ......................................................................105

Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>circRNA</td>
<td>Circular RNA, RNA that has been covalently linked at its 5’ and 3’ end</td>
</tr>
<tr>
<td>NLuc</td>
<td>Nanoluciferase</td>
</tr>
<tr>
<td>IVT</td>
<td>In vitro transcription</td>
</tr>
<tr>
<td>HPLC</td>
<td>High Perform Liquid Chromatography</td>
</tr>
<tr>
<td>PIE</td>
<td>Permuted Intron-Exon, autocatalytic form of circularization</td>
</tr>
<tr>
<td>RppH</td>
<td>RNA Pyrophosphohydrolase</td>
</tr>
<tr>
<td>CRC</td>
<td>Complement-reverse complement</td>
</tr>
<tr>
<td>NT</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>Kb</td>
<td>Kilobases</td>
</tr>
<tr>
<td>IRES</td>
<td>Internal ribosomal entry site</td>
</tr>
<tr>
<td>ψ</td>
<td>Pseudouridine</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methyl Cytidine</td>
</tr>
<tr>
<td>m6A</td>
<td>N6-Methyladenosine</td>
</tr>
<tr>
<td>2TU</td>
<td>2-thio Uridine</td>
</tr>
<tr>
<td>BLI</td>
<td>Bioluminescence Imaging</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>IVIS</td>
<td>In vivo imaging system</td>
</tr>
<tr>
<td>L+</td>
<td>Linear mRNA that has been capped and tailed</td>
</tr>
<tr>
<td>L-</td>
<td>Linear mRNA that have no additional modification (no cap, tail, or NT mods)</td>
</tr>
<tr>
<td>UNTR</td>
<td>Untransfected</td>
</tr>
</tbody>
</table>
Chapter 1: Synthesizing Circular RNA

Contributions: All experiments were conceived of by Ellese Carmona and Michael Goldberg and performed by Ellese Carmona. Ashley Deonarain aided in CRC motif optimization experiments. Chun Gwon Park aided in HPLC optimization.

1.1: Introduction

Discovery of endogenous circular RNA and initial characterizations

Prior to 1990, knowledge related to the existence of circular RNA (circRNA) was limited to the genomes of certain plant viroids (Sanger, 1976), a small satellite virus (Kos, 1986) and, separately, specific noncoding regions found in lower unicellular eukaryotes and archaea (Cech, 1982; Kjems, 1988). As sequencing methodology had not yet been fully developed, scientists utilized alternative experimentation to confirm their RNAs were in fact covalently linked at the 5’ and 3’ ends, such as electron microscopy, stringent sequential nested PCR reactions, phosphodiesterase, and 5’-3’ labeling resistance assays.

The first evidence that circRNA existed in higher eukaryotes, including humans and rats, was in a study that described the phenomenon of “scrambled exons” derived from the DCC gene where its transcripts contained exons in an order different from their sequential placement in the genome (Nigro, 1991). It was quickly realized that these scrambled products were circularized RNA (Cocquerelle, 1992; Cocquerelle, 1993; Capel, 1993) and initially dismissed as aberrant splicing products with no cellular function. But, even within the context of these early studies there were common characteristics noted that are still synonymous with circRNA to this day, including a propensity to be derived from genes with long introns that contain inverted repeats or regions of
complementarity, localization primarily in the cytoplasm despite their lack of poly(A) tails, and a much longer half-life than their linear counterparts.

It was not until the advent of next-generation sequencing that the ubiquitous nature of circRNA expression was realized and even then, they were initially overlooked. In most cases sequencing samples are poly(A)+ enriched pools or most sequencing algorithms are designed to select for reads that reflect the order of the reference genomic sequence. Based on either criteria circRNA reads would have been excluded from the analysis (reviewed in Jeck & Sharpless, 2014 and Wilusz & Sharp, 2018). Although methodology for non-polyA pools to be enriched and analyzed had been developed (Suzuki, 2006), it was not until a study attempting to identify transcripts derived from chromosomal rearrangements that the actual extent of circRNAs in humans was realized (Salzman, 2012). This study found thousands of scrambled reads that were derived from 1 or more exons, poly(A) negative, and exonuclease resistance providing strong evidence that these reads were derived from circRNAs. Through the subsequent sequencing-based studies, it was shown that circRNAs are pervasive among all branches of life, including worms and flies, (Ivanov, 2015; Westholm, 2014), mammals (Jeck, 2013; Memczak, 2013; Salzman, 2012; Salzman, 2013), plants (Barret, 2015), bacteria (Broadbent, 2015), and archaea (Danan, 2012). Furthermore, the global analysis of circRNAs providing additional broad characterization, like most circRNAs are derived from 1 or more exons, most circRNAs are not conserved across species, and the majority of circRNAs are expressed at significantly lower levels than their linear counterparts despite their reported stability.

With the thousands of putative circRNAs that have been identified across all systems of life the conclusion that these molecules were just aberrant splicing products began to be questioned. As such, a burst of studies aiming to determine how these circRNAs are generated and what
possible function they may serve in the cell has been initiated. Although this area of study is still in its infancy much has already been learned about this novel class of RNAs.

**Biogenesis of circRNA**

A back-splicing event characterize the general mechanism behind circRNA biogenesis. Meaning that instead of an upstream 5’ splice site (ss) interacting with a downstream 3’ss the interaction is reversed; a downstream 5’ss interacts and splices with an upstream 3’ss. Initial genomic characterizations analyzing sequences known to generate circRNA showed that the flanking intronic regions often contained areas of high complementarity through motifs like inverted repeats or Alu repeat elements. These motifs would have the effect of looping out the intron and bringing distant splice sites near to each other, which would presumably aid in disordered splicing. As additional evidence for these intronic regions playing a role in circularization, when these repeats sequences are deleted the corresponding circRNAs were no longer generated (Jeck, 2013).

Beside the cis-elements that might promote circularization, particular trans-factors have been shown to aid in this mechanism as well. Quaking (QK-1), an alternative splicing factor, has been shown to promote circularization by binding to flanking intronic region that would bring the distal splice sites close together and that this factor plays a role in generating circular RNAs specifically during human epithelial to mesenchymal transition (Conn, 2015). Conversely, DHX9, an RNA helicase, binds explicitly to dsRNA and has been shown to inhibit circRNA biogenesis (Li, 2017).

The lack of a consensus sequence necessary for circularization and the variety of cis and trans-acting factors that have been identified to aid or inhibit circRNA generation suggests that synthesis of these molecules is dynamic and reliant on the current state of the cell, referring to its
stage in development or even its stage in cell division. (Guo, 2014; Ye, 2017). As more work sheds light on the mechanisms behind circRNA synthesis, the potential for useful information that could be applied to developing circRNA as a therapeutic tool increases as well.

**Current knowledge of circRNAs’ functions in the cell**

Of the thousands of putative circRNAs identified, only a small handful (<1%) have an associated function. The most established functional role for circRNA at this point is as a microRNA sponge that can aid in the regulation of translation. This function was first identified in the circRNAs Sry (found in mice) and CDR1as (found in humans), with each found to contain 16 and 70 binding sites for miR-138 and miR-7, respectively (Hansen, 2013; Memczak, 2013). Further studies again showed the abundance of circRNA, specifically CDR1as, in neuronal tissue and were the first to show a phenotypic effect when the circRNA was knocked out in mice (Piwecka, 2017).

A separate function for circRNA as translatable transcripts has recently been described in two settings: human (Legnini, 2017) and flies (Pamadurti, 2017), offering a divergent finding from many previous sequencing studies that could not detect target circRNAs associated with ribosomes. CircRNAs lack both the necessary components for canonical cellular translation -- a 5’ cap and 3’ poly(A) tail -- suggesting an alternative mechanism of translation. Indeed, both studies provided evidence that each circRNA contained motifs with IRES activity. Additionally, N6-Methyladenosine (m6A) nucleotide modifications have recently been shown to be a novel way for eukaryotes to undergo translation in a cap-independent manner. While synthetic RNAs have been shown to translate in this manner (Yang, 2017) no studies yet have shown whether endogenous circRNA use this mechanism to translate protein as part of their function.
Additional studies have begun to provide evidence for circRNA’s regulatory function within plant nuclei (Conn, 2017) and potentially through cell-to-cell communication for those localized in exosomes (Preusser, 2018). As new data emerges, those circRNAs with more apparent cellular function will continue to be revealed. However, at this time it is unclear if the studies reported thus far are the exception or the rule to circRNA function.

*circRNAs’ Role in Disease*

The accumulation or misregulation of circRNA in specific tissues has been associated with disease progression (cancer, cardiovascular disease, Alzheimer’s) as well as aging. In the case of cancer, particular genomic translocations can generate oncogenic fusion circRNAs that promote transformation and resistance to therapy (Guarnerio, 2016). Like other RNA classes, it appears that in disease states circRNA levels can also become dysregulated and methods are currently being developed to utilize particular circRNAs as a prognostic tool or biomarker for associated diseases. As an example, there is a clinical trial that has been initiated aiming to use the circRNA MICRA as a prognostic tool for cardiac disease (Vausort, 2016).

*circRNA: A New Therapeutic Tool?*

Characterizations of circRNA will undoubtedly continue at a rapid pace but even what has been gleaned so far provides support for the potential therapeutic utility of circRNA in a variety of modalities. By far, its most attractive quality is its notable stability over linear RNA. Furthermore, its capacity to translate protein and also provide regulatory function makes the potential applications for this molecule immense. In order to test the efficacy of circRNA in a therapeutic setting, we must be able to generate exogenous circRNA at the bench efficiently. To that end, we studied the existing methodologies for RNA ligation and their associated advantages and limitations.
Chemically-synthesized circRNA

Chemical reactions utilizing cyanogen bromide can be used to generate the phosphodiester bond necessary to link the 5’ and 3’ end of RNA molecules (Dolinnaya, 1991). Separately, solid-phase synthesis has also been used to generate short circular RNAs. However, unlike DNA oligos, the 2’OH present in RNA give rise to the possibility of a 2’-5’ phosphodiester bond and that likelihood increases as the RNA being synthesized gets longer (Micura, 1999). Even then normal synthesis for these two forms of chemical ligation are only suitable for RNAs less than 50 nucleotides in length and, moreover, ligation efficiency is only 15% (Micura, 1999). While this method may be useful for the study of short RNAs, we could think of no feasible manipulation that would allow for this method to be adaptable for synthesizing large circular RNAs. As such, this is the only methodology we did not attempt during the duration of this project.

Permuted Intron Exon (PIE) ligation

Group I Introns describe a particular type of autocatalytic RNA circularization found to occur in lower eukaryotes, bacteria, and higher plants. This mechanism involves an intronic region of RNA with a highly specific sequence and secondary structure that enables a two-step transesterification reaction that only requires a GTP donor group and cations (usually Mg2+) to produce a covalently-linked intron RNA circle (Kruger., 1982). In the early 1990s, two groups independently manipulated the Group I introns found in Anabaena (Puttaraju, 1992) and T4 bacteriophage (Ford, 1993) to synthesize exonic circular RNA, a method that is now known as permuted intron exon (PIE) ligation. By permuting the intron such that it flanked the exon, they were able to generate an RNA circle comprised of the exon instead of the intron. This method was limited to very small RNAs until manipulated by Perriman and colleagues (Perriman, 1998; Perriman, 2000) to generate rolling circle transcripts of eGFP. However, these studies were
performed by transforming DNA vectors containing the PIE-eGFP sequence into *E. coli*; thus, the circRNA was not generated until after transcription of the DNA vector inside bacterial cells. As such, it was unclear how much of the RNA was circularized inside the cell. A recent study that optimized the sequence content and cell-free ligation conditions allowed for very efficient ligation of large RNAs using this PIE method (Weeselhoeft, 2018).

**Enzymatic ligation (T4 DNA Ligase, T4 RNA Ligase 1, T4 RNA Ligase 2)**

T4 DNA ligase is capable of ligating double-stranded regions of nicked DNA, RNA, or DNA:RNA hybrids. In the case of RNA circularization, a DNA splint is designed to be complementary to the 5’ and 3’ end of the RNA molecule. The specificity of the splint is essential as imperfect complementarity or just one missing nucleotide prevents ligation (Lohman, 2011). The ligation efficiency of DNA ligase is dependent on the suitable design of the splint and must be designed and optimized empirically for each new RNA varying in their 5’ and 3’ ends (Moore, 2000).

T4 RNA Ligase 1 is suitable for RNA ligations in which substrates’ ends are single-stranded. Additionally, this particular ligase exhibits preference over certain nucleotide combinations which effects ligation efficiency. RNA Ligase 1’s nucleotide preference is as followed: 3’ end - A > G ≥ C > U and 5’ end - pC > pU > pA > pG (Bain, 1992; England, 1978; Romaniuk, 1982). As the conventional and robust method for *in vitro* transcription involves T7 polymerase, which requires a 5’ pG nucleotide to initiate, the efficiency of an RNA substrate to be ligated by RNA ligase I can only be manipulated by design at the 3’ end, preferentially with an adenosine nucleotide. Studies have shown that only one free nucleotide at either end of the molecule is sufficient for T4 RNA Ligase I functionality (Nilsson, 1994).
T4 RNA Ligase 2 prefers nicked double stranded RNA as its substrate. It functions similarly to DNA ligase in that an oligonucleotide splint (Abe, 2015) can be used to facilitate end joining, although DNA ligase is more commonly used. It is unclear which of these three ligases enzymes is more efficient as most studies do not include this information.

**Design considerations for synthesizing therapeutic circRNA**

As the studies utilizing synthetic circRNAs are minimal, it is unclear which production method is the most efficient, capable of ligating long RNAs, and easily adaptable for large-scale synthesis. GFP-encoded circRNA have been successfully synthesized in bacteria using the PIE method of ligation (Perriman, 1998; Perriman, 2000), although ligation efficiencies were not specified in these studies. Chemical ligation is designed explicitly for generating short oligoribonucleotides (>50 NT), so testing of this method was not pursued in the context of this study. Enzymatic ligation using T4 RNA ligase 1 showed ~50% ligation efficiency with an RNA ~350 nucleotides (NT) in length but with a specific caveat that optimal 5’, 3’ end positioning is necessary to reach this level of ligase activity (Perrault, 1995). Splint-mediated ligation has been used to circularize RNAs roughly one kb in size in a cell-free system (Chen, 1995), although specific efficiencies were not indicated in these studies either. Based on this limited information, we first aimed to compare the PIE, DNA ligase, and RNA Ligase I-mediated ligation methodologies head-to-head to determine which was the most efficient on a per reaction basis while also establishing stringent confirmatory and purification assays post-ligation.

**1.2: Results**

*Enzymatic ligation is more efficient than splint-mediated and autocatalytic ligation methods*

The permuted intron exon (PIE) method of circularization was extensively tested in our initial studies. However, no circular product was ever detected despite multiple optimization
attempts. The amount of circular product generated is conventionally determined by running the ligation reaction on a denaturing PAGE gel where circular RNA migrates slower than its linear form (Perrault, 1995). By quantifying the intensities of the linear and circular bands, the ligation efficiency of a given reaction can be determined. We attempted extensive optimization studies varying the concentrations of the two components necessary for the auto-catalytic, transesterification reactions to occur (GTP and Mg$^{2+}$), but with every attempt only a single band representing the full-length, unspliced transcript was ever detected (data not shown). In retrospect, our insert may have been too long to be efficiently spliced out. The intronic region for Group 1 Introns is usually 300-500 NT in length, and we had initially attempted splicing with a much larger transcript, as has been reported in the literature (~1 kb). However, in these studies, the larger circRNAs were generated in bacterial cells rather than a cell-free system, which may have negatively affected autocatalysis. With the inability to detect circular product from the PIE method, we were left to compare the two forms of enzymatic ligation: T4 RNA Ligase I-mediated (ssRNA) and DNA ligase-mediated (splint) circularization.

To confirm and compare ssRNA ligation and splint ligation, we initially synthesized randomly-generated RNAs ranging from 400, 500, and 600 nucleotides (NT) in length by in vitro transcription (IVT). Equal amounts of starting products were exposed to either a T4 RNA ligase I or DNA ligase (+splint) using industry-standard reaction conditions (described in detail in the Chapter 1 Methods section). Ligation efficiencies were determined by comparing linear and circular band intensities after ligation products were run on a denaturing Urea PAGE gel (Figure 1A). Running linear RNA not exposed to either ligase showed an absence of the slower-migrating (circular) band. We were able to see putative circular bands from both types of enzyme reactions,
with ssRNA ligation appearing slightly more efficient than splint-mediated ligation for this batch of RNAs.

As subsequent confirmation that the “slower-migrating” bands were in fact circRNA, the post-ligation products were exposed to a 3’ → 5’ exonuclease, RNase R (its processivity is illustrated in Figure 1C), and the digested products run on a PAGE gel under the same denaturing conditions (Figure 1B). The majority of product left over after exonuclease treatment of the ssRNA ligation reactions was the putative circular band (Lane 1, 2, 3). The splint reactions did not show this enrichment (Lane 4, 5, 6), suggesting that the slower-migrating bands from these reactions were not circRNA. Although speculative, we inferred that the bands present in these samples may have been nicked RNA or perhaps RNA with residual splints still annealed at the 5’-3’ junction.

Figure 1: T4 RNA ligase I-mediated ligation is more efficient than splint-mediated ligation. (A) Randomly-generated RNA of incremental sizes (400, 500, 600 NT) were generated by in vitro transcription (IVT). The resulting IVT products were exposed to either ssRNA Ligase I or DNA Ligase (mediated by a 40 nucleotide DNA splint complementary to the 5’ and 3’ ends of each RNA). The non-ligase treated IVT RNA (lanes 2, 3, 4), ssRNA Ligase I-treated products (lanes 5,
**Figure 1 (continued):** 6, 7) and DNA ligase-treated products (lanes 8, 9, 10) were run on a 6% 8M Urea PAGE gel at 150V for 2 hours. (B) Ligation products after exposure to RNase R (37°C for 1 hour). (C) Diagram depicting the circular products that are generated from the indicated ligases. RNase R is an exonuclease with 3’ → 5’ processivity. Since circular RNA is covalently linked at its 5’-3’ end it is resistance to RNase R digestion. (D) The percentage of circular product post-ligation was determined using Image J-based gel quantification. As a comparison, the amount of product recovered after RNase R digestion was quantified as well. The similarities in ImageJ and RNase R quantifications for the T4 ssRNA ligase I products confirm accurate approximations of ligation efficiency. NT: nucleotides.

