

Translational implications of off-frame stop codon optimisation in RNA therapeutics

by

Charles Garrett

**Submitted in fulfilment for the degree of MSc by Research
specialising in Genetics**

**University of Kent
School of Natural Sciences**

February 2026

Word Count: 11,997

DECLARATION

I declare that the wording of this report is completely my own. No part of the report has been copied from scientific journals, websites or other sources.

Abstract

N1-methyl-pseudouridine (m1 Ψ) is a modified nucleoside used in therapeutic mRNAs to reduce immune responses; however, its use leads to a higher rate of random frameshift events during translation. Inclusion of stop codons in the +1 reading frame (+1 FSS) is proposed as a mitigation strategy against increased aberrant peptide products arising from frameshift events. The efficacy of mRNA vaccines can be increased through a reduction in off-target antigens. Emerging applications of mRNA therapeutics to genetic disease and cancer require higher dosage levels and treatment frequency; potential side effects could arise from increased burden of aberrant proteins, otherwise manageable by cellular processes at lower concentrations. This study investigates expression levels and kinetics of a Nanoluciferase reporter protein in response to differing levels of +1 FSS. Experimental sequences range in +1 FSS content from zero to the maximum possible given for the same amino acid sequence. Expression results indicated a ~20% reduction in NanoLuc expression at 100% +1 FSS content relative to the commercial sequence, aligning with a higher content of sub-optimal codons resulting from inclusion of +1 FSS. Increased expression was observed at 25% +1 FSS, suggesting a role of improved secondary structure in the findings, whether due to robustness of the structure or accessibility to ribosome complexes. Prolonged protein expression was observed in the 100% +1 FSS sequence at consistently lower expression rates, contradicting estimates from predictive tools. To explain these results, we also studied alterations in secondary structure arising from inclusions of +1 FSS through differential scanning fluorimetry. The findings suggest that in applications requiring sustained production of therapeutic proteins, such as the aforementioned proposed cancer treatments, +1 FSS can be a viable feature of sequence design, particularly when de-optimisation is considered for increasing mRNA half-life. Improved translation efficiency resulting from +1 FSS inclusion at low levels may aid broader applications, but further research across diverse amino acid sequences would strengthen evidence.

List of Contents

1.0 Introduction.....	9
1.1 mRNA Developments.....	10
1.2 mRNA primary structure	10
1.2.1 Modified nucleosides	11
1.2.1.1 Frameshifting events	12
1.2.2 Codon optimisation	14
1.2.2.1 Off-frame stop codon inclusion	16
1.3 mRNA secondary structure.....	17
1.3.1 Predictive secondary structure models	19
1.4 Project significance	20
2.0 Materials and Methods.....	21
2.1 mRNA preparation	21
2.1.1 Template and vector design	19
2.1.2 Template amplification	22
2.1.3 Agarose gel electrophoresis for DNA	22
2.1.4 Cloning vector preparation.....	23
2.1.5 Bacterial transformation.....	23
2.1.6 DNA extraction	24
2.1.7 Sequence validation	24
2.1.8 <i>In vitro</i> transcription	25
2.1.9 mRNA quantification and quality analysis.....	26
2.2 mRNA expression assay processes	26
2.2.1 Human tissue culture propagation.....	26
2.2.2 NanoLuc expression assays	27
2.3 mRNA secondary structure analyses.....	28
2.3.1 RNAfold secondary structure predictions	28
2.4 mRNA DSF assays.....	28
2.5 Statistical Approach	29

3.0 Results	30
3.1 mRNA preparation	30
3.1.1 Template amplification	32
3.1.2 Vector preparation for <i>in-vitro</i> transcription	33
3.1.3 Bacterial transformation	34
3.1.4 Clone analyses	35
3.1.5 DNA analysis	35
3.2 <i>In-vitro</i> transcription	37
3.2.1 mRNA analysis	37
3.3 NanoLuc expression assays	39
3.3.1 24-hour expression	39
3.3.2 72-hour expression	41
3.4 Predictive mRNA secondary structure analyses	44
3.4.1 RNAfold predictions	44
3.5 mRNA DSF assays	49
3.5.1 Pure samples	49
3.5.2 Buffered samples	53
4.0 Discussion	55
5.0 Conclusions	60
6.0 References	61
7.0 Appendix	70

List of Figures

Figure 1 – mRNA primary structure.....	11
Figure 2 – Chemical structures of uridine and modified variants	12
Figure 3 – tRNA profile of HEK 293	15
Figure 4 – Predicted secondary structures of the moderna mRNA-1273 vaccine	19
Figure 5 – Sequence map of integrated stop codons in the +1 reading frame.....	22
Figure 6 – Graphical map of recombined NanoLuc vector	31
Figure 7 – Gel electrophoresis of amplified template DNA	32
Figure 8 – Gel electrophoresis of vector backbone amplicon.....	34
Figure 9 – Gel electrophoresis of extracted plasmid DNA.....	36
Figure 10 – Graphical comparison of transcribed mRNA capillary electrophoresis	38
Figure 11 – Normalised NanoLuc luminescence 24 hours post-transfection	40
Figure 12 – Periodic NanoLuc luminescence 24-72 hours post-transfection.....	42
Figure 13 – Predicted secondary structures of sequence 1 (0% +1 FSS)	45
Figure 14 – Predicted secondary structures of sequence 2 (25% +1 FSS)	46
Figure 15 – Predicted secondary structures of sequence 3 (50% +1 FSS)	46
Figure 16 – Predicted secondary structures of sequence 4 (75% +1 FSS)	47
Figure 17 – Predicted secondary structures of sequence 5 (100% +1 FSS)	47
Figure 18 – mRNA DSF melting curves comparing dye concentrations	50
Figure 19 – mRNA DSF melting curves comparing experimental sequences	51
Figure 20 – mRNA DSF melting curves comparing change in fluorescence in experimental sequences.....	52
Figure 21 – mRNA DSF melting curves comparing experimental sequences under buffered conditions.....	53

List of Tables

Table 1 – Nanodrop readings of extracted DNA	36
Table 2 – Capillary electrophoresis quantification of transcribed mRNA	38
Table 3 – Spectrophotometry quantification of transcribed mRNA prior to 24h expression	39
Table 4 – Spectrophotometry quantification of transcribed mRNA prior to 72h expression	42
Table 5 – RNAfold predicted secondary structure parameters without pseudouridine adjustments	44

List of Abbreviations

+1 FSS - +1 frameshifted stop codon

AO – Acridine orange

ASOs – Antisense oligonucleotides

CAI – Codon adaptation index

DSF – Differential scanning fluorimetry

IVT – *in-vitro* transcription

LNPs – Lipid nanoparticles

Ψ - Pseudouridine

m1Ψ – N1-methyl-pseudouridine

NanoLuc – Nanoluciferase

siRNA – Small interfering RNA

UTR – Untranslated region

1.0 Introduction

RNA therapeutics is a branch of modern medicine utilising both natural and engineered functions of RNA to alter or direct protein expression in target cells. A number of classifications exist; some exploit functions of smaller sequences e.g. small interfering RNA (siRNA) and antisense oligonucleotides (ASOs) to silence and regulate genes, whilst mRNA treatments involve transfection and subsequent translation of an encoded protein from a whole sequence. At the time of this writing, six siRNA therapeutics have been approved for administration (Xiao *et al.*, 2025) alongside three mRNA therapeutics in the UK (Comirnaty, Spikevax and Kostaive). Due to the COVID-19 pandemic and subsequent emergency authorisation, mRNA therapeutics as vaccines have been administered in greater numbers than siRNA. The potential applications of therapeutic mRNAs include, but are not limited to, protein replacement, antibody therapies, gene therapies and in vaccination (Žak and Zangi, 2025); each with highly variable requirements for dosage and treatment duration.

The use of *in vitro* transcription (IVT) to synthesise mRNA allows for customisation of the molecule through structure and encoded protein, enabling reactive treatment of a wide range of diseases and conditions including cancer (Liu *et al.*, 2023) in addition to prophylactic treatments as exemplified in SARS-CoV-2 vaccines (Creech *et al.*, 2021). Conceptually, any condition arising from protein deficiency has the potential to be treated with an mRNA-based therapeutic (Vavilis *et al.*, 2023). Currently the only mRNA therapeutics approved in the UK are spike-encoding COVID-19 vaccines; the potential applications listed above are still at pre-clinical and early clinical stages.

1.1 mRNA Developments

Development efforts through the 1990s and into the 2000s improved foundational elements of mRNA therapeutics; delivery vehicles advanced from liposomes to lipid nanoparticles (LNPs) whilst untranslated regions (UTRs) were incorporated into the mRNA primary structure, both of which increased localised target expression (Sahin *et al.*, 2014). With increasing treatment viability, interest and investment has risen and the process of designing new treatments has become increasingly economic for industry and academia alike (Jackson *et al.*, 2020); development timeframes have equally shortened as the base of scientific knowledge grows and manufacturing technologies and infrastructure improve. Despite these improvements, such as algorithms attempting to optimise design (Zhang *et al.*, 2023), there is a need for empirical data to complement *in silico* developments, ensuring efficacy and safety *in vivo*.

1.2 mRNA primary Structure

The primary structure of mRNA consists of the coding sequence and untranslated regions (UTRs) of a given gene. In therapeutic mRNAs the structure consists of a 5' cap structure, 5'-UTR, open reading frame (ORF), 3'-UTR and a poly(A) tail (von der Haar *et al.*, 2023); this structure is outlined in Figure 1 below. The 5' cap structure is bound co-transcriptionally and is crucial to designation as a native mRNA, ribosome recruitment and prevention of exonuclease cleavage (Ramanathan *et al.*, 2016); engineered versions of native m⁷G caps such as CleanCap AG (m⁷GTP)(TriLink Biotechnologies) are used in therapeutic mRNAs to closely mimic the type 1 structure of m⁷G whilst offering post-translational capping *in-vitro* (Mandell *et al.*, 2025). The 5'-UTR bridges the cap and the open reading frame consisting of nucleobases unrestricted from the need to serve as a triplet, the 5' UTR facilitates ribosome scanning to initiate translation and contributes functional structures (Leppek *et al.*, 2018). The ORF encodes an amino acid sequence through nucleotide triplets, beginning with a start codon AUG (Methionine) and ending with one of three stop codons, UAA, UAG, or UGA (Hellen, 2018). The 3'-UTR can contain regions that bind regulatory miRNAs and proteins to alter RNA stability and translation, these contribute to the formation of secondary structure across the entire molecule and can determine cellular localisation of the molecule (Hong and Jeong, 2023). Finally, the poly(A) tail comprises a chain of adenine monophosphates at the 3' end of the mRNA molecule, typically ~100-nt in length for use in therapeutic mRNAs (von der Haar, 2023).

The poly(A) tail serves to protect the molecule from degradation and plays a crucial role in translation; the length of poly(A) tails in mammalian mRNAs is typically 200 bp but observations of regular shortening and lengthening of the tail present a dynamic process to alter translation parameters (Passmore and Collier, 2021).

Mature eukaryotic mRNAs contain these same regions as outlined in Figure 1, but therapeutic mRNAs allow, through design, the modification of each element to suit the intended function; from the type of cap structure, content from 5' UTR to 3' UTR and length of poly(A) tail.

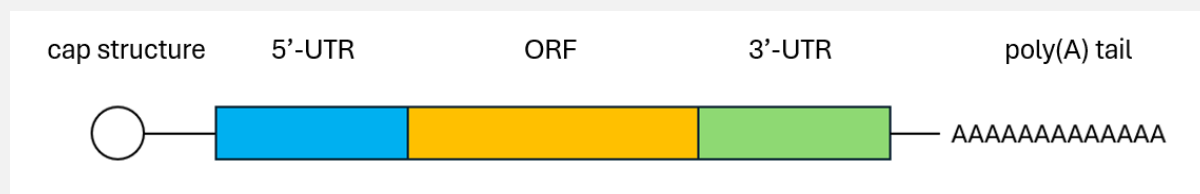


Figure 1. mRNA primary structure. Depicted from left to right are the 5' cap (circular), 5'-UTR (blue), open reading frame (ORF)(orange), 3'-UTR (green) and poly(A)-tail (As). (adapted from von der Haar, 2023).

The *in-vitro* production of therapeutic mRNA mimics that of natural processes in eukaryotes whilst bypassing intron-containing pre-messenger RNA, otherwise removed in the nucleus in preparation of export to the cytoplasm (Palazzo and Lee, 2018).

1.2.1 Modified Nucleosides

Modified nucleosides occur naturally as deviations from the standard adenine, uracil, guanine and cytosine typically found in RNAs; in eukaryotes they are found across mRNA and tRNAs alike, with N⁶-Methyladenosine (m⁶A) found to play a potential role in regulating the obesity-associated and fat-mass protein (FTO), the list of identified natural modifications is over 100 (Roundtree *et al.*, 2017). Synthetic modified nucleosides are engineered, often as analogues of natural counterparts, to achieve specific functionality in RNAs, such as N1-methyl-pseudouridine (m1Ψ) engineered from the naturally occurring pseudouridine (Ψ), inspired by observations of Ψ evading toll-like receptors (Karikó *et al.*, 2008).

To avoid recognition as a foreign sequence and subsequent immune response, therapeutic mRNAs are synthesised with a synthetic modified nucleoside, N1-methyl-pseudouridine (m1Ψ) in place of uridine. The structure of the molecule as outlined in Figure 2 (below) shows the exposed hydrogen at C5 in uridine is replaced with a methyl group at N1 in m1Ψ, this change to steric configuration

induced in the mRNA major groove effectively evades recognition by toll-like receptors TLR7 and TLR6 (Young and Kanneganti, 2025).

In addition to immune evasion, the use of modified nucleosides alters core features relevant to mRNAs and their function, namely the rate of translation and secondary structure (von der Haar *et al.*, 2023), both of which contribute to overall protein expression for a given dose. Higher translation efficiencies have been reported in mammalian cells from m1 Ψ -containing mRNA in addition to improved biological stability, translating for an extended duration (Karikó *et al.*, 2008).

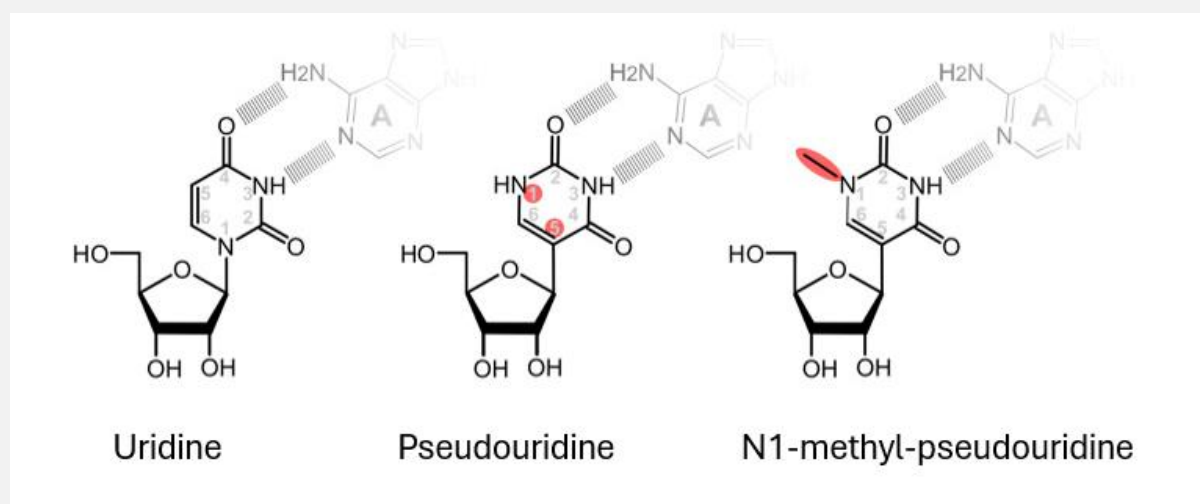


Figure 2. Chemical structures of uridine, and modified variants. Depicted from left to right are the structures of uridine, pseudouridine (Ψ , modified) and N1-methyl-pseudouridine (m1 Ψ , modified) (adapted from von der Haar, 2023).

1.2.1.1 Frameshifting events

A frameshift event occurs when a ribosome is displaced from its ordinary reading frame of triplets directly downstream of a start codon, either missing nucleotides in the forward direction (+1, +2 etc.) or reading the same nucleotides in a reverse frameshift (-1, -2 etc.) (Demo *et al.*, 2021). The resultant product of frameshifts, termed aberrant proteins, contain ordinary amino acids in the chain up to the frameshift location, where amino acids are drastically altered. Frameshifts occur naturally as a part of programmed ribosome frameshifting (PRF), wherein conserved evolved structures in viral RNA induce a -1 frameshift to produce an alternate product (Kelly *et al.*, 2020); PRF is also observed more rarely in humans with the PLEKHM2 gene displaying a +1 frameshift essential to the overall gene function (Loughran *et al.*, 2025).

Frameshift occurrence is significantly higher at known slippery sites, the codon pairs AAAGGU, AAACGG and UUUUGG were identified by Mulroney *et al.*, 2023 as +1 frameshift inducing, with the first 4 nucleotides being determinant; these motifs all contain a homopolymeric codon in the first position.

In 2023, Mulroney *et al.* (2023) confirmed that +1 frameshift errors were induced through the m1 Ψ replacement in Firefly Luciferase-encoding mRNA sequences; these occurred in approximately 10% of proteins produced from the modified IVT mRNA transfected into cells. Following ribosome slippage, out-of-frame codons are translated downstream resulting in an aberrant protein. The composition of an aberrant protein is dependent on the position of frameshift event, and the distance to the nearest stop codon in the new +1 reading frame. Not all studies have demonstrated +1 frameshift, with Rozman *et al.*, 2026 unable to produce statistically significant amounts of +1 frameshift in Sars-CoV-2 spike-encoding mRNAs.

Cellular effects of aberrant proteins vary from increased proteasome burden to potential amyloid disease through aggregation (Pinheiro *et al.*, 2021), although no side effects from the SARS-CoV-2 vaccine have been directly attributed to frameshift products.

In the context of using therapeutic mRNAs as vaccines, it has been shown that unintended peptide products elicit off-target T cell responses, detracting from potential host immunity to the in-frame antigen (Mulroney *et al.*, 2023), thus reduced instances could produce a proportionate increase in treatment efficacy.

1.2.2 Codon optimisation

Redundancy in the genetic code combined with nucleotide level customisation, enables codons within a coding sequence (CDS) to be optimised across all reading frames. Codon choice determines numerous downstream parameters, most pertinently for the analysis done here protein expression, molecular stability, slippery sites and stop codons in the +1 frame. Optimisations can be made manually, however trade-offs between desirable outcomes prevent perfect optimisation; programs have therefore been developed to automate this process, directing optimisation towards specific outcomes within set limits, <60% GC content for example (Vostrosablin *et al.*, 2024).

Minimisation of known slippery site sequences is an initial step to reduce frameshift events, however occurrences are still observed at a reduced rate (Mulroney *et al.*, 2023). Since the predominant slippery sites consist of a phenylalanine-encoding UUU codon followed by a U or a C, replacing UUU codons with their isocoding UUC codons removes the most effective frameshift sites (Mulroney *et al.*, 2023). In addition, stop codons can be intentionally placed in the +1 reading frame with no consequence to the structure of intended proteins; this attempts to truncate any aberrant protein but therein lies a key conflict between parameters, where codons optimal for protein expression differ from those with embedded stop codons.

Protein expression levels are affected by a wide range of cellular conditions, but in relation to mRNA design, codon usage bias is most significant at the species level (Parvathy *et al.*, 2021). The most frequently used codons are those decoded by the most abundant tRNAs. Due to their abundant cognate tRNAs these codons are typically decoded the fastest (Hedayioglu *et al.*, 2025), which is most congruent with the high ribosome flow required for high expression levels. Incorporation of less frequently used codons for an organism implicates reduced expression rates. The codon adaptation index (CAI) is used to quantify how similar codons within a given sequence are to those optimally expressed (Sharp and Li, 1987). The effect of mRNA modification affects tRNA responses differently, depending on the presence of anticodons attempting to pair with modified bases, in addition to flexibility of the anticodon loop and ability to conform to the modification (Monroe *et al.*, 2022).

Since tRNA abundance also varies between human tissue types alongside age and gender (Dittmar *et al.*, 2006), therefore rules applicable to optimisation in a singular cell type may not be universal across all human cell types; the exact parameters of this variation are still poorly understood. Figure 3 (below) shows the distribution of tRNA content in a model HEK 293 cell, with the highest content for each amino acid being the target of an optimisation model purely directed towards abundance.

