

Force spectroscopy of the frameshift signal from West Nile virus reveals multiple folding pathways and structural heterogeneity

by

Matthew Thomas J Halma

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Physics
University of Alberta

© Matthew Thomas J Halma, 2019

Front Matter

Abstract

Programmed ribosomal frameshifting (PRF) represents an important mechanism for translational genetic recoding, especially in viruses. The components of a PRF stimulator have been well characterized, though accounting for the variation in the frameshift stimulating efficiency has thus far been elusive. Frameshift efficiencies at known PRF sites vary from a few percent to 70-80%, and several studies have been undertaken to determine what distinguishes a high efficiency PRF site from a low efficiency PRF site via structural characterization of the stimulatory structure. Observations suggest that conformational plasticity, the ability of a certain sequence to adopt multiple conformations, is correlated with frameshift efficiency. We examine a very high efficiency (70%) PRF stimulatory structure responsible for the NS1' frameshift in West Nile virus (WNV) to determine its characteristics. We find a high degree of structural plasticity and heterogeneity; the PRF signal exhibits multiple different starting states and unfolds via two main pathways. Furthermore, we characterize the structures involved in these pathways, and find that they correspond to predicted structures using bioinformatic predictions and SHAPE analysis. Moreover, we suggest a new operational metric of conformational plasticity, one that obviates two existing problems with the previous method for defining conformational plasticity, namely the requirement to specify a native state, and the insensitivity to multiple conformations. Additionally, we extend this definition to be force dependent, and find that the value of this conformational plasticity metric in the force range of ribosomal stalling correlates highly with frameshifting efficiency. These results may elucidate the process of frameshifting by illustrating the relationship between conformational plasticity within a specific force range and frameshift efficiency. In addition, the characterization of a high efficiency frameshift signal allows for a better understanding of the structural dynamics underlying frameshifting.

Preface

This thesis includes work performed by Matthew TJ Halma and Dustin B Ritchie on the structural characterization of the West Nile virus frameshift signal in Chapter 4. Synthesis of experimental samples was initially performed by D.B. Ritchie, and later by both D.B. Ritchie and M.T.J. Halma. The initial research idea belongs to Michael T Woodside and D.B. Ritchie, though M.T.J. Halma expanded on it. Measurements were done by M.T.J. Halma with assistance from D.B. Ritchie, Data interpretation was performed by M.T.J. Halma with assistance from D.B. Ritchie and M.T. Woodside.

Dedication

“What transforms this world is — knowledge. Do you see what I mean? Nothing else can change anything in this world. Knowledge alone is capable of transforming the world, while at the same time leaving it exactly as it is. When you look at the world with knowledge, you realize that things are unchangeable and at the same time are constantly being transformed.”

~ Yukio Mishima, *The Temple of the Golden Pavilion*

Acknowledgements

Many thanks to my supervisor Michael T. Woodside, only with the support of himself and his laboratory was this project possible.

I would like to thank Dustin B. Ritchie, who taught me how to prepare samples and how to use the experimental apparatus, in addition to being a friend and conversation partner.

Thank you to Krishna, Dushanth, Russel and Noel, who all generously shared their knowledge of optical trapping with me.

Thank you to Craig, Chunhua, Lindsay, Meng, Rafayel, Andrew, Shubhadeep, Supratik, Uttam, Sookpichaya, Negar and Tonia for their feedback and advice.

Many thanks to my recent friends Bryana, Kelly, Ian, Joe, Shereif, and Kiko who demonstrated to me novel modes of being. Thank you to my old friend Brendan Paiha, who has known me longer than anyone besides family. Thank you to my friend Nick Buzinsky whose friendship I greatly value.

Thank you to my many great teachers: Mr. and Mrs. Swen, Mr. Vuch, Mr. Terakita and Rocky Wilson who encouraged me during primary and secondary school. Thank you to my many professors at the University of Alberta.

Thank you to the many workers who maintain the buildings and utilities at the University of Alberta and the wider city of Edmonton. These components are prerequisites to any sort of research.

Thank you to my Mother and Father, Rita and Clarence Halma, who have supported me all throughout my education.

Thank you to my siblings: Geralis and her husband Andrew; Erin and her husband Brad; Kevin and his girlfriend Michela.

Table of Contents

<i>Front Matter</i>	<i>ii</i>
Abstract	ii
Preface	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
1. Introduction	1
2. Frameshifting	6
2.1 Introduction to translation	6
2.2 Translational recoding	9
2.2.1 Frameshifting assays	10
2.2.2 Programmed ribosomal frameshifting (PRF) function.....	11
2.2.3 The mechanism of frameshifting	14
3. Biology and evolution of the WNV frameshift signal	20
3.1 Structure of the WNV frameshift signal	20
3.2 Frameshift in WNV: Biological implications	22
3.3 Flaviviral evolution	25
3.4 Epidemiology of WNV	27
4. Methods	31
4.1 Single-molecule force spectroscopy	31
4.1.1 Sample Preparation	33
4.1.2 Slide Preparation	35
4.1.3 Optical Tweezers Measurements	35
4.2 Analysis	42
4.2.1 Properties of biopolymers	42
4.2.2 Properties of states	45
5. Novel Analysis Methods	50
5.1 State Assignment	50
5.1.1 Rip detection	50
5.1.2 Moving window contour length fits	52
5.1.3 Whole trajectory categorization	54

6. <i>Single Molecule Force Spectroscopy measurements of a high-efficiency frameshift stimulatory structure from West Nile virus</i>	56
6.1 Introduction	56
6.2 Results	56
6.2.1 The WNV PRF signal forms the short pseudoknot predicted by Atkins and Firth	65
6.2.2 A short hairpin forms at the 3' end of the short pseudoknot	66
6.2.3 A set of double hairpins forms	67
6.2.4 Evidence for the long pseudoknot (PKL) assignment.....	69
6.2.5 The West Nile Virus frameshifting pseudoknot exhibits high heterogeneity in the force range of ribosomal stalling.....	74
6.3 Discussion	76
7. <i>An improved metric for correlating structural heterogeneity of frameshift signals with PRF efficiency</i>	77
7.1 Introduction	77
7.2 Results	78
7.3 Discussion	82
8. <i>Conclusion</i>	84
8.1 Future Work	85
9. <i>Literature Cited</i>	89

List of Tables

Table 1: Observed and proposed states in wild-type data.....	59
Table 2: Table of all calculated contour lengths.	63
Table 3: Average unfolding force for specific transitions in WT data.....	67
Table 4: Comparison of structure predictions and experimental measurements.	71

List of Figures

Figure 1: Three protein sequences are possible from the same mRNA sequence.	2
Figure 2: Replication of HIV is reliant on there being a roughly 1:20 to 1:10 ratio of gag-pol to gag proteins in HIV-1.	3
Figure 3: The genetic code for translating triplet mRNA codons into protein.	6
Figure 4: Translational initiation.....	7
Figure 5: The structure of the ribosome.....	7
Figure 6: The process of elongation.....	8
Figure 7: The frameshift example from the HIV gag-pol gene.....	9
Figure 8: Luminescent activity of firefly luciferase (Fluc) and renilla luciferase (Rluc) scales linearly with number	10
Figure 9: Plasmid genome for dual-luciferase activity	11
Figure 10: Programmed ribosomal frameshifting allows for the conservation of protein N-termini, with modular C-termini.....	12
Figure 11: Models of -1 ribosomal frameshifting differ based on the translational step the PRF event occurs at	15
Figure 12: Elements of programmed ribosomal frameshifting.....	15
Figure 13: Frameshifting efficiency correlates with conformational plasticity, as defined by the fraction of molecules alternatively folded	17
Figure 14: Energetic predictions suggest ribosomal frameshifting occurs where the energy of the -1 frame becomes less than the 0 frame energy.....	17
Figure 15: The ribosome moves in discrete sized movements, given by the length of a codon.	20
Figure 16: West Nile virus exhibits high structural plasticity and has a high frameshift efficiency.	21
Figure 17: Depiction of the WNV polyprotein in the 0 frame (top) and -1 frame (bottom).....	23

Figure 18: Phylogenetic reconstruction of family flaviviridae	26
Figure 19: Cases of WNV from 1999 to 2017	27
Figure 20: Deaths from West Nile Virus in the USA from 1999 to 2017.....	28
Figure 21: Cases of WNV in Canada from 2002 to 2017.....	29
Figure 22: Atomic force microscopy of single biomolecules.	32
Figure 23: Force/torque spectroscopy by magnetic tweezers.	32
Figure 24: Single-molecule force spectroscopy of nucleic acid construct.....	32
Figure 25: Oxygen scavenging reactions.	34
Figure 26: Flow cell schematic.	35
Figure 27: Schematic of position sensitive device (PSD).....	37
Figure 28: Deflection of incident light by a sound wave.....	39
Figure 29: Power spectral density plot for bead trapped in x (pulling axis) and y (perpendicular to pulling axis).....	41
Figure 30: Worm-like chain model.....	43
Figure 31: Unfolding of a single molecule	45
Figure 32: DNA hairpin consisting of a 20 bp stem with 4 nt loop.....	46
Figure 33: Fit to multiple branches of the WNV frameshift signal using the worm-like chain (WLC) model proposed by Marko and Siggia	47
Figure 34: The Murine Leukemia Virus (MLV) read-through pseudoknot.....	48
Figure 35: A classic experiment showcasing the higher unfolding forces of tertiary structure as opposed to secondary structure	49
Figure 36: Rip detection schema.....	51
Figure 37: Decision boundary for rip detection.....	52
Figure 38: Trace of contour length vs extension for a collection of 20 FECs.	53