ImageJ was used to calculate ligation efficiencies of the pre-RNase R treated samples in Figure 1A by quantifying the circular band’s signal intensity divided by the sum signal intensity of the linear and circular bands (Figure 1D, left panel). We also quantified the amount of RNA remaining after RNase R digestion relative to input amount (Figure 1D, right panel). Comparable ligation efficiencies were obtained using both ImageJ and RNase R calculation methods for the RNAs treated with RNA ligase I but not the splint reactions.

Based on these initial experiments, we determined that the splint-mediated reactions were generated very little circular product. After RNase R exposure, we observed no enrichment of the circular band indicating that the putative circular bands in these reactions may not be true circRNA. In retrospect, the original splint design was sub-optimal, which likely explains why we saw little to no circular product in these experiments. However, it did provide evidence that the splint method may not be the best for large-scale production as poor splint design or difficult RNA sequences could seriously hinder ligation efficiency. We thus moved forward with working to increase the efficiency of the RNA ligase I.

*Uniform end-positioning of the 5’-3’ RNA ends via a CRC motif significantly enhances ssRNA ligation efficiency*

It has been noted that in order for T4 RNA ligase 1 to function well, the 5’ and 3’ end of the RNA molecule must be in close proximity to each other, and there must be a certain level of freedom at the 5’ and 3’ terminal nucleotides (Perrault, 1995). One can imagine that the likelihood
that both sets of criteria would be met likely decreases as an RNA molecule increases in length, which we thought might explain why we saw such a reduction in ligation efficiency as we increased the size of our RNA templates. Furthermore, analysis of the secondary structure of the RNAs used in Figure 1 showed that there was substantial variability at the 3’ and 5’ ends of these molecules (Supplemental Figure S1), which may have also contributed to the ligation efficiencies we observed.

We thus aimed to design a motif that we could append to either end of RNA that would promote optimal positioning well-suited for single-stranded RNA ligation. This motif would require a stretch of complementary nucleotides with a melting temperature above physiological levels to ensure the two ends are efficiently annealed when inside cells. Secondly, the terminal ends would need to be free of base-pairing to prevent any steric hindrance that could block the ligase from its function. If this motif functions well it would provide advantages over the existing method of RNA ligation as sequence-specific splint optimization and secondary structure considerations would not cause hindrances to workflow.

Our original design of this sequence, which we named a complement-reverse complement (CRC) motif, contained a 20-nucleotide length of complementary nucleotides at the 5’ and 3’ ends of molecule (called the “complementary regions”) followed by 10 free and random nucleotides at either end of the molecules (called the “non-complementary regions”) (Figure 2A). We initially tested the effect that this motif had on T4 RNA Ligase I efficiency by generating a small panel of RNAs (300, 400, 500 nucleotides in length) synthesized with or without a CRC motif. We used the RNAfold secondary structure predictor program to verify the RNA was positioned in the desired conformation at its 5’ and 3’ ends (Figure 2B). The same RNAs without the appended CRC motif showed substantial variation at their end positions as well as the freedom of their
terminal nucleotides. We enzymatically circularized these RNAs using T4 RNA ligase I and compared ligation efficiencies by again measuring band intensities on denaturing PAGE gels (Figure 2C). For each RNA the addition of the CRC motif increased the amount of circular product by at least 20% compared to their non-CRC motif containing counterpart, with the 300 NT CRC RNA reaching virtually 100% ligation efficiency.

**Figure 2: CRC motif enhances single-stranded RNA ligation efficiency.**

(A) A diagram of the complement-reverse complement (CRC) motif. It is comprised of 2 parts: the complementary region responsible for bringing the two ends of the RNA molecule together and the non-complementary region that provides the single-stranded, free ends necessary for ssRNA T4 Ligase functionality. Above the diagram is a linear representation of an mRNA containing the CRC motif, with the black boxes representing the non-complementary region and the blue boxes representing the complementary regions of the CRC motif. (B) A panel of randomly generated RNAs (300, 400, and 500 NT in size, from left to right) were generated without (top panel) or with (bottom panel) the CRC motif. Black arrows point to ligation point. (C) Each RNA was ligated using T4 RNA Ligase I and run on a 6% TBE Urea gel for 2 hours at 150V. White arrows indicate the circular product; blue arrows indicate the non-ligated, linear RNA. Below the gel, are the % circular product values for each construct, quantified by ImageJ. (D) A secondary set of RNAs ranging from 300 – 1300 NT was generated with or without a CRC motif. This panel was generated to be identical at their first and last 100 NT to try and produce similar 5’-3’ secondary structures at their ends irrespective of the presence of a CRC motif. RNase R protection assay was performed after ligation reactions in triplicate. (E) The optimized ligation method was used to circularize mRNA ranging from 1-4 kb; reactions were digested in triplicate and the
average remaining product after digestion quantified. Linear RNAs were assayed as well to verify that RNase R could efficiently digest RNAs of this size.

Our initial results suggested that end position played a more significant role in ligation efficiency than RNA size. To confirm this notion, we generated RNAs ranging from 300 to 1400 NT in size but with the additional design of having both flanking 100 nucleotides be identical at either end. We used the RNAfold predictor to confirm that this overlap in sequence resulted in similar end positions for each RNA in the panel, even with the CRC motif (Supplemental Figure S2). We hypothesized that in this panel ligation efficiency would be similar irrespective of RNA size. This 300-1400 NT RNA panel was also generated with the CRC motif appended. As predicted the non-CRC motif containing RNAs all showed similar, albeit low levels of circular product. This larger panel showed a doubling of circular product (~averaging 60%) for all constructs containing a CRC motif (Figure 2D). Using our final optimized CRC motif and ligation protocol we were able to ultimately show efficient ligation of RNA up to 4 kb in size (Figure 2E). These results provide support for uniform end joining of RNA’s 5’ and 3’ ends to enhance T4 RNA 1 ligation efficiency.

**Optimization of CRC length and nucleotide composition**

The original CRC motif was designed with very generalized criteria. We posited we could increase ligation efficiency further by determining the optimal length and sequence composition of this motif. This optimization was also necessary to confirm the motif did not have a adverse effect on circRNA functionality, specifically in the context of translation efficiency and potential immune recognition (both topics are covered extensively in Chapter 2 and 3, respectively). Briefly, short hairpins in the 5’ UTR of mRNA can inhibit canonical translation and double-stranded RNA is a known activator of cellular immunity under certain conditions. Thus, aside from optimally
enhancing ligation efficiency, it was also necessary to verify that there were no adverse effects associated with incorporating this motif.

A panel of nanoluciferase-encoded RNAs was designed with different CRC motifs. We hypothesized that RNAs with a longer “complementary” regions would increase ligation efficiency as the higher Tm would ensure the majority of molecules conformed the desired end positions. Although it has been reported that RNA Ligase I needs as little as one free terminal nucleotide to function, it is not known what the preferred length of free ends is for this enzyme. We inferred that if the “non-complementary” region were too long, it would reduce ligation efficiency. With these considerations in mind, the CRC motifs in this panel contained “complementary regions” that were 10, 20, or 30 NT in length and “non-complementary regions” that were 10, 15, or 20 NT in length. The non-complementary regions were comprised either fully of adenosines (denoted by an “A” in figures) or a random assortment of non-base paired nucleotides (Figure 3A). Head-to-head comparison of ligation efficiencies showed that a 20 nucleotide long “complementary region” provided maximal contributions to ligation efficiency as constructs with 30 NT complementary regions showed no added benefit to ligation efficiency (Figure 3B). With regards to the non-complementary region, the RNAs comprised of a random assortment of free ends showed a reduction in ligation efficiency as the length of these regions increased. This same trend was not observed from the constructs containing adenosine stretches in their non-complementary region. The ligation efficiencies of each RNA in the CRC panel was determined by Image J measurements of band intensities for three independent batches of RNA; the averaged efficiencies are depicted in Figure 3C.
Figure 3: CRC motif’s functionality is enhanced with optimal sequence length and composition.

(A) A panel of ~1 kB long, nanoluciferase-encoded RNAs were generated that varied only in their CRC motif length and composition. Complementary (Comp.) regions were 10, 20, or 30 nucleotides long. Non-complementary (non-comp.) regions were comprised of a random assortment on nucleotides (10, 15, 20 – black bars) or completely comprised of adenosines (10A, 15A, 20A – white bars), except for the 5’ GGG sequence that every RNA contained due to the requirements T7 polymerase. (B) Each RNA in the panel underwent ligation and 500 ng of each product was run on an 8M Urea 6% PAGE gel. (C) Average ImageJ quantifications of 3 independent batches of RNA that were each run on an 8M Urea 6% PAGE gels. (D) Each ligated RNA in the panel was RNase R-treated and transfected into Hep3B cells (100 ng/well complexed to Lipofectamine 2000 according to the manufacturer’s specifications, Life Technologies). Cells density: 10,000 cells/well in a flat-bottomed 96-well plate. Luciferase expression was measured after cell lysis and substrate addition (Promega). Data is graphed relative to the first RNA in the panel. (E) The RNAs containing a poly(A) non-complementary region were transfected into Hep3B cells as in Fig.1D and interferon-beta levels were measured 8 hours post-transfection by qPCR. Induction was quantified relative to untransfected cells.
As mentioned, confirming the lack of adverse effects this motif has on circRNA was crucial to verify. We compared protein expression of circRNA with or without a CRC motif in various cells lines (Figure 3D). Although most motifs showed little effect on protein expression, there was a particular portion of the panel that enhanced translation efficiency, specifically the RNAs containing a 10 NT long complementary region a non-complementary region containing only adenosines. These same constructs showed very little induction of IFNβ in HeLa cells compared to constructs containing longer CRC motifs (Figure 3E). We concluded that a shorter CRC motif comprised of an adenosine non-complementary region was best suited for use in our circRNA constructs as these characteristics enhanced ligation without negatively effecting translation or immunity.

**CRC motif does not induce immunity or affect translation efficiency of circRNA**

With our optimized CRC motif, we carried out extended confirmatory assays to ensure that its incorporation did not hinder circRNAs functionality in vitro or in vivo. Using the same nanoluciferase reporter protein, we transfected various cell lines and primary cells with linear or circular forms of mRNA with or without a CRC motif and measured protein expression by luciferase assay (Figure 4E) and IFNβ induction by qPCR (Figure 4F). The linear RNA in these experiments were capped and tailed while the circRNA contained an optimized IRES (described further in Chapter 2). In every cell line tested no difference in protein expression was detected when the CRC motif was included in circRNA. We did, however, see a slight reduction in protein translation when the CRC motif was incorporated into cap-dependent, linear RNA. This mirrors the work by Kozak that reported the importance of unstructured 5’ UTRs in the context of cap-dependent translation. Little to no difference in IFNβ induction was detected from circRNAs with or without a CRC motif.
Figure 4: Optimized CRC motif does not induce immunity or affect translation efficiency of circRNA in vitro or in vivo.

(A) Schematic of the components encoded in the circRNA used for CRC characterization studies (diagramed in linear form for clarity). (B) 3.5 pmoles of mRNA (~2.5 – 5 µg), complexed to TransIT (Mirus), was injected intravenously into BALB/c mice (N=2). The mock treatment group was injected with TransIT alone (not shown). 6 and 24 hours post-injection, mice were injected with 10 µg of furimazine substrate diluted in PBS and imaged 5 minutes post-injection (1 min exposure, medium binning, f-stop: 2). Representative images from the 24-hour time point are depicted. (C) Average radiance levels for both time points were quantified using Caliper Living Images Software. (D) To verify that equal starting levels of mRNA were present in the target tissue, mice were sacrificed after the 24-hour time point, liver and spleen harvested, and qPCR performed on the resultant cDNA, using NLuc-specific primers. cT values were normalized to beta-actin and plotted relative to the Spleen, ΔIRES(+CRC) sample. (E) The +IRES, +/- CRC constructs were also assayed in vitro to confirm that the presence of the CRC motif did not hinder functionality across multiple cell lines (HeLa, HEK293T, HepG2, PBMCs). Protein levels were
Figure 4 (continued): determined by nanoluciferase assay 24 hours-post transfection. No signal was detected from the untransfected cells (data not shown). (F) IFNβ expression was monitored in HeLa cells upon transfection of linear or circRNA with or without a CRC motif. Poly I:C was used as a positive control for IFNβ induction. Samples were isolated 12 hours post-transfection. RNA levels were normalized to actin and quantified relative to untransfected cells (UNTR).

The effect the CRC motif had on circRNAs functionality was tested in vivo using the test constructs outlined in Figure 4A. Bioluminescence imaging (BLI) was performed 6 and 24 hours post tail vein injection of mRNA (Figure 4B). Most signal, as expected, was detected in the liver and spleen as has been established in previous studies, but no difference between circRNAs with or without a CRC motif was detected at either time point (both time point radiance levels are quantified in Figure 4D). A mutated IRES was used as negative control in these experiments. To ensure equal levels of RNA were entering the specified tissue and that the motif did not hinder tissue entry, total RNA was isolated from the liver and spleen of mice after the 24-hour imaging session and target RNA detected by qPCR (Figure 4C). Reflecting the signal intensity from the BLI images, more NLuc RNA was detected in the spleen than in the liver. There was no significant difference at 24 hours for all three circular RNAs in either tissue even with the translation-deficient ΔIRES circRNA. These in vitro and in vivo data provide support for the use of the CRC motif in circRNA to enhance ligation efficiency without the concern for potential adverse effects to its functionality.

RppH is a novel and efficient method for obtaining a 5’ monophosphate end

All forms of enzymatic ligation (ssRNA and splint-mediated) require a 5’ monophosphate end to undergo ligation, which means two phosphates must be removed from the linear, triphosphate RNA that is generated by in vitro transcription. This has been done previously by two method: (1) Guanosine monophosphate (GMP) is added in access to the IVT reaction at a 4:1 ratio compared to GTP such that statistically 75% of the resultant product would have a GMP end or
(2) a two-step reaction which first remove all 3 phosphate groups (via a variety of commercially available phosphatases: calf intestinal-CIP, Antarctic-AP, or recombinant shrimp alkaline phosphatase-rSAP) followed by treatment with a kinase (conventionally polynucleotide kinase, PNK) which adds a single phosphate back to the 5’ end of the RNA. Neither method was particularly efficient in our hands. The GMP:GTP method reduced the amount of IVT product generated by more than half and the resultant mRNA contained large amounts of byproduct (Figure 5A). Every iteration of the 2-step method leads to low levels of final circular product, averaging around ~15% ligation efficiency (Figure 5B).

Two new commercially-available enzymes derived from E. coli, RNA Pyrophosphohydrolase (RppH) and Apyrase, were said to be able to generate a 5’ monophosphate end in a single reaction with great efficiency. We tested the efficiency of RppH and Apyrase against the two existing methods and found that RppH was by far the superior method for obtaining 5’ monophosphate ends. Further testing using XRN-1, a 5’ → 3’ exonuclease that specifically targets 5’ monophosphate ends (Figure 5C), showed that RppH and Apyrase were capable of cleaving the γ and β phosphates with almost 100% efficiency (Figure 5D). This optimization was one in a series of successful attempts to enhance the ligation efficiency through updated techniques and newly available reagents.
Figure 5: RppH is the superior enzyme for obtaining the 5’ monophosphate end necessary for enzymatic ligation.

(A) GMP incorporation during in vitro transcription resulted in <25% of the normal product generated in standard IVT reactions and the resultant RNA was of low quality, made evident by the smears present in gel capillary electrophoresis runs (Agilent, TapeStation) and PAGE gels. (B) Ligation efficiency as a readout for monophosphate efficiency was also compared using the various two-step options and newly optimized one-step options. The RppH and Apyrase-treated RNA produced more than double the circular product as any of the two-step treated RNAs. (C) Diagram depicting the methods of generating 5’ monophosphate ends. XRN-1 is 5’→3’ exonuclease that explicitly targets monophosphate ends. (D) The efficiency of the newly available options (RppH and Apyrase) for generating monophosphate ends was tested by treating 500 NT RNA with the respective enzymes, followed by XRN-1 digestion. The extent to which the RNA is degraded reflects the efficiency of the monophosphate enzyme. Linear RNA, which contains a triphosphate end and not suitable for XRN-1 digested was included as a negative control.

**HPLC methodology for purifying circRNA**

HPLC purification of RNA has shown to remove impurities that arise during *in vitro* transcription that hinder expression through their induction of the many RNA-sensing immune receptors in the cell (Kariko, 2011; Weissman, 2013). By far the most significant advancements in circRNA functionality in this project came with the incorporation of HPLC purification into our circularization methodology. Before the incorporation of this rigorous purification step, protein expression was low and immune response high both *in vitro* and *in vivo*. However, we were able
to create an RNA HPLC methodology to isolate pure, circRNA; the optimized conditions of this method are shown in **Figure 6A-B**. As a reference, the HPLC chromatogram of a commercial RNA ladder (Century, Invitrogen) is shown as an example for the capacity of HPLC to separate a broad range of RNAs (**Figure 6C**). The extent to which HPLC can remove impurities can be observed by re-running the isolated samples on the HPLC column, which shows a single clean peak with all non-specific peaks efficiently removed (**Figure 6D**). Although covered more in Chapter 3 we analyzed the extent of to which HPLC effected translation efficiency and immune induction using an EMCV-GFP RNA and measured GFP expression and in MFI by flow cytometry (**Figure 6E**). We also determined IFNβ induction compared to a positive control (poly I:C) and an optimized, commercial eGFP mRNA (Trilink), which showed a ~3-fold reduction in IFNβ levels compared to non-HPLC purified samples. (**Figure 6F**). This data and the experiments presented further in Chapter 3 provide strong support for the necessity of HPLC purification for circRNA to function optimally.

**Figure 6**: HPLC purification is an effective means of increasing RNA functionality.
Figure 6 (continued): (A) Optimized sequence (including flow rate, gradient, buffer concentrations are shown). (B) Optimized solvent ratio gradient optimized for large RNA isolation. (C) The ability of this method to separate a wide range of RNAs was confirmed by running century ladder (Invitrogen) on the RNAsep column. (D) Example chromatograms of what RNA peaks like pre and post-HPLC purification. Although covered more thoroughly in Chapter 2 and 3, HPLC leads to higher translation efficiency (E) and reduced immune induction compared to non-HPLC purified RNAs (F). Translation efficiency was measured by transfecting EMCV-eGFP circRNA (+/- HPLC) into HEK293 cells and GFP expression (FITC+ Cells and MFI) by flow cytometry 24 hours post-transfection. IFNβ induction was measured 12 hours post-transfection in HeLa cells. Values were normalized to beta-actin and calculated relative to untransfected cells (UNTR).