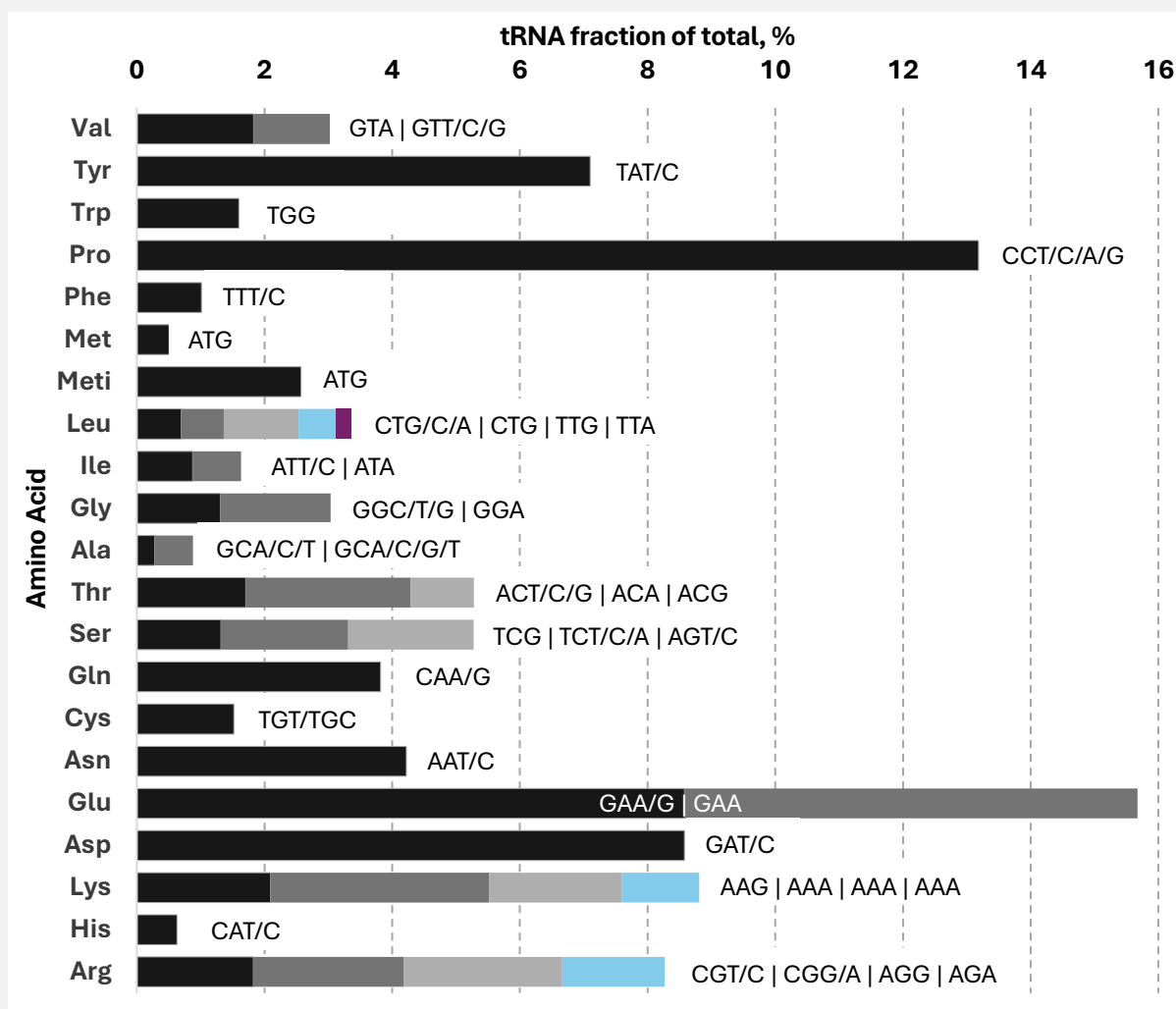


Figure 3. tRNA abundance profile of HEK 293. tRNA microarray results depict tRNA concentrations per tRNA isoacceptor alongside their cognate codon, grouped by amino acid. Values comprise the mean of n=4 biological replicates (adapted from Polte *et al.*, 2019).

1.2.2.1 Off-frame stop codon inclusion

In the canonical genetic code, translation is terminated by the three stop codons UAA, UAG, and UGA (Hellen, 2018). Inclusion of stop codons in the +1 reading frame requires pairing of specific codons, where the first ends in either UA or UG and second begins with A/G or A respectively, producing the patterns NUAANN, NUAGNN and NUGANN. These patterns are compatible only with certain dipeptides, which contain M, I, L, or V in the first position, and A, E, G, I, K, M, R, S, T, or V in the second position. The amino acid sequence of the protein cargo intended to be encoded within an mRNA therefore limits available stop codon placements. Moreover, because in mammals G-ending codons are generally preferred over A-ending codons, the use of NUAANN and NUAGNN patterns reduces codon optimisation because the first codon becomes A-ending. Moreover, patterns encoding +1 FSS also contradict codon pair preferences such as NNGCNN, NNCANN and NNUNCN patterns (Tats *et al.*, 2008). Thus, the more stop codons are present in the +1 reading frame the more negatively this is expected to affect expression levels.

Whilst the molecular stability and functional half-life of mRNA is in part attributable to UTR sequence content (Kim *et al.*, 2022), the CDS composition and resultant secondary structure is equally implicated (Mauger *et al.*, 2019). In addition to affecting expression rates, codon substitutions can affect secondary structure by changing structural features, any stable stem will be disrupted if incompatible pairing is introduced i.e. non-adherence to Watson-Crick base pairing or the additional tolerated pairings in RNAs (Ananth *et al.*, 2026), introducing possible loops and hairpins or altering those ordinarily present. The impact of substitutions will vary depending on amino acid sequence and specifics of substitutions such as location, original codons, and replacement codons.

1.3 mRNA secondary structure

The secondary structure of mRNA arises from self-hybridisation of the molecule, where exposed hydrogen bond donors and acceptors from each nucleoside bind, broadly in Watson-Crick base pairings, to form intramolecular structures (Vicens and Kieft, 2022). Unlike in DNA where sense and antisense strands complement to form an organised linear structure, the requirement to form triplets within the ORF and the molecule self-hybridising forces mismatches in non-structural RNAs. The result of these factors generates intermittent pairings, where short regions (relative to the overall nt-length) find stable stem configurations whilst others remain unpaired; a series of sub-structures arise from the positions of stems and exposed bases, driven broadly by non-Watson-Crick pairings (Vicens and Kieft, 2022).

To outline key sub-structures, the most basic structure is a stem of paired bases of varying length. Hairpin loops form the end of some stems, they can form stable structures with one naturally occurring example being the cUUCGg tetraloop hairpin (Nozinovic *et al.*, 2009). Internal loops form within stems where incompatible bases exist, whilst bulge loops present where an asymmetry is found in regular pairing, multiloops form branches in the molecule, where multiple stems meet imperfectly (Georgakopoulos-Soares *et al.*, 2022). G-quadruplexes are an additional notable structure, although considered tertiary or higher-order, these comprise highly stable structures within RNA molecules, arising from interactions between guanines across different crossing sections of the molecule (Rhodes and Lipps, 2015).

Secondary structure is sensitive to the level of single-nucleotide variations, wherein deviation from a wild-type sequence can destabilise well-defined structures and increase ensemble diversity at low-nt lengths (Moss, 2018). Since codon optimisation involves a series of nucleotide alterations, the resultant secondary structures will vary from the original sequence.

The stability of mRNA is generally inversely proportional to the nt-length (Chheda *et al.*, 2024) and is sensitive to the level of single-nucleotide variations, wherein deviation from a wild-type sequence can destabilise well-defined structures and increase ensemble diversity at low-nt lengths (Moss, 2018). Ensemble diversity is one of contributing factors to decreased stability at high nt-lengths, increasing exponentially as lengths increase (Martin, 2014) as more sub-optimal structures become probable in an increasingly complex molecule.

The specific biological stability of RNA has also been linked to the ratio of paired to unpaired bases, reflecting the stabilising effect of stems and cleavage exposure in large, unpaired loops (Wayment-

Steele *et al.*, 2021). Thermal stability can also be used as a proxy for biological stability, reflecting the ratio of paired to unpaired bases throughout an RNA (Chheda *et al.*, 2024).

As a ribosome translocates across an mRNA during translation, secondary structure is unwound as a single strand is read; hence actual *in-vivo* secondary structure is non-static, beyond the multitude of possible structural configurations, the quantity and position of ribosomes present on a molecule further affect the structure at a given moment (Wu and Bazzini, 2023).

1.3.1 Predictive secondary structure models

Since the mRNA secondary structure is highly complex and not static in nature, computational tools have been developed such as RNAfold (Hofacker, 2003) to predict various parameters for a given RNA sequence under variable conditions such as temperature and salt concentration. Energy parameters are assigned to each nucleobase, with alternates available for some modified nucleosides (Yamamura *et al.*, 2025), and through a partition function, are assigned base pairing probabilities between base *i* and *j*. From these base pairing probabilities, structures can be predicted in addition to their energetic state, across the ensemble of possible structures (Hofacker, 2007).

RNAfold outputs both a minimum free energy (MFE) and centroid structure graphically, these represent the most energetically efficient and most likely observed *in vivo* respectively (Ding *et al.*, 2005). Figure 4 (below) provides an example of these outputs as the relatively complex predicted secondary structures of the Moderna mRNA-1273 vaccine.

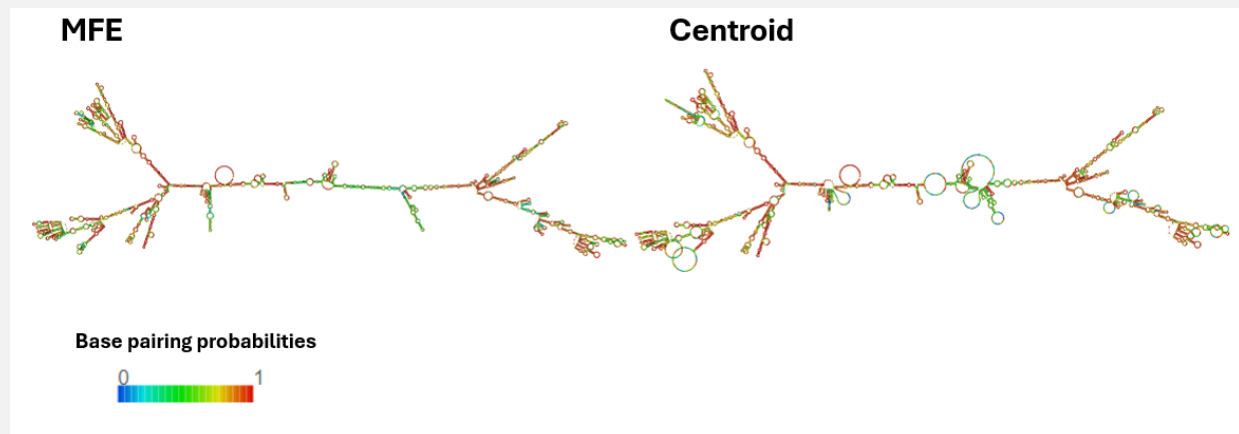


Figure 4. Predicted secondary structures of the Moderna mRNA-1273 vaccine. Depicted are the minimum free energy (MFE)(left) and centroid (right) predicted mRNA secondary structures for the SARS-CoV-2 spike protein encoding sequence used in the Moderna vaccine (Appendix 3). Colours indicate that the base pairing state (paired or unpaired) displayed in the structure is actually observed in any individually occurring structure, ranging from 0% (blue) to 100% (red) (sequence from Jeong *et al.*, 2021) (adapted from RNAfold output).

1.5 Project significance

Truncation of aberrant proteins produced by m1Ψ-induced frameshifts aims to reduce their potentially harmful effects, presence of any off-frame stop codon prevents run-off translation of large aggregate-prone polypeptides. These data will aid bases for optimising distribution patterns of off-frame stop codons across the CDS, ensuring accordance with expression and stability requirements of a given treatment. Whilst translational effects will vary by protein, one specific relationship is exemplified.

In sequence optimisation of RNA therapeutics, data pertaining to tissue-specificity of codon usage (particularly codons involved in stop codon integration) can improve targeting and the accuracy of expected protein expression levels in target tissues. Correct dosage and localisation of transfection constitute improved overall safety of mRNA treatments. Requirements for improved targeting delivery methods are reduced as tissue specificity is optimised; expression data from this experiment can be linked to HEK293 tRNA abundance profiling to aid modelling.

Investigations into links between the extent of hybridisation in mRNA and duration of expression seeks to tie molecular stability to the former; statistical correlation could aid implementation of accurate durations of expression in treatments, from intentionally short bursts of expression to those with a longer half-life at lower expression levels.

The study seeks to achieve the following:

- Evaluate effects on protein expression in transfected mRNA following introduction of stop codons in the +1 reading frame to the coding sequence at varying frequencies.
- Determine presence of links between the extent of mRNA secondary structure and longevity of protein expression *in vivo*.
- Test a simple, resource-efficient method of differentiating mRNA secondary structure through melting in solution with intercalating dyes, termed differential scanning fluorimetry (DSF).

Materials and Methods

2.1 mRNA preparation

The subsequent section covers steps undertaken to produce mRNA for expression analysis in transfected cells and secondary structure thermodynamic analysis; it follows the process from DNA template design through to the production of the *in-vitro* transcribed mRNA.

2.1.1 Template and vector design

Insert template DNA, supplied by Genscript Biotech, was designed manually using Serial Cloner (Serial Basics) from an initial 664 bp sequence encoding NanoLuc luciferase (NanoLuc) that only used the most frequently used codons for each amino acid (Sequence 1, 0%)(Appendix 1). Based on this starting sequence, codons were adjusted to introduce maximal levels of +1 FSS, then incrementally reduced at random to the original +1 FSS content producing the series 100%, 75%, 50%, 25%, 0% +1 FSS (Appendix 1); Figure 5 (below) indicates +1 FSS locations for each of these sequences. Whilst +1 FSS would ideally be placed at the 5' end of a sequence (such that translation is terminated at the earliest opportunity), a random approach to selecting insertions was chosen to best analyse overall changes to structure and translation. All sequences encoded an identical amino-acid sequence for the NanoLuc protein, and 5'- and 3'-untranslated regions were kept unaltered. The cloning vector was designed in a prior experiment using Serial Cloner (Serial Basics) to include a selective marker (KanR kanamycin resistance gene), T7 promoter site, origin of replication (ColE1) and UTRs, both 3' (bases 823-963) and 5' (bases 106-151) (Appendix 5) flanking the insertion site, illustrated in figure 6 below.

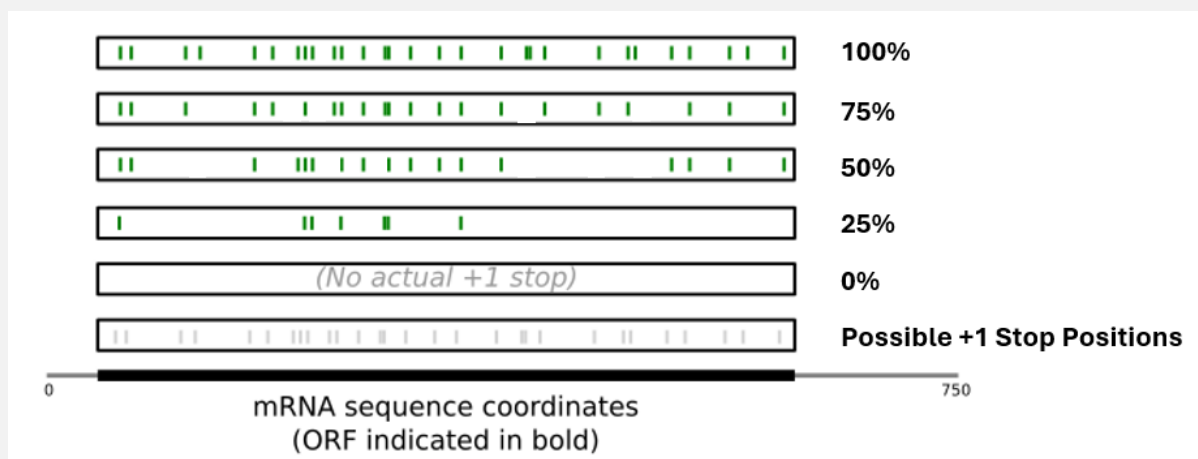


Figure 5. Sequence map of integrated stop codons in the +1 reading frame. Distributions of stop codons for the range 0, 25, 50, 75, 100% (upper five rows) in five sequences of NanoLuc Luciferase (NanoLuc) are shown in relation to all possible positions (row 6) and the open reading frame (ORF) (row 7). (von der Haar Laboratory, 2025).

2.1.2 Template Amplification

Insert templates for experimental sequences 1-5 (Appendix 1) were amplified in a PCR reaction containing 25 μL 2x PrimeSTAR Max master mix, 1 μL forward primer 'mod_utr_5' (54.3 nmol) (5'GGAATAAGAGAGAAAAGAGAGTAAG3'), 1 μL reverse primer 'mod_utr_3' (64.4 nmol) (5'GGAACAAAGGAACCTTTAAATAGAAATT3'), 1 μL template DNA (50 ng μL^{-1}) and 22 μL dH₂O for a total volume of 50 μL . Reaction conditions consisted of initial denaturation at 98 °C for 10 seconds prior to 30 cycles of denaturation at 98 °C for 10 seconds, annealing at 51 °C for 15 seconds and extension at 72 °C for 20 seconds before final extension at 68 °C for 10 seconds.

2.1.3 Agarose gel electrophoresis for DNA

All gel electrophoresis was carried out under uniform conditions; agarose gels were prepared uniformly using 1% w/v agarose, 50 mL 1x TAE buffer, and 5 μL Ethidium Bromide (10 mg mL^{-1}). DNA samples were migrated at 80 V for 40 minutes in 1x TAE buffer. Migrated samples were imaged under ultraviolet light (peak λ = 465 nm) (sourced from the Benchmark Smartblue Illuminator, Sigma Aldritch) using an iPhone 11 Pro Max camera.

2.1.4 Cloning vector preparation

The plasmid backbone was reused from a prior experiment and originally sourced from GenScript (Appendix 5, exc insert sequence). It contained a Lamp2b gene so PCR was conducted to amplify the backbone without the prior CDS. Primers were designed to amplify from UTR regions away from the prior CDS, retaining the prior poly(A) tail, as illustrated in figure 6, below. The reaction contained 25 μL 2x PrimeSTAR Max master mix, 1 μL forward primer 'HEK_lin_f' (5'GCTCGCTTTCTTGCTGTCCAATTT C3') (30.5 nmol), 1 μL reverse primer 'HEK_lin_r' (5'GGTGGCTGTTATATTTCTTCTTACTCTTC3') (32.1 nmol), 1 μL plasmid DNA (target 50 ng μL^{-1}) and 22 μL dH₂O for a total volume of 50 μL . Reaction conditions consisted of initial denaturation at 98 °C for 10 seconds prior to 30 cycles of denaturation at 98 °C for 10 seconds, annealing at 51 °C for 15 seconds and extension at 72 °C for 20 seconds before final extension at 68 °C for 30 seconds.

Plasmids were recombined via the Gibson Assembly Protocol (New England Biolabs, E5510) consisting of 2 μL template, 3 μL plasmid backbone from the above PCR reaction, and 15 μL Gibson master mix, incubated for 1 hour at 50 °C in a 2 mL Eppendorf tube. Gibson assemblies were performed for each of the 5 sequences (0 – 100% +1FSS).

2.1.5 Bacterial transformation

Cultures of *E. coli* Top10 were made competent and transformed in accordance with Hanahan, 1983. LB (Luria-Bertani) broth was prepared using 10.0 g Tryptone, 5.0 g Yeast extract, 10.0 g NaCl and water up to 1 L, per litre made. Single colonies were added to 5 mL LB broth and shaken overnight at 30 °C, 600 μL of glycerol was added to the culture to achieve a 15% final concentration, 1 mL aliquots were taken and stored at -80 °C. 250 mL LB broth was placed in a 1 L conical flask and autoclaved for 2 hours at 127 °C. In a 1 L duran bottle, 1L of CCMB80 buffer was prepared using 0.98 g KOAc (10 mM), 11.8 g CaCl₂ · 2H₂O (80 mM), 4.0 g MnCl₂ · 4H₂O (20 mM), 2.0 g MgCl₂ · 6H₂O (10 mM), 100 g glycerol (10%) and water up to 1L. pH was adjusted using HCl to reduce to 6.4. The buffer was stored at 4 °C.

1 L LB broth was inoculated with the seed stock and incubated at 30 °C until an OD_{600} of 0.3 was achieved. Cells were harvested via centrifugation at 3000 rpm at 4 °C for 10 minutes, with supernatant discarded alongside any excess broth. Cells were resuspended in 80 mL CCMB80 buffer that had been kept on ice, and placed back on ice for 20 minutes once added to cells. A further

centrifugation step was carried out, and cells were instead suspended in 10 mL CCMB80 broth; 50 μ L of which was added to 200 μ L LB broth and the OD_{600} tested to be within the range 1.0 - 1.5. Aliquots were taken once more and stored at -80°C .

9 μ L DNA (post-Gibson Assembly) was placed in an Eppendorf tube on ice alongside an aliquot of competent cells for thawing. After approximately 20 minutes, 100 μ L competent cells were added to the DNA and incubated on ice for a further 30 minutes. A water bath was preheated to 42°C ; the competent cell-DNA mixture was heat shocked for 60 seconds before returning to ice. 1 mL LB broth was added to the mixture and shaken at 37°C for 45 minutes. A further centrifugation step was performed at 3000 rpm for 5 minutes, and 800 μ L supernatant removed before resuspension in the remaining 200 μ L.

Cells were then plated onto LB agar plates supplemented with 50 mg mL^{-1} kanamycin. Plates were incubated at 37°C for 16 hours. Isolated colonies were cultured in 2 mL LB broth for a further 16 hours whilst; cultures for confirmed sequences inoculated with a single colony in 10 mL LB broth, incubated for 16 hours.

2.1.6 DNA extraction

DNA was extracted from cultures using the QIAprep Spin Miniprep Kit and protocol (QIAGEN N.V.); 2 mL/10 mL cultures were centrifuged (8000 rpm, 3 min) and pellets resuspended in 250 μ L P1 buffer (50 mM Tris CL, 10 mM EDTA, 100 mg mL^{-1} RNase A) before lysis in 250 μ L P2 buffer (200 mM NaOH, 1% w/v SDS) for 5 minutes and neutralisation in P3 buffer (3 M sodium acetate). Lysed solutions were centrifuged (13,000 rpm, 10 min) and 800 μ L supernatant placed in a QIAprep 2.0 spin column before further centrifugation (8,000 rpm, 90 sec); supernatant was removed before 700 μ L buffer PE (proprietary) was centrifuged (8,000 rpm, 60 sec) through the matrix, and again to clear. Samples were eluted in 40 μ L buffer EB (10 mM Tris-CL) and stored at -20°C .

2.1.7 Sequence validation

In order to validate sequences, plasmid DNA was diluted to $50\text{ ng }\mu\text{L}^{-1}$; 15 μ L was added to an 1.5 mL safe-lock tube alongside 2 μ L primer (10 μM). Tubes were labelled and sent to Eurofins.

2.1.8 *In vitro* transcription

DNA was linearised prior to IVT in a restriction reaction consisting of 6 μL Bsp QI (New England Biolabs, R0712S), 30 μL DNA, 8 μL r3.1 buffer (New England Biolabs, B6003S), and 36 μL dH_2O , incubated for 1 hour at 50 °C. Bsp QI is a type IIS restriction enzyme, cleaving outside of the recognition sequence (GCTCTTC(1/4)), we used Bsp QI to cleave a clean end to the poly(A) whilst ensuring reliable targeting of the site.

The recognition site and cleavage site can be denoted as follows:

5' A A A G G A A G A G C 3'
 3' T T T T C **CTTCTCG** 5' at positions 979-990, green indicating the recognition sequence and underlining indicating cleaved bases, the poly (A) tail ends at the beginning of the initial 5' end.

IVT was carried out in accordance with the Takara Pro IVT kit and protocol (Takara Bio Inc.); whilst the protocol includes alternative DNA preparatory steps, those undertaken in prior steps were equivalent in outcome ($>50 \text{ ng } \mu\text{L}^{-1}$ DNA yield) and used reagents and equipment available in the laboratory.