Figure 39: Trace of contour length vs force for a collection of 20 FECs.	53
Figure 40: A maximum likelihood trajectory assignment of a force extension curve with three branches.	55
Figure 41: Single molecule force spectroscopy of the West Nile Virus frameshift signal	57
<i>Figure 42: Representative force-extension curves (FECs) of the full-length West Nile Virus (WNV) NS1' frameshift signal.....</i>	<i>59</i>
Figure 43: Transition map of WNV states during force ramp.	60
Figure 44: Disruption of structures by oligonucleotides.....	62
<i>Figure 45: Novel Structures formed in presence of oligo 1.....</i>	<i>63</i>
Figure 46: Force-extension curves in the presence of oligo 1.	64
Figure 47: Force-extension curves in the presence of oligo 2.	64
Figure 48: Oligo 1 abolishes PKS and maintains S1.	65
Figure 49: Support for the existence of a hairpin at the 3' end of PKS.....	66
Figure 50: Potential modifications of SLL structures such that RHP is formed.....	68
Figure 51: Some modifications to originally proposed structure.....	70
Figure 52: The stem loop pathway.....	72
Figure 53: The long pseudoknot pathway.....	73
Figure 54: The West Nile virus frameshift signal is highly structurally plastic, exhibiting multiple starting conformations.	74
Figure 55: State occupancy for the WNV PRF signal... ..	75
Figure 56: Force-dependent conformational plasticity for WNV PRF signal.....	79
Figure 57: Force dependent entropy for 10 frameshifting pseudoknots.	80
Figure 58: Relationship between zero force SE and FSE.	81
Figure 59: Relationship between SE averaged in the range of 11-15 pN and FSE.....	82
Figure 60: Flavonoid compounds exhibiting anti-Japanese encephalitic virus effects.....	86

Figure 61: Binding site of JEV shown with daidzin bound. Source: ⁷⁷ (CC BY).....	87
Figure 62: Frameshift stimulatory sequence in the Japanese Encephalitic virus (JEV) serogroup, which includes West Nile virus (WNV).....	88

List of Symbols

5' to 3': direction in which translation occurs

A: aminoacyl site of ribosome

aa: amino acid

AFM: atomic force microscopy

AOD: acoustic-optical device

CPCV: cacipacore virus

DENV: dengue virus

E: exit site of ribosome

EF-G: elongation factor G

ESI-MS: electrospray ionization mass spectrometry

EWLC: extensible worm-like chain

FRET: Förster resonance energy transfer

GTP: guanine triphosphate

hepC: hepatitis C

HIV: human immunodeficiency virus

HSV: herpes simplex virus

IRES: internal ribosome entry site

ITC: isothermal titration calorimetry

JEV: Japanese encephalitic virus

MLV: murine leukemia virus

MOPS: 3-(N-morpholino)propanesulfonic acid

mRNA: messenger RNA

MT: magnetic tweezers

MVEV: Murray Valley encephalitic virus

ncRNA: non-coding RNA

NMD: nonsense mediated decay

NMIA: N-Methylisatoic Anhydride

nts: nucleotides

NW: New world

OT: optical tweezers

OW: old world

P: peptidyl site of ribosome

PCBP: poly(C) binding protein

PKL: pseudoknot long

PKL-: pseudoknot long partial

PKS: pseudoknot short

PRF: Programmed Ribosomal Frameshifting

PRRSV: porcine reproductive and respiratory syndrome virus

PSD: position-sensitive detector

PTC: premature termination codon

RD: rip-detection

RHP: residual hairpin

RNase: ribonuclease

RSV: rous sarcoma virus

SARS-CoV: severe acute respiratory syndrome coronavirus

SD: Shine Dalgarno

SHAPE: Selective 2'-hydroxyl acylation analyzed by primer extension

SL: stem-loop

SLL: stem-loop long

SLL-: stem-loop long partial

smFRET: single-molecule Forster resonant energy transfer

SMFS: Single-molecule force spectroscopy

ssRNA: single stranded RNA

tRNA: transfer RNA

VEEV: Venezuelan Equine encephalitis virus

VMD: Visual Molecular Dynamics

WLC: worm-like chain

WNND: West Nile neuroinvasive disease

WNV: West Nile virus

WT: wild type

YAOV: Yaoundé virus

YFV: yellow fever virus

1. Introduction

Within the framework of the central dogma of molecular biology is the one gene-one enzyme hypothesis, proposed by George Beadle and Edward Tatum in a widely-cited 1941 paper, stating that a single gene codes for a single enzyme¹. In a Kuhnian fashion, this framework was first modified to the ‘one gene-one polypeptide hypothesis’, and upon mounting contrary evidence, even this modified form may be deemed an oversimplification.

There have been many challenges to this view, emerging from the greater understanding of gene regulation in biology. For example, messenger RNAs (mRNAs) transcribed from DNA may be spliced differently in order to produce different proteins². Another way to obtain multiple gene products from a single sequence is from the recoding of *translation*, e.g. via programmed ribosomal frameshifting (PRF)³⁻⁶. It is the concept of PRF that motivates the work in this thesis.

One consequence of the genetic code being composed of three nucleotide (nt) codons, whereby the ribosome translates three mRNA nucleotides (nts) to a single amino acid (aa), is that there exist multiple possible reading frames for each given sequence (Figure 1). Programmed ribosomal frameshifting (PRF) is a type of non-canonical translation, whereby the mRNA after (3' of) the frameshift is translated in an alternate reading frame (RF) than occurs during normal translation. The amino acid sequence prior to (5' of) the frameshift site is conserved between the normally translated and frameshifted protein, but the amino acid sequence of the frameshifted protein following (3' of) the frameshift may be drastically different from the protein produced from zero frame translation. This coding degeneracy in the mRNA sequence may be leveraged by the organism to allow for dual-coding of a given mRNA transcript, thereby expanding the coding capacity of a fixed genome. This is highly advantageous for viruses, which have energetic⁷⁻⁹ and spatial constraints¹⁰⁻¹³ on genome length.

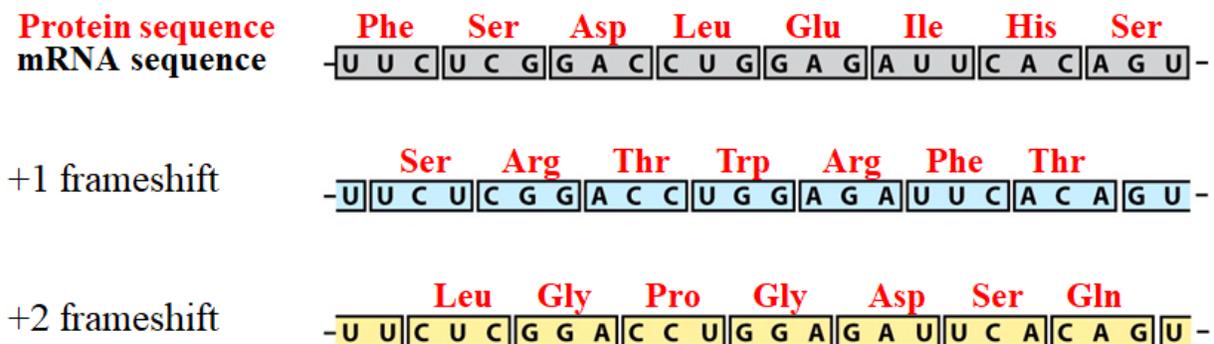


Figure 1: Three protein sequences are possible from the same mRNA sequence.

Frameshifting occurs spontaneously at low rates throughout the translated genome, at rates of $<3 \times 10^{-5}$ per codon^{14,15}. While most translational mutations are deleterious^{16,17}, some frameshift mutations may prove advantageous to the organism¹⁸. Certain frameshifts appear to be ‘programmed’, as they occur at higher rates (10^{-1} per codon¹⁹⁻²¹), appear at specific sites and appear to confer adaptive value to the organism. Such rates are high enough to have biological significance, as in the case of the human immunodeficiency virus (HIV) gag-pol polyprotein^{22,23}, a protein which plays essential roles in virion (viral particle) assembly, including packaging of genomic RNA²⁴. The PRF causes a shift in reading frame during translation, such that the protein sequence following the frameshift is different from the zero-frame product. In order for proper virion formation, a ratio of 1:10 to 1:20 gag-pol to gag must be expressed; achieved via programmed ribosomal frameshifting (Figure 2)²⁵⁻³⁰. While PRF may shift the ribosome into many frames (-2^{31-33} , -1^{34} , $+1^{35-40}$, $+2^{41}$, $+5/+6^{42}$, and $+50^{43,44}$), this thesis focuses exclusively on -1 PRF.

Of course reliance on PRF is not restricted to HIV, many viruses have dual coding regions achieved via PRF⁴⁵, and in fact PRF is used by all kingdoms of life^{18,46-53}. The fact that PRF is ubiquitous suggests some adaptive value, as it has been suggested that it has emerged independently in multiple different lineages⁵⁴⁻⁶⁰.

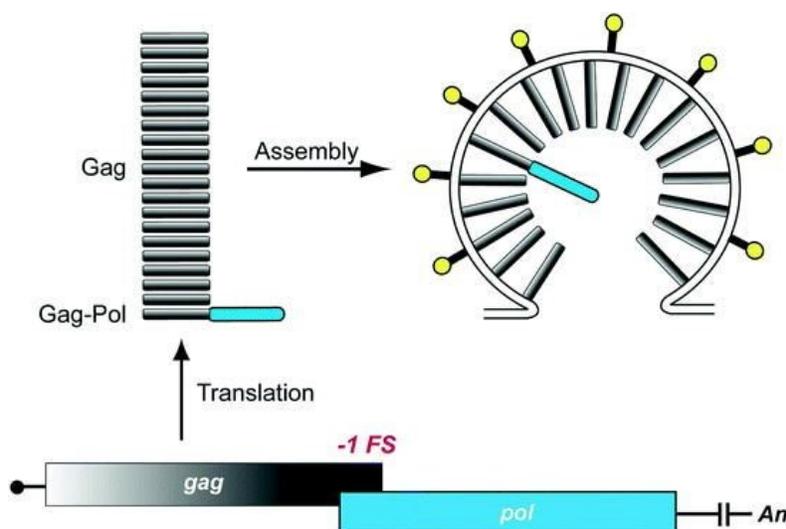


Figure 2: Replication of HIV is reliant on there being a roughly 1:20 to 1:10 ratio of gag-pol to gag proteins in HIV-1. Source: ⁶

The site of frameshifting is a heptanucleotide (7-nt long) slippery sequence of sequence N NNW WWH (spaces represent 0 frame)⁶¹, where N is any base, W is A or U, and H is A, C or U⁶². A frameshift stimulatory element is located 1-15nt downstream (3') of the slippery sequence⁶³. While stimulatory elements are usually pseudoknots (PKs)^{19,21,64,65}, they may also be stem-loops (SLs)⁶⁶⁻⁶⁸. These stimulatory elements respond to cellular signals, which alter the frequency of frameshifting, thereby altering protein expression. PRF is responsive to a wide variety of elements such as ions^{69,70}, polyamines^{58,71-76}, flavones⁷⁷, antibiotics^{78,79}, other synthetic compounds⁸⁰⁻⁸⁸, cellular proteins^{31,89-93}, RNA interactions⁹⁴⁻⁹⁸, the co-translationally folding polypeptide chain^{76,99}, the concentrations of canonical transfer RNAs (tRNAs) in the cytoplasm (cellular media)^{32,100-113}, as well as mutant tRNAs¹¹⁴⁻¹¹⁶. This ability to change the frequency of frameshifting based on a wide variety of elements suggests a possible role for programmed ribosomal frameshifting as a mechanism of gene regulation. Similar cases, whereby metabolite-sensitive riboswitches alter gene regulation have been widely studied^{4,63,89,117-119}.