We found that even with the highly-sensitive HPLC method, ligation reactions containing both linear and circular RNA still co-eluted from the HPLC column as a single peak (Figure 6D, top chromatogram, and Figure 7C, top chromatogram). We assumed that this was due to linear and circular RNA being identical in size and charge; the only system sensitive enough to separate the two RNA conformations was low percentage denaturing PAGE gels as shown in previous figures and studies. We hypothesized that if there was a robust way to alter the linear RNA through their unique 5’ or 3’ end availability we could more easily identify and separate these two populations (Figure 7A). We found that by exposing the ligation reaction to a Poly(A) enzyme, we could efficiently add 100-200 nucleotides to the free 3’ end of linear RNA, which occurs with total efficiency. This was shown by tailing a linear mRNA sample and seeing a complete shift in RNA size (Figure 7B). When ligation reactions were treated with a poly(A) enzyme, there was a population that was resistant to the adenylation; this represented our circular product. When running the ligation products with or without a poly(A) tail on the RNAsep HPLC column, we observed a single and double peak, respectively (Figure 7C). Functional testing on the isolated peaks confirmed that the first peak, representing poly(A) resistant RNA was indeed circRNA (Figure 7D-E).

To summarize, we have shown that we can efficiently utilize HPLC not only to remove detrimental impurities in our samples, but we could also use this methodology as a verification
and isolation scheme for obtaining circRNA. This methodology works for a wide range of RNA sizes with only minor adjustment needing to be made for very large RNA (>2000 NT) or RNA containing significant nucleotide modifications (Supplemental Figure S3).

**Figure 7: Poly(A) resistance assay can confirm circularization and aids in HPLC isolation.**

(A) Post ligation the reaction contains a mixture of linear and circular RNA as a strategy for confirming circularization and isolating the circRNA product. The ligation reaction is exposed to a 3’ poly(A) enzyme that capable of efficiently adding 100-200 nucleotides of adenosines to the linear RNA only. The difference in size can be detected by gel capillary electrophoresis (Agilent TapeStation) (B) and separated by HPLC peak isolation (C). To verify the identity of each peak isolated NLuc-encoded circRNA with or without a functional IRES was poly(A) tailed and each peak purified by HPLC (D-E). The translation capacity of each peak was measured by nanoluciferase assay after transfection into HepG2 cells. The “CircRNA” peak without a functional IRES showed no translation capacity (E).

**Final optimizations to T4 RNA ligase 1 circularization methodology**

At the start of this project the existing literature had only shown T4 ssRNA ligase to generate sufficient levels of product for RNA of 350 nucleotides in size, and even in that study only ~50% efficiency was achieved. Here we show that RNA of similar size can be generated at 100% efficiency and moreover larger RNA up to 4 kB can be generated using our optimized methodology, outlined fully in Figure 8A. This was possible through manipulations intrinsic to
RNA via the incorporation of CRC motifs and updates to the reagents used during the synthesis process.

Although briefly mentioned throughout Chapter 1, the stringent quality control assays necessary to confirm circularization and its proceeding purity were performed for every new batch of RNA used in experimentation (Figure 8B). Representative data using constructs with two different coding sequences (NLuc and eGFP) are shown as examples throughout Figure 8. First and foremost, RNase R digestion is performed to remove the majority of linear RNA (Figure 8C). Gel capillary electrophoresis (TapeStation, Agilent) was performed to ensure correct degradation of linear RNA, a more sensitive assay than agarose gels (Figure 8D). After HPLC purification, resultant products are used as the template for outward-oriented PCR (Figure 8E) and the resultant products sequenced to ensure proper end ligation (Figure 8F).

Figure 8: The optimized outline for RNA ligase I enzymatic ligation and subsequent confirmatory assays. (A) A step-by-step outline of T4 ssRNA ligase I enzymatic ligation. Post-ligation there are several steps, outlined in (B), that must be taken to remove residual, linear RNA and impurities. (C) The first step is RNase R digestion, which under the correct conditions results in >90% circular RNA
Figure 8 (continued): purity. (D) To ensure the remaining product is intact circular RNA gel capillary electrophoresis (TapeStation, Agilent) is performed on purified products, with linear RNA (+/- RNase R) run as a control. (E) Outward-oriented PCR is performed using divergent primers that span the 5'-3' ligation junction to confirm circularization. (F) The resultant product is cloned into a PCR cloning vector (NEB) and sequenced to verify accurate 5'-3' covalent linkage.

1.3: Discussion

Here, we present an optimized methodology for T4 RNA ligase-I mediation RNA circularization as well as stringent confirmation and purification steps. This methodology allows for very large RNAs (up to 4 kB) to be efficiently ligated. This efficiency is dependent on the uniform endjoining of the 5’ and 3’ ends of RNA which has been accomplished through the incorporation of our CRC motif. We were able to confirm in vitro and in vivo that this motif does not affect circRNA functionality (e.g., translation or immune activation). We also developed a new assay for confirming and separation circRNA through 3’ targeted poly(A) tailing which allows for size distinctions to arise between circRNA and residual linear RNA in the ligation reaction.

Although the CRC motif provided significant enhancements to ligation efficiency and does not appear to effect circRNA functionality, its presence in circRNA has made certain aspects of verification and purification difficult. For example, attempts to visualize or quantify circRNA using probes spanning the 5’-3’ junction were unsuccessful due to the double-stranded, hairpin-like nature of the CRC motif, which prevented efficient binding of the sequence-specific probes. Additionally, many studies design their templates to ligate in the middle of the coding sequence (CDS) so that expression could be a readout for successful ligation. However, placing the CRC motif in the middle of a CDS was not possible as it would disrupt the sequence needed to make the correct protein. Even the conventional RNase H nicking assay used to confirm circularization was difficult to interpret because of the CRC motif. Despite these setbacks, we were able to
develop a pipeline of assays that allowed us to confirm ligation efficiency (outlined in Figure 8) confidently.

A recent study (Weeselhoeft, 2018) showed that incorporating homology arms, similar to our CRC motifs, also enhance the efficiency of PIE-mediated ligation. We did preliminary experiments to compare the efficiency of this new PIE method to the optimized Ligase 1 method presented here (Figure 9A). We were able to confirm the efficiency and ease by which this alternative method can generate circRNA as no additional enzymes are needed after in vitro transcription to generate the ligated product. In more advanced stages of therapeutic development, this method would appear to be more desirable on a large-scale basis as the time and cost are significantly lower than that of any enzymatic ligation methods. We did find issues with the detection methods (e-gel system) used in this study, which was not sensitive enough for efficient separation of linear and circular bands especially when using the PIE method of ligation where multiple intermediates and byproducts are generated. As such, ligation efficiencies were challenging to calculate in these preliminary experiments (Figure 9B-C). We also found that modified nucleotide incorporation during in vitro transcription reduced PIE ligation efficiency (Figure 9D). The intronic region’s positioning and secondary structure are crucial to its autocatalytic activity, so it may be that the modifications alter this positioning, preventing the autocatalytic reactions from occurring normally. In the subsequent chapters, I will present the preliminary data generated in response to this Weeselhoeft manuscript as it very closely mirrors the flow of our work and reflects the field’s most recent developments of circRNA for therapeutic utility.
Figure 9: Comparison of optimized enzymatic ligation to newly developed PIE method.

(A) EMCV-NLuc and CVB-NLuc encoded mRNAs were generated with the optimized PIE intronic regions as in (Weeselhoeft, 2018) or with our CRC motif, and autocatalytic or enzymatic ligation reactions were performed, respectively. (B-C) Ligation efficiency for CRC versus PIE method for the sets of constructs was tested. (D) The effect that modified nucleotides had on ligation efficiency was tested. E: EMCV IRES, C: CVB IRES, White circles: Linear RNA, Blue Circles: Circular RNA.
Chapter 2: Properties Required for Robust Translation of circRNA

Contributions: All experiments were conceived of by Ellese Carmona and Michael Goldberg and performed by Ellese Carmona.

2.1: Introduction

Canonical, eukaryotic translation requires processed, mature mRNA containing a 5’ cap and a 3’ poly(A) tail. As circRNA lacks molecular “ends” an alternative route of translation was required. Here we briefly describe the current state of knowledge regarding eukaryotic translation including canonical, IRES, and m6A-mediated forms translation. With this information, we can competently strategize our design criteria for synthesizing robustly translatable circRNA.

Mechanism of eukaryotic translation

Eukaryotic translation is broken down into three parts: initiation, elongation, and termination. The initiation step involves the recruitment of multiple complexes, including the 43S pre-initiation complex and the EIF4F complex, which localize mainly to the 5’end of the mRNA. The exception to 5’ localization occurs with the poly(A) binding protein (PABP) which binds both the 3’ poly(A) tail and the EIF4G scaffold that is part of the EIF4F complex resulting in a pseudo-circle at the start of translation. Once the 43s complex has associated with mRNA, it scans the 5’ portion region to find the start codon, associates the full ribosomal complex, and begin the elongation step of protein synthesis. (Gebauer, 2004; Hinnebusch, 2014; Sonenberg, 2009)

Marilyn Kozak did extensive work that provided evidence for the scanning mechanism of eukaryotic translation and identified the consensus sequence present upstream of eukaryotic start codons necessary for translation (Kozak, 1984; Kozak; 1986; Kozak; 1987). Interestingly, her work provided evidence for the scanning method of translation one way by attempting to initiate translation on a circular RNA, which she found to be incapable of translating protein. She also
observed that most eukaryotic mRNA tended to have fairly unstructured 5’ UTRs and when highly structured motifs, like hairpins, were included in this region canonical translation was inhibited (Kozak, 1989). Taken together it is clear that circRNA lacks the conventional motifs necessary for canonical, eukaryotic translation.

**IRES-mediated translation**

The existence of Internal Ribosomal Entry Sites (IRES) was first discovered within the Picornavirus family (Pelletier, 1988; Jang, 1988). Despite viruses like EMCV and poliovirus appearing unfit for eukaryotic translation due to their highly structured 5’ region, lack of a 5’ cap, and more, they were still able to efficiently hijack cellular translation machinery pointing to an alternative mechanism of translation that could occur in eukaryotic systems. Additional studies that inhibited essential components necessary for canonical translation initiation showed that these viruses were still able to efficiently translate their protein (Canaani, 1976; Baretta 1997; Gingras 1996). The critical experiment that confirmed internal recruitment of the ribosome came with from studies that utilized bicistronic constructs where the first coding sequence is reliant on cap-dependent translation while the second should only be translated if it sequence dividing the two coding sequences is capable of internal ribosomal binding. (Pelletier, 1988; Jang, 1988). Additionally, Chen and Sarnow (1995) used circRNA containing an IRES to show that translation could occur in circRNA when these specific motifs were incorporated.

The intrinsic and extrinsic factors necessary for internal ribosomal entry do not appear to be conserved from one IRES to another and no consensus sequence conferring IRES activity has been identified. However, small portions derived from verified full-length IRES have been shown to maintain their IRES activity (Dorokhov, 2002). These synthetic IRES have no distinct motifs despite their apparent activity. On a separate note, the secondary structure has repeatedly been
shown to be crucial to IRES activity. Studies that made small deletions or even single nucleotide mutations IRES sequences resulted in severe reductions in translation capacity (Trono, 1988; Svitken, 1985). Additionally, it was found that each IRES utilize a unique set of initiation factors some which are involved in canonical translation and other that are not, known as IRES-transacting factors (ITAFS).

**Discovery of non-canonical eukaryotic translation**

Like viral IRES the existence of cellular IRES was initially contested, but extensive work has shown persistent evidence for this cellular mechanism of non-canonical translation. Interestingly, what has been gleaned from cellular IRES requirements are distinct from viral IRES. For example, many of the first cellular IRES motifs were comprised of non-contiguous, unstructured sequences similar in sequence to rRNA (Yang, 1997; Huez 1998; Stoneley, 1998). The complementarity to ribosomal RNA appears to facilitate ribosomal binding to these cellular IRES and can confer IRES activity with as little as nine nucleotides of complementarity (Chappell, 2000).

The utilization of these cellular IRES appears to be associated with specific cellular states like cell division, stress, and apoptosis (Joahnnes, 1999; Pyronnet, 2000). All three of these distinct cellular states induce the global down-regulation of protein synthesis and yet there are specific genes that are still translated many of which have been shown to harbor IRES activity. These cellular IRES along with many viral IRES have been shown to remain active during states of global protein synthesis inhibition (Honda, 2000; Huang; 1991) Overall, this information provides support for the utilization of IRES in the context of circRNA as it bypasses the need molecular ends and may continue to function during cellular stress of cell division unlike canonical translation.
**m6A-mediated cellular translation**

Distinct from the characteristics and dependencies associated with cellular or viral IRES is a new mechanism by which eukaryotic translation appears to initiate that is associated with a specific nucleotide modification: N6-methlyadenosine (m6A) (Meyer, 2015; Zhou, 2015). This modification has now been shown to allow for translation to occur completely independent of the EIF4F complex (Coots, 2017) and with no known requirements related to secondary structure or ITAFs. High throughput sequencing capable of identifying these methylation sights has shown an enrichment for this nucleotide modification specifically in the 5’ UTR region upstream of many start codons. Furthermore, enrichment of methylation at these sites appears to occur during cellular stress, which suggests a mechanism of regulatory expression for the genes that play a role in responding to cellular stress. Interestingly, a study recently showed that synthetic circular RNAs could generate protein when the motif necessary for m6A methylation, called the RRACH motif, was incorporated into its sequence. (Yang, 2017). Overall, this newly identified mechanism appears to be a suitable potential candidate for enabling protein synthesis in therapeutic circRNA.

**Utilizing translation insights to design optimal therapeutic circRNA**

There is much to consider when deciding which motif might be the best option for ensuring therapeutic levels of protein translation in the context of circRNA. A viral IRES that requires little to none of the conventional translation initiation factors might be desirable, as is the case for EMCV, HCV, and CRPV IRES (Pestova, 1996; Pestova, 1998; Jan 2002). However viral IRES with their characteristically high secondary structure might contain motifs that can be recognized by the RNA-sensing immune receptors within the cell. Cellular and synthetic IRES, as well as the m6A-mediated form of translation, tend to be shorter and less complex in their requirements to undergo translation. However, it is unclear how potent these particular IRES are at translating
protein across multiple cell types and systems, whereas viral IRES may have adapted mechanisms to ensure their efficient translation across multiple systems.

The work presented in this chapter focuses on characterizing the relative translational capacity conferred to circRNA when varying internal ribosomal entry sites (IRES) or m6A motifs are incorporated into its sequence. As there are thousands of IRES that have now been identified each with variations in their requirements to undergo translation successfully, we initially chose a pool of candidates that, based on the literature, we believed would: (1) robustly translate protein across multiple systems, (2) not be hindered by the circularity of the RNA and (3) not be hindered by the supplementary modifications that may ultimately be included in circRNA design (i.e., modified nucleotides).

2.2: Results

IRES panel inclusion considerations

The IRESsite database was established as a centralized depository of validated and putative sequences with IRES activity (Mokreis, 2010). Of the thousands of IRES listed, we ultimately chose 18 motifs to include in our panel (Table 1). The criteria we used to narrow down this list of candidates was based in part on IRES size (< 600 NT) and the breadth of literature that existed which verified IRES activity. The size of IRES tested ranged from 9 to 583 NT. Although large IRES like EMCV have been confirmed to function in the context of a circular RNA, it was unclear whether the shorter RNAs would do so as well. The panel contained a fairly even mixture of viral, cellular, and synthetic IRES. Fortunately, by virtue of our circular construct design, we had devised a system that is well suited for testing IRES potency as long as circRNA purification was stringent. Secondary confirmatory assays to ensure that our circRNA was generating protein through IRES-mediated translation were performed (Thompson, 2012). Details on how this library was
constructed can be found in the methods section of this paper. Briefly, Gibson Assembly was used
to synthesize each IRES and NEBuilder used to clone the IRES sequence into a nanoluciferase-
encoded vector. Clones were sequenced to confirm correct synthesis. PCR adding the T7 promoter
sequence to each construct was used to verify the correct size of each construct and then used as
template in the in vitro RNA synthesis reactions.

**Table 1: IRES Panel & Source Information**

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Species</th>
<th>Size (NT)</th>
<th>IRESite ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMCV-R</td>
<td>Viral</td>
<td>Encephalomyocarditis Virus</td>
<td>583</td>
<td>353</td>
</tr>
<tr>
<td>CRPV-IGR</td>
<td>Viral</td>
<td>Cricket Paralysis Virus</td>
<td>192</td>
<td>40</td>
</tr>
<tr>
<td>HCV TYPE 1A</td>
<td>Viral</td>
<td>Hepatitis C Virus</td>
<td>385</td>
<td>98</td>
</tr>
<tr>
<td>HCV TYPE 1B</td>
<td>Viral</td>
<td>Hepatitis C Virus</td>
<td>341</td>
<td>112</td>
</tr>
<tr>
<td>CSFV</td>
<td>Viral</td>
<td>Classical Swine Fever Virus</td>
<td>442</td>
<td>359</td>
</tr>
<tr>
<td>GBV-B</td>
<td>Viral</td>
<td>Hepatitis GB Virus -B</td>
<td>445</td>
<td>153</td>
</tr>
<tr>
<td>XIAP</td>
<td>Cellular</td>
<td>Homo sapien</td>
<td>162</td>
<td>486</td>
</tr>
<tr>
<td>C-MYC</td>
<td>Cellular</td>
<td>Homo sapien</td>
<td>352</td>
<td>635</td>
</tr>
<tr>
<td>N-MYC</td>
<td>Cellular</td>
<td>Homo sapien</td>
<td>320</td>
<td>72</td>
</tr>
<tr>
<td>L-MYC</td>
<td>Cellular</td>
<td>Homo sapien</td>
<td>52</td>
<td>89</td>
</tr>
<tr>
<td>RBM3</td>
<td>Cellular</td>
<td>Homo sapien</td>
<td>22</td>
<td>569</td>
</tr>
<tr>
<td>APAF1</td>
<td>Cellular</td>
<td>Homo sapien</td>
<td>56</td>
<td>417</td>
</tr>
<tr>
<td>GTX</td>
<td>Cellular</td>
<td>Mus musculus</td>
<td>9</td>
<td>204</td>
</tr>
<tr>
<td>NDST41</td>
<td>Cellular</td>
<td>Mus musculus</td>
<td>421</td>
<td>267</td>
</tr>
<tr>
<td>GAAA(X20)</td>
<td>Synthetic</td>
<td>x</td>
<td>80</td>
<td>614</td>
</tr>
<tr>
<td>POLYA(X80)</td>
<td>Synthetic</td>
<td>x</td>
<td>80</td>
<td>x</td>
</tr>
<tr>
<td>PPT19(X4)</td>
<td>Synthetic</td>
<td>x</td>
<td>76</td>
<td>613</td>
</tr>
<tr>
<td>KMI1</td>
<td>Synthetic</td>
<td>x</td>
<td>58</td>
<td>476</td>
</tr>
</tbody>
</table>

**EMCV IRES robustly translates protein across multiple systems**

A list of the IRES included in our library and specifications of each is contained within
**Table 1**, and a full annotation of each sequence can be found in Supplemental Table S1. Using
our circularization protocol, we synthesized and purified circRNA for each construct. We
transfected equimolar levels of each RNA first into Hep3B cells (**Figure 10A**). Only two of our
20 constructs generated protein to significant levels: EMCV and HCV, both viral-derived IRES.
We tested the IRES library across multiple cell lines and while the absolute levels varied from cell line to line the relative relationship between constructs remained the same (Supplemental Figure S4). We did additional testing of these two hits in the context of variable 3’UTRs and found that HCV appeared to be much more context dependent, whereas EMCV, irrespective of CDS or 3’ UTR was able to potently translation protein.