The IVT protocol consists of thawing of reagents at room temperature (10x enzyme mix on ice) before preparing reactions in an appropriate container (1.5 mL Eppendorf Microtube); inputs per reaction comprised 2 μL 10x transcription buffer, 2 μL 10x enzyme mix, 2 μL 10x adenosine-triphosphate (ATP)(100 mM), 2 μL 10x cytidine-triphosphate (CTP)(100 mM), 2 μL 10x guanosine-triphosphate (GTP)(100 mM), 2 μL 10x uridine triphosphate (UTP)(100 mM), 1.6 μL CleanCap Reagent AG (TriLink BioTechnologies, N-7113)(8 mM), and 1 μg DNA. Since the volume of 1 μg DNA is dependent on concentration, a variable amount of nuclease-free water (New England Biolabs, B1500S) was added for a total volume of 20 μL . Reagents were mixed and incubated at 37 °C for 2 hours, then a further 15 minutes with 4 μL DNase I to halt the reaction. In IVT reactions for our experimental sequences, 2 μL 100 mM N1-methyl-pseudouridine (m1 Ψ) (New England Biolabs, N0431S) substituted uridine-triphosphate (UTP)(100 mM) at this step. Sequences transcribed purely for secondary structure analysis were prepared using the standard UTP.

mRNA was recovered from the reaction solution via lithium chloride (LiCl) precipitation; Takara Bio Inc. provides their Lithium Chloride Precipitation Solution within the kit. The solution was thawed at room temperature; 30 μL was added to reactions alongside 30 μL nuclease-free water then held at -20 °C for 30 minutes. Samples were centrifuged at 4 °C at 13,000 rpm for 15 minutes before a wash step with 1 mL 70% ethanol and a repetition of the centrifugation. Finally, supernatant was removed with care taken to not disturb any formed pellet of mRNA, air dried and resuspended in 30 μL nuclease-free water. mRNA samples were stored at -70 °C.

2.1.9 mRNA quantification and quality analysis

The 4150 TapeStation System (Agilent Technologies) was used to analyse mRNA samples, prepared in accordance with the manufacturer protocol (Agilent Technologies). The protocol entailed thawing the sample buffer at room temperature for 30 minutes, thawing RNA samples and ladder on ice. Flicking and testing the 'RNA ScreenTape Device' (consumable electrophoresis capillaries) in the instrument to ensure it is not past expiry and which capillaries were spent. To prepare RNA for electrophoresis, 1 μ L sample mRNA or 1 μ L ScreenTape Ladder was loaded alongside 5 μ L sample buffer into a 10-tube strip, sealed and vortexed at 2000 rpm for 1 minute. Samples were then centrifuged briefly before denaturing at 72°C for 3 minutes and placing on ice for 2 minutes. Strips were loaded into the instrument and labelled accordingly.

Data was analysed using the proprietary TapeStation software to produce electropherogram, showing concentration graphs across the full nucleotide-length range, indicating the quality of the mRNA through the ratio of any peaks to material at lower nucleotide lengths. Despite the increased cost of using this method, it provides confirmation that observed concentrations comprise viable mRNA free of contaminants, of the correct molecular weight.

Prior to any experimental work with mRNA samples, concentrations were remeasured via UV spectrophotometry (A_{260}) using a NanoDrop ND-1000 (Thermo Fisher Scientific Inc.). 1 μ L mRNA was placed onto a cleaned sensor and concentration measurement taken from the NanoDrop 1000 software (v3.8.1).

2.2 mRNA expression assay processes

2.2.1 Human tissue culture propagation

Cultures of HEK 293 were propagated in T25 tissue culture flasks containing 6 mL Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific Inc.) supplemented with 10% Fetal bovine serum (FBS) (Thermo Fisher Scientific Inc.) and incubated at 37°C with 5% CO₂. Cultures were passaged every ~72 hours (once high confluence was achieved), involving the following steps: washing cells with 3 mL phosphate-buffered saline (PBS), cell dissociation with 2 mL trypsin (held at 37 °C for 3 minutes), dilution through addition of 4 mL DMEM and transfer to a 15 mL Falcon tube before centrifugation at 200 x g for 5 minutes. Supernatant was discarded and the pellet was

resuspended in 5 mL DMEM, with 1 mL transferred to a new T25 (1:5 split) and supplemented with 5 mL DMEM. Each solution was warmed to 37 °C prior to use.

2.2.2 NanoLuc expression assays

In preparation of an assay, excess media was removed and cells dissociated with 2 mL trypsin, incubated at 37 °C for 3 minutes, and resuspended in 6 mL DMEM before counting a sample 500 µL using the Vi-CELL cell viability analyser (Beckman Coulter Life Sciences). In the 24-hour assay, wells of a 24-well plate were seeded with 1.5×10^6 cells each, immersed in 800 µL DMEM growth media and incubated overnight to increase confluence.

Prior to transfection, a two-part solution was prepared with each containing 100 µL Opti-MEM (Thermo Fisher Scientific Inc.) alongside either 1 µL Lipofectamine 2000 (Thermo Fisher Scientific Inc.) or 250 ng mRNA per test well. Parts were mixed equally and incubated at room temperature for 20 minutes and added dropwise to the prepared wells. Plates containing transfected wells were incubated at 37 °C with 5% CO₂ until measured.

To measure the luminescence, and by proxy, the quantity of NanoLuc present in cells in each well, Nano-Glo Luciferase Assay Substrate and Buffer (Promega Corporation) were mixed at room temperature in a ratio of 1:50 respectively for a total volume of 100 µL per well. When using the 24-well plate for the 24-hour expression assay, 24 µL Substrate was added to 1.2 mL Buffer for 100 µL applications to 12 wells. For the 72-hour expression assay, 4.8 µL Substrate was added to 240 µL Buffer for 20 µL applications to 12 wells, repeated twice. Luminescence measurements were directly measured using the FLUOstar Omega microplate reader (BMG Labtech). The sequence was an intracellular version of Nanoluciferase, not secreted.

In contrast to the singular 24-hour expression measurements prior, the 72-hour assay was conducted in a 96-well plate. To reflect the reduced total volume of 200 µL, each well input was reduced to 20% of those in the 24-hour assay, with 3.0×10^5 cells seeded per well in 160 µL DMEM growth media, transfection with 50 ng mRNA alongside 40 µL Opti-MEM and 0.2 µL Lipofectamine 2000.

2.3 mRNA secondary structure analyses

2.3.1 RNAfold secondary structure predictions

Each of the 5 experimental sequences were inserted into the 'Sequence Input' field of the RNAfold Webserver (accessible at rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi), with no deviation from default settings. An option is available under 'Modified Bases' -> 'Apply energy corrections for modified bases' that when ticked, the substitution can be made for U representing uridine with P representing pseudouridine, although no predictions were made using this modified energy parameter. There is currently no option available for N1-methylpseudouridine.

The web tool outputs the minimum free energy (MFE) structure, centroid structure and free energy in each arrangement (kcal mol^{-1}) alongside the ensemble free energy (kcal mol^{-1}), diversity and frequency of the MFE frequency within the ensemble. A submission was made for each experimental sequence (0-100% +1 FSS), with values recorded alongside image files for the predicted MFE and centroid structures.

2.4 mRNA DSF Assays

A series of melts were performed using the CFX Connect Real-Time PCR Detection System (Bio-Rad Laboratories Inc.), utilising the functions of qPCR to both excite at a given wavelength and detect emissions; readings could be taken across a designated temperature gradient (40 – 90 °C, later 35 – 95 °C). This method follows the principles of DSF assays carried out by Silvers *et al.*, 2015.

Acridine orange (AO) (Sigma Aldrich, 235474) was used as an intercalating dye, bound to experimental mRNA sequences in solution to differentially fluoresce when bound to single stranded and double stranded mRNA. Acridine orange has an excitation maxima at 460 nm, and emission maxima bound to dsRNA at 560 nm (Darzynkiewicz *et al.*, 2004), both within instrumental ranges.

The first test melt to determine the correct dye concentration contained well inputs comprising 26 μL dH₂O containing AO at 300 nM, 3 μM , 30 μM and 300 μM , 3 μg mRNA (sequence 5, 100% +FSS) diluted in 4 μL dH₂O for a total volume of 30 μL per well. n=2 samples per dye concentration. A temperature range of 40 – 90 °C was used for this melt at a ramp rate of $0.125\text{ }^{\circ}\text{C s}^{-1}$, measured every 4s at each 0.5 °C increment; the same ramp rate and measurement interval was used for each melt.

In the melt comparing curves for different sequences, well inputs comprised 26 μL dH_2O containing 30 μM AO and 4 μL dH_2O containing 3 μg mRNA for each sequence 1, 2, 4, 5 (0, 25, 50 and 100% +1 FSS). The melt was conducted across a temperature gradient of 35 – 95 $^{\circ}\text{C}$ with a ramp speed of , with fluorescence measured every . Insufficient mRNA was available for sequence 3 (75% +1 FSS).

To mimic cytosolic conditions, a solution was prepared containing 50 mM phosphate buffer (pH 6.5) and 1 mM Mg^{2+} , phosphate buffer was prepared from 96.75 mM disodium heptahydrate and 153.2 mM monosodium dihydrate in a 2 mL volume with dH_2O (in line with Chheda *et al.*, 2024). 30 mM AO was added before the third melt.

Well inputs for the third melt comprised 26 μL buffered solution (50 mM phosphate buffer, 1 mM Mg^{2+} , 30 μM AO), and 4 μL dH_2O containing 3 μg mRNA for each sequence 1, 2, 4, 5 (0, 25, 50 and 100% +1 FSS).

In each melt following the concentration test, both using pure water and a buffered solution, a singular melt was conducted per sequence due to the quantity of experimental mRNA available; this was a resource split between DSF and transfection assays with the latter taking priority.

Data was collected using the CFX Connect Maestro Software v2.3 (Bio-Rad Laboratories Inc.), with exports to .csv processed in Microsoft Excel. For each melt, raw RFU (relative fluorescence units), and derivative values were produced in separate documents, in addition to a plate view denoting well labels.

2.5 Statistical Approach

In the design of the 24-hour luminescence assays, three biological repeats were conducted in addition to three technical repeats. Values were normalised between biological repeats to ensure sequence-dependent results remained unaltered between repeats; this was conducted by determining the ratio between mean values in sequence 1 between each assay, and multiplying values for the entire assay by the ratio; hence normalising all to the average of sequence 1 assay two. To determine normality of the dataset, a Shapiro-Wilk test was performed. Followed by One-way ANOVAs, conducted between sequence 1,2 and 5. A Tukey's test was finally conducted as good practice to reduce possibility of a type I error (Zaiontz, 2026). In the 72-h luminescence assays standard errors were calculated for technical repeats and did not contain biological repeats (one was conducted). Beyond these experiments into the DSF assays, specific datapoints are referenced and ratios calculated, but not statistically significant conclusions drawn.

3.0 Results

3.1 mRNA preparation

Our experimental mRNA for use in subsequent assays was produced via *in-vitro* transcription (IVT), using DNA templates containing experimental sequences. Templates were amplified, then integrated into the cloning vector via Gibson assembly. The cloning vector included a T7 promoter for transcription initiation, KanR kanamycin resistance gene, ColE1 origin of replication (for replication in *E. coli* Top10). After cloning and extraction, our circular vectors were linearised via BSP QI. Features are illustrated and annotated in Figure 6, below. In order to be translatable in cells, RNAs require a 5' cap structure and poly(A) tail, here we used an m⁷GTP cap (CleanCap AG, Trilink Technologies) (bound co-transcriptionally) and an encoded 26-nt length poly(A) tail. Bases for uracil were replaced with N1-methylpseudouridine in the IVT solution where appropriate.

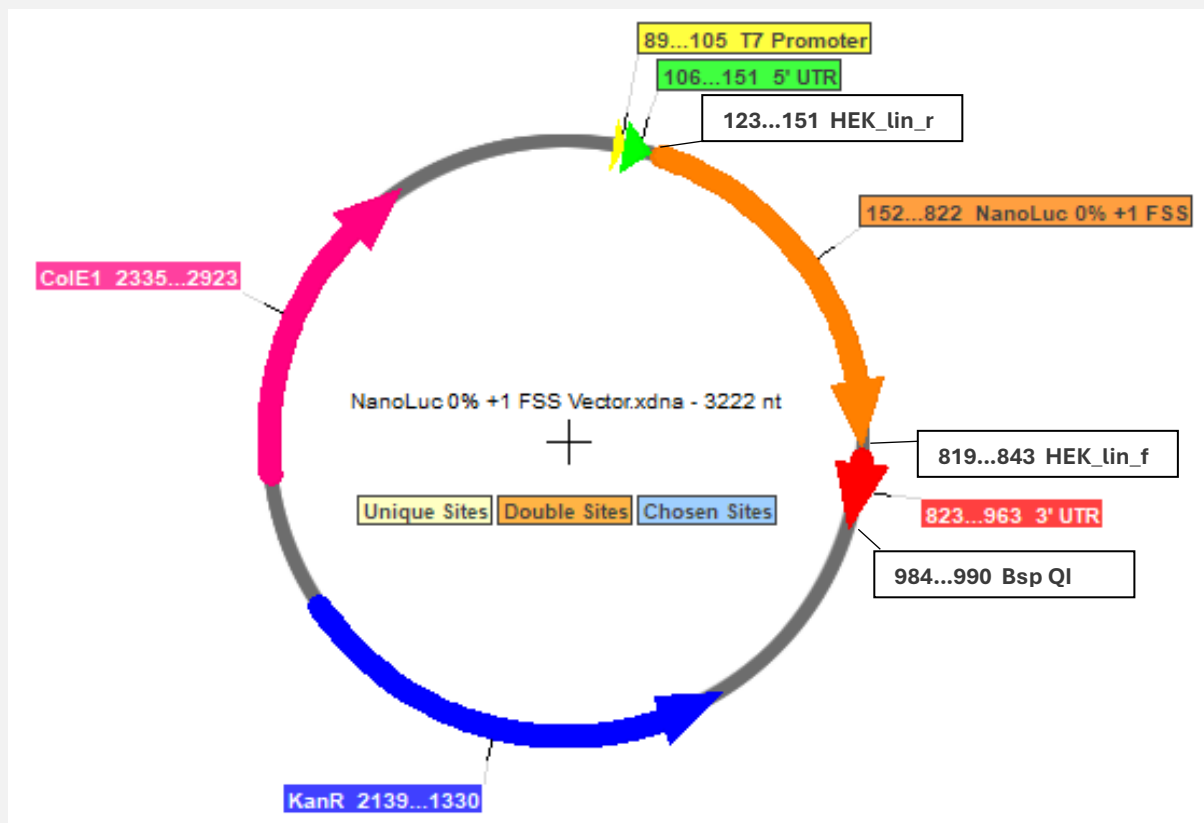


Figure 6. Graphical map of recombined NanoLuc vector. Graphical map of circularised vector post-recombination with an NanoLuc sequence (0% +1 frameshifted stop codons shown); included are the T7 promoter, ColE1 origin of replication, KanR kanamycin resistance gene, 5'UTR, 3' UTR, primer binding sites HEK_lin_r and HEK_lin_f and restriction site Bsp QI. Image exported from Serial Cloner 2.6 (Serial Basics) and additionally annotated inside Microsoft Word.

3.1.1 Template amplification

Template DNA for each NanoLuc ORF was delivered in a dry form, requiring resuspension to a molarity of 100 nM in 10 mM Tris-EDTA (TE). In order to be useful for *in vitro* transcription the ORFs needed to be combined with functional untranslated regions (UTRs) and a T7 promoter sequence. We chose a cloning strategy in which PCR-amplified NanoLuc ORFs replaced the ORF in existing template vectors already containing the T7 promoter and UTRs, as used in the Moderna COVID-19 vaccine (Jackson *et al.*, 2020). The strategy required PCR amplification of the NanoLuc ORFs, linearisation of the vector backbone without the original ORF which we also achieved by PCR, and recombining the two PCR products by Gibson assembly.

We initially amplified the NanoLuc ORFs PCR and PCR products were then validated via gel electrophoresis. All samples presented a strong band at the expected lengths of 671 bp (Fig. 7, below). Samples for sequence 3 and 4 displayed an additional band ~700 nt, the origin of which was unknown.

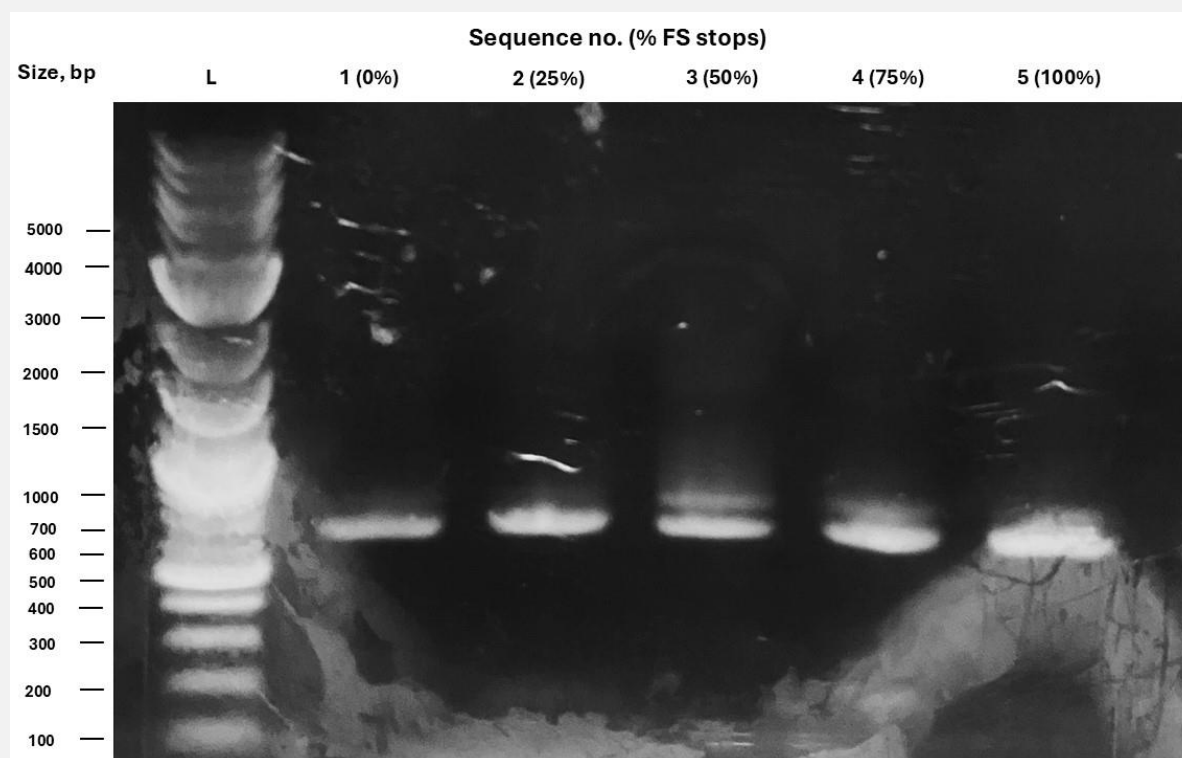


Figure 7. Gel electrophoresis of amplified template DNA. Lane contents left to right: 1 kb Plus DNA Ladder (New England Biolabs), sequences 1-5, (percentages shown indicate the relative content of stop codons in the +1 reading frame from an unaltered NanoLuc sequence up to the maximum possible). Lane inputs comprise 5 μ L DNA and 1 μ L loading dye (New England Biolabs).

3.1.2 Vector preparation for *in-vitro* transcription

The vector backbone comprised a T7 promoter, 5' UTR, 3' UTR, KanR resistance gene, ColE1 origin of replication and a 26-nt poly(A) tail, with an expected overall length of 2551 nt (shown in Figure 6, above). The T7 promoter initiates transcription in the *in vitro* transcription reaction, UTRs flanking the coding sequence complement and complete those on the insert sequence. The KanR resistance gene encodes kanamycin kinase, providing a selectable marker in cloning whilst the ColE1 origin of replication facilitates this replication within *E. coli*.

Initially, our vectors were circularised, containing a LAMP 2b gene between UTRs. The vector backbone was amplified via PCR using primers binding at each end of the CDS, 'HEK_lin_f' forward primer binding at positions 819-843 and 'HEK_lin_r' binding at positions 123-151 as illustrated in figure 6, above. Gel electrophoresis was performed the PCR product with the resulting band observable in Figure 8 (below) resides near the 2000 nt ladder marker indicating linearised DNA of the expected length.

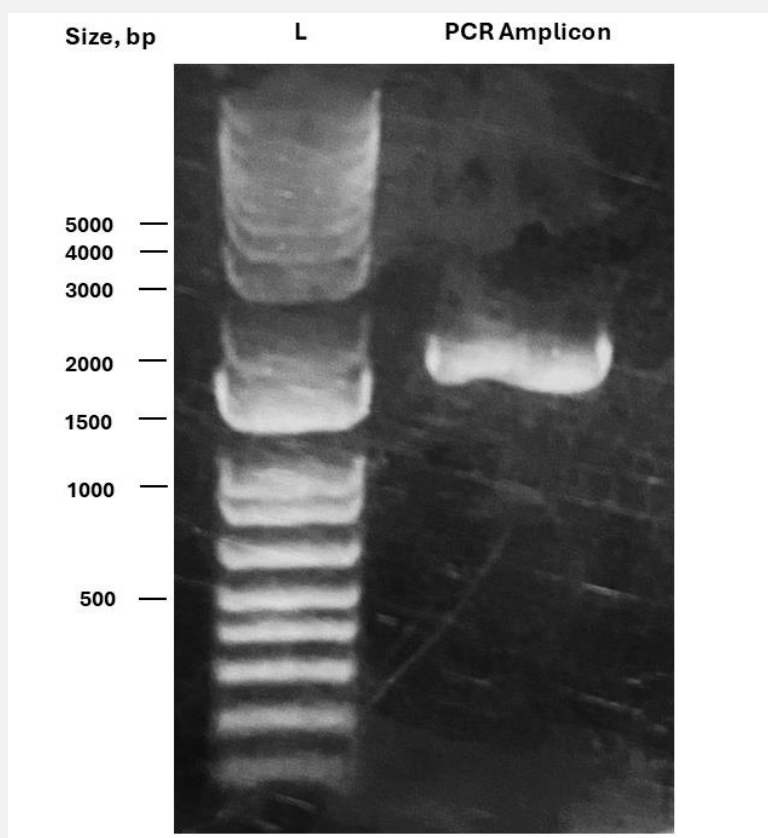


Figure 8. Gel electrophoresis of vector backbone amplicon. Lane contents: 5 μ L Invitrogen 1 Kb Plus DNA Ladder (Thermo Fisher) (left) and 10 μ L vector backbone PCR product (right). Both lanes were loaded with 1 μ L 10X BlueJuice Gel Loading Buffer (Thermo Fisher).