This role for PRF may be an important mechanism in eukaryotes and bacteria, but the effects of frameshifting have been most widely studied in viruses⁴⁵. Viruses often translate the genome as a single

polyprotein gene, which is post-translationally cleaved into multiple distinct proteins^{120–123}. While this means that proteins are expressed at equimolar ratios normally, PRF may alter this equimolar ratio to one more beneficial for the virus. Cases of PRF has been independently observed in many distinct viruses⁴⁵, for which the disruption of PRF causes an attenuation of the virus *in vitro*^{26,28,77,97,98,124–135} and *in vivo*^{135–138}, suggesting adaptive value. Given that PRF is tunable, the frequency of frameshifting may change over the course of infection to optimize for the present stage of infection¹³⁹. Given the ubiquity and apparent usefulness of PRF for viruses, these findings suggest PRF as a potential anti-viral target, an approach explored by several reviews^{3,140,141}.

The many roles of PRF underscore the potential significance that a better understanding might bring. Different models have been proposed for what occurs nanomechanically during -1 PRF; the models differ based on what stage in translational elongation the frameshift occurs⁶. Some studies show a positive correlation between pseudoknot unfolding force and frameshift efficiency^{68,142–148}, operationalized as the fraction of ribosomes that slip into the -1 frame, measured via dual luciferase reporter¹⁴⁹. Meanwhile, other studies show no relationship between pseudoknot stability and frameshift efficiency^{21,65,145,150,151}. Several studies have proposed a link between pseudoknot conformational plasticity, or the tendency of a molecule to form alternate structures, and frameshift efficiency^{21,83,139,152}. This hypothesis motivates our study of the West Nile virus (WNV) frameshifting pseudoknot, as it represents both a highly efficient frameshifting stimulator¹³⁹, and has multiple predicted conformations^{139,153}, a potential sign of high structural plasticity.

The second section (introduction being the first) will provide an overview of frameshifting and explain several models of frameshifting and its biological implications. The third section will provide details regarding the biology and evolution of the PRF signal in WNV. The fourth section will provide an overview of force spectroscopy methods and data interpretation. The fifth section will go over novel analysis methods. The sixth section provides the experimental results for the structural characterization of WNV. The seventh section will explore the relationship between conformational plasticity and frameshift

efficiency for ten pseudoknots total. The final section will discuss the implications of this work, and discuss potential future work.

2. Frameshifting

2.1 Introduction to translation

Translation is the process by which the ribosome converts the information encoded in a messenger RNA (mRNA) sequence into a protein chain. The codon sequence determines the amino acid sequence of the protein (Figure 3). Codons are sets of three consecutive nucleotides, each coding for one out of the 22 possible amino acids¹⁵⁴.

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Figure 3: The genetic code for translating triplet mRNA codons into protein. Source: OpenStax College, Biology, accessed via Khan Academy (CC BY 3.0)

An mRNA transcript is read by the ribosome in the 5' to 3' direction, which corresponds to synthesizing the protein from its N-terminus to its C-terminus. There are differences in ribosomal binding sites (RBSs) between prokaryotes and eukaryotes¹⁵⁵⁻¹⁵⁷, owing to different ribosome structures^{158,159}. In prokaryotes, the ribosome binds at the Shine-Dalgarno (SD) sequence (5'-AGGAGG-3')¹⁶⁰, whereas in eukaryotes, the ribosome binds at either the 5' end^{161,162} or rarely an internal ribosome entry site (IRES)^{162,163}.

The ribosome will actually begin the process of translation when it encounters a start codon (AUG)¹⁶⁴, which codes for a methionine. Before translation begins, the ribosome is not yet assembled, consisting of the initiator tRNA (codon AUG) bound to the small ribosomal subunit. When the initiator tRNA encounters the start codon, it binds, and recruits the large ribosomal subunit to attach to the small ribosomal subunit so that translation may begin (Figure 4).

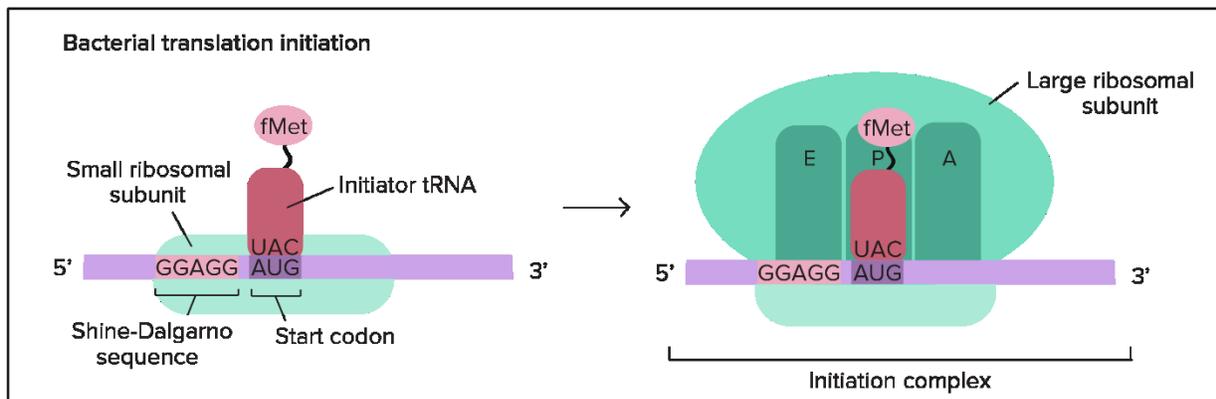


Figure 4: Translational initiation. Source: Khan Academy (CC BY 4.0)

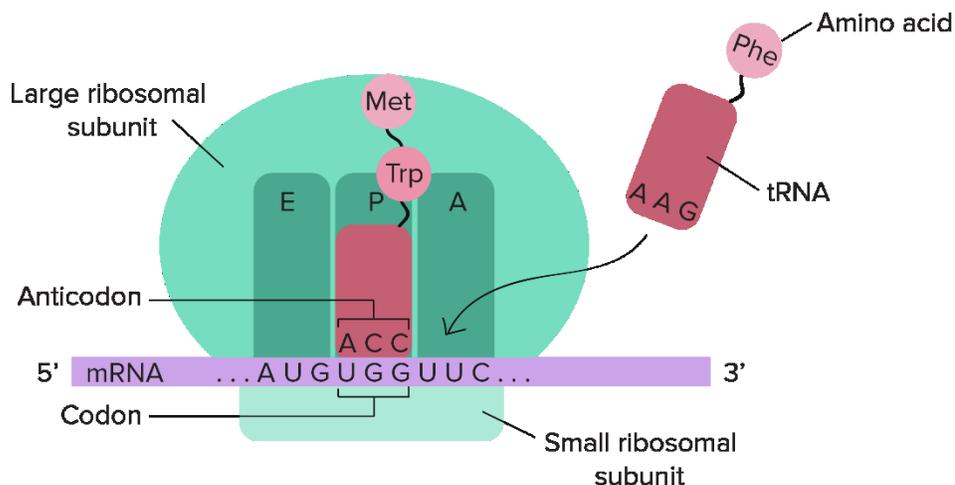


Figure 5: The structure of the ribosome. Source: Khan Academy (CC BY 4.0)

The ribosome contains three sites where tRNAs bind. In 5' to 3' order, these are the exit (E), peptidyl (P) and aminoacyl (A) sites. Immediately following initiation, the initiator tRNA is located in the

P site. With the initiator tRNA in the P site, a new tRNA enters the A site, binding to the codon (Figure 5). The process of adding amino acids to a growing polypeptide chain is known as *elongation* and PRF occurs during this translational stage (Figure 6). After the A site tRNA is docked, the first methionine binds to the amino acid associated with the A site tRNA. When this occurs, the polypeptide chain is now attached to the A site tRNA, and the P-site tRNA lacks an amino acid group. Following this, the ribosome moves forward by one codon (roughly 1.7nm¹⁶⁵), the tRNA previously in the P-site moves into the E-site, where it releases from the mRNA. Likewise, the A-site tRNA moves into the P-site, and the system is now ready for another round of elongation which will proceed until a stop codon (UAA,UAG, or UGA) is encountered.