Figure 10: EMCV IRES produces robust levels of translation across multiple systems. (A) The sequence content of each construct in our IRES reporter library contained the nanoluciferase (NLuc) coding sequence, a poly(A)x50 3’ UTR, and the optimized CRC motifs. The only difference in sequence in each construct was the IRES inserted between the 5’ CRC and the NLuc CDS. The predicted secondary structure for the EMCV-NLuc RNA is pictured. Each circular construct was generated and purified as outlined in Figure 8. (B) The resultant purified circRNA library was transfected into various cell lines at equimolar levels complexed to Lipofectamine 2000 (Hep3B results are shown). Protein expression was measured 24 hours post-transfection by luminescence assay. The experiment was repeated three times in each cell line tested (Supplemental Figure S4: Hep3B, HepG2, HEK293T, H1299 cells). (C) Secondary confirmation of IRES activity was performed in the cell-free, rabbit reticulocyte lysate translation system. The cap-binding translation initiation factor necessary for canonical translation, EIF4E, can be sequestered by the addition of access cap analog (ARCA, NEB), thereby inhibiting cap-dependent
Figure 10 (continued): translation. (D) Protein expression in the retic lysate system was measured from canonical linear NLuc RNA containing a cap, poly(A) tail, and a non-IRES (human insulin 5’ UTR, 61 NT) and the EMCV-NLuc circRNA. (E) Expression was also measured using a different CDS, eGFP, and expression compared to a commercially available eGFP mRNA (Trilink) with the industry modifications (5’ cap, poly(A) tail, Pseudouridine, and 5mC nucleotide modifications).

To confirm IRES activity, we performed additional protein expression tests in the rabbit reticulocyte lysate in vitro translation system, where we could push translation to a cap-independent mode of translation (Figure 10B). We did this by adding access ARCA cap (NEB), which sequesters the initiation factor necessary for cap-dependent translation, EIF4E. In this setting, EMCV-encoded circRNA outperformed linear RNA in terms of protein expression and was unaffected by the presence of excess cap (Figure 10C). EMCV retained its potency when an alternative CDS was used to monitor protein expression. Even when compared to commercially available, optimized, linear eGFP RNA (Trilink), EMCV-eGFP circRNA showed higher protein expression (Figure 10D).

Although most of our work focused on the 5’UTR content to promote translation efficiency we also found that the 3’ UTR even in the context of circRNA could affect translation capacity. We selected a smaller panel of 3’ UTRs to test in the context of EMCV-NLuc mRNA to see what effect this motif had on translation. The 3’UTR of alpha and beta globin are most conventionally used therapeutic RNA, with a twice-repeated beta globin (HbB1x2) 3’UTR generally considered the most stabilizing (Sahin, 2014). However, we wanted to avoid any sequences with any known motif recognized by endonucleases to circumvent the possibility of circRNA degradation. Previous work has shown that both alpha and beta globin carry endonuclease motifs despite their characterized stability. Other work showed that a 20-mer derived from the chicken elastin 3’ UTR conferred significant stability to mRNA. Finally, in the context of both canonical and IRES-mediated translation, the presence of a poly(A) tail has an additive and even synergistic effect on
protein translation. In the end, we tested HbBx1, a thrice repeat of the chicken elastin motif (60 NT), and a 50 NT stretch of adenosines (poly(A)x50) in the context of the 3’ UTR to determine their effect on circRNAs translation efficiency.

Most noticeably, poly(A)x50 had the most significant impact on protein expression in the context of circRNA (Figure 11B). This is not surprising as previous work had shown that although not necessary for functionality EMCV IRES-mediated translation is enhanced by the presence of a poly(A) tail. With cap-dependent translation, the poly(A) tail aids in the recruitment of translation initiation factors, which may be why this addition boosted translation so much when incorporated into our EMCV-dependent circRNA. For all future studies, we incorporated a 50 NT poly(A) tail after the CDS stop codon to enhance protein expression further.

Figure 11: 3’ UTR motifs can significantly affect the translation of circRNA.
EMCV-NLuc RNA was generated with variable 3’ UTRs: a 50x poly(A) stretch, the conventionally used HbBx1, a thrice repeated 20-mer derived from the chicken elastin 3’ UTR which has been shown to confer stability, or no 3’ UTR at all. Linear (A) or circular forms (B) of these mRNAs were transfected into HEK293T cells and protein expression measured by luminescence assay over a 3-day time course.

**Neither modified nucleotide incorporation nor circularization affect cellular entry**

It was unclear whether the size, modifications, or secondary structure variations in each RNA could affect entry into the cells and be behind the variations in translation efficiency observed in our luciferase reporter assays. To address this concern, we investigated differences in the cellular
entry of our test RNA by measuring RNA levels at an early time point in the experiment. We regularly do this by performing qPCR from samples within the first few hours on transfection or injection. For example, we transfected linear or circular RNA with different modified nucleotides incorporated into their sequence during *in vitro* transcription. (Figure 12A). These results indicated that as long as equimolar amounts of RNA were transfected NT modifications nor circularization resulted effected cellular entry.

![Figure 12: Neither circularization nor modified nucleotide incorporation affects cellular entry of RNA.](image)

(A) During *in vitro* transcription EMCV-NLuc mRNA was generated with no nucleotide modifications, 100% 5mC, 100% Pseudouridine, or 50% 5mC + 50% Pseudouridine. A portion of the batch was circularized according to our optimized protocol. Both linear and circular forms of each mRNA was transfected using Lipofectamine 2000 into HEK293T cells. 4-6 hours post-transfection cells were washed thoroughly and total RNA isolated using the Cell-to-Ct kit (Life Technologies). qPCR was performed to compare the relative levels of NLuc present in the cells at this early time point. (B) In a separate experiment, RNA was synthesized with 25% Cy5-UTP (Trilink). The resultant IVT product was either left as linear product (L-), capped and tailed (L+), or circularized (Circ). All RNAs were purified by HPLC including the control eGFP mRNA (Trilink). These RNAs were transfected into HepG2 or HEK293T cells as previously described. The Celigo imager/plate reader was used to image and quantify the number of Cy5+ cells 6 hours
**Figure 12 (continued):** post-transfection in both cell types. (C) Representative images for circRNA-transfected HepG2 cells are shown. The 3rd image shows untransfected (UNTR) cells.

This was confirmed in a secondary set of experiments by utilizing fluorescently labeled RNAs, generated by incorporating 25% Cy5-UTP (Trilink) into their sequences during *in vitro* transcription. Cy5-labeled RNAs were transfected into cells and quantified using Celigo-based fluorescent microscope/cytometer, which allows for automated fluorescent quantifications and imaging. These data showed that linear and circular forms of RNA, as well as capped and tailed RNA, showed no difference in cellular entry across multiple cell lines (Representative images are seen in Figure 12C, and quantifications of the fluorescent images are shown in Figure 12B). A full quantification of GFP+ and Cy5+ cells from these experiments can be found in Supplemental Figure S5. From these experiments, we could rule out variations in starting levels of RNA as a reason for variations in protein expression.

**RRACH motif produces little protein expression in vitro**

m6A is the most abundant eukaryotic RNA modification and the only modification found in all forms of cellular RNA (Li, 2014; Wei, 1975). The consensus motif RRACH (R = G or A, H = A, C, or U) is a known m6A methylation site and when located in the 5’ UTR has been shown to increase translation efficiency and promote cap-independent translation (Csepany, 1990; Harper, 1990). Moreover, it was shown that the presence of this motif in synthetic circRNA was capable of translating protein in the absence of an IRES (Yang, 2017). We thus wanted to compare this mode of cap-independent translation to EMCV IRES-mediated translation, which had been our top hit in our initial IRES panel studies. Moreover, we wanted to test whether these two forms of translation could be combined to additively or perhaps synergistically enhance translation efficiency in our circRNA.
We generated circular RNA constructs with 2, 1, or 0 RRACH motifs upstream of the nanoluciferase start codon. We generated a construct with two scrambled RRACH motifs as well (Denoted as “N” in figures). We compared these construct’s translation efficiency to EMCV-NLuc circRNA with or without m6A modified nucleotides (50% m6A) directly incorporated during the in vitro transcription step of synthesis (Figure 13A). We found that circRNA containing RRACH motifs produced very little protein relative to the unmodified EMCV-NLuc circRNA (Figure 13B).

Work related to m6A has shown that m6A-mediated translation is induced by cellular stress. To see if we could induce methylation and enhance m6A-mediated protein expression we exposed transfected cells to cellular stress via heat shock and assayed for changes in protein expression, but no difference was observed in any of the constructs tested compared to non-heat shock controls (data not shown). Interestingly, we found that EMCV-NLuc circRNA that had m6A incorporated directly into its sequence completely ablated protein translation, despite confirmation assays showing correct size and synthesis (Figure 13C). As we knew that protein ablation was not due to incorrect synthesis or inability of the modified circRNA to enter the cell, we next wanted to compare the effect m6A incorporation had on cap-dependent linear RNA to IRES-dependent circRNA. As in the literature, m6A did not hinder or ablate protein translation in cap-dependent linear mRNA (Figure 13D).

Taken together, it appeared that the m6A modification was specifically inhibiting IRES-mediated translation. This result was unexpected and forced us to rethink our circRNA design criteria as modified nucleotides are ubiquitously used in mRNA therapy studies and were initially going to be utilized in our circRNA to enhance its functionality. Before completely ruling out the
use of all modified nucleotides in circRNA we first wanted to verify that all of the commonly used modified nucleotides (i.e., 5mC and Ψ) had the same effect on IRES activity.

Figure 13: m6A is not a potent inducer of protein translation in circRNA.
(A) NLuc circRNA was generated with 2, 1 or 0 “RRACH” motifs directly upstream of the start codon. A construct with 2 scrambled “RRACH” motif was generated as well (denoted by the letter “N”). Additionally, our EMCV-NLuc construct was generated with unmodified nucleotides or with 50% m6A incorporated during in vitro transcription. (B) These constructs were transfected into HEK293T cells and assayed for protein expression by luminescence assay 24 hours post-transfection. (C) To confirm correct construction and quality, gel capillary electrophoresis (Agilent, TapeStation). (D) Canonical linear mRNA (capped and tailed) was generated with or without m6A modifications (50% incorporation during in vitro transcription) along with the previously tested circular EMCV-NLuc construct (+/- 50% m6A). Expression in HEK293T cells was measured 24 hours post-transfection.

Modified nucleotides hinder IRES-mediated translation and ligation efficiency

We next wanted to test whether other commonly used modified nucleotides natively found in endogenous eukaryotic RNAs and commonly used in other mRNA therapeutic studies had the same effect on IRES translation. Despite equal RNA entry into the cell, we observed that every
modification incorporated into EMCV circRNA ablated protein translation (Figure 14A). This was in direct opposition to the effect modified nucleotides had on canonical, linear RNA where a boost in translation is normally observed, matching well with the existing literature (Kariko, 2008; Anderson, 2010; Kormann, 2011; Mays, 2013). We tested the effect that modified nucleotides have on selected synthetic and cellular IRES from our original IRES panel as well (Supplemental Figure S4).

Figure 14: Incorporation of modified nucleotides into IRES-driven circRNA ablates translation capacity.
Figure 14 (continued): (A) Indicated NT modifications were incorporated into cap-dependent linear mRNA or IRES-driven EMCV-NLuc circRNAs and assayed for protein expression by luminescence assay 24 hours post-transfection in HEK293T cells. Nucleotide modifications included: 100% 5mC, 100% Pseudouridine, or 50% 5mC + 50% Pseudouridine combined. (B) The same EMCV-encoded circRNAs were injected into mice (as previously described, 5 µg complexed to TransIT and injected intravenously) and bioluminescence imaging performed 8 hours post-injection. At the time of imaging 10 µg of furimazine substrate diluted in 200 uL of PBS was injected I.P. into each animal and images take 5 minutes post-substrate injection. Radiance quantification were both in BLI experiments were calculated using Caliper Live Animal Imaging Software (C, E). (D) In a separate animal experiment, cap and tailed non-IRES linear mRNA and EMCV circRNA was synthesized with or without 50% pseudouridine and injected into mice (N=2) at equimolar levels as in 13B.

We also tested whether this phenomenon was observed in vivo by injecting EMCV-NLuc circRNA with or without the indicated nucleotide modifications intravenously into mice (Figure 14B, quantified in Figure 14D). Again, we saw little to no signal above background from all of the modified circRNAs. Additional control studies with modified cap-dependent linear RNAs confirmed that this ablation only occurred in the context of IRES-driven circRNA but, as in the literature, boosted cap-dependent mRNA translation (Figure 14E, quantified in Figure 14F).

**Translation is extended in cells transfected with circular RNA**

Modified nucleotides have become synonymous with therapeutic protein-encoded mRNAs and are considered a requirement for therapeutic RNAs to function well in vitro and in vivo (Warren, 2010; Kormann, 2011; Mays 2013, Zangi, 2013). However, one group provided evidence that UTR selection, codon optimization, and HPLC purification could bypass the need for modified nucleotides (Thess, 2015). As our HPLC-purified, EMCV circRNAs had consistently shown high levels of protein expression in various systems in vitro and in vivo we wanted to test next how their expression kinetics compared to the industry standard modified linear mRNAs.
Figure 15: CircRNA extends protein translation \textit{in vitro}.

(A) 1 µg of Linear EMCV-NLuc mRNA (Linear -), circular EMCV-NLuc mRNA (Circular), and linear capped & tailed mRNA (Linear+) was used as template in rabbit reticulocyte lysate \textit{in vitro} translation reactions (Life Technologies). Reactions were incubated at 30 °C and at each time point 15 µL as removed from the reaction. 5 µL aliquots were added to a white 96-well plate containing 45 µL of PBS per well. NLuc substrate (diluted 1:50, Promega) was added to each well and luminescence measured after 5 minutes. The same circular and linear+ mRNAs were transfected into HEK293T cells and assayed for protein expression each day over a 3-day time course. (C) To see how long expression could be detected, HEK293T cells were transfected with the same RNAs described in 14A. However, in this experiment transfected cells were split at each time point before being assayed. For example, on Day 1 each well was split 1:2. Half of the cells were assayed the other half re-plated for further assaying. This was carried out on Day 1, 3, 6, and 9.

Protein kinetics derived from 3 different forms of mRNA was tested in the retic lysate \textit{in vitro} translation system (Figure 15A). In this setting, by hour 4, protein expression from linear mRNA without a cap or tail (Linear -/-) had plateaued. While linear mRNA with a cap, tail and NT modifications (Linear +/-) continued to increase at a similar trend as circRNA, its relative rise in protein expression were not as high as what was detected from circRNA. Comparatively, when protein expression kinetics were measured in cells the overall trends were somewhat similar.
Linear cap, tailed, and modified mRNAs peaked at Day 1 while circRNA peaked at Day 2.

In cells, when comparing EMCV-NLuc mRNA in linear (without cap or tail) or circular form the first time point frequently shows modified linear mRNA outperforming circRNA (Figure 15C). However, by Day 3 circRNA surpassed the modified linear RNA. Frustratingly, we noticed that even by Day 2 most cell lines cells were overgrown and beginning to die, so we attributed the reduction in protein expression more to do with this characteristic over the inherent stability of the given RNA. Even after the repeated splitting of transfected cells, which allowed us to perform the luciferase time course out to day 9, we could detect protein expression from circRNA-transfected cells.

2.3: Discussion

From these studies were able to show EMCV IRES is a potent inducer of protein expression in the context of circRNA. Although low levels of protein could be detected with the incorporation of a RRACH motif, the EMCV IRES generated protein at least an order of magnitude higher. Weak protein expression was observed by all cellular IRES tested. The differences in potency between viral and cellular IRES at translating protein could be explained by the diversity of environments each encounter. Viral IRES would benefit from being able to efficiently hijack and translate protein from any cell it may infect whereas cellular IRES appear to function in response to specific cellular states like stress or apoptosis, so its robust translation may not be desirable to the health of the cell. This is, of course speculative as the m6A and cellular IRES mechanisms of translation are still be elucidated.

Additionally, we were able to show that modified nucleotides cannot be incorporated into circRNA when it is dependent on an IRES for translation. Every modified nucleotide incorporated
into circRNA ablated protein translation *in vitro* and *in vivo*. We hypothesize that this loss of function is due to changes in secondary structure which viral IRES needs to maintain their function. It could also be possible that the modifications are affecting the binding affinity of certain ITAFs necessary for IRES translation. Although unsuccessful, we attempted to synthesize circRNA from two separate IVT reactions: an unmodified piece encoding the EMCV IRES and a modified piece containing the protein coding sequence. We predicted that this type of construct would have rescued the EMCV IRESs’ translational capacity in circRNA. Alternative approaches for generating this construct should still be pursued as modified nucleotides have also been shown to be beneficial in reducing cellular immune activation caused by the introduction of exogenous RNAs.

In Wesselhoft 2018, they showed that the Coxsackievirus B3 (CVB3) IRES, a viral IRES not conventionally used, was able to potently induce protein translation at levels higher than the EMCV IRES. We did not include the CVB IRES in our original panel due to the imposed size limitations we original implemented (<600 NT, the CVB IRES is over ~750 NT). In our preliminary experiments, we also found that CVB was more potent at inducing translation that EMCV in the cell lines tested (**Figure 16B**). Interestingly, the ablation in protein translation caused by the incorporation of modified nucleotides was also observed with the CVB IRES (**Figure 16A**). Finally, we did preliminary testing *in vivo* and found that the differences between EMCV and CVB IRES translation potency were less pronounced, but CVB did produce slightly higher levels of protein compared to EMCV in this setting as well (**Figure 16C**, quantified in **Figure16D**).