With the now linearised and amplified vector backbones, DNA inserts for each sequence (1-5, 0-100% +1 FSS) were recombined using Gibson assembly to form the final NanoLuc vector; Figure 8, above, depicts the desired vector sequence, containing sequence 1 (0% +1 FSS) as an example.

3.1.3 Bacterial transformation

Cultures *E. coli* Top10 were transformed with recombined plasmids from sequences 1-5. Three samples of 1 mL of culture were spread onto three LB agar plates containing 50 mg mL⁻¹ kanamycin spread, in addition to a sample containing only 3 μ L vector (linear, without a CDS insert) on two plates; the additional sample served as a reference to assess self-circularisation rates. After incubation overnight at 37 °C on LB agar, 15-20 colonies were observed per plate containing experimental transformants, and 8 colonies in the vector-only control. Colony morphologies were homologous, exhibiting a circular form, slightly raised elevation and average diameter of 1 mm.

3.1.4 Clone analyses

To ensure isolated clones contained only the intended sequence, DNA was extracted from 2 mL cultures and subjected to further analysis. Four samples were sequenced for each experimental sequence and results returned from Eurofins were aligned in Serial Cloner. Results for sequences 1, 2 and 5 (0, 25, 100% stop codons) returned one or more sequences with 100% alignment. Sequences 3 and 4 were omitted from further experiments as no significant sequence similarity was found (<40% alignment) and the stop codon contents of 50 and 75% being intermediate values were deemed non-essential to the overall project aims. Multiple attempts were made to identify the cause of invalidated plasmids, including reparation from the stage of Gibson Assembly, but each sequencing result returned invalid.

3.1.5 DNA analysis

To prepare sufficient DNA for *in vitro* transcription, bacterial colonies containing sequence-verified DNA were inoculated into 10 mL of LB broth and extracts were prepared using standard plasmid miniprep kits, yielding final DNA eluted in 40 μ L buffer EB. Nanodrop readings were performed to measure DNA purity and final concentrations as displayed in Table 1 below. We aimed for concentrations of 150-200 ng μ L⁻¹ to enable *in vitro* transcription in accordance with protocol, however, samples 1 and 2 measured approximately 1/3 of the desired concentration with an A_{260}/A_{280} ratio below that of pure DNA. The A_{260}/A_{230} ratios being >2.0 for each sequence indicated absence of any significant salts or other contaminants that would otherwise impede transfection efficiency.

Additionally, gel electrophoresis was performed on the plasmid DNA extracts to analyse samples qualitatively. All bands shown in Figure 9, below, are of similar size and contain no contaminants, although incomparable to nucleotide values for linear DNA shown in the ladder lane. Omitted sequences 3 and 4 were included in the electrophoresis as part of investigating sequence invalidation.

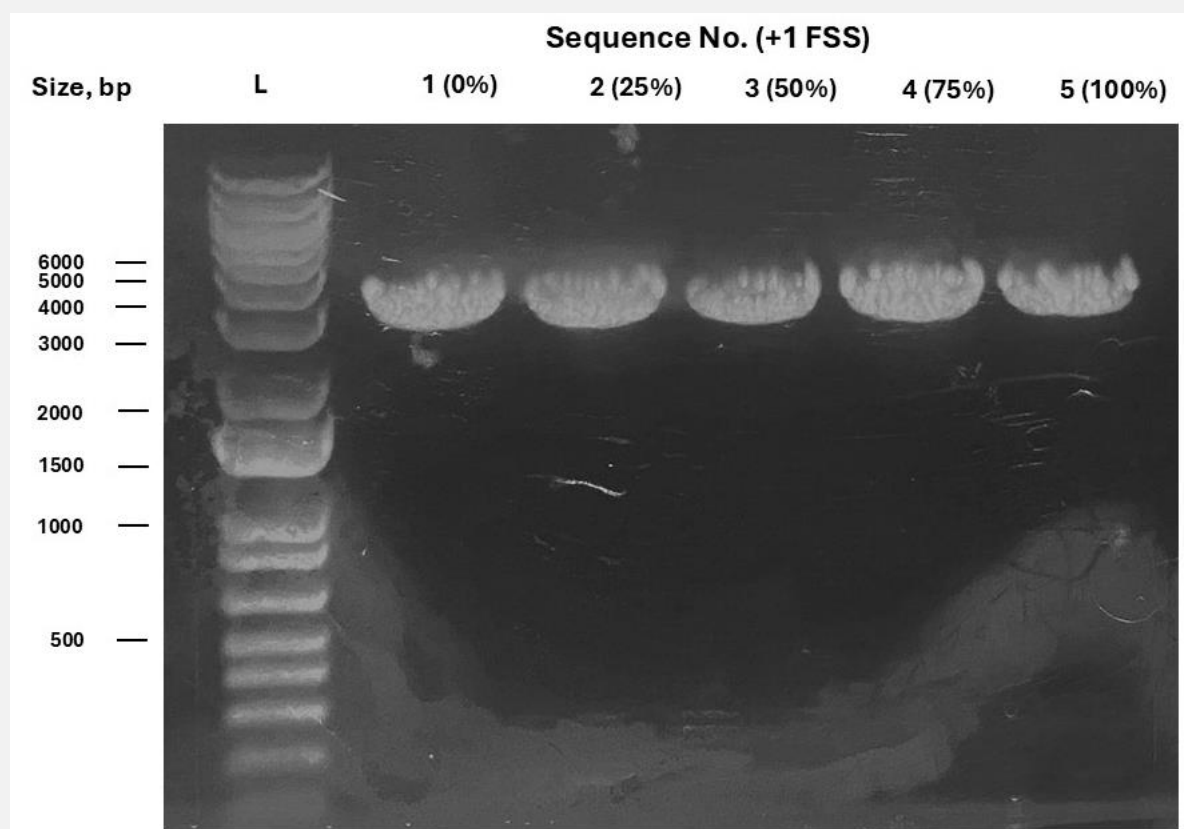


Figure 9. Gel electrophoresis of extracted plasmid DNA. Lane contents left to right: Invitrogen 1 Kb Plus DNA Ladder (Thermo Fisher), sequences 1-5, (percentages indicate +1 frameshifted stop codon content relative to the commercial sequence). Lane contents comprise 5 μ L DNA and 1 μ L 10X BlueJuice Gel Loading Buffer (Thermo Fisher). Sequences 3 (50%) and 4 (75%) were omitted from further study due to sequence invalidation.

Sample number	Concentration (ng μ L ⁻¹)	A_{260}/A_{280} ratio	A_{260}/A_{230} ratio
1	64.9	1.43	2.52
2	77.7	1.53	2.38
5	144.8	2.01	2.63

Table 1. Nanodrop readings of extracted DNA

^a Samples were analysed using a nanodrop cassette reader.

^b Volumes of 1 μ L DNA per well alongside 5 μ L sample buffer.

^c Purity was determined using A_{260}/A_{280} ratios (~1.8 indicating pure DNA).

3.2 *In-vitro* transcription

In preparation of *in vitro* transcription (IVT), plasmid DNA was extracted as specified above in section 2.1.6 and linearised via restriction with BspQI restriction enzyme. The sequence succeeding the poly(A) tail was engineered to contain the BspQI recognition sequence (as outlined in section 2.1.7) and would thus cleave the tail at the final adenine. Restriction reactions were performed for experimental sequences 1, 2 and 5. Restriction efficiency was not directly verified, however the reaction time at 1 hr provided both sufficient time for complete digestion against a singular site whilst minimising risk of STAR activity (non-specific cutting of DNA).

Since an input of 1 µg DNA was required for IVT, and concentrations of 64.9, 77.7 and 144.8 ng µL⁻¹ were achieved for sequences 1, 2 and 5 respectively. Volumes of 15.4, 12.87 and 6.90 µL (respectively) were used as inputs to the IVT mixture, each made up to 20 µL with nuclease-free water in accordance with section 2.1.7.

3.2.1 mRNA analysis

Viable *in vitro* transcribed mRNA was required to perform analyses on each of the experimental sequences. For sufficient uptake into Lipofectamine complexes, and by extension, successful transfection into mammalian cells, the mRNA concentration should exceed ~200 ng µL⁻¹. mRNA concentrations and purities were verified using the 4150 TapeStation System and TapeStation Software 5.1 (Agilent Technologies). Concentration values (Table 2, below) indicated samples 1 and 5 required resuspension in reduced volumes (67 and 25 µL respectively) to achieve the 200 ng µL⁻¹ target whilst sample 2 required a 4x dilution.

The graphical view displayed within the TapeStation Software (Fig. 10, below) expresses the concentration of nucleic acid present in a sample at a given nucleotide length in the y-axis. By contrasting known lengths against lower nucleotide lengths in the x-axis we can infer the integrity of each mRNA sample. Conformal sharp peaks are present at the 900-nt x-axis point, indicating the *in vitro* transcription products were of expected size. A small shoulder is present to the left of each peak in all samples; these reside in the 300-750-nt region and are proportional to the peak size between samples and most likely represent low levels of degradation products.

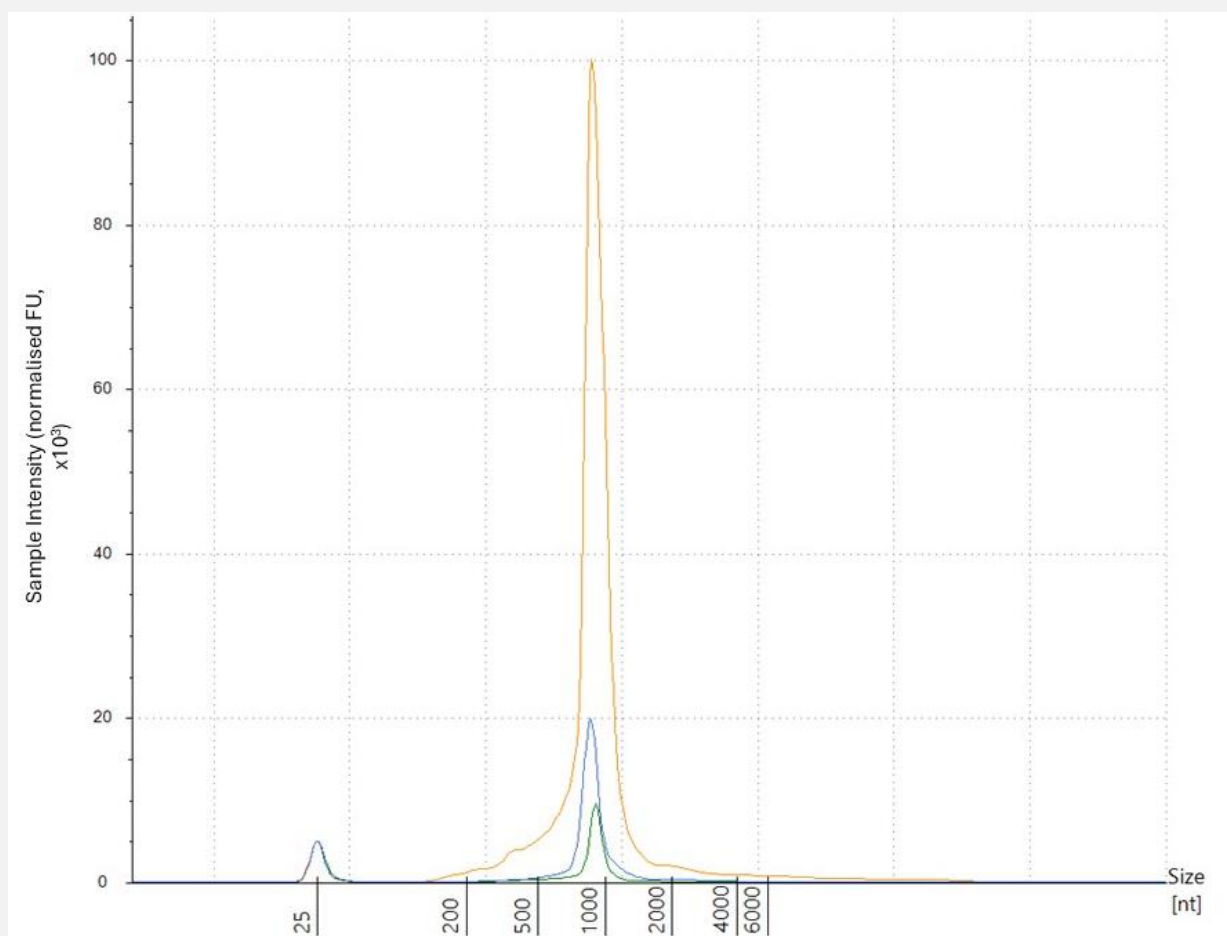


Figure 10. Graphical comparison of transcribed mRNA capillary electrophoresis. Graphical output from the Agilent TapeStation Software 5.1 (Agilent Technologies). Samples 1, 2 and 5, shown in blue, yellow and green respectively, are derived from capillary electrophoresis of 1 μ L mRNA alongside 5 μ L sample buffer. A 25-nt marker is present within the sample buffer.

Sequence number (% +1 FSS)	Concentration (ng μ L ⁻¹)
1 (0)	135
2 (25)	817
5 (100)	50

Table 2. Capillary electrophoresis quantification of transcribed mRNA.

^a Samples were analysed using the 4150 TapeStation System and TapeStation Software 5.1 (Agilent Technologies).

^b Volumes of 1 μ L mRNA per well alongside 5 μ L sample buffer.

3.3 NanoLuc expression assays

To investigate the implications of implementing stop codons in the +1 frame through use of codons with lower usage bias, the sequence variants were formulated with lipofectamine 2000 and transfected into cultured cells. The effect on translational efficiency in our mRNA sequences was then measured in a single time-point Nanoluciferase assay, with a subsequent time-course assay then carried out to determine differences in the biological stability of our mRNA sequences. Transfection was performed on Human Embryonic Kidney 293 cells (HEK 293), chosen as a model human cell line with an established tRNA profile (Polte *et al.*, 2019) (Fig. 3, above). The tRNA profile enables comparisons between the known differences in average tRNA concentration for each substituted codon pair, and the subsequent effect on both translational efficiency and longevity.

3.3.1 24-hour expression

Sequence number (% +1 FSS)	Concentration (ng μL^{-1})
1 (0)	112
2 (25)	738
5 (100)	48

Table 3. Spectrophotometry quantification of transcribed mRNA prior to 24h expression.

^a Samples were analysed using the NanoDrop ND-1000 (Thermo Fisher Scientific Inc.).

^b Concentrations derived from A_{260} .

^c Volumes of 1 μL mRNA per measurement.

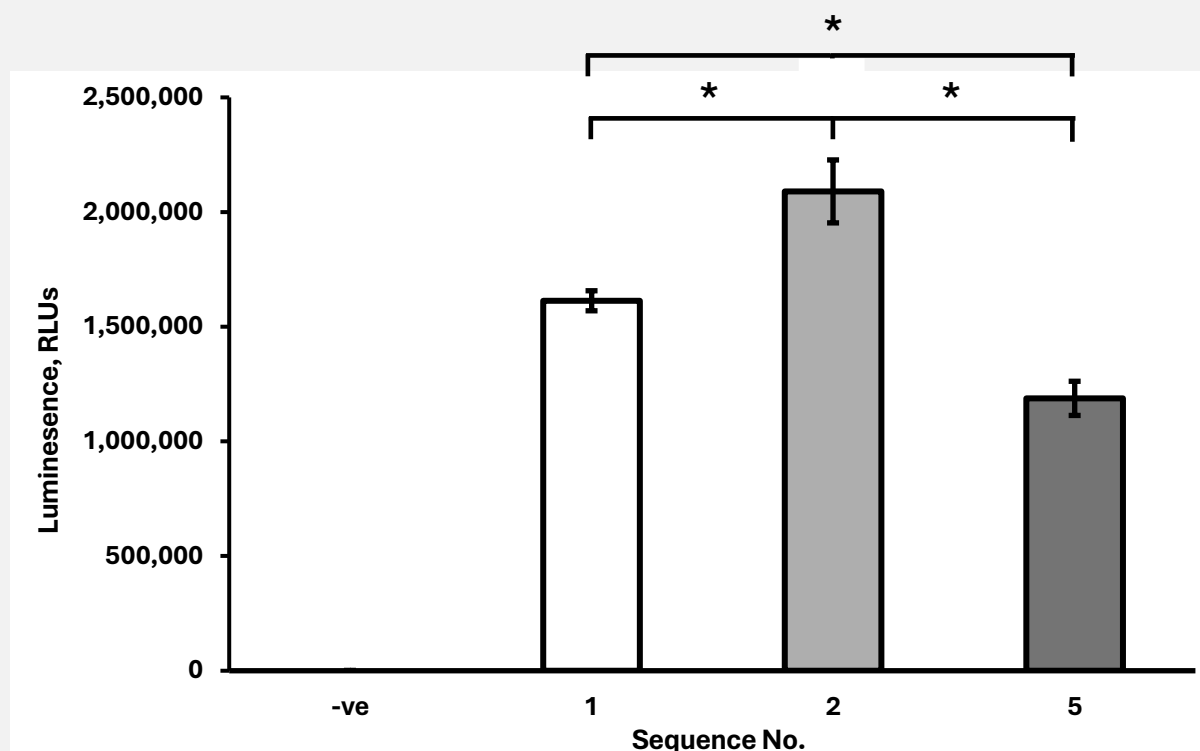


Figure 11. Normalised NanoLuc luminescence 24 hours post-transfection. HEK 293 cells were transfected with mRNA encoding NanoLuc. Sequences 1, 2 and 5 contained 0, 25 and 100% +1 FSS, relative to the commercial sequence. n=9 luminescence readings were taken per sequence (three technical replicates across 3 biological replicates) and normalised to the sequence 1 mean of the second replicate (source data, Appendix 2). Readings were produced by the FLUOstar Omega microplate reader (BMG Labtech). Error bars indicate $\pm$ standard error. Statistically significant differences ($p < 0.05$) signified by *.

Concentrations of mRNA, as shown in Table 3 above, were confirmed prior to transfection to calculate volumes required for each sequence. To equal 2.25 μ g mRNA (250 ng across 9 wells), volumes were calculated to 20.0, 3.0. and 46.8 μ L for sequences 1, 2 and 5 respectively.

Luminescence measurements were taken 24 hours post-transfection using the FLUOstar Omega microplate reader (BMG Labtech), measuring intracellular Nanoluciferase i.e. non-secreted. A total of 9 wells were analysed per mRNA sequence, comprising triplicate wells across three biological replicates. Values were normalised by the ratio of the mean in sequence 1 triplicates in the second technical repeat (since they were highest amongst sequence 1) relative to that of technical repeats one and three; the calculations were as follows:

Sequence 1 mean, assay 1 = 502,258, assay 2 = 1,613,257, assay 3 = 886394.

Normalisation factor for assay 1 = $1613257/502258 = 3.212$

Normalisation factor for assay 3 = $1613257/886394 = 1.82$

Therefore RFU readings in each assay were multiplied by these factors (appendix 2). This accounts for any variability between biological repeats. Due to a high gain setting, two readings for sequence 2 in replicate 2 saturated the plate reader ($> 2.0 \times 10^6$ RLUs) and were thus omitted from analysis. The included value for sequence 2 in the third technical repeat (3,114,020 RFU) were below this threshold pre-normalisation (1,711,000 RFU).

Normalised mean values are displayed in Figure 11, above; sequence 1 (0% additional +1 FSS) gave a baseline of 1.61×10^6 RLUs, increasing $\sim 30\%$ to 2.10×10^6 RLUs in sequence 2 (25% +1 FSS) and reducing $\sim 27\%$ to 1.18×10^6 RLUs in sequence 5 (100% +1 FSS). The negative control confirmed no detectable luminescence in non-transfected wells.

To determine normality of the dataset, and thus the correct statistical test to apply, a Shapiro-Wilk test was performed and found each set to be normal ($p < 0.05$). One-way ANOVAs conducted between sequence 1, 2 and 5 produced $p < 0.05$. A further Tukey's test was conducted as good practice to reduce possibility of a type I error; p-values of 1.3×10^{-3} , 4.4×10^{-3} and 3.78×10^{-3} were obtained comparing sequence 1 to 2, 1 to 5 and 2 to 5 respectively, indicating statistically significant differences between luminescence values for each sequence. Relative to the negative control, p-values of 2.85×10^{-14} , 7.22×10^{-15} and 6.75×10^{-11} were calculated for sequences 1, 2 and 5 respectively.

3.3.2 72-hour expression

Whilst measuring over a 24-hour period provided a baseline expression level for each experimental sequence, in order to understand the expression kinetics (and infer the biological stability of secondary structures), periodic measurements were required. Three measurements with an interval of 24 hours were chosen as the third measurement would occur at 72 hours, before significant confounding variables would arise from reduced cell viability in an aged medium with less reduction in expression being reliably attributed to mRNA degradation.

Concentrations of mRNA, as shown in Table 4 below, were confirmed prior to transfection to calculate volumes required for each sequence. To equal 450 ng mRNA (50 ng across 9 wells), volumes were calculated to 1.94, 2.39 and 3.51 μL for sequences 1, 2 and 5 respectively.

Sequence number (% +1 FSS)	Concentration (ng μL^{-1})
1 (0)	232
2 (25)	188
5 (100)	128

Table 4. Spectrophotometry quantification of transcribed mRNA prior to 72h expression.

^a Samples were analysed using the NanoDrop ND-1000 (Thermo Fisher Scientific Inc.).

^b Concentrations derived from A_{260} .

^c Volumes of 1 μL mRNA per measurement.