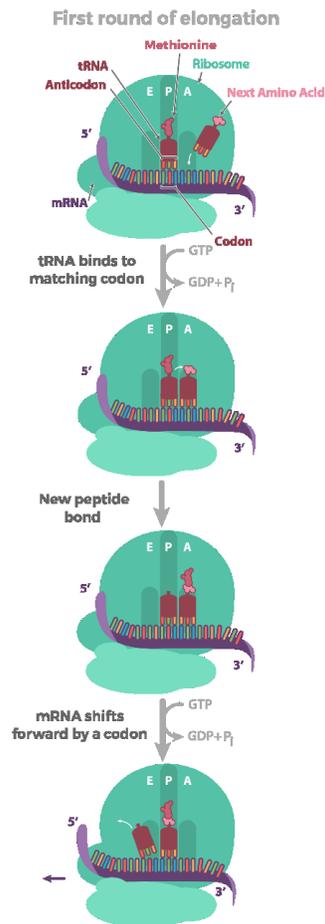


Figure 6: The process of elongation. Source: Khan Academy (CC BY 4.0)

Stop codons are bound by release factors, as opposed to tRNAs. When the release factor binds to the A site, it catalyzes the release of the polypeptide chain from the P-site tRNA, as well as the release of the ribosome from the mRNA transcript. Oftentimes, PRF will redirect the ribosome to a different stop codon, altering the length of the protein.

2.2 Translational recoding

One consequence of the genetic code is that there exist three possible reading frames for any given mRNA sequence, typically producing vastly different protein sequences. Translational frameshifts occur when the ribosome undergoes a shift in reading frame (Example shown in Figure 7). This is a distinct phenomenon from transcriptional frameshifts, where the actual mRNA transcript is affected. A translational frameshift is an alternate way of reading a given RNA sequence, rather than a change in the sequence itself.

***gag* reading frame**

--- Leu — Gly — Leu — Arg — Leu — Thr — Asn — Leu Stop
 5'---C U A G G G C U C C G C U U G A C A A A U U U A U A G G G A G G G C C A --- 3'

***pol* reading frame**

--- C U A G G G C U C C G C U U G A C A A A U U U A U A G G G A G G G C C A ---
 Ile — Gly — Arg — Ala ---

Figure 7: The frameshift example from the HIV gag-pol gene. The creation of a certain polyprotein requires a ratio of approximately 10:1 gag(0 frame) to gag-pol (-1 frame) ratio, which is reflected by the experimental frameshift percentage of 5-10%¹⁶⁶. Source: ¹⁶⁷

Frameshift errors compose a subset of the total errors in translation, accounting for an error rate of about 10^{-5} per codon¹⁶⁸⁻¹⁷³. While most mutations are deleterious¹⁷⁴, some of these frameshift mutations may be leveraged towards allowing for the expression of multiple proteins from a single genetic sequence. Occasionally, serendipitous frameshifts may produce a beneficial product from a pre-existing coding sequence^{175,176}. PRF is ubiquitous, it is estimated that the human genome consists of at least 1% dual-coding regions^{63,177}. Additionally, estimations suggest that 10% of genes in all eukaryotic genomes contain a -1 programmed ribosomal frameshifting (PRF) signal¹⁷⁸.

2.2.1 Frameshifting assays

Frameshift percentage is measured by comparing the amounts of the proteins in each frame. Since the protein product in the alternate frame will differ from those of the normal reading frame, the frameshift efficiency is calculated as the ratio of frameshift proteins to total proteins, usually via a dual-luciferase reporter^{149,179}. Dual-luciferase measurements use luminescent measurements in order to measure the gene expression of the proteins which catalyze the luminescent reaction. There is a linear relationship between luminescent activity and the concentration of these catalysts, firefly luciferase and renilla luciferase¹⁸⁰ (Figure 8).

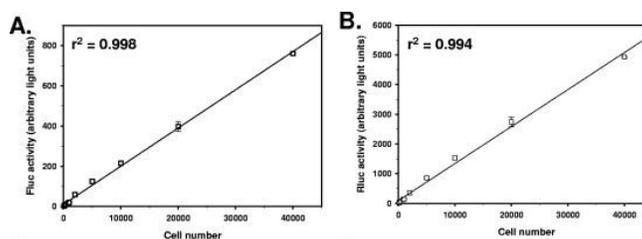


Figure 8: Luminescent activity of firefly luciferase (*Fluc*) and renilla luciferase (*Rluc*) scales linearly with number. Source: ¹⁸⁰

With an easily quantifiable measure of protein copy number known, it is possible to calculate the frameshift percentage by placing a frameshift sequence in between the two luciferase genes such that the second gene, is only expressed in the alternate reading frame (Figure 9). The frameshift percentage is calculated then as ratio of the number of ribosomes that respond to the recoding signal to the total number of ribosomes that reach the recoding signal¹⁴⁹. Translational frameshift errors occur spontaneously at a rate $<3 \times 10^{-5}$ per codon^{14,15 168–173}. Frameshifting occurs at high frequency at many sites that have been characterized^{181,182}, where rates can exceed 10^{-1} per codon^{20,21}.

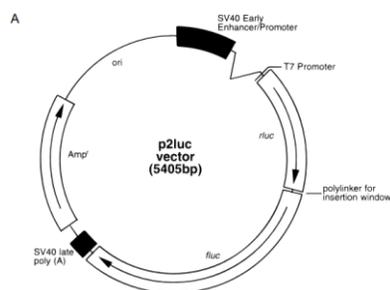


Figure 9: Plasmid genome for dual-luciferase activity. Ribosomes bind at the T7 promoter and translate until the stop at SV40. The frameshift sequence is inserted into the insertion window. The insertion window is engineered such that Fluc is only expressed in the alternate reading frame. Source: ¹⁴⁹ (CC BY-NC 4.0)

2.2.2 Programmed ribosomal frameshifting (PRF) function

Initially, frameshift sites were found via the discovery of frameshifted gene products¹⁸³, being first discovered in viruses⁵, where phage proteins^{183–185} were shown to be expressed out of frame. These products were inessential for the propagation of the virus^{183,186,187}, but frameshift sites essential for viral propagation were discovered shortly afterwards^{188,189}. Bioinformatic methods, which search for sequence elements common to frameshift signals^{63,182,190–193}, have expanded this repertoire. Additionally, another method, ribosomal profiling¹⁹⁴, has been applied to find frameshift sites^{99,195}. These methods have found PRF in all kingdoms of life^{18,196,197}, performing important and varied functions^{198–200}. One factor contributing to the prevalence of PRF in viruses two main constraints on genome length; a limited capsid volume^{10–13,140}, and energetic constraints^{7–9}. The energetic cost of reproduction scales with genome length⁷, and a lower energetic cost per virion (viral particle) is beneficial as it allows the virus to produce more virions within a given energy budget. When PRF signals are mutated such that the frameshift is ablated; viruses are greatly attenuated. This applies to Venezuelan Equine Encephalitis Virus (VEEV)¹³⁶, West-Nile Virus (WNV)^{137,138,201}, the severe acute respiratory syndrome coronavirus (SARS-CoV)^{201,202}, the human immunodeficiency virus (HIV)^{80,124,125,203}, Japanese Encephalitic virus (JEV)^{135,138}, herpes simplex (HSV)²⁰⁴, hepatitis C (hepC)²⁰⁵ and others.

One implication of frameshifting is that the N-terminus of the protein is conserved between the 0-frame and -1-frame products. C-modular proteins may be produced via PRF, whereby a protein with a useful N-terminal domain may be modified into several different proteins by changing the C-terminus. A recently described frameshift, producing a copper transporter and chaperone from the same gene (Figure 10), is an excellent example of this phenomenon, as the N-terminal copper binding domain was conserved between the two products⁹⁹. The -1 frame product (chaperone) terminated almost immediately after the copper binding domain, whereas the 0 frame product produced a much longer protein (transporter) (Figure 10).

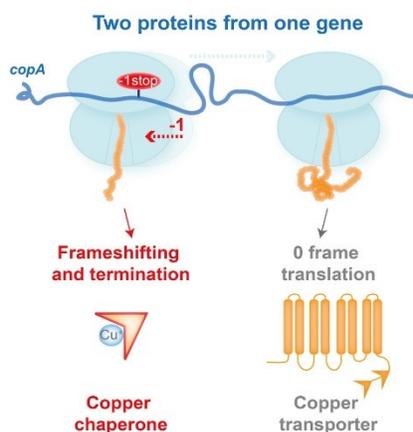


Figure 10: Programmed ribosomal frameshifting allows for the conservation of protein N-termini, with modular C-termini. This example is of a copper transporter and chaperone from the same gene, where the N-terminal copper binding domain is conserved. The proteins differ from their C-termini; where one terminates soon after the N-terminal copper binding domain (-1 frame; chaperone) the other adds a C-terminus which results in a functional copper transporter. Source:⁹⁹

Besides the creation of novel proteins, another advantage of PRF is that it allows for gene regulation in cases where the frameshift efficiency varies depending on cellular conditions, suggesting a riboswitch-like role^{117,118} in gene regulation^{4,63,89,119}. Control over gene expression is a mechanism by which cells may respond to their environment by changing the expression of certain genes to best suit the current environment. In many RNA viruses, a single polyprotein gene is expressed, and then later cleaved

into individual proteins with equimolar ratios^{120,123}. PRF allows the relative concentrations to vary to a level more optimal for viral replication^{137,139}. Furthermore, the optimal protein ratio may change with different stages of the infection cycle, and a PRF signal could potentially respond to cellular signals by altering its FS efficiency³.

PRF signals also respond to protein interactions^{31,92}, bound oligonucleotides^{206,207}, antibiotics⁷⁸ and even to interactions with the translational polypeptide chain^{76,99}. These mechanisms allow PRF to regulate gene expression, and may be an underappreciated mechanism of gene regulation²⁰⁸. For the case of gene regulation by oligonucleotides, it has been found that antisense oligonucleotides may competently promote frameshifting with efficiencies up to 40%²⁰⁶, presenting a potential mechanism of gene regulation by non-coding RNA (ncRNA)²⁰⁹, in addition to presently known mechanisms^{210,211}.