Compared to this recently published study we confirmed the potency of viral IRES like those derived from EMCV and CVB. We found that both IRES lose their functionality in the
presence of modified nucleotides. Despite this, we have shown through both of our \textit{in vitro} protein kinetic studies that circRNA can extend protein translation.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure16.png}
\caption{Modified nucleotides have similar ablation effect on CVB IRES. (A) CVB-NLuc circRNA was generated with one of the indicated modified nucleotide and transfected into various cell lines. Expression from the given construct was measured 24 hours post-transfection by luciferase assay as previously described. (B) Comparison of protein expression potency between EMCV-NLuc versus CVB-NLuc circRNA was measured in HEK293T cells 24 hours post-transfection. The indicated modified nucleotides were incorporated at a 50\% ratio. (C) Comparison of EMCV and CVB IRES was tested using two primary cells: PBMCs and primary mouse hepatocytes. PBMCs were transfected with TransIT-mRNA and Hepatocytes with Lipofectamine 2000. (D) Comparison of EMCV and CVB IRES was measured in mice injected intravenously with indicated mRNA and expression measured by BLI 8 hours post-injection.}
\end{figure}
Chapter 3: Examination of circRNA Stability and Immune Induction

**Contributions:** All experiments were conceived of by Ellese Carmona & Michael Goldberg and performed by Ellese Carmona.

### 3.1: Introduction

Here we detail the known sources of RNA instability, the strategies that have been utilized to overcome this instability in various therapeutic settings and finally how more robust modifications could be applied, primarily through RNA circularization, to enhance the overall efficacy of RNA as a therapeutic tool.

**2’ OH: the intrinsic source of RNA instability**

Deoxyribonucleic Acid (DNA) and Ribonucleic Acid (RNA) differ only by the presence of a hydroxyl (OH) group at the 2’ sugar moiety of RNA, which is absent in DNA. This single hydroxyl group is the primary intrinsic reason for RNA’s instability. Because of this 2’OH, RNA is unable to position itself into the stabilizing double helices that DNA forms causing every RNA to have a unique secondary structure derived from its sequence and internal homology. Furthermore, this hydroxyl group is highly reactive and susceptible to nucleophilic attacks that can lead to self-cleavage. These types of reaction are much more likely to occur in alkaline conditions, so it is crucial to keep RNA out of high pH conditions to prevent its spontaneous degradation. Storage solutions to maintain these neutral to high pH conditions are regularly used especially with RNA meant for therapeutic testing. In the therapeutic setting where RNA can be exposed to serum or the cellular cytoplasm, it must be ensured that RNA is protected from degradation. However, in these settings pH is relatively neutral, so the source of RNA instability most likely arises through separate mechanisms, like through cell-mediated mRNA decay.

**Cellular mRNA turnover mechanisms**
The machinery involved in eukaryotic mRNA turnover plays vital role in controlling gene expression, ensuring foreign RNA destruction, and destroying aberrant mRNA (Parker, 2004). Normal mRNA turnover is typically initiated by a deadenylase that explicitly targets and degrades the poly(A) tail. This ultimately leads to decapping and can results in 5’ → 3’ digestion by exonucleases like XRN-1. Aside from 5’ or 3’ end targeting a third mechanism utilizing endonucleases can also mediate RNA destruction. However, this form of decay has been shown to function primarily in the context of sequence-specific interactions, like AU-rich elements, or through miRNA or siRNA targeting (Dodson, 2002). Cells also have mechanisms that allow for recognition of aberrant transcripts known as nonsense-mediated decay and nonstop decay. A premature stop codon or no stop codon at all is quickly recognized by the cell and distinct mechanisms activated to ensure the immediate destruction of the aberrant transcripts. As circRNA does not have a 5’ or 3’ end, it should not be endangered by the exosome. However, care must be taken to exclude binding motifs that are recognized by known endonucleases or miRNAs.

**Immune-based sources of RNA instability**

Aside from molecular or cellular intrinsic mechanisms of RNA decay, there are a large family of innate immune receptors in eukaryotic cells that specifically recognize and destroy foreign RNA, including *in vitro* transcribed RNA. This was also a major hindrance to the development of synthetic RNA in the therapeutic setting initially. Little expression or therapeutic efficacy was observed due to the destruction of the introduced RNA due mostly to the molecules described here.

TLR 3, 7, 8 are the three toll-like receptors (TLR) within the TLR family that specifically recognize pathogen-associated recognitions patterns (PAMPs) specific to RNA. Evolved to defend against viral pathogens, these receptors also potently recognize and bind to *in vitro* transcribed
RNA. When bound to their cognate RNA motif a downstream signaling cascade is activated that induces the expression of interferon and proinflammatory cytokines. Although these receptors have been shown to bind to IVT RNA, they are expressed primarily on professional immune cells (Hayashi, 2003; Kadowski 2001; Krug, 2001). So, induction of an innate immune response in cell culture and particular animal tissues most likely has less to do with the TLR receptors and more likely to do with the cytoplasmic receptors expressed ubiquitously in mammalian cells: like those included in the RIG-like Receptor (RLR) family.

Table 2: Mammalian RNA-Sensing Immune Receptors

<table>
<thead>
<tr>
<th>Name</th>
<th>Expressed In/On</th>
<th>Cellular Localization</th>
<th>Target</th>
<th>MoA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLR3</td>
<td>Immune Cell; not neutrophils or pDCs</td>
<td>Endosomal</td>
<td>dsRNA (∼40 NT) poly I:C</td>
<td>IFN / NFkB induction</td>
</tr>
<tr>
<td>TLR7</td>
<td>Immune Cell; B cells, Mono/φ, pDCs</td>
<td>Endosomal</td>
<td>ssRNA, R-848 U-rich ssRNA</td>
<td>IFN / NFkB induction</td>
</tr>
<tr>
<td>TLR8</td>
<td>Immune Cells: Mono/φ, cDCs</td>
<td>Endosomal</td>
<td>ssRNA, R-848 GU-rich ssRNA</td>
<td>IFN / NFkB induction</td>
</tr>
<tr>
<td>RIG-I</td>
<td>Ubiquitous</td>
<td>Cytoplasmic</td>
<td>Short dsRNAs ppp-5’ dRNA LMW poly I:C</td>
<td>IFN / NFkB induction Caspase-mediated apoptosis</td>
</tr>
<tr>
<td>MDA5</td>
<td>Ubiquitous</td>
<td>Cytoplasmic</td>
<td>Long dsRNA (&gt;2000 NT) Bind as little as 15NT</td>
<td>IFN / NFkB induction Caspase-mediated apoptosis</td>
</tr>
<tr>
<td>PKR</td>
<td>Ubiquitous</td>
<td>Cytoplasmic</td>
<td>Hairpins Short dsRNA</td>
<td>EIF2α phosphorylation</td>
</tr>
<tr>
<td>OAS</td>
<td>Ubiquitous</td>
<td>Cytoplasmic</td>
<td>dsRNA</td>
<td>RNase L activation</td>
</tr>
</tbody>
</table>

The RLR family is comprised of 3 receptors: RIG-I, MDA5, and LGP2. They each contain a DEAD box and CARD domains, except for LGP2 which only contains a DEAD box. This family of cytoplasmic receptors appears to be expressed ubiquitously at low levels until their activation. The CARD domains present in RIG-I and MDA5 are responsible for signaling cascade that can
activate either a type I interferon response or caspase-mediated apoptosis. Although distinct, RIG-I and MDA5 both appear to recognize and be activated by different forms of double-stranded RNA and contain helicase domains which bind dsRNA and aid in its unwinding. These two receptors are likely to be the main culprits in foreign RNA recognition destruction as RIG-I was recently shown to specifically target exogenous circRNA in a recent study (Chen, 2017).

PKR and OAS are distinct from TLRs and RLRs in that their mechanism of action appears to directly target RNA either through inhibition of translation or global destruction of mRNA, respectively. PKR activation leads to the phosphorylation the translation initiation factor eIF2a which prevents its associated with the translation initiation complex, preventing translation. Separately, OAS has a direct function on global mRNA degradation in that when activated, it phosphorylates and activates RNase L, a potent endonuclease capable of globally degrading cellular RNA which during viral infection is the cells attempt to destroy the invading viral RNA. The products of RNase L are recognized by the RLR family which serves as a feedback loop in the cell’s response to foreign RNA. PKR and OAS can also be induced and activated by the other RNA-sensing receptors and have functions separate from inducing an adaptive response or activating programmed cell death.

**Strategies for bypassing immune recognition and inherent instability**

A major advancement in RNA therapy was work pioneered by Kariko which showed that the lack of modified nucleotides, which occur naturally in endogenous RNA, in IVT RNA was the primary source of the sensing by the various RNA immune receptors in the cell. Almost 100 different modified nucleotides occur naturally in endogenous RNA, most heavily in noncoding RNAs like tRNA and rRNA. It was also found that there was a correlation between less evolved species NT modification prevalence and the ability for that species’ RNA to induce a cellular
immune response. For example, bacterial RNA which is not heavily modified potently induced RNA-sensing receptor, while total cellular RNA does not. Furthermore, it was found that when cellular RNA was fractionated into cytoplasmic, nuclear or mitochondrial RNA the mitochondrial RNA, which is the least modified fraction of RNA in the cell and thought to be derived for a symbiotic bacterium, was the only RNA fraction that significantly induced RNA immunity.

Separately, it was found that byproducts that arise during in vitro transcription were a main source of immune activation, not the correct template being synthesized (Kariko, 2005). These byproducts were efficiently removed by HPLC (Kariko, 2011). Additional work has provided some evidence for alternative options to modified nucleotides in the therapeutic setting. This work showed that engineering RNA with specific UTRs as well as codon optimization in a setting by setting basis. In this design scheme, HPLC was still necessary to reduce RNA-based immunity (Thess, 2015).

**Designing circRNA for specific therapeutic settings**

The therapeutic efficacy of a given RNA must be modulated to fit the therapeutic setting. In terms of protein replacement therapies, prolonged protein expression would be desired, and the prevention of RNA-based immune induction should be avoided to prevent subsequent destruction. However, in vaccination settings (i.e., infectious disease or cancer), controlled induction of an immune response might increase the efficacy of such therapeutics. In the vaccination setting the RNA usually encodes an antigen, either pathogen or cancer-associated, and introduced to immune cells like dendritic cells in order to induce a response to that particular antigen. An immune reaction is multi-pronged and requires more than one mode of stimulation to become fully induced typically. In that sense, RNA can act as two different pieces of molecular information. The antigen it encodes which provides the desired immune specificity and the RNA itself which can act as an
adjuvant and secondary stimulant to ensure the immune response is efficient and robust. Here we begin to characterize the relationship between immune induction and stability of circRNA.

3.2: Results

*Modified nucleotides prevent immune recognition of circRNA*

The consensus in the field of RNA therapy is that modified nucleotides are essential to ensure functionality, especially for protein replacement applications. To date, virtually every mRNA-based protein replacement study incorporated modified nucleotides into their constructs to ensure functionality (Kormann, 2011; Mays, 2013; Zangi, 2013; Warren, 2010). Furthermore, strong evidence for a RIG-I specific response specifically induced by foreign circRNA has been proposed (Chen, 2017). Although we have learned through our protein translation studies that modified nucleotides ablate IRES-mediated translation in circRNA we still wanted to confirm that immune induction could be inhibited through the incorporation of modified nucleotides as this information would be useful for studies outside the realm of protein replacement applications.

**Figure 17:** Modified nucleotides prevent immune induction when incorporated into circRNA.
Figure 17 (continued): (A) HeLa cells were transfected with non-HPLC purified linear mRNA (Linear +: capped, tailed, and no IRES) or CircRNA (+EMCV IRES), with or without indicated modifications. IFNβ induction relative to untransfected cells (UNTR) was measured 8 hours post-transfection. Commercial Trilink mRNA (capped, tailed, & modified) and Poly I:C were included as controls. (B) At the same time point, NLuc RNA were measured to affirm equal initial levels of RNA. (C) EMCV-NLuc circRNA transcribed with or without indicated NT modifications were transfected into HeLa cells after RNase R treatment (-HPLC) or after RNase R + HPLC purification (+HPLC). NLuc protein expression was measured 24 hours post-transfection by luminescence assay. eGFP mRNA served as a negative control. (D) Non-HPLC purified samples without NT modifications were transfected into HeLa cells and RIG-I induction measured by qPCR 8 hours post-transfection.

We first confirmed in vitro that modified nucleotide incorporation into circRNA was sufficient to prevent an interferon response even without HPLC purification (Figure 17A). Our results mirrored that which has been shown in the context canonical linear RNA. This stark reduction was observed in all cell lines tested. Again, we confirmed that the immune induction we observed was reflective of equal starting levels of mRNA upon transfection (Figure 17B). HPLC purification combined with modified nucleotide incorporation has been shown to yield superior RNA templates for therapeutic use (Kariko, 2011). We checked to see if HPLC purification could rescue the ablation caused by modified nucleotides. Although HPLC consistently doubled protein expression it did not have any effect on rescuing modified circRNA translation (Figure 17C).

Next, we wanted to compare the extent to which our circRNA induced RIG-I. It has now been proposed that the RIG-I response to exogenous circRNAs occurs due to a lack of bound endogenous proteins, like splicing factors, which normally would prevent recognition by receptors like RIG-1. We found that both linear and circular forms of our RNA (with or without a CRC motif) potently induced RIG-I when modified nucleotides were not incorporated and HPLC purification not performed (Figure 17D). This varies slightly from what has previously been published as the RIG-I receptor was shown to specifically be induced by exogenous circRNA, and less so in linear RNA. As in the Chen paper, we observed no significant induction of MDA5 upon
linear or circular mRNA transfection even when RNAs were introduced unpurified to HeLa cells (data not shown).

**Synthesis and purification methods affect immune recognition**

One major difference between our studies and those previously published is that their RNA was generated using the PIE method of ligation. We speculated that the different ligation and purification methods used between these studies could be a source of these variable results. Support for this hypothesis came from previous experiments performed during the optimization phase of our ligation protocol in which we worked to identify the best purification method for our circRNA. One approach we attempted was a poly(A) tailing treatment of our the circRNA reactions followed by a negative selection of circRNA using an oligo(dT) column (**Figure 18A**). Although relatively efficient, we found that this purified product was more potent at inducing interferon beta in cells than even RNase R-treated circRNA (**Figure 18B**). This provided evidence to us that there were byproducts present in the reaction that act as potent inducers of immunity that that required removal. It can also be posited that these byproducts may be distinct from one ligation method to another which is what we focused our attention on next.

![Figure 18: Purification method can modulate circRNA immune induction](image)

**Figure 18: Purification method can modulate circRNA immune induction**

(A) An early strategy for removing remnant linear RNA from ligation reactions was to poly(A) tail the reaction which would selectively tail linear RNA at its free 3’ end. The sample was then run on an oligo(dT) column and the poly(A)- flow through collected for downstream use. (B) This
**Figure 18 (continued):** product’s immune induction, through IFNβ measurements, was compared to RNase R-treated and HPLC purified circRNAs when transfected into HeLa cells, 8 hours post-transfection compared to untransfected cells (UNTR).

To determine if different synthesis methods were behind the variable immune responses observed we generated circRNA using our RNA Ligase I method, the generic splint-mediated method, and the recently optimized PIE method for generating large RNAs (Weeselhoeft, 2018). All three ligation products were RNase R-treated before transfection in HeLa cells. In this setting circRNA generated from splint-mediated and PIE-mediated ligation methods had higher RIG-I induction than ligase I-generated circRNA (**Figure 19A**). When these same circRNAs were HPLC-purified there was a 2 to 3-fold reduction in IFNβ levels compared to the non-purified versions (**Figure 19B**).

circRNAs that are generated by Ligase 1 or splint-mediated ligation elute from the HPLC column primarily as a single peak (like **Figure 6D**) with additional small peaks representing non-specific product that arise during synthesis. In contrast, the PIE-generated circRNA elute several distinct peaks representing the target circRNA and the spliced-out intronic byproducts (**Figure 19C**). We collected the target circRNA peak as well as the intronic “byproduct” peak and transfected these isolated samples, along with the linear form of the same RNA, into HeLa cells and measured IFNβ induction (**Figure 19D**). As we expected the isolated byproducts were much more immunostimulatory than the actual target circRNA, after HPLC purification.

We found that even after RNase R digestion this “byproduct” peak was still observed in the resulting HPLC chromatograms. Even if digested PIE products looked “clean” on Tapestation or denaturing agarose gels these samples still product a strong “byproduct” peak when run the HPLC column, suggesting that conventional methods for confirming purity may not be sufficient. In our experience HPLC or PAGE gels extractions are the only methods sufficient to remove the
impurities necessary to significantly reduce RNA-based cellular immunity when modified nucleotides aren’t incorporated.

Figure 19: RIG-I is potently induced irrespective of mRNA form but more so by PIE-generated circRNA.

(A) We synthesized EMCV-NLuc circRNA using the 3 methods of ligation: RNA Ligase I, splint mediated, or PIE method. All 3 RNAs were RNase R digested and transfected into cells (without HPLC purification). RIG-I induction was measured 8 hours post-transfection in HeLa cells. Poly I:C used as a positive control. The same forms of RNA with and without HPLC purification were transfected again and IFNβ induction measured in HeLa cells, 8 hours post-transfection. (C) Representative chromatogram of EMCV-NLuc circRNA post-ligation reaction. The first peak is the intronic byproduct, the larger major peak represents final circularized product. Each peak was isolated separately for downstream analysis. (D) Linear and circular PIE mRNA, pre- and post-HPLC purification were transfected into HeLa cells. The byproduct was transfected at 25% of the total amount of circRNA transfected. IFNβ was measured 8 hours post-transfection relative to untransfected cells (UNTR).

From our ligation product immunity comparisons, we consistently observed that circRNA generated by RNA ligase I resulted in the lowest relative immune induction as measured by IFNβ. We next wanted to see if this reduction was as stark in primary dendritic cells. Dendritic cells work well in this system due to their inherent proficiency at up-taking foreign products and furthermore
they have been shown to be potently activated by exogenous RNA since they express all known endosomal and cytosolic RNA-sensing immune receptors (Kariko, 2005).

**Figure 20: HPLC-associated reduction in immune induction is observed in primary cells.**

RIG-I (A) and IFNβ (B) mRNA levels were measured in DCs by qPCR 24 hours post-transfection of EMCV-eGFP circRNA +/- HPLC purification (normalized to beta actin, relative to untransfected cells, UNTR). (C) Before qPCR sample isolation, supes (~100 µL) were collected and 50 µL of undiluted sample was used to detect IFNβ cytokine induction by the same primary DCs transfected with the indicated circRNAs that had all been HPLC purified. (D) The remaining 50 µL of supes were used to detect Epo protein secretion by ELISA.