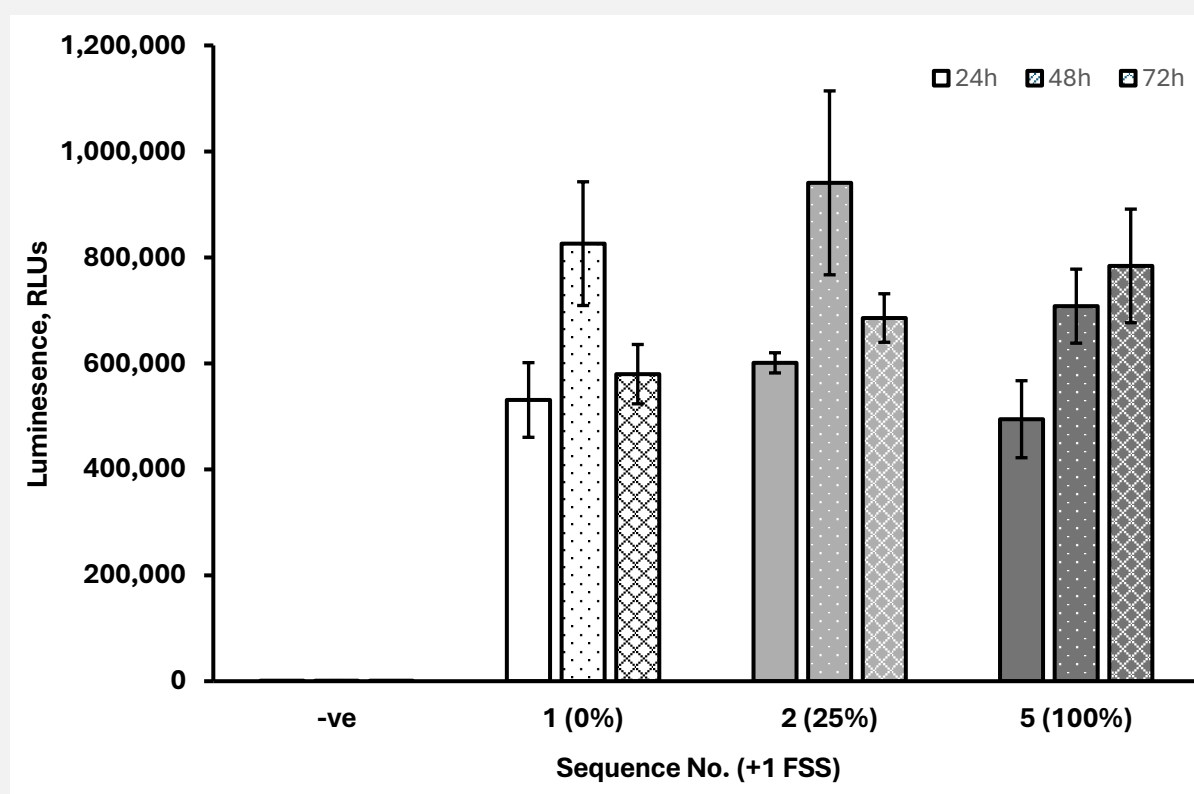


Figure 12. Periodic NanoLuc luminescence 24-72 hours post-transfection. HEK 293 cells were transfected with mRNA encoding NanoLuc luciferase. Sequences 1, 2 and 5 respectively contained 0, 25 and 100% +1 FSS, relative to the commercial sequence. $n=3$ luminescence readings were taken per sequence per timepoint comprising three technical replicates. Readings were produced by the FLUOstar Omega microplate reader (BMG Labtech). Error bars indicate $\pm$ standard error. Source data in appendix 3.

Similarities can be observed between sequences 1 and 2 (0 and 25% +1 FSS) (Fig. 12, above) with both showing a 55% increase between 24-hour and 48-hour luminescence, from 5.30×10^5 and 8.25×10^5 to 6.0×10^5 and 9.40×10^5 RLU respectively. Both then saw a 30% reduction to 5.79×10^5 and 6.85×10^5 at the 72-hour point. Measurements of sequence 2 were consistently higher than sequence 1 across each period. Luminescence of sequence 5 (100% +1 FSS) also saw an increase of 43% from 4.94×10^5 to 7.07×10^5 RLU at the 48-hour mark; and increased a further 10% to 7.83×10^5 RLU at 72-hours. No normalisation of these results was carried out since that of the 24-h experiment sought to remove differences between biological repeats; here, the pattern across each timeframe is of interest and only technical repeats were conducted.

Overall, these patterns indicate that introduction of +1 FSS may alter expression levels both by affecting inherent translation efficiency, and by altering mRNA stability (most likely through the effects of sequence changes on secondary structure). For example, constructs 1 and 2 display peak expression at 48 hours whereas sequence 5 displays peak expression at 72 hours (the only outlier from the pattern observed in figure 11), which indicates a longer persistence for this sequence in cells. Due to the singular biological experiment, these indications are not conclusive evidence.

3.4 Predictive mRNA secondary structure analyses

In order to assess how introduction of +1 FSS may have affected secondary structure, we conducted computational analyses using the RNAfold program on the Vienna RNA web server (Hofacker, 2003). We then also attempted to develop experimental assays to quantify secondary structure.

3.4.1 RNAfold predictions

Prediction of secondary structures for each experimental sequence was performed using RNAfold to provide both quantitative and qualitative data to draw relationships with empirical expression data. Each sequence was submitted to the RNAfold web server (accessible at rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi). Outputs for each experimental sequence consisted of the minimum free energy (MFE) structure (the most stable structure a sequence can obtain), centroid structure (the structure which has most base pairs in common with other members of the structural ensemble) alongside the frequency of these structures in the ensemble, the ensemble free energy, and the ensemble diversity (indicating how many base pairs differ in average between two randomly chosen structures in the ensemble). An important distinction is to be made between the centroid free energy and ensemble free energy where the former is derived from a determinant structure and latter from a statistical distribution of multivariate pairings. Although not experimentally investigated, sequences 3 and 4 are included within these computational analyses.

Sequence number (% +1 FSS)	Minimum free energy (kcal mol ⁻¹)	Centroid free energy (kcal mol ⁻¹)	Ensemble free energy (kcal mol ⁻¹)	Ensemble Diversity
1 (0)	-230.00	-146.32	-241.27	201
2 (25)	-221.10	-165.84	-231.72	186
3 (50)	-200.90	-160.00	-210.22	196
4 (75)	-196.00	-144.80	-208.42	176
5 (100)	-185.90	-147.80	-199.43	147

Table 5. RNAFold predicted secondary structure parameters *without* pseudouridine adjustments

^a Sequences were submitted to the RNAfold webserver accessible at rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi.

^b Temperature defaulted to 37 °C and salt concentration 1.021 M.

^c Energy parameters defaulted to Turner, 2004.

By default RNAFold calculates folding energies based on unmodified uridine although it has an option to calculate energies based on pseudouridine instead. Since base pairing energies for the N¹-methylpseudouridine used in our transcripts are actually closer to those of uridine than pseudouridine (Kierzek *et al.*, 2014), and only parameters for pseudouridine present within the ViennaRNA package (from which the RNAFold webserver runs), predictions were processed with uracil energy parameters.

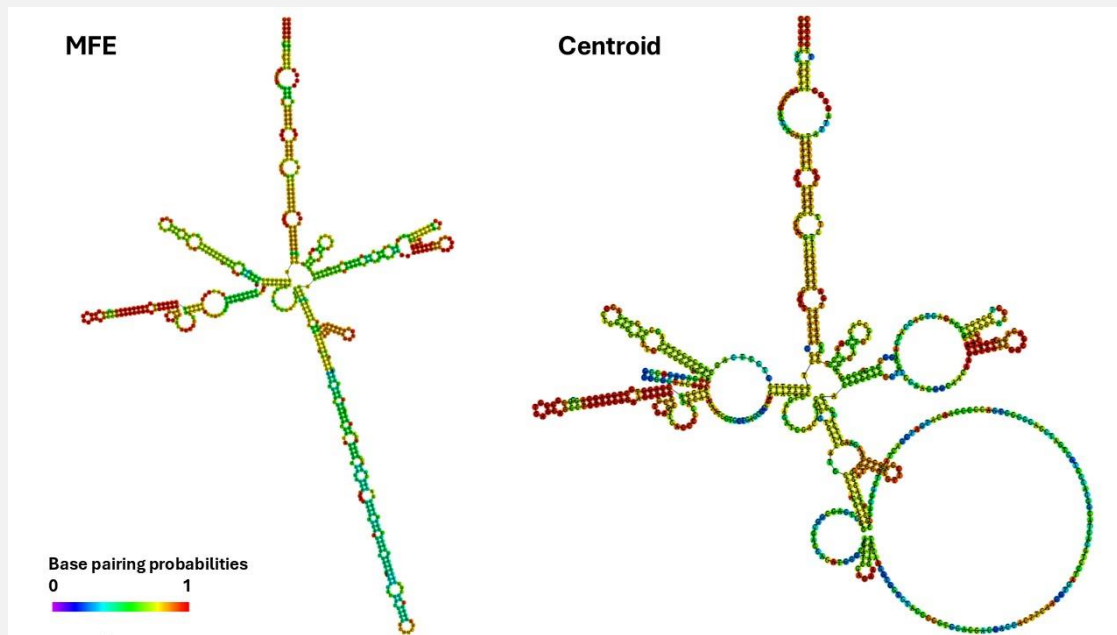


Figure 13. Predicted secondary structures of sequence 1 (0% +1 FSS). Experimental NanoLuc mRNA sequences were submitted to the RNAfold webserver with default settings (temperature 37 °C and salt concentration 1.021 M). Sequence 1 (Appendix 1) contains no additional +1 frameshifted stop codons relative to the commercial NanoLuc sequence. Left: Minimum free energy (MFE) secondary structure. Right: Centroid secondary structure. (Adapted from RNAfold output).

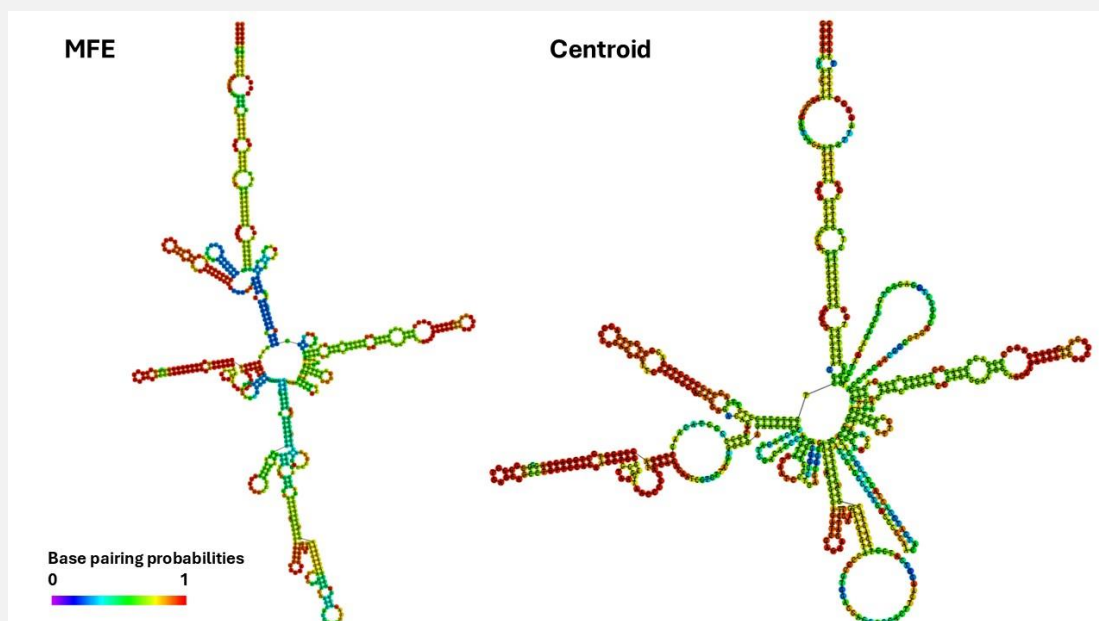


Figure 14. Predicted secondary structures of sequence 2 (25% +1 FSS). Experimental NanoLuc mRNA sequences were submitted to the RNAfold webserver with default settings (temperature 37 °C and salt concentration 1.021 M). Sequence 2 (Appendix 1) contains 25% additional +1 frameshifted stop codons relative to the commercial NanoLuc sequence relative to the maximal possible. Left: MFE secondary structure. Right: Centroid secondary structure. (Adapted from RNAfold output).

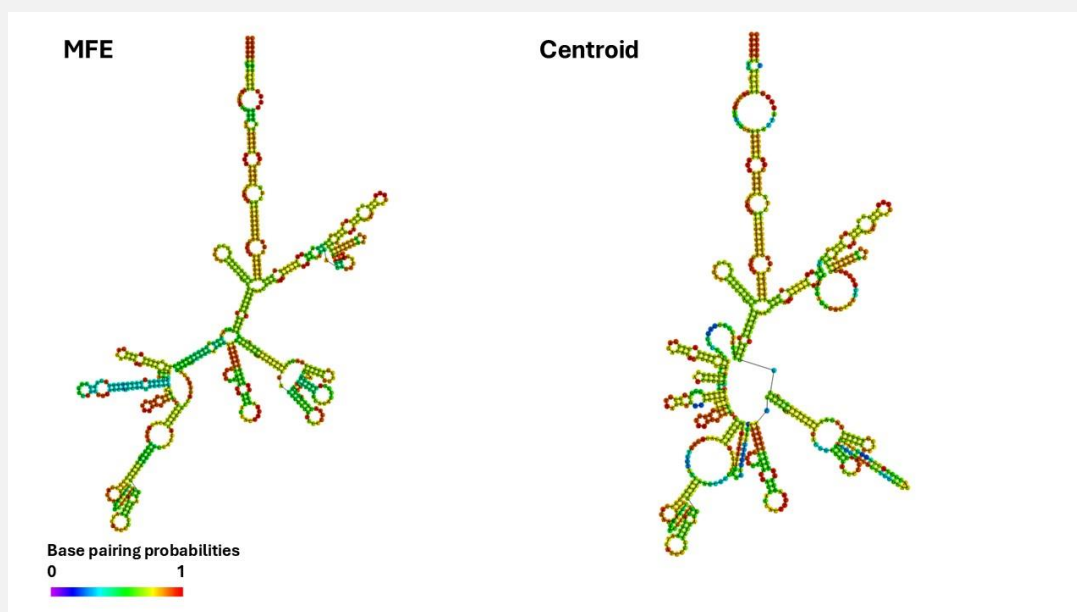


Figure 15. Predicted secondary structures of sequence 3 (50% +1 FSS). Experimental NanoLuc mRNA sequences were submitted to the RNAfold webserver with default settings (temperature 37 °C and salt concentration 1.021 M). Sequence 3 (Appendix 1) contains 50% additional +1 frameshifted stop codons relative to the commercial NanoLuc sequence relative to the maximal possible. Left: MFE secondary structure. Right: Centroid secondary structure. (Adapted from RNAfold output).

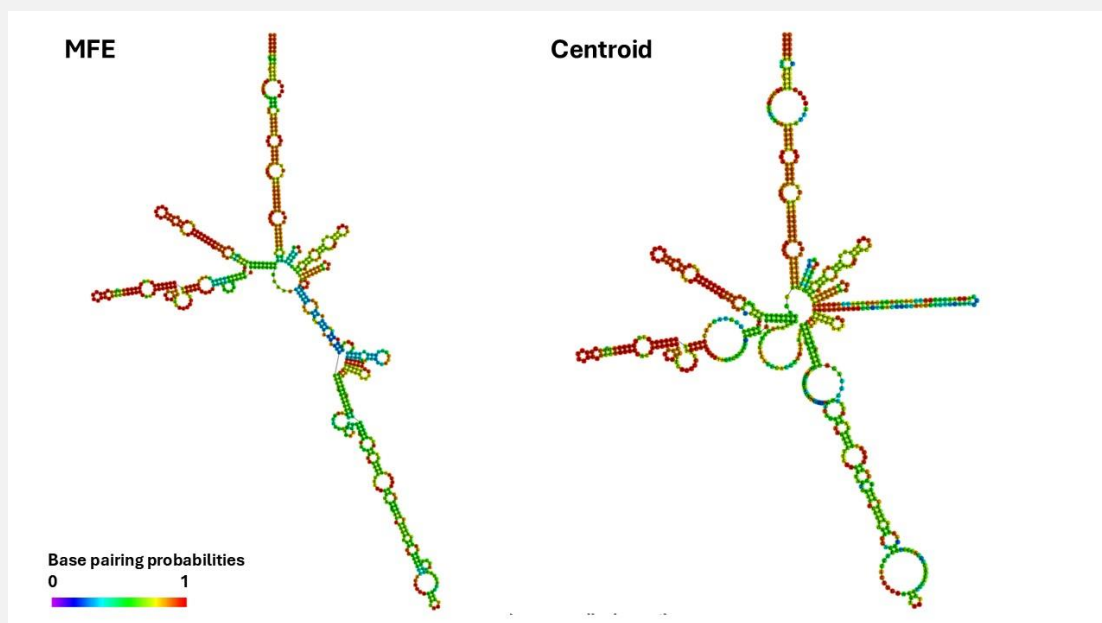


Figure 16. Predicted secondary structures of sequence 4 (75% +1 FSS). Experimental NanoLuc mRNA sequences were submitted to the RNAfold webserver with default settings (temperature 37 °C and salt concentration 1.021 M). Sequence 4 (Appendix 1) contains 75% additional +1 frameshifted stop codons relative to the commercial NanoLuc sequence relative to the maximal possible. Left: MFE secondary structure. Right: Centroid secondary structure. (Adapted from RNAfold output).

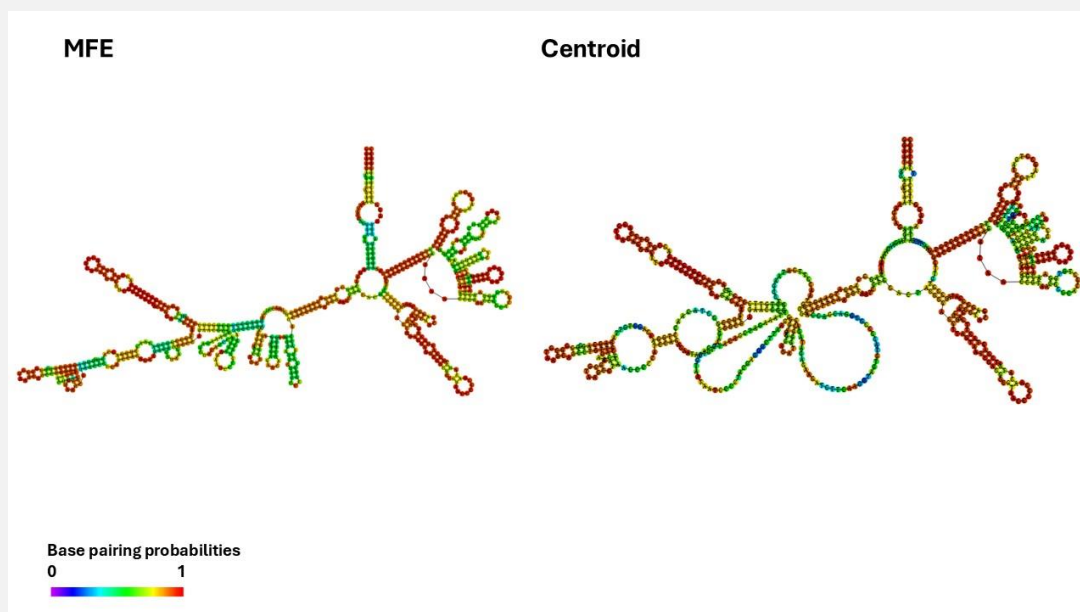


Figure 17. Predicted secondary structures of sequence 5 (100% +1 FSS). Experimental NanoLuc mRNA sequences were submitted to the RNAfold webserver with no changes to default settings (temperature 37 °C and salt concentration 1.021 M). Sequence 5 (Appendix 1) contains maximal additional +1 frameshifted stop codons relative to the commercial NanoLuc sequence. Left: MFE secondary structure. Right: Centroid secondary structure. (Adapted from RNAfold output).

Results displayed in table 5 (above) show increasing minimum free energies with increasing +1 FSS sequence content for MFE and ensemble calculations, from -230.0 to -185.9 kcal mol⁻¹ and -241.27 to -199.43 kcal mol⁻¹ respectively from sequence 1 to 5. The trend for centroid free energies differs where a decrease is observed between sequence 1 and 2, from -146.32 to -165.84 kcal mol⁻¹ before increasing to -147.80 kcal mol⁻¹ in sequence 5.

Qualitatively, structures generated for sequence 5 (Fig. 17, above) are most distinct from 1 through 4 (Fig. 13, 14, 15 and 16, above) as few elongated stems regions appear, instead displaying frequent short stems in parallel, joined by interior loops. Increasing similarities between the MFE and centroid structures can also be seen with increasing +1 FSS.

Colouration of bases observed in each structure represents the probability of being paired in the arrangement at a temperature of 37 °C and salt concentration 1.021 M. Strong differences in secondary structure are observed across each sequence, and between MFE and centroid structures. Sequence 5 (Fig. 17, above) also shows the highest confidence in base pairing at >90% (red) for approximately 35% of bases in both arrangements

Interestingly, the introduction of +1 FSS appears to reduce the predicted ensemble diversity (ED) of the different constructs. The ensemble diversity measures the average numbers of base pairs that differ between two randomly chose structures of the structural ensemble (Hofacker, 2003). Generally, higher ensemble diversity indicates that a sequence adopts a larger number of distinct structures with significant probability, whereas an ED value of 0 indicates that a sequence only adopts a single, stable fold. High ensemble diversity also implies that a sequence converts more easily between distinct structures, i.e. that base pairs are on average shorter lived. Shorter lived base pairs could facilitate ribosome progression through regions of secondary structure, whereas longer lived base pairs would more likely inhibit RNA decay. This is consistent with the observation that sequence 5 has the lowest ED value (Table 5, above), but the highest functional half-life (Fig. 12, above)

3.5 mRNA DSF assays

As a complementary experiment to the NanoLuc expression longevity assay, thermodynamic properties of the experimental sequences were investigated via differential scanning fluorimetry (DSF), a melt curve using intercalating dye bound to mRNA. In principle, thermal unfolding begins through denaturation of the molecule, unfolding initially at the weakest points of self hybridisation and progressing through more stable structures such as stems and hairpin loops. Reductions in self-hybridisation would be observed through decreased fluorescence of the intercalating dye (only fluorescing when bound), with steeper decreases as the stable structures decay. The expected result being to differentiate mRNA with stable structures from those without via the melting curve; a resource- and time-efficient method to provide a proxy for biological stability.