PRF may also be a mechanism regulating the level of mRNAs in a cell via the nonsense-mediated decay (NMD) pathway^{3,18,141,178,199,200,212-215}. This pathway detects early translational stops by the ribosome, termed premature termination codons, and decays the associated mRNA, limiting the production of truncated proteins, which may prove deleterious to the organism²¹⁶⁻²¹⁸. Out-of-frame products tend to be shorter (99% of predicted -1 PRF products are 30 codons or shorter²¹⁹), as stop codons are highly represented in these alternate opening reading frames (ORFs)²²⁰⁻²²⁴. Additionally, over 95% of computationally predicted -1 PRF events direct the ribosome to premature termination codons^{3,199}. As short products trigger the NMD pathway, PRF may be a mechanism in down-regulation of certain mRNAs and their corresponding gene products^{200,219}.

Frameshifting efficiency may also be regulated by the presence of ligands²²⁵. In SARS-CoV, the presence of a novel ligand inhibits frameshifting and is subsequently a possible anti-viral compound^{81,83}. Another example exists where transactivation of frameshifting (activation by a host protein) occurs in the encephalomyocarditis virus, drastically increasing *in vitro* PRF efficiencies from 0% to 70% via binding to the stimulatory structure⁸⁹. In porcine reproductive and respiratory syndrome virus (PRRSV), the nsp2 gene has two alternate products, corresponding to the presence of a -1 PRF (nsp2N)³¹, and a -2 PRF

(nsp2TF)³³, again trans-activated by a host protein⁹⁰. Additionally, it was discovered that a class of antibiotics, the macrolides, induce ribosomal frameshifting, which in turn activate macrolide resistance genes⁷⁹, an intriguing mechanism of antibiotic resistance. One implication of PRF efficiency being tunable by cellular conditions is that compounds taken exogenously (i.e. drugs) may influence PRF and subsequent gene regulation. A similar example exists in riboswitches, where gene expression is regulated by metabolites^{82,226–246}.

The potential for compounds to alter PRF has been studied in the context of anti-viral agents. Since viruses use PRF to maintain optimal stoichiometric ratios between proteins³, drugs that alter PRF can attenuate the virus⁴⁵. Such an approach has been studied in the human immunodeficiency virus (HIV)^{26,27,80,166}, the rous sarcoma virus (RSV)^{247–249}, West Nile virus (WNV)¹³⁷, Japanese encephalitic virus (JEV)^{138,153}, severe acute respiratory coronavirus (SARS CoV)^{81,83,250}, *Saccharomyces cerevisiae* L-A totivirus^{251,252} and many others (reviewed in⁴⁵). Many such viruses are attenuated when their PRF site is targeted by mutagenesis or binding ligands⁴⁵. Binding ligand-based attenuation has been studied within SARS-CoV^{81,83} and JEV⁷⁷, presenting a possible pharmaceutical means of targeting viruses.

2.2.3 The mechanism of frameshifting

Minus one frameshifting has been described as a process whereby a mechanical arrest of the ribosome by a stimulatory structure produces a backwards (5') motion of the ribosome by one single nucleotide, causing a change in reading frame. Several models have been proposed to explain why this occurs and how the components of a frameshift signal (slippery sequence, spacer and stimulatory structure) work together with the ribosome to induce this process. These models differ on which stage of translational elongation the frameshift occurs at (*Figure 11*). The most notable models are the integrated²⁵³ and 9Å models²⁵⁴, where frameshifting occurs during accommodation; the simultaneous slippage model²⁵⁵, where the frameshift occurs between accommodation and peptidyl transfer; the dynamic model^{256,257}, where frameshifting occurs either during the formation of hybrid state

intermediates, or during translocation; the mechanical model^{258,259}, occurring during translocation; and the three tRNA model²⁶⁰, occurring during the next round of elongation.

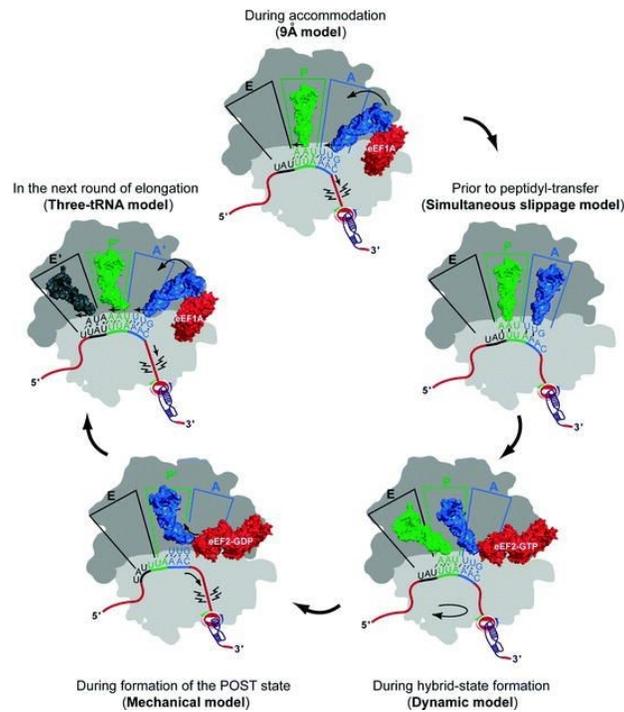


Figure 11: Models of -1 ribosomal frameshifting differ based on the translational step the PRF event occurs at. Source:⁶

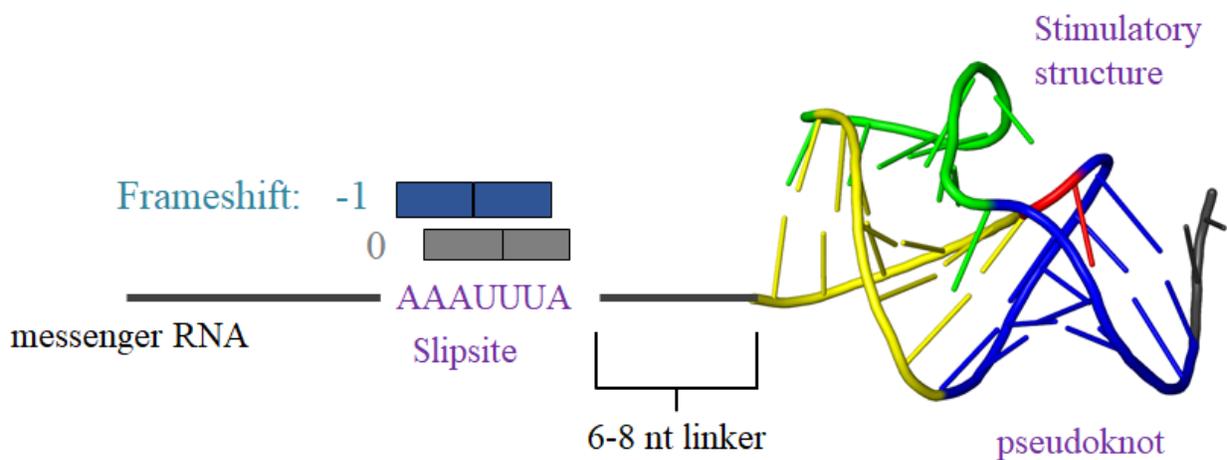


Figure 12: Elements of programmed ribosomal frameshifting. PRF requires a heptameric (7nt) slippery sequence of the form NNN WWW H, where N is any base, W is A or U, and H is A, C or U. In addition, there must also be a stimulatory structure, which tends to be 6-8 nt downstream from the slippery site.

The putative role of the stimulatory structure in stimulating frameshifting is to arrest the motions of the ribosome. Source:¹⁶⁷

Despite their differences, the models do share fundamental characteristics, namely that the downstream structure causes the ribosome to pause during elongation with tRNAs bound to the slippery site. The reason that the slippery site is constrained to be in the form *N NNW WWW* (spaces reflect 0-frame) is because this sequence pattern allows for re-pairing of the tRNA to the -1 frame. Some slippery sequences which stimulate frameshifting do not conform to this motif²⁶¹. For the -1 frame tRNAs, the only base pair mismatches occur at the first tRNA base, so the -1 frame looks like *NNN WWW* *H*, where red text shows a mismatch. Often, however, due to N being the same as W in many functional slippery sequences (4 out of 22 possible)³, and these being over-represented¹⁸¹, there is no mismatch at the first red *N*. Additionally, many frameshifts contain instances whereby *H* and *W* are equivalent, thereby removing the second mismatch. Mutational studies have revealed that heptameric U repeats (i.e. 7 Us in a row. U=N=W=H) can stimulate low levels of frameshifting even without a stimulatory structure⁶¹.

The stimulatory structure provides a barrier to the elongation of the ribosome, pausing it over the slippery sequence. This region is located at 3.3-4.3nm from the slippery site, the sum of the spacer length (3.3nm²⁰) and the distance to the barrier of the pseudoknot (~1-2nm^{21,145,262}). This is an additional reason why the possible spacer lengths are usually within a range of 5-8nt²⁶³. Such a mechanism has been suggested in studies predicting frameshifting based on the free energy difference between 0 and -1 states reversing in this region²⁰ (*Figure 14*). Due to steric constraints, only single stranded RNA (ssRNA) may enter the ribosomal entry tunnel²⁶⁴⁻²⁶⁷. As such, any secondary or tertiary structure ‘roadblocks’ must be unwound if the ribosome is to continue elongation^{67,268-274}. This unwinding occurs via two ribosomal mechanisms^{275,276}, and the ribosomal dynamics during unwinding has been investigated in many studies^{152,165,277-294}.