In this cell-type HPLC purification lead to a reduction in both IFNβ and RIG-I expression relative to non-purified RNAs (Figure 20A-B). We also attempted to measure IFNβ cytokine secretion from these cells however all HPLC-purified RNAs failed to induce enough cytokine production to be detectable by ELISA (Figure 20C). This was not due to failed transfection as cytokine induction was detected from poly I:C transfected samples and a portion of the same supes were used for detecting Epo expression, which was successfully detected (Figure 20D). It is clear that immune characterization based on RNA levels alone are not a sufficient proxy for a potent immune response. Although a single gene may be strongly induced it is unclear whether that...
induction will lead to an actual robust cellular immune response. Additional testing such as through cytokine detection or cellular toxicity are necessary to full characterize the extent to which HPLC-purified circRNAs are truly immunostimulatory.

It appeared from the data presented thus far that HPLC purification might be sufficient to reduce immunity-induced destruction of our circRNA. This was supported not only by the immune reduction consistently observed from the qPCR data but also from the improved cellular viability when HPLC-purified circRNAs were utilized. We consistently observed that by Day 3 cells transfected with unmodified, non-HPLC purified circRNAs were mostly dead, irrespective of the cell lines tested (Figure 21). This cellular death could be equally prevented by either the incorporation of modified nucleotides or by HPLC purification of circRNA. There was no significant difference in cell death between PIE or Ligase-1 generated circRNA if they were HPLC purified. However, unpurified PIE circRNA resulted in higher cellular toxicity compared to non-purified Ligase-1 circRNA. As the cell lines became overconfluent, usually by day 3, cell death occurred at an increasing rate irrespective of NT modification, mRNA form, or purification method. It was difficult to generate reproducible data after this time point so most in vitro kinetic assays were only extended to 3 days post-transfection with the exception of our splitting assays show in Figure 15C. We next wanted to characterize the extent to which purified circRNA extended RNA kinetics in vitro.
Figure 21: HPLC is sufficient to reduce cellular toxicity in circRNA-transfected cells.
HeLa cells were transfected with circRNA generated by PIE method after RNase R Digestion alone or after RNase R + HPLC purification. The same sample preparation for circRNA generated by the Ligase I method were transfection as well. Cells were stained with Dapi for 5 minutes prior to fluorescence microscopy imaging (Celigo Cytometer and Imager). Images were taken 1 and 3 days post-transfection. Blue cells are dead.
Stability conferred by circularization

We first tested circRNA stability in different cell-free systems including serum (Figure 22A-B) and rabbit reticulocyte lysate (Figure 22C). Serum contains a mixture of RNases presumably with both exo- and endo- nuclease activity, so we inferred that the extent to which circRNA enhanced stability in this system would be minimal. Surprisingly, we saw that linear RNAs were extremely sensitive to degradation in serum as we were unable to detect intact linear RNA even after just a few minutes in serum. In stark contrast, circRNA persisted in serum for up to 4 hours. The eventual degradation of circRNA could have been due to changes in serum pH over this time frame or the latent activity of endonucleases.

Figure 22: Circularization confers protection against degradation in serum or cell lysate.
A 500 NT long RNA was generated in 3 forms: linear (no cap or tail), linear (with cap and tail), or circular. 5 µg of each RNA was aliquoted into 6 equal samples that contained PBS alone (A) or 1% FBS diluted in PBS (B). At each indicated time point the RNA was purified. At the last time point all RNA samples were run on an agarose gel (45 min 100 V) and imaged by UV shadowing. (C) Equimolar levels of Linear+ or circular NLuc RNA were added to rabbit reticulocyte lysate reactions. At each time point 5 µL (technical triplicates, 15 µL total per time point) of each sample were used as template in cell-to-cT reactions to generate cDNA. All cDNA samples were used as template in qPCR reactions using inward-oriented NLuc primers alongside serially diluted NLuc samples, which were used to generate a standard curve to calculate copy number per well for each sample. mRNA levels are shown relative to each samples first time point.

We next studied circRNAs stability in rabbit reticulocyte lysates. These lysates are known to contain many of the exosome nucleases which are primarily exonucleases with 3’ → 5’ processivity. We predicted that circRNA would be even more stable in this setting. Indeed, while
there was a steady reduction in the amount of linear mRNA over a 24-hour incubation there was no detectable reduction in circRNA levels for a full 24 hours. Linear mRNAs are generated with a poly(A) enzyme after transcription resulting in a long 100-200 NT long tail. This may explain why linear RNA was able to persist for as long as it did in this system.

The results from our cell-free system experiments provided support for circRNAs predicted resistance to exonuclease degradation but were limiting in scope due to the lack of RNA sensing receptors that could potential target and destroy circRNA in these settings. Aside from cytokine induction there are immune mechanisms that could destroy or prevent translation. For example, if OAS recognizes and binds to our circRNA its activation would lead to RNase L activity which has capacity to globally degrade cytoplasmic RNA. Additionally, PKR activation would lead to the direct inhibition of translation via Eif2a phosphorylation. So aside from immune mediated cell death there are internal cellular mechanism for destroying or inhibiting the function of exogenous circRNA.

To begin testing cellular kinetics of our circRNA we transfected EMCV-eGFP circRNA in HEK293T cells and tracked RNA levels over a 3-day time course by RT-PCR using circ-specific (outward-oriented) and target specific (both linear and circular RNA) primers (Figure 23A). Commercial linear eGFP and untransfected cells were used as controls. Remarkably, we observed that with optimal commercial linear mRNA there an observable day-by-day reduction in RNA levels. However, this reduction was not observed from our circRNA. Our outward-oriented primers confirmed the successful transfection of intact circRNA which were not generate amplicon in either untransfected or linear RNA samples. Additionally, beta-actin was used as an internal control to confirm correct RNA isolation, cDNA synthesis, and equal template input. So, although
circRNA induced elevated expression of RNA immune genes *in vitro* even after HPLC purification, it appeared that purified circRNA was able to evade immune-based destruction.

We also did preliminary experiments to track RNA kinetics *in vivo* (More extensive experimentation is presented in Chapter 4 results). We injected linear or circular eGFP mRNA into mice, intravenously. 1, 2, and 3 days post-injection we sacrificed the animals and isolated total RNA from their livers, which were then used as template for qPCR detection of the target RNA injected. We found that even by Day 1 there was roughly a 4-fold difference between linear and circular RNA levels detected in this tissue (Figure 23B-C). Up to Day 3 circRNA seemed to persist whereas by Day 2 linear RNA levels had dropped by more than 70% compared to Day 1. When this data is plotted relative to each RNAs Day 1 levels to difference in kinetics becomes even more pronounced (Figure 23D). Taken together, these results provide support for the enhanced stability conferred by circularization of RNA.

![Figure 23: CircRNA persists *in vitro* and *in vivo* significantly longer than linear RNA.](image-url)

63
**Figure 23 (continued):** (A) Hek293T cells were transfected with commercial, linear eGFP mRNA (Trilink) or our EMCV-eGFP circRNA. At 6, 24, 48, and 72 hours post-transfection cells lysates were collected, and cDNA synthesized (Life Tech, Cell-to-Ct). RT-PCR was performed on samples using inward-oriented, outward-oriented, and actin-specific primers. Samples were run for 20 cycles and 2 µL diluted into TBE loading dye and run on a 10% TBE gel at 150 V for 45 minutes. The resultant bands for each primer set are shown in A. Untransfected cells were used a negative control. (B-C) The same eGFP mRNAs were injected into mice complexed to TransIT (Mirus) at equimolar levels (N=15). TransIT along was injected into mice as a negative control (Mock). At each time point mice were sacrificed, and livers collected. Total RNA was isolated from liver tissues via Trizol extraction. cDNA was synthesized from µg of total RNA for each sample using SuperScript IV (Thermo Fisher). Sample were diluted 1:4 and 2 µL of cDNA used as template in qPCR reactions. To measure absolute levels, copy number/well was determined by standard curve using template NLuc vector DNA. (D) Relative levels compared to each RNA types DNA 1 levels was plotted.

### 3.3: Discussion

Here, we have shown that the ligation methodology used to synthesize circRNA can have variable impacts on cellular immune activation. Although it appears that modified nucleotides are the best option for preventing immune induction caused by foreign RNA, the incorporation of these modified nucleotides ablate IRES activity necessary for circRNA translation, as shown in Chapter 2. So, alternative methods for tamping down cellular immunity was necessary. We were able to sufficiently reduce immune induction by HPLC purification of our circRNA.

RNA Ligase 1-mediated ligation is the only method that doesn’t introduce or produce byproducts during the ligation step in synthesis, which may explain why this method consistently induced immune genes to a lesser extent than the other two ligation methods. However, it has been shown that many of the immune-stimulatory byproducts arise before the ligation step, during *in vitro* transcription (Kariko, 2004). These byproducts even at very small amounts have the capacity to potently induce cellular immunity. The incorporation of modified nucleotides during *in vitro* transcription may also incorporate into these byproducts rendering them less immune-stimulatory, which may explain why their presence in non-HPLC purified samples do not induce immunity. As we could not efficiently generate modified circRNA using the PIE method due to reduced ligation...
efficiencies (Figure 9D) we could not confirm whether their incorporation lead to the same reduction in immune activation as seen with linear RNA and enzymatically-generated circRNAs.

Finally, we showed evidence for circRNAs enhanced stability in numerous settings by tracking RNA kinetics in serum, cell lysate, cell lines, and mouse tissue. It appears from these studies that the lack of modified nucleotides in circRNA is not affecting its functionality. However, further characterization in vivo needed to be performed to verify this claim.

Chapter 4: Therapeutic Efficacy of circRNA in vivo

Contributions: All experiments were conceived of by Ellese Carmona and Michael Goldberg and performed by Ellese Carmona. Abigail Deloria aided in in vivo experimentation. Chun Gwon Park aided in IVIS analysis. Vera Greiner aided in experimental design.

4.1: Introduction

A growing body of work has shown mRNA-based therapeutics are a promising new potential class of drugs, despite the longstanding belief that RNA was too unstable to be efficiently utilized for such applications. The modalities in which mRNA therapy can be applied are vast and include cancer immunotherapy, infection disease vaccines, protein replacement therapy, genetic engineering and genetic reprogramming (reviewed in Sahin, 2014). Due to RNAs perceived instability this molecule was mostly overlooked for more stable alternatives like DNA or protein-based templates. While circumstantially successful, both DNA and protein come with their own set of limitations in the therapeutic setting. Here we will discuss the unique advantages of using RNA over DNA and protein, the achievements that have occurred using mRNA thus far, limitations that still need to be overcome, and the rationale behind our proposed improvement to this molecular therapy.
Unique advantages of RNA

The limitations associated with DNA (viral or vector)-based therapies are low gene transfer efficiency, inefficient entry into the nucleus, and the threat of detrimental genomic integration. Of note, a particular clinical trial used a retroviral-based gene therapy to cure X-linked Severe Combined Immunodeficiency (X-SCID). Although initially successful, this treatment lead to genomic integration within the oncogenic gene, LMO2, resulting in patient death (McCormack, 2002). Significant effort has been made to prevent this type of detrimental side effect from occurring again and when fully developed this modality will undoubtedly serve as a curative tool for applicable diseases. Indeed, the first viral (AAV)-based therapy has recently been approved by the FDA for the treatment of retinal dystrophy caused by a hereditary mutation in the human retinal pigment epithelium-specific protein, RPE65. However, permanent or extensive expression of therapeutic proteins, like with the RPE65 treatment, are not always desirable. For example, in the setting of genetic reprogramming where mRNA can encode the necessary transcription factors to induce pluripotency, the permanent or long-lived expression of these factors would be undesirable. This is just one of the characteristics that make mRNA an attractive modality over DNA or protein-based therapies in particular settings.

Protein therapy has shown successes for a variety of extracellularly-functioning proteins, most notably with the achievement of recombinant insulin for the treatment of diabetes. However, this modality comes with its own set of limitations including its very short-lived activity, its inability to supplement for proteins that function intracellularly, its extensive production time and costs, and finally its lack of endogenous post-translation modifications. The latter limitation has caused issues in pre-clinical and clinical studies specifically in setting where repeated doses were necessary to elicit a therapeutic effect (Gao, 2004; Casadevall, 2002). In these studies, the
emergence of neutralizing antibodies has been observed and caused adverse effects to patients in the clinical study. Furthermore, it is difficult to predict what glycosylation types induce this immune response and in vivo studies are not a good indicator of the potential for this response to occur due to species to species variability in immune recognition (Kolarich, 2006; Worobec, 2004).

Distinct from DNA and protein, RNA does not suffer from the discussed limitations as there is no threat of genomic integration, expression initiates immediately after entry into the cytoplasm, RNA is transient in nature and its expression can be modulated to the appropriate therapeutic setting, the RNAs can translate proteins that function intracellularly, and since the protein it generates is transcribed endogenously autologous glycosylation would occur reducing the potential for neutralizing antibodies to be generated. Despite these numerous advantages, mRNA was not being intensely developed for therapeutic use until recently. This was due to two significant limitations: its perceived instability and its capacity to potently induce cellular immunity. The mechanisms behind these two limitations are covered extensively in Chapter 3. Suffice it to say that the initial disadvantages associated with RNA have largely been overcome and the door now been opened for mRNAs full development across numerous therapeutic setting.

Development and successes thus far

By far, the cancer vaccine and immunotherapy applications of mRNA therapy have advanced the furthest with multiple studies already reaching clinical trial. In most studies the tumor antigen-encoded mRNA is transfected or electroporated into autologous dendritic cells or T cells which induce anti-tumor immunity when put back into the patient (Boczkowsk, 1996). A separate approach that has shown promise is the direct injection of antigen-encoded mRNA intranodally or intradermally. Promisingly, these studies have shown excellent safety profiles and no off-target
side effects have been observed as of yet. The advancements in this particular area over other modalities like protein replacement are due to these studies not being hindered, but helped, by the immunostimulatory nature of in vitro transcribed, unmodified RNAs. The induction of an immune response is multi-pronged and requires more than one mode of stimulation to become fully induced typically. In that sense RNA can act as two different pieces of molecular information: (1) the antigen that it transcribes which provides the anti-tumor specificity and (2) the RNA itself which can act as an adjuvant and secondary stimulant to ensure the immune response is efficient and robust.

The protein replacement modality of mRNA therapeutics is a newer area of study and as such has not yet reached the clinical trial stage of development. However, with the finding that modified nucleotides could prevent immune induction and the characterization of motifs that aid in RNA stability a burst of exciting and informative studies have been published. The breadth of models therapeutic mRNA have been testing in include: intratracheal spraying of modified SP-B mRNA in SP-B deficient mice to prevent lethal lung disease (Kormann, 2011), intracardial injection of modified VEGF-A mRNA to promote vascular regeneration after myocardial infarction (Zangi, 2013), intratracheal spraying of modified Foxp3 mRNA to treat and prevent allergic asthma (Mays, 2013), and intraperitoneal injections of modified Epo mRNA to induce erythropoiesis (Kariko, 2012). Taken together mRNA has emerged as an exciting new class of drugs, but there are still issues that must be addressed in order for it to reach its full therapeutic potential.

**Potential improvements to RNA therapy**

In many, if not all, mRNA therapy-based studies the authors independently state that mRNA is best suited for situations in which robust, short-term expression is needed due to the
Transient nature of this molecule even in its most stable iteration. Even though the permanent expression is not desired these studies all required continual dosing of the mRNA to maintain therapeutic levels of protein. An mRNA that could be administered less frequently or at smaller doses to reach therapeutic levels is very desirable. To that end, work that has focused on characterizing endogenously expressed circular RNA provide evidence for its potential use in the mRNA therapeutic setting. Its stability, which is on average 2-5 times higher than that of its linear counterparts (Jeck, 2013), could serve as a useful addition to the arsenal of modifications currently used in the context of mRNA therapeutics. Studies characterizing endogenous circRNA and the work presented here provide support for the utility of circRNA in the mRNA therapeutic setting due mainly to its enhanced stability across multiple systems. We hypothesize that circular mRNA can be developed into a new class of long-term mRNA-based therapeutics, which have the potential to be applied to a litany of diseases. Although we can foresee the utility in most if not all therapeutic modalities of mRNA therapy we decided to first test circRNAs utility in the protein replacement setting for our proof of concept in vivo studies.

4.2: Results

Circular RNA extends protein expression longer than linear modified RNA in vivo

We did extensive confirmatory experiments with our initial animal studies to ensure that our delivery and detection methods were optimal as the existing literature for these applications severely limited. This first entailed optimizing injection methods and appropriate complexing agents for intravenous injection into mice (Supplemental Figure S6). In our preliminary animal studies our starting time points were 1-day post injection however we noticed that in early BLI kinetic studies, protein expression after luciferase-encoded mRNA injection tended to peak anywhere between 6-12 hours post-injection (Phua, 2013; Pardi, 2015; Stadler, 2017) with little to
no expression persisting past the 24-hour time point. If these levels were reflective of the amount of RNA still present in the tissue, we could be choosing a time point in which much of the linear RNA has already been degraded. We found that by taking an earlier time point, 4-6 hours post-injection, the relative levels of linear and circular RNA detected in the liver and spleen of injected mice were roughly equal. This allowed us to rule out the notion that the differences in observed signal intensity was due to disparate levels of starting material at the time of injection but were instead reflective of the inherent translational capacity of the RNA being tested. We consistently found that at this time point both linear and circular mRNA were present at relatively equal levels (Figure 24A) in both the liver and spleen. We also were able to confirm that circRNA remained intact in this system by using outward-oriented primers. No amplicon was detected from the tissue of mock-transfected or linear mRNA-transfected animals when using outward-oriented primers (Figure 24B).