3.5.1 Pure samples

Since no specific protocols have been established pertaining to mRNA melts with AO in qPCR instruments, a series of tests were performed to identify suitable concentrations of mRNA, intercalating dye, and overall sample volume. The first test melt (Fig. 18, below) set out to determine the correct ratio of dye. A second test melt containing thiazole orange across the same concentrations failed to produce any discernable changes in fluorescence across the temperature gradient, thus it was determined that acridine orange was to be used over thiazole orange.

The result of a suitable concentration would ideally show a downward trend with increased rates of decline in fluorescence at distinct temperatures.

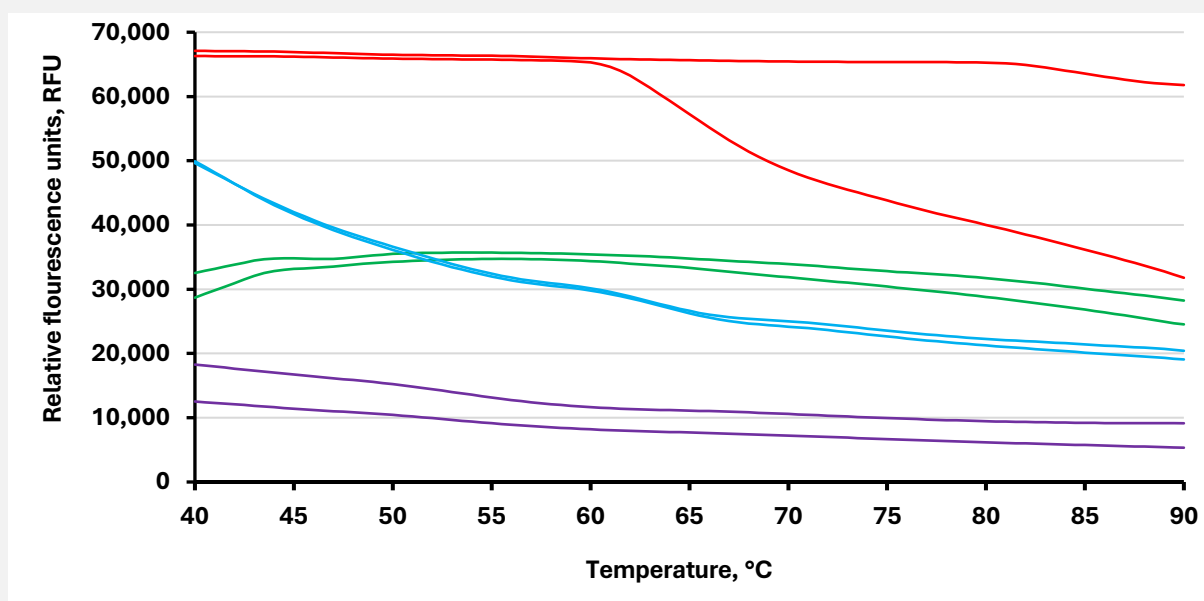


Fig 18. mRNA DSF melt curves comparing dye concentrations. Relative fluorescence units from varying concentrations of acridine orange at 300 μM (green), 30 μM (blue), 3 μM (red) and 300 nM (purple) in solution with 3 μg NanoLuc mRNA (m1 Ψ -containing) (sequence containing maximal +1 frameshifted stop codons) and dH₂O for a total volume of 30 μL . Measured across the temperature gradient 40 - 90 $^{\circ}\text{C}$ (+0.125 $^{\circ}\text{C s}^{-1}$) at 0.5 s intervals. $n=2$ samples were recorded per concentration. Melts were conducted using the CFX Connect Real-Time PCR Detection System and data collected via the CFX Maestro Software v2.3 (Bio-Rad Laboratories Inc.).

Each concentration of acridine orange (AO) exhibited different characteristics across the temperature gradient (Fig. 18, above); initial relative fluorescence units (RFU) were observed to be high for 3 μM (red) at ~65,000, lower for 30 μM (blue) at 50,000 and again for 300 μM (green) at ~30,000, with the lowest being 300 nM at ~15,000 (purple). The 3 μM concentration was the only concentration to differ between replicates, diverging at ~63 $^{\circ}\text{C}$ with one sample dropping to 50% of the steady ~65,000 reading, and the other retaining this reading until 65 $^{\circ}\text{C}$ where a slight drop is observed to ~60,000. Curves for 300 μM remained spaced by ~5,000 RFU, with a steady decline of 10,000 RFU observed in each across the whole gradient, the rate of decline decreased over the gradient. Curves for 30 μM followed a unique trajectory with increases in RFU of 5,000 observed up to ~57 $^{\circ}\text{C}$ before a decline to initial values at the final temperature of 90 $^{\circ}\text{C}$.

The most suitable concentration for DSF with AO was determined to be 30 μM as steady declines were observed in both samples across the whole temperature gradient, with a plateau as the decline reached 30,000 RFU at ~57 $^{\circ}\text{C}$ and sharp decrease to 25,000 RFU at 65 $^{\circ}\text{C}$; both samples continued a decline to 20,000 RFU by the final reading. Concentrations do not produce a linear relationship with

fluorescence intensity, whilst this may be due to dye saturation and potential sequestration of signal at higher concentrations, the cause was not directly investigated.

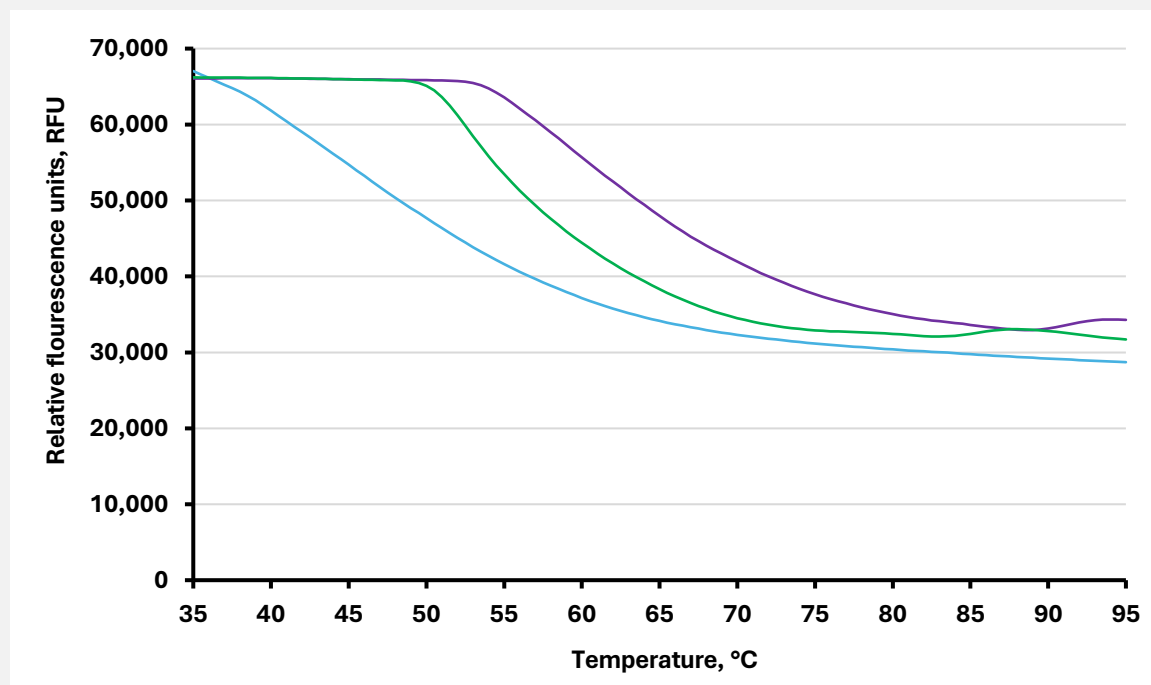


Figure 19. mRNA DSF melt curves comparing experimental sequences. Relative fluorescence units from samples comprising four different sequences of NanoLuc mRNA (m1Ψ-containing) with percentages of +1 frameshifted stop codons; sequence 1 (0%, blue), sequence 2 (25%, purple), sequence 5 (100%, green). Wells contained 3 μg mRNA in 4 μL dH₂O and 30 μM acridine orange in 26 μL dH₂O for a total volume of 30 μL. Measured across the temperature gradient 35 - 95 °C (+0.125 °C s⁻¹) at 0.5 °C intervals. n=1 samples per sequence. Melts were conducted using the CFX Connect Real-Time PCR Detection System using channel 1 (Ex λ = 450-490 nm, Em λ = 515-530 nm) and data collected via the CFX Maestro Software v2.3 (Bio-Rad Laboratories Inc.).

Curves for sequences 2 and 5 (25 and 100% +1 FSS) shown in Figure 19 (above) as purple and green respectively, both displayed a plateau at ~65,000 RFU until ~50 °C where each decrease at differing rates. Sequence 5 displayed the sharpest drop, to 35,000 RFU at ~67 °C whilst sequence 2 displayed the most delayed reduction in fluorescence, holding to ~55 °C before a similar decline to 35,000 RFU at 85 °C. Sequence 1 exhibited the most unique curve, with immediate decline from 65,000 RFU at 35 °C (the same initial fluorescence as other sequences) at a lower rate than that of other sequences, reaching 35,000 RFU at 63 °C and continuing to 30,000 RFU at 80 °C. Since only one repeat was conducted for the experiment, interpretations of the data remain limited to observation.

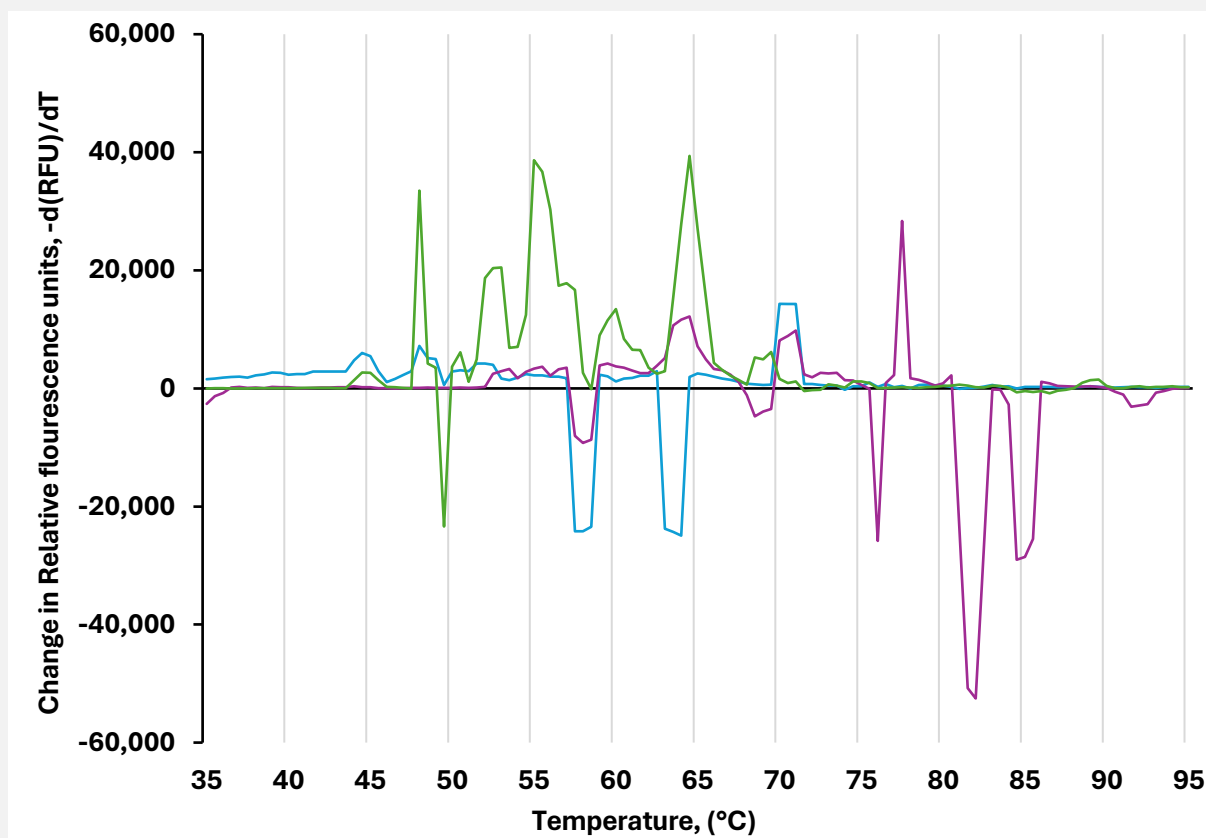


Figure 20. mRNA DSF melt curves comparing change in fluorescence in experimental sequences.

The negative derivative of the relative fluorescence units with respect to temperature ($-d(RFU)/dT$) from samples comprising four different sequences of NanoLuc mRNA (m1 Ψ -containing) with percentages of +1 frameshifted stop codons; sequence 1 (0%, blue), sequence 2 (25%, purple), sequence 5 (100%, green). Wells contained 3 μ g mRNA in 4 μ L dH₂O and 30 μ M acridine orange in 26 μ L dH₂O for a total volume of 30 μ L. Measured across the temperature gradient 35 - 95 °C (+0.125 °C s⁻¹) at 0.5 °C intervals. To reduce noise, data are expressed as a 3-point moving average. n=1 samples per sequence. Melts were conducted using the CFX Connect Real-Time PCR Detection System using channel 1 (Ex λ = 450-490 nm, Em λ = 515-530 nm) and data collected via the CFX Maestro Software v2.3 (Bio-Rad Laboratories Inc.).

To identify potential similarities between specific domains denaturing, a negative derivative was calculated using a 3-point moving average, Figure 20 (above) indicates stronger correlation between peak values in sequence 1 and 2 (purple and blue) than sequence 5 (green), particularly at 58 °C and 72 °C, suggesting a closer relationship between structured domains amongst the low to non-+1 FSS containing sequences against the fully optimised.

3.5.2 Buffered samples

To mimic cytosolic conditions, and thus those which mRNA will be exposed to and form secondary structures under, phosphate buffer (pH 6.5) and Mg^{2+} were included in the AO-containing solution. A further premise of the approach was to determine whether further stabilisation of secondary structures would produce more pronounced decreases in fluorescence.

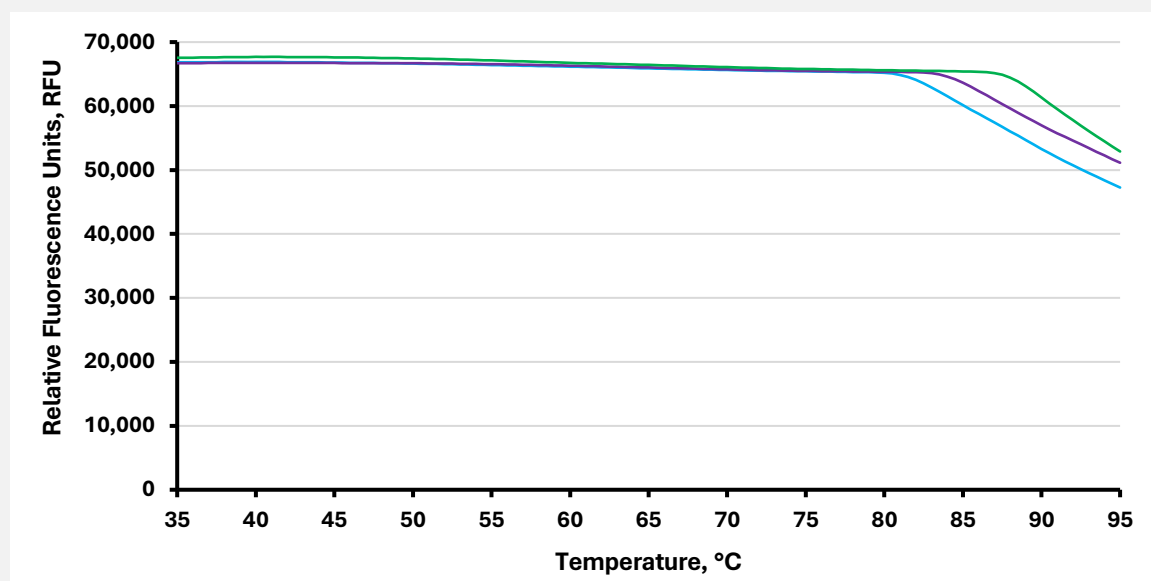


Figure 21. mRNA DSF melt curves comparing experimental sequences under buffered conditions.

Relative fluorescence units from samples comprising four different sequences of NanoLuc mRNA (m1Ψ-containing) with percentages of +1 frameshifted stop codons; sequence 1 (0%, blue), sequence 2 (25%, purple), sequence 5 (100%, green). Wells contained 3 µg mRNA in 4 µL dH₂O and 30 µM acridine orange, 50 mM phosphate buffer (pH 6.5) and 1 mM Mg^{2+} in 26 µL dH₂O, total volume 30 µL. Measured across the temperature gradient 35 - 95 °C (+0.125 °C s⁻¹) at 0.5 s intervals. n=1 samples per sequence. Melts were conducted using the CFX Connect Real-Time PCR Detection System using channel 1 (Ex λ = 450-490 nm, Em λ = 515-530 nm) and data collected via the CFX Maestro Software v2.3 (Bio-Rad Laboratories Inc.).

Curves for each experimental sequence shown in Figure 20 (above) exhibit broadly similar trends, with each sustaining ~65,000 RFU across the majority of the temperature gradient (35 – 80 °C). Sequence 1 (blue) declined first at 80 °C, followed by sequence 2 (purple) at 85 °C and finally sequence 5 (green) at 88 °C; rates of decline were approximately 10,000 RFU per 10 °C for each.

Since only one repeat was conducted for the experiment, interpretations of the data and subsequent conclusions drawn remain limited.

4.0 Discussion

Experiments conducted sought to determine the translational implications of inclusions of stop codons in the +1 reading frame (+1 FSS) through measurements of NanoLuc expression between sequences (0, 25, 75 and 100% +1 FSS), both over 24 and 72 hours, comparisons of predicted secondary structure parameters, and finally through DSF melt assays with and without a buffered solution. Across each of these experiments, primary findings were that introduction of 25% +1 FSS increased expression by ~30% and 100% +1 FSS reduced by ~27%, prolonged expression occurred at 100% +1 FSS, the predicted minimum free energy (MFE) and ensemble free energy broadly increase with +1 FSS, the predicted secondary structures broadly condense with +1 FSS introductions, and that DSF melts produced a delayed melt in the 25% +1 FSS sequence relative to 0%, however since only one repeat was conducted for the DSF experiment, no substantive conclusions can be drawn.

It was expected that introduction of +1 FSS might reduce translational efficacy, because the codons compatible with +1UAA and +1UAG stop codons are A-ending and less preferred than their G-ending counterparts (Nakamura, 2000). Translational efficiency on mRNAs can be limited by translation initiation or translation elongation, and for an elongation-limited mRNA use of non-preferred codons should limit protein production (Chu *et al.*, 2014). If the mRNA is initiation limited, codon changes should not affect protein synthesis. In the 24-hour expression assay we found that introduction of low levels of +1 FSS (25%) increased protein synthesis by ~30% (Fig. 11, above), which cannot be explained purely through mRNA primary structure. Conversely the introduction of high levels of +1 FSS (100%) reduced expression by ~27% (Fig. 11, above), aligning with the expected outcome of non-preferred codons.

It is known that secondary structure also modifies translation (Pelletier and Sonenberg, 1987), so it is possible that modification of secondary structure by +1 FSS overrides the effect of codon usage, at least at the 25% +1 FSS level. The observed increase in expression may also be due to accidental removal of binding sites for downregulating microRNAs, RNA binding proteins or other regulatory factors (Hentze *et al.*, 2018) in the specific 25% +1 FSS sequence, since these were not accounted for during the mRNA design. However, microRNAs typically bind to the 3'-UTR (Shang *et al.*, 2023) which is homologous across our experimental sequences, and in some cases can upregulate expression (Vasudevan, 2011).

Secondary structure is indicated to be a factor in our 72-hour expression assay seeking to quantify biological stability. It is generally considered that the introduction of non-preferred codons, as occurs with increased +1 FSS, reduces mRNA stability as the higher GC content of preferred codons is correlated with increased folding energies (Metkar *et al.*, 2023). In our experiment, qualitative analysis of NanoLuc expression at 24-hour intervals, between timepoints 24 h and 72 h post-transfection, show prolonged expression in sequence 5 (100% +1 FSS) relative to both sequence 1 and 2 (0 and 25% +1 FSS respectively) (Fig. 12, above) albeit at a lower initial rate of expression.

These observations clearly contradict the expected outcome but could be due to the fundamental differences between thermodynamic and biological stability (Ross, 1995). RNA binding proteins can also serve to protect mRNA from degradation (Ross, 1995), sequence 5 (100% +1 FSS) has a higher AU content, and thus more potential binding sites for AU-rich binding-element proteins (Dolicka *et al.*, 2020). The reduced rate of expression implies extended ribosome engagement in sequence 5 (100% +1 FSS) which has been correlated with improved biological stability (Radhakrishnan and Green, 2016).

We had similar expectations of negative correlation between the introduction of +1 FSS and RNAfold predicted secondary structure parameters, being the minimum free energy (MFE), centroid free energy, ensemble free energy and ensemble diversity. However, yielded values for each sequence show MFE and ensemble free energy increasing (indicating reduced thermodynamic stability) and ensemble diversity lowering with +1 FSS (Table 5, above). Interestingly, a reduction in centroid free energy was produced for sequences 2 and 3 (25 and 50% +1 FSS respectively) relative to the 0% +1 FSS sequence, potentially giving rise to observed differences in expression.

As a preface to the interpretation of RNAfold graphical structure results, no energy parameter adjustments we made to account for our experimental mRNA sequences used in expression assays containing N¹-methyl-pseudouridine (m1Ψ), but since this option is not available within RNAfold and m1Ψ binds more similarly to uridine than pseudouridine (Ψ) (Kierzek *et al.*, 2014), the standard energy parameters for uridine were used. The graphical structures produced by RNAfold (Fig. 13-17, above), whilst qualitative, showed a general trend with increasing +1 FSS toward a more focused around central junctions, with shorter and more frequent hairpin loops associated with biological stability (Ross, 1995). A large, circular domain is also observed in the centroid structure of sequence 1 (0% +1 FSS) (Fig. 13, above), presenting a potentially vulnerable feature.