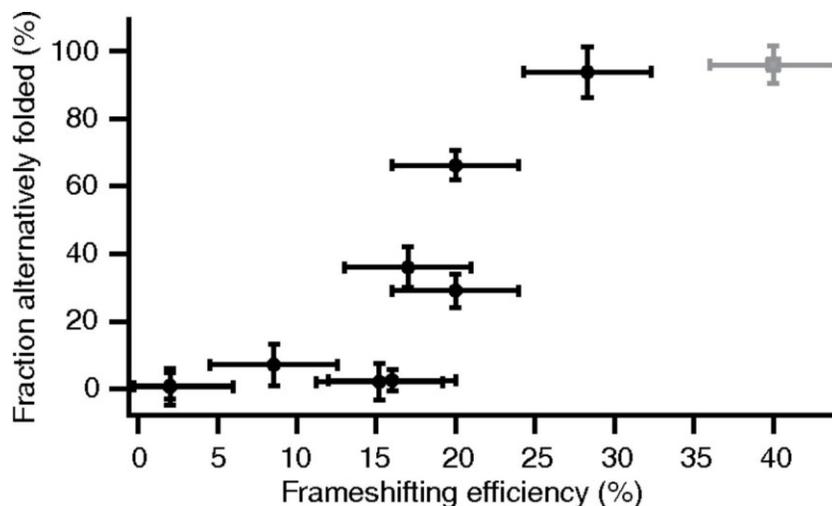


Figure 13: Frameshifting efficiency correlates with conformational plasticity, as defined by the fraction of molecules alternatively folded. Source:²¹

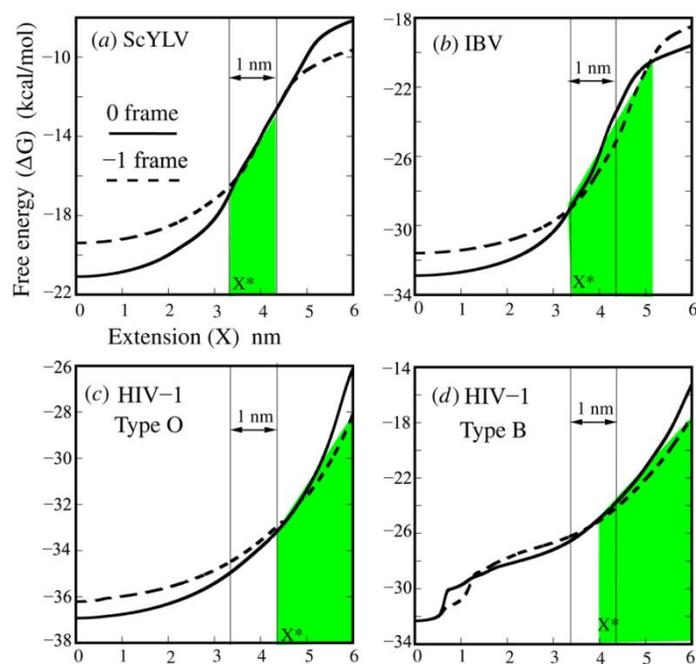


Figure 14: Energetic predictions suggest ribosomal frameshifting occurs where the energy of the -1 frame becomes less than the 0 frame energy. This occurs in a window between 3.3 and 4.3nm. Source:²⁰

2.2.3 Ribosomal Mechanics

The ribosome gains the energy required for motion via the hydrolysis of guanine triphosphate (GTP) by elongation factor G (EF-G)²⁹⁵. A ribosome encountering a blockage must first unwind the messenger RNA before it may proceed. The ribosome uses two distinct and active mechanisms for unfolding messenger RNA (mRNA) during translation²⁷⁵. The first mechanism reflects a common mechanism in molecular motors, known as the ‘thermal ratchet’, which leverages thermal fluctuations towards motion in a single direction by preventing backwards motion^{296–298}. In the case of unfolding mRNA, the ribosome stabilizes the open state, thereby biasing the mRNA towards unfolded²⁷⁵. In the second mechanism, the ribosome translocates by actively applying force to the closed state, mechanically unfolding it. An analysis of prokaryotic ribosomal translocation speeds found an exponential dependence of speed on opposing force, with a zero-force rate of 2.9 codons/s (3 nm/s) and moves in discrete steps with characteristic distance of 1.4 nm²⁹⁹ or 1.7nm¹⁶⁵, depending on the study; these distances are comparable to the distance of 1.48 nm between A- and P- site codons³⁰⁰. The eukaryotic ribosome also moves in discrete steps of three codons, with an average elongation rate of 5.6 codons/s^{301,302}.

The prokaryotic ribosome has been observed to stall at 13 ± 2 pN[‡]¹⁶⁵. Encountering mechanically stable structures causes the ribosome to pause, as the structure obstructs the ribosomal entry tunnel¹²⁶⁸.

‡ Another study has inferred the eukaryotic ribosome stalling force to be 26.5 ± 1 pN³⁰³, though this study is more unreliable than the study producing the value of 13 ± 2 pN for the prokaryotic ribosome. The study of the eukaryotic value monitors translation of mRNA bound with antisense oligonucleotides of varying binding stabilities. From this, the stalling force is inferred, from the exponential dependence of a rate on ΔG . This value is less trustworthy, coming from a less direct observation of ribosomal stalling than the 13 ± 2 pN value, which uses an optical tweezers setup to directly measure the change in ribosome velocity with increasing force.

Studies of the RNA polymerase motor may be applicable to the stall force of the ribosome, though they would be expected to be higher. A 1998 study by Wang et al. produced an RNA polymerase (RNAP) stalling force of 25 ± 2 pN (mean \pm SEM) for solution concentrations of 1mM triphosphate (NTP) and 1 μ M pyrophosphate (PP_i). For alternate solution conditions of 1mM NTP and 1 μ M PP_i, the stall force was 23 ± 2 pN³⁰⁴. A 2003 study by Neuman et al. observed halting of RNAP only in the presence of forces above 27pN³⁰⁵. However, some studies have shown a much lower stall force. A 2000 study by Davenport did not observe transcription by RNAP above 15pN³⁰⁶. Another previous 1995 study by Yin et al. produced a stall force of 12 ± 1 pN for both reversible and irreversible stalls³⁰⁷.

Presently, our best estimate of the ribosomal stalling force is the 13 ± 2 pN figure, given that the experimental methodology involved applying a force to the ribosome and observing the drop in translation with increased opposing force¹⁶⁵.

Stall duration does not correlate with frameshift efficiency, though stalling appears to be a feature of frameshifting^{268,269,308} and frameshifted ribosomes pause ten times longer than non-frameshifted ribosomes³⁰⁹. Furthermore, the ribosome stalling force of 13 ± 2 pN corresponds to a mechanical work value of $19 \text{ pN} \cdot \text{nm} = 5 \text{ k}_B\text{T}$ when one takes the product of the stalling force and the step distance (1.48 nm).

Other studies of ribosomal dynamics have been undertaken. One recent study used Förster resonance energy transfer (FRET)³¹⁰, which is capable of accurately measuring the distance between two dyes³¹¹. These dyes can be attached at different locations of molecules of interest in order to determine the conformational dynamics of a system^{312,313}. With the knowledge that PRF efficiencies are correlated with conformational plasticity^{21,83}, a 2018 study by Wu et al. used FRET to uncover the translocation kinetics of ribosomes undergoing ribosomal frameshifting¹⁵². They classify different conformations of the ribosome-mRNA-tRNA complex and assign them to different FRET levels, finding a significant association between the durations of certain conformations and -1 PRF efficiency.

Another approach is to examine the structural characteristics of PRF stimulatory structures absent ribosomes. Several studies have studied the role of mRNA tension in frameshifting, by mutational studies of known frameshift stimulatory structures. These studies suggest a correlation between the unfolding force of mRNA pseudoknots and frameshift efficiency^{142,143}. Another computational study found a strong positive correlation between the average unfolding force in a range of spacer extension from 3.3 to 4.3 nm²⁰. Another study suggests that frameshift efficiency bears little correlation to mechanical parameters, though a correlation was found between unfolding rates around 24 pN with frameshift efficiency¹⁴⁵.

Other studies find no such relationship between stimulatory structure mechanical stability and frameshifting efficiency^{21,65,150,314}. A strong relationship between the conformational plasticity of pseudoknots and their frameshift efficiency has been suggested through several studies^{21,83,152} (Figure 13). While a clear mechanism eludes current understanding, several studies have closely examined the translocational dynamics of the ribosome at a frameshift site^{152,315}, revealing back and forth motions

(Figure 15). It has been hypothesized that conformational heterogeneity induces fluctuations in mRNA tension when the stimulatory structure (Figure 12) switches conformation; thereby causing the ribosome to shift reading frame²¹. This relationship motivates the study of the structural characteristics of high efficiency frameshifting pseudoknots.

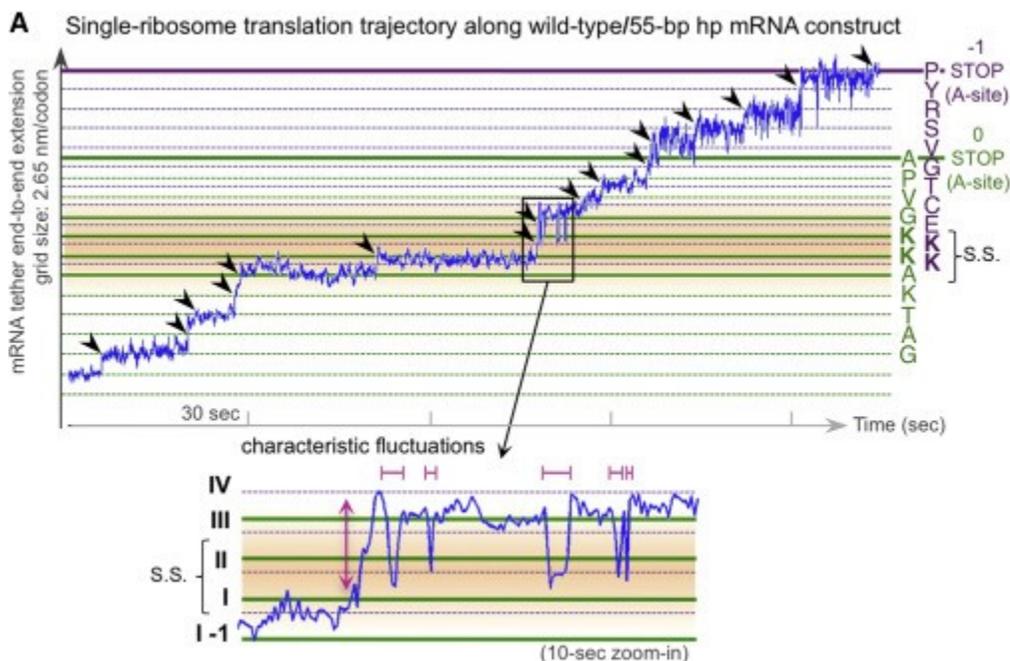


Figure 15: The ribosome moves in discrete sized movements, given by the length of a codon (1.7nm). At the slippery site, excursions on the order of 0.5s occur before resolution into the -1 frame. Source:³¹⁵.