We sacrificed animals once we could no longer reliably detected protein signal from mice injected with circRNA, which in this experiment occurred on Day 20 (Figure 24F, quantifications in Figure 24G). We were able to detect circRNA in both tissues at Day 20, but not linear RNA. We were also able to generate primers that worked to detect mouse IFNα from the isolated animal tissues. Samples at the 6 hour and Day 20 time points for both liver and spleen were used to measure difference in IFNα induction relative to each group’s mock-transfected tissue levels (Figure 24E). We observed a 15-fold increase in IFNα for animals transfected with circRNA over mock, specifically in the spleen but not the liver. By Day 20 all IFNα immune levels had dropped back down to baseline levels despite the persistence of circRNA at that same time point, suggesting a transient immune response that did not lead to the destruction of circRNA.
Figure 24: Initial characterization and confirmation of circRNA persistence in vivo

3.5 pmoles of mRNA (~2.5 - 5 µg), complexed to TransIT (Mirus), was injected intravenously into BALB/c mice (N=6). Mock treatment group was injected with TransIT alone (N=2). 6hr and 20 days post injection mice were sacrificed (3 mice per time point), livers and spleens harvested, and total RNA isolated for downstream qPCR analysis. Inward oriented (red; A, C) and outward-oriented primers (blue; B, D) were used to amplify all target RNA or circRNA alone, respectively. Mouse IFNα primers were used to measure IFNα induction in all samples relative to each time point and tissues mock-transfected levels (E). 6 hours, 5, and 20 days post-injection mice were each injected with 10 µg of furimazine substrate diluted in 200 µL PBS and imaged 5 minutes post-injection (1 min exposure, medium binning, f-stop 2) using Xenogen BioImager (F). Quantification of radiance levels at each time point was determined using Caliper Living Images Software (G). (-): Mock-transfected, (+Mod): with cap and tail, 50% 5mC – 50% Ψ). Red lines in C and D indicate limit of detection.

We performed more substantial repeats of the BLI kinetic studies once we had confirmed optimal injection and detection methods (Figure 25). In these protein kinetics studies, we doubled the amount of mRNA injected (~7 pmoles) and included additional linear RNA controls. The
Linear- mRNA used was identical in sequence to the circRNA but left unligated. Because of the presence of the EMCV IRES, Linear- mRNA showed similar levels of protein expression at the first time point, however this expression was quickly lost most likely due to the instability conferred by its lack of 5’ cap or poly(A) tail. Canonical linear RNA containing a 5’ cap and a poly(A) tail with or without modifications were also tested. The previously noted increase in translation capacity when modified nucleotides are incorporated into cap-dependent mRNA can be observed by the differences in signal intensity at the 6-hour time point.

Figure 25: CircRNA translates protein robustly *in vivo* and provides extended protein expression compared to modified linear mRNA.
**Figure 25 (continued):** 7 pmoles of mRNA (~5-10 µg), complexed to TransIT (Mirus), was injected intravenously into BALB/c mice (N=3). Mock treatment group was injected with TransIT alone (N=2). 6 hours, 3, 6, 10, 15, and 20 days post-injections mice were injected with 10 µg of furimazine substrate diluted in PBS and imaged 5 minutes post-injection (1 min exposure, medium binning, f-stop 2). The linear- and circular RNA are identical in sequence and vary only in the status of their 5’-3’ end linkage; the linear form has ends that are accessible for exonuclease-mediated degradation, while the circular form’s 5′ and 3′ ends have been covalently ligated. The linear+ mRNAs do not contain an IRES but the conventional additions necessary for canonical, cap-dependent translation (5’ cap and poly(A) tail). These two mRNAs only differ in the presence of modified nucleotides incorporated into their sequence. (-): no cap or tail, (+): with cap and tail. NSD: No signal detected.

All previously published work utilizing BLI to track mRNA kinetic *in vivo* show signal intensity peaking between 6-12 hours, dropping considerably by Day 1 and reaching undetectable levels by Day 3. Excitingly, signal had only slightly dropped in circRNA-injected mice at Day 3. Signal could be detected in both Linear+ mRNA-injected treatment group but at roughly a log lower level of intensity that circRNA. Minute levels of signal could still be detected in our Linear+ samples at Day 10 which is a vast improvement over what has previously been published. We attribute the increase in stability of our Linear+ RNAs to the inclusion of HPLC purification and the optimized UTR elements present in these constructs. The time in which circRNA could be detected was twice as long as either of linear RNA samples, with signals persisting out to Day 20. The radiance quantification for this experiment can be found in **Supplemental Figure S7**.

**Circular enhances therapeutic efficacy of RNA therapy in vivo**

An established method for measuring the performance of therapeutic mRNA *in vivo* specifically in the context of protein replacement was recently published (Mahiny, 2016). In this study, the authors injected Erythropoietin (Epo)-encoded mRNA, a cytokine that induces hematopoiesis, into mice and measured Epo protein levels in the blood and resultant changes in reticulocyte %, with a predictable response occurring over a 7-day time course. This simple readout
and the ability to synthesize the same RNA used in their study make it a desirable proof-of-concept model to compare the functionality of our circRNA to the industry standard.

We first verified the correct synthesis and protein expression capacity of our Epo-encoded RNAs in multiple cell lines (Supplemental Figure S8) prior to the initiation of in vivo testing. As supplemental work to our NLuc in vivo studies we performed more robust RNA kinetics and immunity-based studies with the newly synthesized Epo-encoding constructs to determine if any variations were observed from one coding sequence to another.

**Figure 26: Circularization extends persistence of RNA encoding Epo in vivo.**

(A) 3.5 pmoles of Linear-, Linear+ Mod, or circRNA encoding Epo complexed to TransIT was injected intravenously into BALB/c mice (N=18). At 6hr, D1, D3, D5, D7, and D10 three mice from each treatment group were scarified and total RNA isolated from their liver and spleen. qPCR was performed to detect the changes in RNA levels over this 10-day time course. Levels were normalized to beta actin and made relative to their respective treatment’s Day 5 levels. (B) Whole blood was collected at the 6-hour time point via cardiac puncture. Plasma was isolated from the blood samples and used to detect circulating IFNα by ELISA assay. (C-D) Total RNA from the first and last time point for both liver (C) and spleen (D) was analyzed for IFNα induction by qPCR. Each sample was normalized to beta actin and calculated relative to each time and time points mock treatment group.
We characterized the differences in RNA kinetics of Epo circRNA and linear RNAs over a 10-day time course (Figure 26A). By Day 10 no linear RNA was detected in animal tissue as Ct values had reach or surpassed the Ct values detecting mock-injected animal tissues. However, the circRNA was still detectable, albeit at levels roughly 50 times lower than its 6hr time point. We were able to detect variations in circulating IFNα, 6 hours post injection (Figure 26B). As expected the treatment groups that did not contain nucleotide modifications has the highest level of IFNα induction. However, the treatment group that received unmodified linear RNA had roughly ten times more circulating IFNα than mice injected with circRNA. We also measured IFNα induction at the first and last time points (6hr and 10 days post-injection) collected in both the liver (Figure 26C) and the spleen (Figure 26D). Similar to previous experiments we found that IFNα was more potently induced in the spleen compared to the liver. Unmodified linear and circRNA induced IFNα 2-4 times more than modified linear RNA in the spleen. However, by Day 10 this immune induction had gone back down to baseline.

Figure 27: Physiological response derived from Epo-encoding circRNA. (A) 3.5 pmoles of mRNA was injected intravenously into BALB/c mice (N = 4). A single mouse from each treatment group was sacrificed and blood isolated via cardiac puncture at 6, 24, 72, and 96 hours post-injection. Serum was isolated from these samples and used to detect Epo protein in the blood by ELISA. Recombinant Epo protein (rEpo) was used as a positive control, 2 µg/mouse injected I.P. (B) 3.5 picomoles of mRNA were injected intravenously into BALB/c mice (N = 4). Blood draws (<20 µL) were taken at 6 hours, 1, 3, 5 and 7 days post-injection and used to determine
Figure 27 (continued): reticulocyte levels at each time point. Linear mRNA had been capped, tailed, and generated with modified nucleotides (50% 5mC/ψ). rEpo: recombinant Epo protein. Mock: TransIT alone. Data represented as mean plus standard deviation****p < .0001.

When linear mRNA (cap and tailed, +modified nucleotides) or circRNA were injected intravenously into BALB/c mice, we found that circRNA produced detectable levels of protein in mouse serum for up to 72 hours after a single injection (Figure 27A). Both linear mRNA and recombinant Epo protein (rEpo) were only detectable up to 48 hours post-injection. Reticulocyte counts in mouse blood were determined using an established flow-based assay (ReticCounter, BD Biosciences) after a single injection of linear or circRNA (Figure 27B). rEpo or transfection reagent alone (mock) were used as positive and negative control, respectively. At the first time point every treatment group had roughly identical levels of reticulocyte counts (~5%). By day 1, only mice injected with rEpo showed slightly elevated reticulocytes counts. By day 3 rEpo had peaked and linear and circRNA treatment group had begun to show elevated reticulocyte levels. By day 5, and maintained at day 7, the circRNA treatment group had surpassed reticulocyte levels present in the linear mRNA treatment group. By day 10 all treatment group had returned to baseline reticulocyte levels (data not shown).

4.3: Discussion

Here we provide additional evidence that the immune induction brought on by unmodified nucleotides is transient and does not lead to the destruction of target RNA. Our qPCR data provides evidence for circRNAs persistence in vivo over modified linear RNA, which has not been shown until now. This extended circRNA persistence mirrors the prolonged protein expression observed from both NLuc and Epo circRNA compared to their modified linear mRNA counterparts as measured by BLI and Epo ELISA, respectively. Although modest, these results provide the first evidence for circRNAs capacity to outperform modified linear mRNA at inducing a physiological
response \textit{in vivo} as shown in by the monitoring of circulating reticulocytes after a single Epo mRNA injection.

If time had permitted, the next \textit{in vivo} experiments would have involved dosing and multi-dose studies to determine if the enhancements provided by circRNA are more pronounced outside the context of a single injection. Furthermore, comparing circRNAs protein kinetics from different administration routes (intradermal, intramuscular, subcutaneous, etc.) would be very informative. Although limited in scope, these studies provide support for the continued development of circRNA in the context of mRNA therapeutics.

\textbf{Future Directions}

\textit{circRNA’s utility in the field of mRNA therapeutics}

Our improvements to T4 RNA ligase 1 mediated circularization methodology now allows for the efficient ligation of RNAs up to 4 kB whereas previous to the start of this study the largest RNAs generated using this method were roughly 350 NT. Extensive comparisons of the updated PIE method, shown briefly here, need to be carried out to determine which methodology is better suited for the therapeutic setting. It may be that both could function well in different setting, especially as our data suggests that methodologies results in disparate immune response if HPLC purification is not carried out.

It seems apparent that the ablation of IRES translation we observed is due to alterations in its secondary structure which play a role in its functionality. These modifications might also affect recognition and binding of ITAFs. Thus far, all IRES tested showed a complete loss of activity when these NT modifications are incorporated. Interestingly, there was one modification that didn’t have as big of an effect on inhibiting IRES translation: Cy5-UTP. This modification is much bulkier than standard modifications and yet it did not have the same inhibitory effect as m6A or
5mc, weakening the hypothesis that alterations to secondary structure are the cause of IRES activity ablation. Additional studies that shed light on this mechanism of inhibition would be desirable.

Our work shed light on the role impurities present in RNA samples may play in modulating the extent of RNA activation. Although there is no doubt that the exogenous RNA itself can be recognized by RNA sensing receptors our work suggests that the intact RNA alone is not a potent inducer of cellular immunity, but instead the byproducts generated during the synthesis process are the culprit for the potent immune induction associated with in vitro transcribed RNAs. Although circRNAs, generated either by enzymatic or non-enzymatic methods, will undoubtedly provide utility in the mRNA therapeutic setting it may be that certain modalities prefer one synthesis method over another based in part on the extent by which immune activation is desired.

It will also be exciting to see the characterization of circRNAs functionality outside the scope of the Epo model, for example in cancer vaccine or genetic reprogramming models. Both modalities will likely have unique specificities regarding translation and immunity, however, both could potentially be bolstered by the enhanced stability conferred by circRNA. Finally, this methodology will allow for the efficient ex vivo synthesis of endogenous circRNAs, which could help in functional and characterization studies as more than 99% of these putative molecules do not currently have a known function.

**The potential utility of circRNA beyond the mRNA therapeutics modality**

circRNA has potential utility in both RNAi and mRNA-based therapeutics serving either as a regulatory or protein-coding element, respectively. In this study we focused on developing the criteria necessary for circRNAs use in the mRNA therapeutic setting, which meant characterizing properties that conferred robust translation, reduced immunity, and enhanced stability in vitro and
in vivo (covered in Chapter 2-4). At least some of the information garnered from this study will allow us to better understand and develop the design criteria necessary for circRNA’s use in RNAi-based applications. A recent study was the first to engineer synthetic circRNA to function as a microRNA sponge (Jost, 2018).

![Figure 28: Circular RNA recapitulates its proposed endogenous function as a molecular sponge in vitro](image)

(A) An RLuc-encoded vector containing mir20 binding sites in its 3’UTR was co-transfected with linear or circular RNA also containing mir20 binding sites into HEK293T cells, which endogenously express mir20. An Fluc-vector was transfected as well to quantify relative expression to RLuc. Control sponge RNAs containing CXCR4 binding sites (HEK293T cells do not express CXCR4 microRNAs) were also synthesized in linear or circular form. Luminescence of RLuc and Fluc were measured 24 hours post-transfection using the Dual Glo Assay Kit (Promega). (B) Control experiments in which the RLuc vector contained CXCR4 binding sites shows no effect in repression or induction with any treatment type, as expected. (C-D) Day 2 levels are shown for pRLuc-Mir20 (C) and pRLuc-CXCR4 (D) transfected HEK293T cells as well.

Preliminary results where our circRNA is used in this regulatory setting showed promise. In early experimentations, we used a Dual-Glo (Promega) luciferase assay (as published in: Ebert,
2007) to compare the functionality of circRNA sponges to linear and DNA sponges. circRNA over a two-day experiment performed slightly better than linear RNA sponges and DNA sponges in preventing mir20 repression of RLuc (Figure 28A, C). This repression was not observed when these sponges contained non-cell specific binding site for CXCR4 (Figure 28B, D). While circRNA showed comparable efficacy to linear RNA as a microRNA sponge, this experimentation was performed during an early iteration of our circRNA synthesis methodology. If we were to repeat these experiments now using our optimized synthesis and purification methodology, we are certain that circRNA would function significantly better than linear RNA and provide improved recapitulation of its proposed endogenous function.

The recent publications utilizing engineered circRNAs are surely just the first in a litany of studies working towards the development of circRNA as a therapeutically relevant agent. Interestingly, microRNA sponges and protein-coding mRNAs are not the only proposed applications for therapeutic circRNA, other potential effector functions include: an agent for boosting innate immunity, circRNA aptamers, and trans-regulators of transcription or splicing (reviewed in Holdt, 2018). It will be very interesting to see how circRNAs utility develops alongside the exponential growth of the mRNA therapeutics field and the more established RNA (regulatory) therapeutic modality.
Methods

Cell Lines and Transfections

*Cell Lines*

HEK293T, HeLa, HepG2, H1299, Calu6, B16, EO771, and Hep3B cells (ATCC) were cultured at 37 °C and 5% CO2 in Dulbecco’s Modified Eagle’s Medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (Gibco) and 1% penicillin/streptomycin (Sigma). Cells were passaged every 2–3 days and were transfected in 96-well plate format at ~80% confluency.

Purification method for each experiment is indicated within the figure and could include RNase R digestion, HPLC purification, or both. Each cell was transfected using Lipofectamine 2000 (Invitrogen) complexed to no more than 100 ng of RNA according to the manufacturer’s instructions. For all RNA datasets, equimolar quantities of each RNA (equivalent to between 40 and 80 ng dependent on size). For experiments where protein expression was assessed at multiple time points, media was fully removed and replaced at each time point.

*Primary cells*

For assays using immature DCs, PBMCs were isolated from whole blood. Monocyte attachment over 3 days was used to select for DC population. The cells were exposed to GM-CSF (50 ng/ml) and IL-4 (100 ng/ml) to stimulate differentiation to DCs. For these cells, TransIT was always used for transfection instead of Lipofectamine 2000, which showed significant toxicity in this cell type. TransIT (Mirus) was complexed to RNA using 0.25:0.25 µL of Boost and TransIT reagent was used per well.
RNAsynthesis

Permuted Intron Exon (PIE) Ligation

A SpeI-digested vector (Supplemental Table 1, PIE_St7s3sENBIRES5sMS) containing the permuted Group 1 intron of T4 Bacteriophage, a T7 promoter, and the nanoluciferase coding sequence (CDS) was used as template in in vitro transcription (IVT) reactions (High Yield T7 RNA Synthesis Kit, NEB) according to the manufacturer’s instructions but with the following supplementations: final GTP concentration of 10, 15, or 20 mM and excess Mg2+ (2, 5, 10 mM). With every iteration of GTP:Mg2+ ratios used over repeated reactions, no splicing product was ever detected, just a single RNA product appropriate to the size of the whole T7-initiated template.

T4 ssRNA Ligase Enzymatic Ligation

This method was adapted and optimized from Perrault et. al. (1995). DNA templates for in vitro transcription were generated using primers against a vector like, pUC19_NLucPEST, with a forward primer containing the T7 promoter consensus sequence and Q5 Hot Start 2x Master Mix (NEB). 500-1000 ng of purified DNA template was used in each IVT reactions (High Yield T7 RNA Synthesis Kit, NEB) according to manufacturer’s instructions. To obtain monophosphate 5’ end necessary for splint or T4 ssRNA ligation, IVT product was treated with RNA Pyrophosphohydrolase (RppH, NEB). Buffer 4 was used in the ligation reaction instead of the provided 10x T4 RNA ligase Buffer. Final reaction conditions contained: 10% DMSO, 2mM ATP, and 10 Units of RNaseOUT (Life Technologies) in a total volume of 200 µL. RNA was incubated with RNase R at 1:1 ratio of µg RNA to units of enzyme (Epicentre, 20 Units/µL) for 1 hour at 37 degrees. GeneJet RNA CleanUp Purification Kit (Thermo Fisher) was used to isolate RNA after reaction was completed.
Quality Control & Purification Assays

*Outward-oriented PCR*

mRNA was generated using Superscript IV (Thermo Fisher) for animal and reticulocyte lysate in which Trizol extract was used to isolate RNA. The resultant cDNA was used as template in downstream RT-PCR reactions. The PCR amplicon was cloned into a parent vector using the PCR Cloning Kit (NEB). Positive bacterial clones were grown overnight, mini-prepped, and submitted for Sanger sequencing (GeneWiz). Inward and outward primers can be found in Supplemental Table S2.

*Sequencing of Splice Junction*

The amplicon from resultant outward-oriented PCR reaction was cloned into a parent vector using the PCR Cloning Kit (NEB). Positive bacterial clones were grown overnight, mini-prepped, and submitted for sanger sequencing (GeneWiz). Inward and outward primers can be found in Supplemental Table S2.