Whilst the predicted graphical structures are not truly representative of the mRNA used in NanoLuc expression assays, the trend of structural features does appear to align with observed biological stability in the 72-hour expression assays. Predicted reductions in ensemble diversity with increased

+1 FSS contradict the general observation that deviation from wild-type sequences typically increases ensemble diversity (Moss, 2018), but since NanoLuc is an engineered reporter gene the commercial sequence did not evolve naturally (Hall *et al.*, 2012). Reduced ensemble diversity supports the centroid structure of sequence 5 (100% +1 FSS) (Fig. 17, above) as being more stable despite the contradiction of free energies.

Our final experiment line sought to differentiate the extent of secondary structure within each of our mRNA samples through differential scanning fluorimetry (DSF), involving melting the mRNA with an intercalating dye, in our experiment we used acridine orange (AO). We expected that the concentration of the dye would be determinant in successful resolution of differences between samples, but had no prior indication of suitable concentrations at the time. Our initial test as shown in Figure 18 (above) identified 30 mM AO as suitable when paired with 3 µg mRNA (m1Ψ-containing) due to a steadily decreasing fluorescence curve with a defined peak in decline at ~65 °C. Just months following completion of experiments, it was published that AO was deemed unfit for differential scanning fluorimetry on the basis of insufficient derivative values during melting (Blackford *et al.*, 2025). The study used a model 10-mer dsRNA heteroduplex (but did not specify the exact sequence); their findings show a comparable curve to that of 30 µM in Figure 18 (above), with a slight peak in fluorescence decline earlier at 50 °C relative to ours at ~62 °C, the difference in nt-length may account for this discrepancy.

In our melt using the 30 mM AO to compare experimental sequences, we expected to observe differences based on those observed by Silvers *et al.*, 2015, with more stable structures displaying a melting event at a higher temperature. Each sequence displayed a unique melting profile (Fig. 19, above) although no more than one distinct melting event was observed within the curve for a given sequence. Sequence 2 (25%) melted at a higher temperature relative to sequences 1 and 5 (0 and 100% +1 FSS respectively), aligning with observations across both NanoLuc expression assays and predicted structure analysis. Despite the assessment of AO suitability Blackford *et al.*, 2025, the differences observed between sequences appears to constitute valid data for contrasting the thermodynamic stability of each sequence, aided by findings in derivative data from Figure 20 indicating conserved stabilised structures between sequences. However, since only one repeat of the experiment was conducted, and mRNAs were suspended in pure water, they are not representative of expected secondary structures *in vivo* (Chheda, 2024),.

The final DSF assay attempted to mimic cytoplasmic conditions in relation to pH and ion content through pH stabilisation with phosphate buffer (pH 6.5) and the introduction of 10 mM Mg^{2+} ions. We expected to observe more distinct melting events through stabilisation of the mRNA (Chheda *et*

al., 2024). Observations found in Figure 21 (above) show that no reductions in fluorescence occurred until ~85 °C, with a brief melting period beyond with no discernible characteristics between sequences. Findings by Blackford *et al.*, 2025 corroborate that stabilisation of the RNA duplex weakens resolution of melting events, although no direct comparison can be made with our experiment due to methodical differences in RNA sequence and solution composition.

To briefly cover observations made during the mRNA preparation phase, mRNA concentrations did not align with those of template DNA extracts. Although differences were accounted for in preparation of IVT and reactions carried out under identical conditions. The result implicates confounding variables present at the stage of DNA preparation e.g. contamination, in storage i.e. yield loss during freeze-thawing, during IVT e.g. incomplete mixture of reagents, or in purification e.g. disturbance of a centrifugated pellet. These differences could be remedied via lithium chloride precipitation at a small loss to overall yield.

The implications of these findings on the design of therapeutic mRNAs will differ depending on the specific amino acid sequence encoded and intended use of the mRNA. In cancer applications, the prolonged duration of translation in sequences fully optimised for +1 FSS may be acceptable given the high doses required, coinciding with the desirable minimisation of aberrant proteins. Where high expression is desired at the cost of possible aberrant protein synthesis, inclusions of +1 FSS at lower levels (~25% of possible) may achieve this, particularly if the original sequence is initiation-limited and already optimised for preferred codons through the codon adaptive index (CAI). Potential alterations to secondary structure should also be considered, and whilst graphical representations of predicted structures appear to align with observed biological stability, free energies have not shown to be indicative of expression duration with our non-modified energy parameters.

A number of limitations were observed during conductions of this experiment line; firstly the testing of one amino acid sequence in one human tissue type greatly restricts the conclusions that can be drawn, particularly the amino acid type being a reporter gene, not a typical therapeutic product-encoding sequence, and the applicability of translation in HEK-293 cells to cells in a patient for instance. This is compounded by the fact that only 3 of the 5 experimental sequences were tested in luminescence assays.

Next, the experimental design and subsequent statistical approaches greatly limited the ability to draw statistically significant results beyond the 24-hour expression assay. It would have aided conclusions to confirm that the full +1 FSS optimised sequence displayed prolonged expression, instead observations were made that provide an indication for future experimentation. Of course, these experiments can be redesigned and conducted with improved parameters to be able to draw

those conclusions. Notably, having more periodic measurements within the 72-hour expression assay, in addition to measurements beyond this timeframe would improve estimates of expression kinetics and by proxy, mRNA stability.

In the DSF assays, again the number of repeats limited the ability to draw conclusions, but the negative derivative results did display correlations above expectations. If future work were to be undertaken with DSF, the work in trying to mimic cytosolic conditions and produce melting curves consistent with expected mRNA structures, would yield more pertinent data to draw conclusions from.

5.0 Conclusions

To conclude, we sought to determine the translational implications of inclusions of stop codons in the +1 reading frame (+1 FSS) to influence decisions in the optimal design of N1-methylpseudouridine(m1Ψ)-containing therapeutic mRNAs. The need for +1 FSS arises as use of m1Ψ has been implicated in frameshifting events at known (and random) sites within the open reading frame, generating aberrant proteins. Optimising for +1 FSS involves the inclusion of non-preferred codons, which was expected to reduce translational efficiency and reduce biological stability. Our findings, through experiments on differentially +1 FSS optimised sequences of Nanoluc Luciferase found that despite deviations from preferred codons, low levels of +1 FSS inclusion (25%) led to increased 24-hour expression in HEK 293 and similar biological stability in terms of expression duration to the unmodified sequence, whilst high levels of +1 FSS inclusion produced prolonged expression across 72-hours. Parameters of secondary structures were predicted via RNAfold, producing free energies broadly misaligned to observed expression durations, but qualitative analysis of the predicted minimum free energy and centroid graphical structures aligned with expression kinetics in the +1 FSS optimised sequence. Finally, differential scanning fluorimetry was used to potentially link thermodynamic stability to observed biological stabilities; whilst melts conducted in pure water broadly aligned with the observed expression kinetics between sequences, an attempt better replicate conditions *in vivo* (and thus realistic secondary structures) could not derive any meaningful differences. From these outcomes, it is suggested that off-frame stop codon optimisation does not directly oppose optimisations for preferred codons or minimum free energy, but can serve as a balance between them whilst reducing the production of aberrant proteins.

References

- Ananth, P., Goldsmith, G. and Yathindra, N. (2026) 'An innate twist between Crick's wobble and Watson-Crick base pairs', *RNA*, 19, pp. 1038-1053.
- Blackford, N., Ogunkunle, V. and Silvers, R. (2025) 'Fluorescent Dyes for Differential Scanning Fluorimetry of RNA' *ChemBioChem*, 26 (20), e202500560.
- Chen, C., Ezzeddine, N. and Shyu, A. (2008) 'Messenger RNA Half-Life Measurements in Mammalian Cells', *Methods in Enzymology*, 448, pp. 325-257.
- Chheda, U., Pradeepan, S., Esposito, E., Strezak, S., Fernandez-Delgado, O. and Kranz, J. (2024) 'Factors Affecting Stability of RNA – Temperature, Length, Concentration, pH, and Buffering Species', *Journal of Pharmaceutical Sciences*, 113 (2), pp. 377-385.
- Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M. and von der Haar, T. (2014) 'Translation elongation can control translation initiation on eukaryotic mRNAs', *EMBO J*, 7;33 (1), pp. 21-34.
- Creech, C., Walker, S. and Samuels, R. (2021) 'SARS-CoV-2 Vaccines', *JAMA*, 325 (12), pp. 1318-1320.
- Darzynkiewicz, Z., Juan, G. and Srouf, E. (2004) 'Differential Staining of DNA and RNA', *Current Protocols in Cytometry*, 30, 7.3.1 – 7.3.16.
- Demo, G., Gamper, H., Loveland, A., Masuda, I., Carbone, C., Svidritskiy, E., Hou, Y. and Korostelev, A. (2021) 'Structural basis for +1 ribosomal frameshifting during EF-G-catalysed translocation', *Nature Communication*, 12, article no. 4644.

Dolicka, D., Sobolewski, C., de Sousa, M., Gjorgjieva, M. and Foti, M. (2020) 'mRNA Post-Transcriptional Regulation by AU-Rich Element-Binding Protins in Liver Inflammation and Cancer', *International Journal of Molecular Sciences*, 11;21 (18), p. 6628.

Ding, Y., Chan, C. and Lawrence, C. (2005) 'RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble', *RNA*, 11, pp. 1157-1166.

Dittmar, K., Goodenbour, J. and Pan, T. (2006) 'Tissue-Specific Differences in Human Transfer RNA Expression', *PLoS Genetics*, 2 (12), e221.

Georgakopoulos-Soares, I., Parada, G. and Hemberg, M. (2022) 'Secondary structures in RNA synthesis, splicing and translation', *Computational and Structural Biotechnology Journal*, 20, pp. 2871-2884.

Hall, M., Unch, J., Binkowski, B., Valley, M., Butler, B., Wood, M., Otto, P., Zimmerman, K., Vidugiris, G., Machleidt, T., Robers, M., Benink, H., Eggers, C., Slater, M., Meisenheimer, P., Klaubert, D., Fan, F., Encell, L. and Wood, K. (2012) 'Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate', *ACS Chemical Biology*, 7 (11), pp. 1848-1857.

Hanahan, D. (1983) 'Studies on transformation of *Escherichia coli* with plasmids', *Journal of Molecular Biology*, 166 (4), pp. 557-580.

Heddiyoglu, F., Hargreaves, E., Kurusamy, S., Thaventhiran, J., Willis, A., Smales, M. and von der Haar, T. (2025) 'Near-cognate tRNAs dominate codon decoding times in simulated ribosomes', *Nucleic Acids Research*, 53 (20), gkaf1164.

Hellen, C. (2018) 'Translation Termination and Ribosome Recycling in Eukaryotes', *Cold Spring Harbor Perspectives in Biology*, 10 (10), a032656.

Hentze, M., Castello, A., Schwarzl, T. and Preiss, T. (2018) 'A brave new world of RNA-binding proteins', *Nature Reviews Molecular Cell Biology*, 19 (5), pp.327-341.

Hofacker, I. (2003) 'Vienna RNA secondary structure server', *Nucleic Acids Research*, 31 (13), pp. 3429-3431.

Hofacker, I. (2007) 'RNA Consensus Structure Prediction With RNAalifold', *Comparative Genomics*, 395, pp. 527-543.

Hong, D. and Jeong, S. (2023) '3'-UTR Diversity: Expanding Repertoire of RNA Alterations in Human mRNAs', *Molecules and Cells*, 46 (1), pp. 48-56.

Jackson, L., Anderson, E., Roupael, N., Roberts, P., Mikhene, M., Coler, R., McCullough, M., Chappell, J., Denison, M., Stevens, L., Pruijssers, A., McDermott, A., Flach, B., Doria-Rose, N., Corbett, K., Morabito, K., O'Dell, S., Schmidt, S., Swanson II, P., Padilla, M., Mascola, J., Neuzil, K., Bennett, H., Sun, W., Peters, E., Makowski, M., Albert, J., Cross, K., Buchanan, W., Pikaart-Tautges, R., Ledgerwood, J., Graham, B. and mRNA-1273 Study Group (2020) 'An mRNA Vaccine against SARS-CoV-2 – Preliminary Report', *New England Journal of Medicine*, 12 (20), pp. 1920-1931.

Jackson, N., Kester, K., Casimiro, D., Gurunathan, S., DeRosa, F. (2020) 'The promise of mRNA vaccines: a biotech and industrial perspective', *npj Vaccines*, 5 (11).

Jeong, D., McCoy, M., Artiles, K., Ilbay, O., Fire, A., Nadeau, K., Park, H., Betts, B., Boyd, S., Hoh, R. and Shouda, M. (2021) 'Assemblies-of-Putative-SARS-CoV2-Spike-Encoding-mRNA-Sequences-for-Vaccines-BNT-162b2-and-mRNA-1273'. Available online at: <https://virological.org/t/assemblies-of-putative-sars-cov2-spike-encoding-mrna-sequences-for-vaccines-bnt-162b2-and-mrna-1273/663>. Accessed 10 January 2026.

- Karikó, K., Muramatsu, H., Welsh, F., Ludwig, J., Kato, H., Akira, S. and Weissman, D. (2008) 'Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability', *Molecular Therapy*, 16 (11), pp. 1833-1840.
- Kelly, J., Olson, A., Neupane, K., Munshi, S., Emeterio, E., Pollack, L., Woodside, M. and Dinman, J. (2020) 'Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SERS coronavirus 2 (SARS-CoV-2)', *Accelerated Communications*, 295 (31), pp. 10741-10748.
- Kierzek, E., Malgowska, M., Lisowiec, J., Turner, D., Gdaniec, Z. and Kierzek, R. (2014) 'The contribution of pseudouridine to stabilities and structure of RNAs', *Nucleic Acids Research*, 42 (5), pp. 3492-3501.
- Kim, S., Sekhwon, S., Shin, W., Ahn, G., Cho, B., Ahn, J. and Kim, Y. (2022) 'Modifications of mRNA vaccine structural elements for improving mRNA stability and translation efficiency', *Molecular & Cellular Toxicology*, 18, pp. 1-8.
- Leppek, K., Das, R. and Barna, M. (2018) 'Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them', *Translation and Protein Quality Control*, 19, pp. 158-174
- Liu, C., Shi, Q., Huang, X., Koo, S., Kong, N. and Tao, W. (2023) 'mRNA-based cancer therapeutics', *Nature Reviews Cancer*, 23, pp. 526-543.
- Mandell, Z., Ujita, A., Henderson, J., Truong, A., Vo, C., Rezvani, F., Abolhassani, N., Lebedev, A., Xu, C., Koukhareva, I., Ramos, S., Broderick, K. and Hudson, B. (2025) 'CleanCap M6 inhibits decapping of exogenously delivered IVT mRNA', *Molecular Therapy Nucleic Acids*, 36 (1), p. 102456.
- Martin, J. (2014) 'Describing the Structural Diversity within an RNA's Ensemble', *Entropy*, 16 (3), pp. 1331-1348.

Mauger, D., Cabral, J., Presnyak, V., Su, S., Reid, D., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M. and McFadyen, I. (2019) 'mRNA structure regulates protein expression through changes in functional half-life', *Biophysics and Computational Biology*, 116 (48), pp. 24075-24083.

Metkar, M., Pepin, C. and Moore, M. (2023) 'Tailor made: the art of therapeutic mRNA design', *Nature Reviews Drug Discovery*, 23, pp. 67-83.

Monroe, J., Eyler, D., Mitchell, L., Deb, I., Bojanowski, A., Srinivas, P., Dunham, C., Roy, B., Frank, A. and Koutmou, K. (2024) 'N1-Methylpseudouridine and pseudouridine modifications modulate mRNA decoding during translation', *Nature Communications*, 15, p. 8119.

Moss, W. (2018) 'The ensemble diversity of non-coding RNA structure is lower than random sequence' *Non-coding RNA Research*, 3 (3), pp. 100-107.

Mulrone, T., Pöyry, T., Yam-Puc, J., Rust, M., Harvey, R., Kalmar, L., Horner, E., Booth, L., Ferreira, A., Stoneley, M., Sawarkar, R., Mentzer, A., Lilley, K., Smales, M., von der Haar, T., Turtle, L., Dunachie, S., Klenerman, P., Thaventhiran, J. and Willis, A. (2023) 'N1-methylpseudouridylation of mRNA causes +1 ribosomal frameshifting', *Nature*, 625 (7993), pp. 189-194.

Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) 'Codon usage tabulated from international DNA sequence databases: status for the year 2000', *Nucleic Acids Research*, 28 (1), p. 292.

Nozinovic, S., Fürtig, B., Jonker, H., Richter, C. and Schwalbe, H. (2009) 'High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA', *Nucleic Acids Research*, 38 (2), pp. 683-694.

Parvathy, S., Udayasuriyan, V. and Bhadana, V. (2021) 'Codon usage bias', *Molecular Biology Reports*, 49, pp. 539-565.

Palazzo, A. and Lee, E. (2018) 'Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs', *Frontiers in Genetics*, 9, p. 440.

Passmore, L. and Collier, J. (2021) 'Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression', *Nature Reviews Molecular Cell Biology*, 23, pp. 93-106.

Pelletier, J. and Sonenberg, N. (1987) 'The involvement of mRNA secondary structure in protein synthesis', *Biochem Cell Biology*, 65 (6), pp. 576-581.

Pinheiro, F., Santos, J. and Ventura, S. (2021) 'AlphaFold and the amyloid landscape', *Journal of Molecular Biology*, 433 (20), e167059.

Polte, C., Wedemeyer, D., Oliver, K., Wagner, J., Bijvelds, M., Mahoney, J., de Jonge, H., Sorscher, E. and Ignatova, Z. (2019) 'Assessing cell-specific effects of genetic variations using tRNA microarrays', *BMC Genomics*, 20 (8), p. 549.

Radhakrishnan, A. and Green, R. (2016) 'Connections Underlying Translation and mRNA Stability', *Journal of Molecular Biology*, 428 (18), pp. 3558-2564.

Ramanathan, A., Robb, G. and Chan, S. (2016) 'mRNA capping: biological functions and applications', *Nucleic Acids Research*, 44 (16), pp. 7511-7526.

Reiser, A., Woschée, D., Mehrotra, N., Krzysztoń, R., Strey, H. and Rädler, J. (2019) 'Correlation of mRNA delivery timing and protein expression in lipid-based transfection', *Integrative Biology*, 11 (9), pp. 362-371.

Ross, J. (1995) 'mRNA Stability in Mammalian Cells', *Microbiological Reviews*, 59 (3), pp. 423-450.

Rozman, B., Broennimann, K., Rajan, K., Nachshon, A., Saha, C., Arazi, T., Mohan, V., Geiger, T., Wollner, C., Richner, J., Westhof, E., Yonath, A., Bashan, A. and Stern-Ginossar, N. (2026) 'N1-Methylpseudouridine directly modulates translation dynamics', *Nature*, 651, pp. 533-541.

Sahin, U., Kariko, K. and Türeci, Ö. (2014) 'mRNA-based therapeutics – developing a new class of drugs', *Nature Reviews Drug Discovery*, 13, pp. 759-780.

Sakata T., Hiroaki H., Oda Y., Tanaka T., Ikehara M. and Uesugi S. (1990) 'Studies on the structure and stabilizing factor of the CUUCGG hairpin RNA using chemically synthesized oligonucleotides', *Nucleic Acids Research*, 18 (13), pp. 3831-3839.

Shang, L., Lee, S., Senavirathne, G. and Lai, E. (2023) 'microRNAs in action: biogenesis, function and regulation', *Nature Reviews Genetics*, 28;24 (12), pp. 816-833.

Sharp, P. and Li, W. (1987) 'The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications', *Nucleic Acids Research*, 15 (3), pp. 1281-1295.

Silvers, R., Keller, H., Schwalbe, H. and Hengesbach, M. (2015) 'Differential Scanning Fluorimetry for Monitoring RNA Stability', *ChemBioChem*, 16 (7), pp. 1013-1119.

Tats, A., Tenson, T. and Remm, M. (2008) 'Preferred and avoided codon pairs in three domains of life', *BMC Genomics*, 9, article no. 463.

Vasudevan, S. (2011) 'Posttranslational Upregulation by MicroRNAs', *WIREs RNA*, 3 (3), pp. 311-330.

Vavilis, T., Stamoula, E., Ainatzoglou, A., Sachinidis, A., Lamprinou, M., Dardalas, I. and Vizirianakis, I. (2023) 'mRNA in the Context of Protein Replacement Therapy', *Pharmaceutics*, 15 (1), p. 166.

Vicens, Q. and Kieft, J. (2022) 'Thoughts on how to think (and talk) about RNA structure', *Proceedings of the National Academy of Sciences*, 119 (17), e2112677119.

von der Haar, T., Mulroney, T., Hedayioglu, F., Kurusamy, S., Rust, M., Lilley, K., Thaventhiran, J., Willis, A. and Smales, M. (2023) 'Translation of *in vitro*-transcribed RNA therapeutics', *Frontiers in Molecular Biosciences*, 10 (11280067).

Vostrosablin, N., Lim, S., Gopal, P., Brazdilova, K., Parajuli, S., Wei, X., Gromek, A., Prihoda, D., Spale, M., Muzdalo, A., Greig, J., Yeo, C., Wardyn, J., Mejlík, P., Henry, B., Partridge, A. and Bitton, D. (2024) 'mRNAid, an open-source platform for therapeutic mRNA design and optimization strategies', *NAR Genomics and Bioinformatics*, 6 (1), iqcae028.

Wayment-Steele, H., Kim, D., Choe, C., Nicol, J., Wellington-Oguri, R., Watkins, A., Sperberg, R., Huang, P., Eterna Participants and Das, R. (2021) 'Theoretical basis for stabilizing messenger RNA through secondary structure design', *Nucleic Acids Research*, 18 (11), pp. 10604-10617.

Wu, Q. and Bazzini, A. (2023) 'Translation and mRNA Stability Control', *Annual Review of Biochemistry*, 92, pp. 227-245.

Xia, X. (2021) 'Detailed Dissection and Critical Evaluation of the Pfizer/BioNTech and Moderna mRNA Vaccines', *Vaccines (Basel)*, 9 (7), p. 734.

Xiao, B., Wang, S., Pan, Y., Zhi, W., Gu, C., Guo, T., Zhai, J., Li, C., Chen, Y. and Wang, R. (2025) 'Development, opportunities, and challenges of siRNA nucleic acid drugs', *Molecular Therapy Nucleic Acids*, 36 (1), p. 102437.

Yamamura, K., Terai, G. and Asai, K. (2025) 'A method for estimating energy parameters of RNAs by differentiating base-pairing probabilities', *NAR Genomics and Bioinformatics*, 8 (4), p. lqaf171.