3. Biology and evolution of the WNV frameshift signal

We will examine the WNV frameshift signal in greater detail in this chapter, providing details regarding its potential utility from a biological and evolutionary standpoint, as well as an overview of the epidemiology of WNV and related flaviviruses, providing a potential motivation for studying PRF in WNV.

3.1 Structure of the WNV frameshift signal

The apparent relationship between conformational plasticity and frameshift efficiency motivates the study of the West Nile virus (WNV) frameshift signal (NY strain, accession number NC_009942³¹⁶),

which features a very high frameshifting efficiency of $\sim 70\%$ ¹³⁹, one of the highest observed (Figure 16). Likewise, the WNV frameshift signal's wild type (WT) sequence is highly structurally plastic, with structural predictions producing three distinct structures: a 61 nt H-type pseudoknot¹⁵³, a set of dual hairpins (38nt and 71nt)¹³⁹, and a 109nt pseudoknot¹³⁹ (Figure 16).

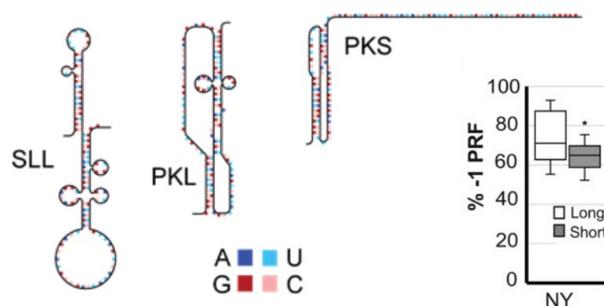


Figure 16: West Nile virus exhibits high structural plasticity and has a high frameshift efficiency. Main figure. Predicted structures from Atkins 2009 (PKS)¹⁵³ and Moomau 2016 (SLL and PKL)¹³⁹. Inset shows frameshift percentage for 129 nt WNV NY long sequence and a 75 nt short sequence. Source: Adapted from ^{139,153}

The methods used for determining the structure of the WNV PRF stimulatory structure have been Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE)³¹⁷ and bioinformatic predictions¹⁵³. The former method is an experimental technique designed to study local backbone flexibility in RNA. The fundamental principle of SHAPE is that flexible RNA bases are more reactive to hydroxyl-selective electrophilic reagents such as N-Methylisatoic Anhydride (NMIA)³¹⁸ or 1-methyl-7-nitroisatoic anhydride (1M7)³¹⁹, forming an adduct on the 2' hydroxyl group of RNA (2'-O-adducts)³²⁰. Once the RNA has been exposed to NMIA or 1M7 and adducts have formed, a 5' radio-labelled complementary DNA primer is annealed to the RNA, and the primer is extended via reverse transcriptase³¹⁷. The 2'-O-adducts disrupt the process of primer extension by reverse transcriptase³²¹, making a primer extension stop more likely near regions of high adduct formation (i.e. higher RNA local flexibility). This creates a range of DNA lengths which can then be separated based on length by high-

resolution gel electrophoresis or deep sequencing^{322–325} and compared to a control that is reverse transcribed in the absence of NMIA. The distribution of stops will reflect the local flexibility of the RNA structure. This information may then be used by secondary structure prediction software to constrain the possible structures³²⁶.

The second method involved the use of the RNA secondary structure predictors RNAfold³²⁷, pknots³²⁸, and manual inspection of possible base pairs³²⁹. Structural homology between conserved sequences of the JEV serogroup, found via alidot^{330,331} was also used to inform predictions, as differences in nucleotide sequence retained secondary structure¹⁵³ and the frameshift products are found within studied members (JEV^{332–335}, MVEV^{334,336,337} and WNV³³⁴).

3.2 Frameshift in WNV: Biological implications

West Nile virus is a positive sense RNA virus^{338–340} with an 11kb long viral genome^{341–344}, coding for a single polyprotein which is cleaved^{120,123,345–349} into three structural (capsid, C³⁵⁰; precursor membrane and membrane, prM/M³⁵¹; envelope, E³⁵²) and seven non-structural³⁵³ (NS1³⁵⁴, NS2A³⁵⁵, NS2B³⁵⁶, NS3³⁴⁹, NS4A³⁵⁷, NS4B³⁵⁸, NS5³⁵⁹) proteins^{360,361}. The cleavage of a single polyprotein by host and viral proteases is common in flaviviruses^{345,362,363}, and consequently produces near equimolar concentrations for all proteins³⁶³. Alterations from this ratio can weaken a virus; for example, the introduction of an internal ribosome entry site (IRES) between the C and prM genes of Murray Valley Encephalitic Virus, a close relative to WNV, changes the molar ratio of C to all other proteins, resulting in greatly reduced virulence³⁶³. This approach has been applied in vaccines for the Chikungunya virus (CHIKV)³⁶⁴ and the Venezuelan equine encephalitis virus (VEEV)^{365–367}.

There is a wide body of evidence supporting the notion of PRF as a function having utility in viruses³⁶⁸. This rise in utility may be due to effects on gene expression^{27,252}, or through the expression of novel proteins⁹⁹, or some combination of the two. For a PRF site, any proteins prior to the frameshift have their proportions increased. In the case of West Nile Virus, the frameshift increases the ratios of the

three structural proteins (capsid, C; envelope, E; premembrane, prM) as well as the mutant NS1', to the last six non-structural proteins (NS2A, NS2B, NS3, NS4A, NS4B and NS5) (Figure 17). Additionally, it is thought that the optimal ratio of proteins may change with stage of virus infection, with the envelope protein in particular having a great impact on viral effectiveness at early stages of infection³⁶⁹. The impact of solution, i.e. cellular, conditions on frameshift efficiency has not been well studied, but could present a mechanism by which protein ratios are tuned for a particular stage of infection³

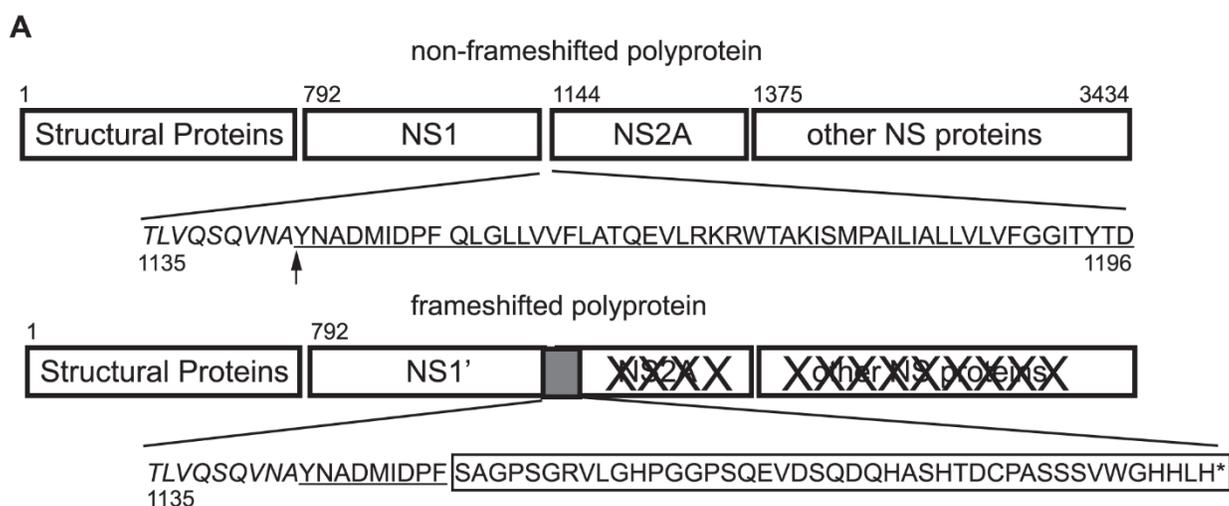


Figure 17: Depiction of the WNV polyprotein in the 0 frame (top) and -1 frame (bottom). The -1 frame polyprotein extends the C-terminus of the NS1 protein, and avoids expressing the proteins after NS1. This increases the ratio of structural proteins and NS1' relative to other non-structural proteins. Source:¹³⁷ (CC BY)

For the case of the NS1' frameshift in WNV, the structural proteins and NS1' (which is a C-terminal extension of NS1) have their ratios increased by this frameshift. The presence of the NS1' frameshift has utility for the virus, evidenced by the reduced pathogenicity of WNV¹³⁸ and JEV^{135,370} when NS1' is removed. It is unsure how PRF achieves this increase, though we discuss the changes it produces at the biological level.

The role of the NS1 protein in flaviviruses has been the subject of much research^{138,354,371,371-373}, and its structure has been characterized through X-ray crystallography³⁷¹. Further research on the role of

NS1 suggests that it is necessary for the formation of the WNV replication complex³⁷⁴. The replication complex is the machinery required to reproduce the components of a virus and package them within the viral capsid^{375,376}.

The mutant NS1' is a 52 amino acid (aa) C-terminally extended form of the NS1 protein, meaning that it is merely an 'extension' of an already existing protein^{138,153,332,333}. Furthermore, studies show that the NS1' protein is localised to the same cellular compartment as NS1, and appears to have the same role in replication as NS1^{354,374}. NS1' has, in addition, been shown to have a higher cellular retention than NS1³⁷⁷, meaning it is secreted less often.

Additionally, NS1' seems to create heat-stable dimers, unique to WNV. JEV^{377,378} and MVEV³⁷⁷, lack these dimers, although NS1/NS1' heterodimers have been observed in WNV³⁷⁷, JEV^{378,379} and MVEV³³⁴. Additionally, NS1' dimers appear to have greater stability at low pH than NS1 dimers, which possess high stability at pH 3.5³⁸⁰, and even maintained partial stability at pH 2.2³⁷⁷.