*RNAsep – HPLC*

Original methodology was adapted from Kariko et. al. (2011). RNAsep column (Concise Separations) was installed into a series 1200 Agilent HPLC system. Column compartment was used to keep column at 75 °C during sequence runs, as was necessary for optimal peak clarity. Sequence was run using Solvent A (10mM Triethyl ammonium acetate (TEAA) and Solvent B (10mM TEAA, 25% Acetonitrile) over a linear gradient ranging from 35 -55 % Buffer B over 20 minutes. The modified method sequence for larger or modified RNAs can be found in Supplemental Figure S3.
**cDNA Synthesis and qPCR**

At each time point, cells were first washed with PBS and samples harvested using the Cell-to-Ct Kit (Life Technologies) according to the manufacturer’s instructions. Power Sybr PCR Master Mix (Life Tech) was used for qPCR reactions and run on a CFX96 480 thermal cycler instrument (BioRad). For each sample, the real-time PCR reaction was performed in triplicate, and the averages of the obtained threshold cycle values (Ct) were processed by comparative Ct method. Gene expression levels were normalized to the expression of the housekeeping gene Actin.

**Protein Expression**

**Luciferase assays**

Nanoluciferase assays were always performed on 96-well white plates to prevent leaching signals from neighboring wells. Half of the media was removed (usually 50 µL), and then a mixture of substrate diluted in lysis buffer was added to each cell (1:50 dilution of substrate into lysis buffer). Samples were pipetted up and down with care taken not to introduce air bubbles. Five minutes after substrate was added, plates were read on a Envision plate reader (Perkin Elmer).

**Dual-Glo**

20,000 HEK293T cells were plated per well in 96-well plates the day before transfection. Cells were transfected with Lipofectamine 2000 in triplicate with PGL3-FLuc, RLuc-Mir20 vector, and linear or circular sponge mRNA. We performed Dual-glo assays (Promega) 24 hours post-transfection on a Perkin Elmer Envision luminometer.

**Animal Studies**

All experiments were performed with approval of the Institutional Animal Care and Use Committee (IACUC) of the Dana-Farber Cancer Institute and Harvard Medical School. Female
BALB/c mice from Jackson Laboratory aged 6–12 weeks were used. No more than 100 µL of blood samples were drawn over a two-week period. Mice were euthanized using the CO2 method of asphyxiation with secondary confirmation of death by cervical dislocation.

**Injections**

RNA was complexed to TransIT-mRNA (Mirus Bio, Madison, WI) according to the manufacturer’s instruction for all animal studies. A ratio of mRNA (2.5 - 10 µg), TransIT-mRNA reagent (10 µL), and Boost reagent (5 µL) in a final volume of 200 µL PBS was used. For complexing different amounts of mRNA the volumes of the reagents and the final volume were scaled proportionally.

**Bioluminescence Imaging (BLI)**

Anesthetic was administered during the substrate injections and imaging session. Furimazine (Promega) was administered intraperitoneally in a volume of 200 µl. After 5 minutes, the bioluminescence was measured by the Xenogen Detection Method. Exposure time ranged from Auto, 1 minute and 5 minutes. However, for each experiment, the exposure remained consistent for all images taken. Radiance quantifications were determined using the Caliper Live Animal Software.

**RNA Extraction & Processing from Animal Tissue**

After organ removal, spleens and livers were cut into smaller pieces using a sterile razor blade and then pushed through a sterile 40-µm cell strainer using a plunger of a syringe. Removal of red blood cells using RBC buffer (Gibco) was used as needed. For total RNA extractions from animal tissue, samples were ground up in 1 mL of Trizol using a handheld homogenizer followed by the addition of 200 µL of chloroform. The resulting top fraction after
centrifugation at 12g for 15 minutes was column purified, and total RNA concentration measured by Nanodrop.

**Epo & IFNa ELISA**

BALB/c mice were injected with TransIT-complexed mRNA (~3.5 pmoles). At indicated time points, animals were sacrificed and whole blood isolation by cardiac puncture into EDTA tubes to prevent coagulation. Plasma was isolated for the detection of Epo or IFNa (both kits from R&D Systems).

**Retic-Counter**

5 µL of blood was drawn from mice and added to 0.5 µL of EDTA (2 mM). Retic-Counter reagent was added to each sample and incubated at room temperature for 30 minutes. Reticulocyte counts were then measured using a Fortessa flow cytometer (BD Bioscience) and quantified using FloJo software. Red blood cells are first selected for based on FSC and SSC. From this population, the percentage of FITC+ (Reticulocytes) are quantified relative to total red blood cell counts.
Supplemental Figure S1: Secondary structure and alignments for 400, 500, 600 NT RNAs. RNAfold predicted secondary structures of 400 (A) 500 (B) and 600 NT (C) RNAs used in Figure 1.1A. Colors refer to base pair probability with green bases having lower probability and red bases having the highest probability. Black arrows point to 5’-3’ junction point. (D) Sequence alignments (Qiagen, CLC DNA Workbench) of the three RNAs. The last 20 nucleotides at the 5’ and 3’ ends of each RNA were identical so that a universal splint could be used for each splint-mediated ligation reaction.
Supplemental Figure S2: Secondary structure of 300-1300 NT RNA Panel of Figure 2E
Generated by the RNAfold Vienna Software.
Supplemental Figure S3: Alternative HPLC Sequences for larger or heavily modified RNAs

(A) The extended gradient shown in the HPLC sequence gradient in necessary for large RNA and or RNAs with bulky modifications. (B-D) These RNAs elute slower than their unmodified form and do not always match up to the RNA ladder used as a reference for peak size.
Supplemental Figure S4: IRES Panel Controls

(A) IRES panel was tested in various cell lines and showed in every line that EMCV was the best option for potent translation efficiency. (B) To verify that modified nucleotide’s ablation of IRES-mediated translation was not coding sequence dependent we assayed for protein expression with an eGFP encoding circRNA by Western blot. As expected linear RNAs with various protein produced correct protein however GFP was only detected from circRNA that contained no modified nucleotides. (C-F) Additional testing of selected IRES panels for the effect modified nucleotides had on their translation efficiency.
Supplemental Figure S5: Celigo Cell Counts
Cell counts from Celigo cytometer for all transfection conditions are shown here ad confirm the expected signal specificity from cells transfected with Cy5-containing (A) and or GFP-encoded mRNA (B). Although initially tested as a way to measure RNA kinetics it became clear that Cy5 molecules stuck around inside cells and was not a good option for tracking RNA kinetics.

Supplemental Figure S6: mRNA in vivo Optimization
BALB/c mice were injected with 5 µg of mRNA complexed to the indicated transfection reagent. All reagents are now from Thermo expect for TransIT which is sold by Mirus. All reagents were complexed to RNA according to manufacturer’s instructions. 24 hours post transfection livers were isolated. One portion use for RNA extraction (Trizol) (B) and the other for protein lysate isolated (A). Protein concentrations were measured by BCA assay (Pierce) and 50 µg of total lysate diluted in 50 µL of PBS used for nanoluciferase assay (Promega).
Supplemental Figure S7: Protein Expression Kinetics in vivo
Radiance levels were measured for each mouse at each time point and the average graphed here. Signal from injection site was subtracted from all animals.

Supplemental Figure S8: Confirmation of Epo Expression in vitro
(A) Construct design of linear and circular Epo mRNAs. Linear mRNA contain motifs necessary for canonical, cap-dependent translation (5' cap and poly(A) tail) as well as the 50% replacement of uridine and cytidine nucleotides with pseudouridine (ψ) and 5-methyl cytidine (5mC), respectively. (B) Newly designed Epo mRNA were confirmed to express detectable levels of protein in vitro prior to in vivo testing.
### Supplemental Tables

**Supplemental Table S1: Vector List**

<table>
<thead>
<tr>
<th>Vector Name</th>
<th>sequence (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>pie_st7s3senb</strong></td>
<td>tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggcagatcgaaaagaatttttaaggtgggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga</td>
</tr>
<tr>
<td><strong>fress5sms</strong> (partial sequence)</td>
<td>tgcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggcagatcgaaaagaatttttaaggtgggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga</td>
</tr>
<tr>
<td><strong>puc19_nluc</strong></td>
<td>tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggcagatcgaaaagaatttttaaggtgggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcgacggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactgtaacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagtaaccatcaacgga</td>
</tr>
</tbody>
</table>

For detailed sequences, please refer to the original source.
Crpv-igr

Hcv type 1a

Hcv type 1b

Csfv

Gbv-b

Xiap

C-my

N-my

L-my

Rbm3

Apaf1

Gtx
gagttggagccgataaagttgagacccattctctgcctgcggtgctttatgacaccaagacatgtttggaagtctttttctgaagagtctttttctgccgcgtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt...
## Supplemental Table S2: qPCR Primer List

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence (5' → 3')</th>
<th>Primer</th>
<th>Species</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Beta Actin</strong></td>
<td>gacgagccccagagcaagagaggg</td>
<td>F</td>
<td>Human/Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>ggtgttggaagttcacaacatgt</td>
<td>R</td>
<td>Human/Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>IL-1β</strong></td>
<td>ggcctcaagaaagaaagaaatc</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>ttctgcttgaagaggtgtga</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>IL-1α</strong></td>
<td>gcttcaagatgtaaggcaaaag</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>tggaccaaaaatgcctgtatat</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>IFNα</strong></td>
<td>ccatctctgtctctcatgag</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>attctctctgacaactctct</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>IFNβ</strong></td>
<td>tgggaggattgcattaccc</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>tggctctgctgtgaagaaga</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>PKR</strong></td>
<td>taaaataatttgccccagaagaagta</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>tggacctgaagcgtgtgc</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>MDA5</strong></td>
<td>accaataacagagccatgc</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>gcgattttctttttcag</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>RIG-I</strong></td>
<td>ctcccggacacagaatgtat</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>ctctctctgtctgtgtttt</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>TNFα</strong></td>
<td>ctctctctcagacagacagag</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>ecaaagtgaactggccacaaga</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>IL-6</strong></td>
<td>tacccccagagaagatcc</td>
<td>F</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>gcacattttggaagttcag</td>
<td>R</td>
<td>Human</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>IFNα</strong></td>
<td>ctctcttagactctcttcga</td>
<td>F</td>
<td>Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>gcctctctgactctctctct</td>
<td>R</td>
<td>Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>Epo-WT</strong></td>
<td>ccacccctgcctcttcactc</td>
<td>F</td>
<td>Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>ecatctctgctgctgctgc</td>
<td>R</td>
<td>Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>Epo-SeqEng</strong></td>
<td>agacactgacgtgacacatc</td>
<td>F</td>
<td>Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>agacgtgctgacagatgcccaga</td>
<td>R</td>
<td>Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>Epo-CO</strong></td>
<td>agggctgagacagccactct</td>
<td>F</td>
<td>Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>egtctgactgcactctgcctc</td>
<td>R</td>
<td>Mouse</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>eGFP</strong></td>
<td>acgcacccagccccacatgcc</td>
<td>F</td>
<td>x</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>aagctgctgtgctgaatctt</td>
<td>R</td>
<td>x</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>Nluc</strong></td>
<td>attctgtggggaactggcga</td>
<td>F</td>
<td>x</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>egatctggccctattgctg</td>
<td>R</td>
<td>x</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>EMCV IRES</strong></td>
<td>acgtctgtgagccaccccttgg</td>
<td>F</td>
<td>x</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>taeggctgagggagccattc</td>
<td>R</td>
<td>x</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>CVB IRES</strong></td>
<td>ggcttcacactctgtctct</td>
<td>F</td>
<td>x</td>
<td>Inward</td>
</tr>
<tr>
<td></td>
<td>egaagcctttcctcaac</td>
<td>R</td>
<td>x</td>
<td>Inward</td>
</tr>
<tr>
<td><strong>EMCV IRES Out</strong></td>
<td>atagacaaaaaggcacacgagggc</td>
<td>R</td>
<td>x</td>
<td>Outward</td>
</tr>
<tr>
<td><strong>CV IRES Out</strong></td>
<td>gtaataaaagggccacaaaaag</td>
<td>R</td>
<td>x</td>
<td>Outward</td>
</tr>
<tr>
<td><strong>Nluc Out</strong></td>
<td>gctccctgtgcttgctgata</td>
<td>F</td>
<td>x</td>
<td>Outward</td>
</tr>
<tr>
<td><strong>eGFP Out</strong></td>
<td>gatactgctgagccagctg</td>
<td>F</td>
<td>x</td>
<td>Outward</td>
</tr>
<tr>
<td><strong>Epo-WT Out</strong></td>
<td>tcggcagctcggctgctgctg</td>
<td>F</td>
<td>x</td>
<td>Outward</td>
</tr>
<tr>
<td><strong>Epo-Seq Eng Out</strong></td>
<td>tcgcacacctcctcctcctcct</td>
<td>F</td>
<td>x</td>
<td>Outward</td>
</tr>
<tr>
<td><strong>Insulin Out</strong></td>
<td>gacagcagtcgtctgtgatg</td>
<td>F</td>
<td>Mouse</td>
<td>Outward</td>
</tr>
</tbody>
</table>
**Supplemental Table S3: General Primer List**

<table>
<thead>
<tr>
<th>Sequence (5' → 3')</th>
<th>Description</th>
<th>Primer</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>taatacgactcaataggggaagaagagaagaagaagagaagatttcacactcgagaaggggaaagaaaga</td>
<td>Universal Forward Primer (400, 500, 600 NT)</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccagatgcctcctactcgaagaacggcgtg</td>
<td>400 NT Primer</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>500 NT Primer</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>600NT Primer</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>cccca</td>
<td>Universal splint for 400-600 NT RNAs</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>ttagacgttgatgcgagctg</td>
<td>Universal Primer for (-CRC Constructs)</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>taatacgactcaataggggaagaagagaagaagaagagaagatttcacactcgagaaggggaaagaaaga</td>
<td>Universal Forward Primer (400, 500, 600 NT)</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcactgtaacagggaccctg</td>
<td>300 NT (-CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccagatgcctcctactcgaagaacggcgtg</td>
<td>400 NT (-CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>500 NT (-CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>cccca</td>
<td>Universal splint for 400-600 NT RNAs</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>ttagacgttgatgcgagctg</td>
<td>Universal Primer for (-CRC Constructs)</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccagatgcctcctactcgaagaacggcgtg</td>
<td>300 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>400 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>500 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>cccca</td>
<td>Universal splint for (+CRC Constructs)</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>ttagacgttgatgcgagctg</td>
<td>Universal Primer for (+CRC Constructs)</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccagatgcctcctactcgaagaacggcgtg</td>
<td>300 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>400 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>500 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>cccca</td>
<td>Universal splint for (+CRC Constructs)</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>ttagacgttgatgcgagctg</td>
<td>Universal Primer for (+CRC Constructs)</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccagatgcctcctactcgaagaacggcgtg</td>
<td>300 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>400 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>tcctcctctctccccccacacagcaggggagctgctcg</td>
<td>500 NT (+CRC) Primer</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>cccca</td>
<td>Universal splint for (+CRC Constructs)</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>ttagacgttgatgcgagctg</td>
<td>Universal Primer for (+CRC Constructs)</td>
<td>R</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>Genebanks</td>
<td>Primer</td>
<td>Location</td>
<td>Vector</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>300 NT (+CRC) Primer</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>500 NT (+CRC) Primer</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>700 NT (+CRC) Primer</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>1000 NT (+CRC) Primer</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>NLuc (+CRC) Primer</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>eGFP (+CRC) Primer</td>
<td>R</td>
<td>pUC57_EMCV-eGFP</td>
<td></td>
</tr>
<tr>
<td>Universal Forward for (+CRC Constructs) F</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4kb Universal</td>
<td>F</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>1kb</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>2kb</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>3kb</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
<tr>
<td>4kb</td>
<td>R</td>
<td>pUC57_EMCV-NLuc</td>
<td></td>
</tr>
</tbody>
</table>

**PPT19-Nluc_2**

F pUC19-NLucPEST

**PPT19-Nluc_1**

F pUC19-NLucPEST
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Synthetic Oligos (5'-3')</th>
<th>Type</th>
<th>Construction Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>gaattctggactaaagccggctttgtgtatatc atggcttcacactcgaaga</td>
<td>PPT19-Nluc_0</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>taatacgactcactataaggggaaagaatgtc gtgctgttaaaaagggaaaagaaggg aagaagggaaaaagggaaaaggg aagaagggaaaaagggaaaaggg aagaagggaaaaagggaaaaggg gaggcccttcgagccgggtacc</td>
<td>PPT19-Nluc_N</td>
<td>F</td>
<td>pUC19-NLucPEST</td>
</tr>
<tr>
<td>rand6-Xho1-mir20_F</td>
<td>F</td>
<td>pRLuc-TK-mir20</td>
<td></td>
</tr>
<tr>
<td>mir20-apal-rand6_R</td>
<td>R</td>
<td>pRLuc-TK-mir20</td>
<td></td>
</tr>
<tr>
<td>Rluc-TK Seq_F</td>
<td>F</td>
<td>pRLuc-TK-mir20</td>
<td></td>
</tr>
<tr>
<td>mir20Sponge_F</td>
<td>F</td>
<td>pU6-27-mir20</td>
<td></td>
</tr>
<tr>
<td>CXCR4Sponge_F</td>
<td>R</td>
<td>pU6-27-cxcr4</td>
<td></td>
</tr>
<tr>
<td>T7-CRC-SIINFEKL_F</td>
<td>F</td>
<td>pUC19-sec-siinfekl-mtd</td>
<td></td>
</tr>
<tr>
<td>CRC-SIINFEKL_R</td>
<td>R</td>
<td>pUC19-sec-siinfekl-mtd</td>
<td></td>
</tr>
</tbody>
</table>
References


Chappell S.A. & Mauro V.P. A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. PNAS 97: 1536-1541 (2000).


Gross J.D. & Wagner G. Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. Cell 115:739-750 (2003).


Kariko K & Muramatsu H. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoded mRNA. Nucleic Acid Research, 1-10 (2011).


Kariko K. & Weissman D. Increased erythropoiesis in mice injected with submicrogram quantities


Kelly B. Industrialization of mAb production technology: The bioprocessing industry at a crossroads. mAbs, 1:443-52 (2009).


Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283-292. (1986).


Krug A. & Hornung V. Toll-like receptor expression reveals CpG DNA as a unique microbial stimulus for plasmacytoid dendritic cells which synergizes with CD40 ligand to induce high amounts of IL-12. European Journal of Immunology 31: 3026-3037.


Pardi N. & Weissman D. Expression kinetics of nucleoside-modified mRNA delivered in lipid nanoparticles to mice by various routes (2015).


Salzman J. & Gawad C. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One 7: e30733 (2012).


Surono A. & Matsuo M. Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. Human Molecular Genetics 8: 495-500 (1999).


Thompson S.R. So you want to know if your message has an IRES? Wiley Interdisciplinary Reviews: RNA. 3: 697–705 (2012).