Young, A. and Kanneganti, T. (2025) 'For "U" the bell tolls: Pseudouridine evades TLR detection', *Molecular Cell*, 85 (17), pp. 23206-3208.

Zaiontz, C. (2026) Real Statistics Using Excel. Accessible at: <https://real-statistics.com> [Accessed 14th January 2026].

Žak, M. and Zangi, L. (2025) 'Clinical development of therapeutic mRNA applications', *Molecular Therapy*, 33 (6), pp. 2583-2609.

Zhang, H., Zhang, L., Lin, A., Xu, C., Li, Z., Liu, K., Liu, B., Ma, X., Zhao, F., Jiang, H., Chen, C., Shen, H., Li, H., Mathews, D., Zhang, Y. and Huang, L. (2023) 'Algorithm for optimized mRNA design improves stability and immunogenicity', *Nature*, 621, pp. 396-403

The data analysis for this paper was generated using the Real Statistics Resource Pack software (Release 9.8). Copyright (2013 – 2026) Charles Zaiontz. real-statistics.com.

Appendix

1. Full Experimental Sequences

Sequence 1 (0% +1 FSS)

1 ggaataagag agaaaagaag agtaagaaga aatataacag ccaccatgaa gtgggtgacc
61 ttcatcagcc tgctgttctt gttcagcagc gcctacagcg tgttcacctt ggaggacttc
121 gtgggcgact ggagacagac cgccggctac aacctggacc aggtgctgga gcagggcggc
181 gtgagcagcc tgttcagaa cctgggcgtg agcgtgacct ccatccagag aatcgtgctg
241 agcggcgaga acggcctgaa gatcgacatc cacgtgatca tcccctacga gggcctgagc
301 ggcgaccaga tgggccagat cgagaagatc ttcaaggtgg tgtaccccggt ggacgaccac
361 cacttcaagg tgatcctgca ctacggcacc ctggtgatcg acggcgtgac cccaacatg
421 atcgactact tcggcagacc ctacgagggc atcgccgtgt tcgacggcaa gaagatcacc
481 gtgaccggca ccctgtggaa cggcaacaag atcatcgacg agagactgat caaccccgac
541 ggcagcctgc tgttcagagt gaccatcaac ggcgtgaccg gctggagact gtgcgagaga
601 atcctggcct gatgatgagc tcgctttctt gctgtccaat ttctattaaa ggttcctttg
661 ttcc

Sequence 2 (25% +1 FSS)

1 ggaataagag agaaaagaag agtaagaaga aatataacag ccaccatgaa gtgggtaacc
61 ttcatcagcc tgctgttctt gttcagcagc gcctacagcg tgttcacctt ggaggacttc
121 gtgggcgact ggagacagac cgccggctac aacctggacc aggtgctgga gcagggcggc
181 gtgagcagcc tgttcagaa cctgggcgta agcgtaaccc ccatccagag aatcgtgtta
241 agcggcgaga acggcctgaa gatcgacatc cacgtaataa tcccctacga gggcctgagc
301 ggcgaccaga tgggccagat cgagaagatc ttcaaggtag tgtaccccggt ggacgaccac
361 cacttcaagg tgatcctgca ctacggcacc ctggtgatcg acggcgtgac cccaacatg
421 atcgactact tcggcagacc ctacgagggc atcgccgtgt tcgacggcaa gaagatcacc
481 gtgaccggca ccctgtggaa cggcaacaag atcatcgacg agagactgat caaccccgac
541 ggcagcctgc tgttcagagt gaccatcaac ggcgtgaccg gctggagact gtgcgagaga

601 atcctggcct gatgatgagc tcgctttctt gctgtccaat ttctattaaa gggtcctttg
661 ttcc

Sequence 3 (50% +1 FSS)

1 ggaataagag agaaaagaag agtaagaaga aatataacag ccaccatgaa gtgggtgacc
61 ttcatcagcc tgctgttctt gttcagcagc gcctacagcg tgttcacctt ggaggacttc
121 gtgggcgact ggagacagac cgccggctac aacctggacc aggtgttaga gcagggcggc
181 gtgagcagcc tgttcagaa cttaggcgta agcgtaacct ccatccagag aatcgtgtta
241 agcggcgaga acggcctgaa gatcgacatc cacgtaatca tcccctacga gggcttaagc
301 ggcgaccaga tgggccagat agagaagatc ttcaaggtag tgtacccctg ggacgaccac
361 cacttcaagg taatcctgca ctacggcacc ctggtgatcg acggcgtgac cccaacatg
421 atcgactact tcggcagacc ctacgagggc atcgccgtgt tcgacggcaa gaagatcacc
481 gtgaccggca ccctgtggaa cggcaacaag ataatcgacg agagactgat caacccgac
541 ggcagcctgc tgttcagagt aaccatcaac ggcgtaaccg gctggagact gtgcgagaga
601 atcttagcct gatgatgagc tcgctttctt gctgtccaat ttctattaaa gggtcctttg
661 ttcc

Sequence 4 (75% +1 FSS)

1 ggaataagag agaaaagaag agtaagaaga aatataacag ccaccatgaa gtgggtaacc
61 ttcatcagcc tgctgttctt gttcagcagc gcctacagcg tgttcacctt agaggacttc
121 gtaggcgact ggagacagac cgccggctac aacctggacc aggtgttaga gcagggcggc
181 gtaagcagcc tgttcagaa cctgggcgta agcgtgacct ccatccagag aatagtgtta
241 agcggcgaga acggcttaaa gatcgacatc cacgtaataa tcccctacga gggcttaagc
301 ggcgaccaga tgggccagat agagaagatc ttcaaggtag tgtacccctg ggacgaccac
361 cacttcaagg taatcctgca ctacggcacc ctggtgatcg acggcgtaac cccaacatg
421 atcgactact tcggcagacc ctacgagggc atagccgtgt tcgacggcaa gaagataacc
481 gtaaccggca ccctgtggaa cggcaacaag ataatcgacg agagactgat caacccgac
541 ggcagcctgc tgttcagagt gaccatcaac ggcgtaaccg gctggagact gtgcgagaga
601 atcttagcct gatgatgagc tcgctttctt gctgtccaat ttctattaaa gggtcctttg
661 ttcc

Sequence 5 (100% +1 FSS)

1 ggaataagag agaaaagaag agtaagaaga aatataacag ccaccatgaa gtgggtaacc
 61 ttcataagcc tgctgttctt gttcagcagc gcctacagcg tttcacctt agaggacttc
 121 gtaggcgact ggagacagac cgccggctac aacctggacc aggtgttaga gcagggcggc
 181 gtaagcagcc tgttcagaa cttaggcgta agcgtaaccc ccatccagag aatagtgtta
 241 agcggcgaga acggcttaaa gatcgacatc cacgtaataa tcccctacga gggcttaagc
 301 ggcgaccaga tgggccagat agagaagatc ttcaaggtag tgtaccccgt ggacgaccac
 361 cacttaaggg taatcctgca ctacggcacc ttagtaatcg acggcgtaac cccaacatg
 421 atcgactact tcggcagacc ctacgagggc atagccgtgt tcgacggcaa gaagataacc
 481 gtaaccggca ccctgtggaa cggcaacaag ataatcgacg agagattaat caaccccgac
 541 ggcagcctgc tgttcagagt aaccatcaac ggcgtaaccg gctggagact gtgcgagaga
 601 atcttagcct gatgatgagc tcgctttctt gctgtccaat ttctattaaa ggttccttg
 661 ttcc

2. 24-hour NanoLuc luminescence (normalised to mean of sequence 1 values in assay 2)

			Sequence (% +1 FSS)		
Assay	Replicate	-ve	1 (0)	2 (25)	5 (100)
1	1	652	1716703	2005843	1414042
	2	1378	1586951	2264227	1410502
	3	726	1536107	1658233	1152243
2	1	594	1641827	2000089	1183569
	2	586	1483356	2000073	1426157
	3	538	1714590	1949365	1204343
3	1	911	1617529	1870960	1138541
	2	1226	1825460	3114020	1035436
	3	888	1396724	2049320	724497
	Average	833	1613250	1949365	1187703
	Standard Error	99	43659	137280	74484
	CI95%	65	28523	89688	48662

^a normalised to the mean of sequence 1 values in assay 2 (bold)

^b values omitted from analysis and graphical representation in red

Pre-normalisation data

			Sequence (% +1 FSS)		
Assay	Replicate	-ve	1 (0)	2 (25)	5 (100)
1	1	203	534384	624389	440170
	2	429	493994	704820	439068
	3	226	478167	516183	358676
2	1	594	1641827	2000089	1183569
	2	586	1483356	2000073	1426157
	3	538	1714590	1949365	1204343
3	1	501	888752	1028000	625572
	2	674	1003000	1711000	568921
	3	488	767431	1126000	398075
Mean	1		502258		
	2		1613257		
	3		886394		

^a values omitted from analysis and graphical representation in red

3. 72-hour NanoLuc luminescence

			Sequence		
Timepoint (h)	Replicate	-ve	1 (0)	2 (25)	5 (100)
24	1	640	562026	589065	421468
	2	697	396241	638139	422125
	3	732	634212	575472	639853
48	1	877	969208	1214000	831337
	2	1039	659478	618673	589490
	3	1075	1051000	989235	702869
72	1	638	491896	693654	608249
	2	814	683715	602423	764991
	3	751	563104	760436	978207
24	Average	690	530826	600892	494482
48	"	997	825895	940636	707899
72	"	734	579572	685504.3	783816
24	Standard Error	27	70445	19032	72685
48		61	116770	173566	69860
72		51	55982	45796	107212

4. Moderna mRNA-1273 vaccine full sequence

1 gggaaataag agagaaaaga agagtaagaa gaaatataag acccggcgc cgccaccatg
 61 ttctgttcc tgggtctgct gccctgggtg agcagccagt gcgtgaacct gaccaccggg
 121 acccagctgc caccagccta caccaacagc ttcaccggg gcgtctacta ccccgacaag
 181 gtgttcgga gcagcgtcct gcacagcacc caggacctgt tctgcccctt cttcagcaac
 241 gtgacctggt tccacgccat ccacgtgagc ggcaccaacg gcaccaagcg gttcgacaac
 301 cccgtgctgc ctttaacga cggcgtgtac ttcgccagca ccgagaagag caacatcatc
 361 cggggctgga tcttcggcac caccctggac agcaagacc agagcctgct gatcgtgaat
 421 aacgccacca acgtggtgat caaggtgtgc gagtccagt tctgcaacga ccccttctg
 481 ggcgtgtact accacaagaa caacaagagc tggatggaga gcgagttccg ggtgtacagc
 541 agcgccaaca actgcacctt cgagtacgtg agccagccct tctgatgga cctggagggc
 601 aagcagggca acttaagaa cctgcgggag ttcgtgttca agaacatcga cggctacttc
 661 aagatctaca gcaagcacac ccaatcaac ctggtgcggg atctgcccc gggcttctca
 721 gccctggagc ccctggtgga cctgcccac ggcatcaaca tcaccgggtt ccagaccctg
 781 ctggccctgc accggagcta cctgaccca ggcgacagca gcagcgggtg gacagcaggc
 841 gcggctgctt actacgtggg ctacctgcag ccccggaact tctgctgaa gtacaacgag
 901 aacggacca tcaccgacg cgtggactgc gccctggacc ctctgagcga gaccaagtgc
 961 accctgaaga gttcaccgt ggagaagggc atctaccaga ccagcaactt ccgggtgcag
 1021 cccaccgaga gcatcgtgc gttcccaac atcaccaacc tgtgccctt cggcgagggtg
 1081 ttaacgcca cccggttcgc cagcgtgtac gcctggaacc ggaagcggat cagcaactgc
 1141 gtggccgact acagcgtgct gtacaacagc gccagcttca gcacattca gtgctacggc
 1201 gtgagccca ccaagtgaa cgacctgtgc ttaccaacg tgtacgccga cagcttcgtg
 1261 atccgtggcg acgaggtgag gcagatcgca cccggccaga caggcaagat cgccgactac
 1321 aactacaagc tgcccgacga cttaccggc tgcgtgatcg cctggaacag caacaacctc
 1381 gacagcaagg tggcgaggaa ctacaactac ctgtaccggc tgttcggaa gagcaacctg
 1441 aagccctcg agcgggacat cagcaccgag atctaccaag ccggctccac cccttgaac
 1501 ggcgtggagg gttcaactg ctacttcct ctgcagagct acggcttcca gccaccaac
 1561 ggcgtgggct accagcccta ccgggtgggt gtgctgagct tcgagctgct gcagcccca
 1621 gccaccgtgt gtggcccaa gaagagcacc aacctgtga agaacaagtg cgtgaacttc

1681 aacttcaacg gccttaccgg caccggcgtg ctgaccgaga gcaacaagaa attcctgccc
1741 tttcagcagt tcggccggga catcgccgac accaccgacg ctgtgcggga tccccagacc
1801 ctggagatcc tggacatcac cccttgacgc ttcggcggcg tgagcgtgat cccccaggc
1861 accaaccacca gcaaccaggt ggccgtgctg taccaggacg tgaactgcac cgaggtgccc
1921 gtggccatcc acgccgacca gctgacaccc acctggcggg tctacagcac cggcagcaac
1981 gtgttcaga cccgggcccgg ttgcctgacg ggcgccgagc acgtgaacaa cagctacgag
2041 tgcgacatcc ccatcggcgc cggcatctgt gccagctacc agaccagac caattacccc
2101 cggagggcaa ggagcgtggc cagccagagc atcatgcct acacatgag cctgggcggc
2161 gagaacacg tgccctacag caacaacagc atgccatcc ccaccaactt caccatcagc
2221 gtgaccaccg agattctgcc cgtgagcatg accaagacca gcgtggactg caccatgtac
2281 atctcgggcg acagcaccga gtgcagcaac ctgctgctgc agtacggcag cttctgcacc
2341 cagctgaacc gggccctgac cggcatcgcc gtggagcagg acaagaacac ccaggaggtg
2401 ttcgccagg tgaagcagat ctacaagacc cctcccatca aggacttcgg cggcttcaac
2461 ttacgccaga tcctgcccga cccagcaag cccagcaagc ggagcttcat cgaggacctg
2521 ctgttcaaca aggtgaccct agccgacgcc ggcttcatca agcagtacgg cgactgcctc
2581 ggcgacatag ccgcccggga cctgatctgc gccagaagt tcaacggcct gaccgtgctg
2641 cctcccctgc tgaccgacga gatgatcgcc cagtacacca gcgccctgtt agccggaacc
2701 atcaccagcg gctggacttt cggcgctgga gccgctctgc agatcccctt cgccatgcag
2761 atggcctacc ggttcaacgg catcggcgtg acccagaacg tgctgtacga gaaccagaag
2821 ctgatcgcca accagttcaa cagcgccatc ggcaagatcc aggacagcct gagcagcacc
2881 gctagcgcgc tgggcaagct gcaggacgtg gtgaaccaga acgcccaggc cctgaacacc
2941 ctggtgaagc agctgagcag caacttcggc gccatcagca gcgtgctgaa cgacatcctg
3001 agccggctgg accctcccga ggccgaggtg cagatcgacc ggctgatcac tggccggctg
3061 cagagcctgc agacctagt gaccagcag ctgatccggg ccgccgagat tcgggccagc
3121 gccaacctgg ccgccaccaa gatgagcgag tgcgtgctgg gccagagcaa gcgggtggac
3181 ttctcgggca agggctacca cctgatgagc ttcccaga gcgcaccca cggagtgtg
3241 ttctgcacg tgacctagt gccgcccag gagaagaact tcaccaccgc cccagccatc
3301 tgccacgacg gcaaggccca cttcccgg gagggcgtgt tcgtgagcaa cggcaccac
3361 tggttcgtga cccagcgga cttctacgag cccagatca tcaccaccga caacacctc
3421 gtgagcggca actgcgacgt ggtgatcggc atcgtgaaca acaccgtgta cgatcccctg
3481 cagcccgagc tggacagctt caaggaggag ctggacaagt acttaagaa tcacaccagc

3541 cccgacgtgg acctgggcca catcagcggc atcaacgcca gcgtggtgaa catccagaag
3601 gagatcgatc ggctgaacga ggtggccaag aacctgaacg agagcctgat cgacctgcag
3661 gagctgggca agtacgagca gtacatcaag tggccctggt acatctggct gggcttcac
3721 gccggcctga tcgccatcgt gatggtgacc atcatgctgt gctgcatgac cagctgctgc
3781 agctgcctga agggctgttg cagctgcggc agctgctgca agttcgacga ggacgacagc
3841 gagcccgtgc tgaagggcgt gaagctgcac tacacctgat aataggctgg agcctcggtg
3901 gcctagcttc ttgcccttg ggctcccc cagcccctcc tccccttct gcacccgtac
3961 cccgtggtc ttgaataaa gtctgagtgg gcggcaaaaa aaaa

(Jeong et al., 2021)

5. Full Cloning Vector Sequence (excluding experimental sequence 1 (0% +1 FSS), highlighted)

```

1  tgcgtttcta caaactctct ctggctaact agagaacca ctgcttactg gcttatcgaa
61  atgctagcgt ttctgcacac ttaagcttta atacgactca ctataaggaa taagagagaa
121 aagaagagta agaagaaata taacagccac cggataaga gagaaaagaa gagtaagaag
181 aaatataaca gccaccatga agtgggtgac cttcatcagc ctgctgttcc tggtcagcag
241 cgcctacagc gtgttcaacc tggaggactt cgtgggcgac tggagacaga ccgccggcta
301 caacctggac caggtgctgg agcagggcgg cgtgagcagc ctgttcaga acctgggcgt
361 gagcgtgacc cccatccaga gaatcgtgct gagcggcgag aacggcctga agatcgacat
421 ccacgtgac atcccctacg agggcctgag cggcgaccag atgggccaga tcgagaagat
481 cttcaagtg gtgtaccccg tggacgacca ccacttaag gtgatcctgc actacggcac
541 cctggtgac gagggcgtga ccccaacat gatcgactac ttcggcagac cctacgaggg
601 catcgccgtg ttcgacggca agaagatcac cgtgaccggc accctgtgga acggcaacaa
661 gatcatcgac gagagactga tcaacccga cggcagcctg ctgttcagag tgaccatcaa
721 cggcgtgacc ggctggagac tgtgcgagag aatcctggcc tgatgatgag ctgcgtttct
781 tgctgtccaa ttctattaa aggttcttt gttcctgagc tcgctttctt gctgtccaat
841 ttctattaaa ggttcctttg ttcctaagt ccaactacta aactggggga tattatgaag
901 ggcttgagc atctggattc tgctaataa aaaacattta ttctattgc actagtaaaa
961 aaaaaaaaaa aaaaaaaaaa aaggaagagc ctagctacct actcgagtct agagggcccg
1021 tttaaaccg ctgatcagcc tcgactctag cttggattct caccaataaa aaacgccggg
1081 cggcaaccga gcgttctgaa caaatccaga tggagtctg aggtcattac tggatctatc
1141 aacaggagtc caagctcaga tcattgtcac gtctgacgct cagtggaacg aaaactcacg
1201 ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc tttaaatta
1261 aaaatgaagt tttaaatcaa gcccaatctg aataatgtta caaccaatta accaattctg
1321 attagaaaaa ctcacgagc atcaaatgaa actgcaattt attcatatca ggattatcaa
1381 taccatattt tgaaaaagc cgtttctgta atgaaggaga aaactaccg aggcagttcc
1441 ataggatggc aagatcctgg tatcggctg cgattccgac tcgtccaaca tcaatacaac
1501 ctattaattt ccctcgtca aaaataaggt tatcaagtga gaaatcacca tgagtgcga
1561 ctgaatccgg tgagaatggc aaaagttat gcatttctt ccagacttgt tcaacaggcc

```

1621 agccattacg ctgctcatca aaatcactcg catcaaccaa accgttattc attcgtgatt
1681 ggcctgagc gaggcgaaat acgcatcgc tgtaaaagg acaattacaa acaggaatcg
1741 aatgcaaccg gcgcaggaac actgccagcg catcaacaat attttcacct gaatcaggat
1801 attcttctaa tacctggaat gctgttttc cggggatcgc agtggtgagt aaccatgcat
1861 catcaggagt acggataaaa tgcttgatgg tcggaagagg cataaattcc gtcagccagt
1921 ttagtctgac catctcatct gtaacatcat tggcaacgct acctttgcca tgtttcagaa
1981 acaactctgg cgcacgggc ttccataca agcgatagat tgtcgacacct gattgccga
2041 cattatcgcg agccattta taccatata aatcagcatc catgttgaa ttaatcgcg
2101 gcctgcagct ttccgttga atatggctca taacaccct tgtattactg tttatgtaa
2161 cagacagttt tattgttcat gatgatata tttatcttg tgcaatgaa catcagagat
2221 ttgagacac gggccagagc tgcatcgcgc gtttcggtga tgacggtgaa aacctctgac
2281 acatgcagct cccggacgct cccgtagaaa agatcaaagg atcttcttga gatcctttt
2341 ttctgcgct aatctgctgc tgcaacaa aaaaaccacc gctaccagcg gtggtttgtt
2401 tgccggatca agagctacca actcttttc cgaaggtaac tggcttcagc agagcgcaga
2461 taccaatac tgttcttcta gtgtagcgt agttaggcca ccactcaag aactctgtag
2521 caccgcctac atacctgct ctgctaattc tgttaccagt ggctgctgcc agtggcgata
2581 agtctgtct taccgggttg gactcaagac gatagttacc ggataaggcg cagcggtcgg
2641 gctgaacggg gggttcgtgc acacagcca gcttgagcg aacgacctac accgaactga
2701 gatactaca gcgtgagcta tgagaaagcg ccacgcttc cgaagggaga aaggcggaca
2761 ggtatccggt aagcggcagg gtcggaacag gagagcgac gagggagctt ccagggggaa
2821 acgcctgga tctttatagt cctgtcgggt ttcgccacct ctgactgag cgtcgattt
2881 tgtgatgctc gtcagggggg cggagcctat ggaaaaacgc cagcaacgcg gcctttttac
2941 ggttcctggc ctttctggt cctttgctc acatattgtc gatctccca tgcgagagta
3001 gggaactgcc aggcatacaa taaaacgaaa ggctcagtcg aaagactggg ctttctgtt
3061 tatctgtgt ttgtcgtga acgctctct gagtaggaca aatccgccg gagcggattt
3121 gaacgttgcg aagcaacggc ccggagggtg gcgggcagga cgccgccat aaactgccag
3181 gcatcaaatt aagcagaagg ccacctgac ggatggcctt tt