This increased stability and cellular retention appears to be a sequence-dependent property of the first 10 aa of the last 20 aa of NS1'. Mutant sequences were formed, truncating the last 20 aa of NS1', and subsequently, tests showed decreased stability and loss of cellular retention. This is likely to be a sequence dependent effect, rather than an effect merely depending on the length of the protein³⁷⁷, suggesting that the frameshift achieves these greater stabilities for the NS1' protein by accessing the codons 'hidden' in the -1 frame at the 5' (N-terminal) end of NS2A.

These properties may not be important for viral pathogenesis however, as 20 aa truncation does not appear to have a large effect on virus pathogenicity, suggesting that the increased neuroinvasiveness in PRF sequences is a property of the frameshift itself, and not of the novel gene product (at least not the last 20 aa). Additionally, NS1' expression in JEV does not appear to enhance virulence in mice³⁸¹. This provides support for the hypothesis that the positive effect of frameshifting emerges from the ability to alter the ratios of expressed proteins, particularly by increasing the ratio of structural proteins and NS1 to

the other non-structural proteins¹³⁷. Accompanying this discussion about the possible effects of the NS1' frameshift at the molecular level, we will also discuss the evolution of the NS1' frameshift.

3.3 Flaviviral evolution

Multiple reconstructions of the phylogenetic tree of flaviviruses have been created^{55,382–386} using bioinformatic approaches^{387–389}, which infer an evolutionary history from known present genomic data. Often, mutation rates are used to predict divergence times^{390–394}; additionally, geographic and historical data may be used to clarify and calibrate certain evolutionary divergences, as in reconstructing ancient human migrations^{395,396}, or more recent viral outbreaks^{397–399}.

In the case of flaviviruses, yellow fever virus^{55,400,401} (YFV) and dengue virus^{55,401,402} (DENV) are thought to have come to the Americas via the slave trade, and historical records of slave voyages are used to clarify the viral phylogenies⁵⁵. Additionally, more recent outbreaks also present data points in constructing a picture of flaviviral evolution, such as the introduction of WNV to New York in 1999⁴⁰³, likely coming from infected birds and mosquitos on international flights^{401,404,405}.

From this historical and genetic data, the phylogenetic tree of flaviviruses was constructed most recently by Mooreau et al. in 2015⁵⁵. Intriguingly, there are distinct nodes on the evolutionary tree whereby the presence of functional PRF is dichotomous, allowing one to observe the attributes separating functional PRF from lack of PRF. Node C of Figure 18 represents such a juncture, as SLEV, which has 69% identity with WNV^{341,406}, does not possess a computationally predicted frameshift stimulatory element^{55,153}. It remains somewhat ambiguous whether or not cacipacore virus (CPCV) induces frameshifting or not, as it possesses a slippery sequence; but no predicted stimulatory structure⁵⁵. All of the clade having node B1 as an ancestor appears to have frameshifting, with the possible exception of the Yaoundé virus⁵⁵ (YAOV), which possesses a slippery sequence and has a predicted stem-loop structure that could potentially stimulate frameshifting, as one does in the HIV frameshift signal^{27,66,166}. For Usutu virus (USUV), Koutango virus (KOUV), Murray Valley encephalitic virus (MVEV), and the closely related Murray Valley encephalitic virus-Alfuy (ALFV), the only evidence for frameshifting comes from

bioinformatic predictions^{19,138,153}. For JEV^{135,138} and WNV^{137,139} (including Kunjin strain (KUNV)), frameshifting has been both shown to occur, through the production of NS1', as well as shown to be associated with neuroinvasiveness.

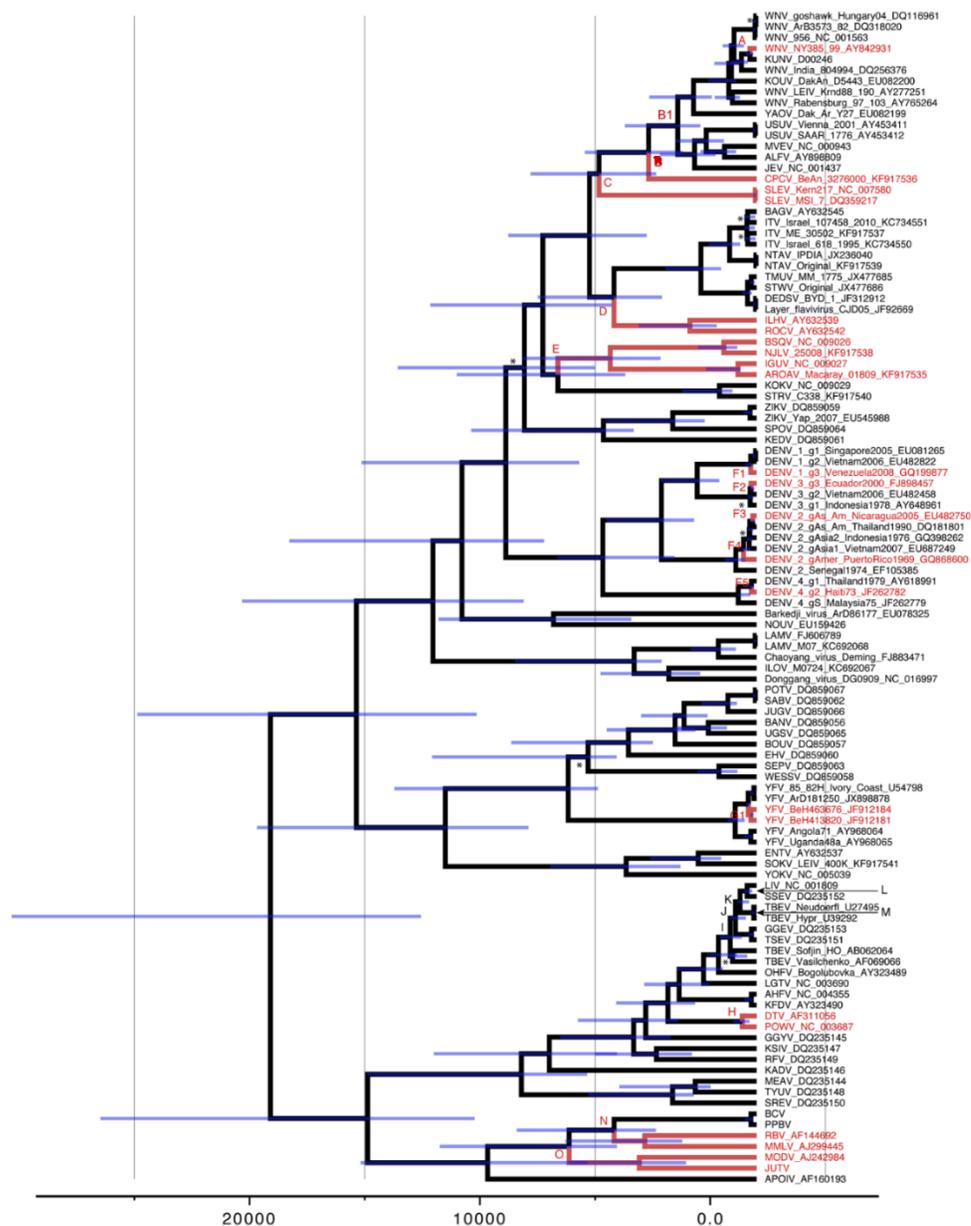


Figure 18: Phylogenetic reconstruction of family flaviviridae. Blue lines show error in node position, red names represent New World (NW) origin, whereas black names represent old world (OW) origin. Source: ⁵⁵ (CC BY)

3.4 Epidemiology of WNV

West Nile virus was discovered in 1937 in Uganda from a case of fever⁴⁰⁷. Early outbreaks of WNV were characterized only by a mild fever⁴⁰⁸⁻⁴¹⁵; neuroinvasive disease associated with WNV was not discovered until a case among elderly Israelis in 1957⁴¹⁵. Subsequent outbreaks of the more serious West Nile neuroinvasive disease (WNND) have occurred in multiple locations since then⁴¹⁶⁻⁴¹⁹. The frequency, severity and range of WNV greatly increased in the mid 1990s⁴⁰⁷. West Nile virus came to North America in 1999, making landfall in New York⁴²⁰⁻⁴²⁴, thought to have come via infected birds or mosquitos on an international flight^{404,420,425,426}. Within three years of the original New York outbreak, the virus had spread to neighboring Mexico and Canada⁴⁰⁷. Between 1999 and 2017 in the USA, WNV has been reported in 48,183 cases, of which 22,999 have been neuroinvasive (Figure 19), and has resulted in 2,163 deaths (Figure 20)⁴²⁷.

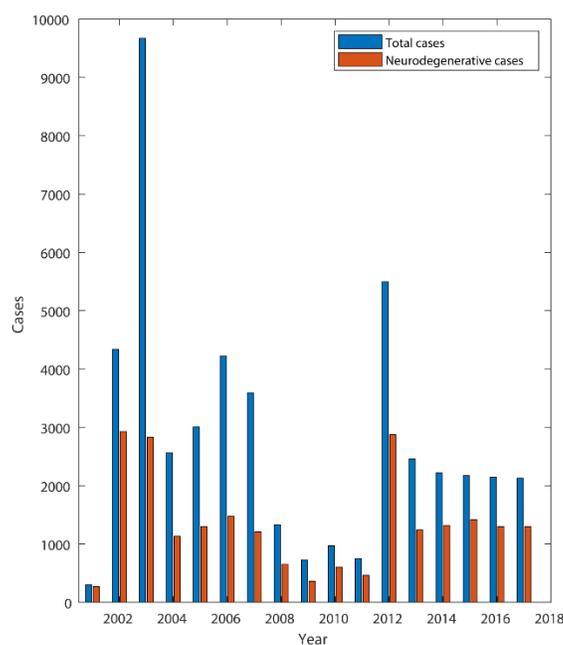


Figure 19: Cases of WNV from 1999 to 2017 (total, blue; neurodegenerative, red). Source: Adapted from⁴²⁷

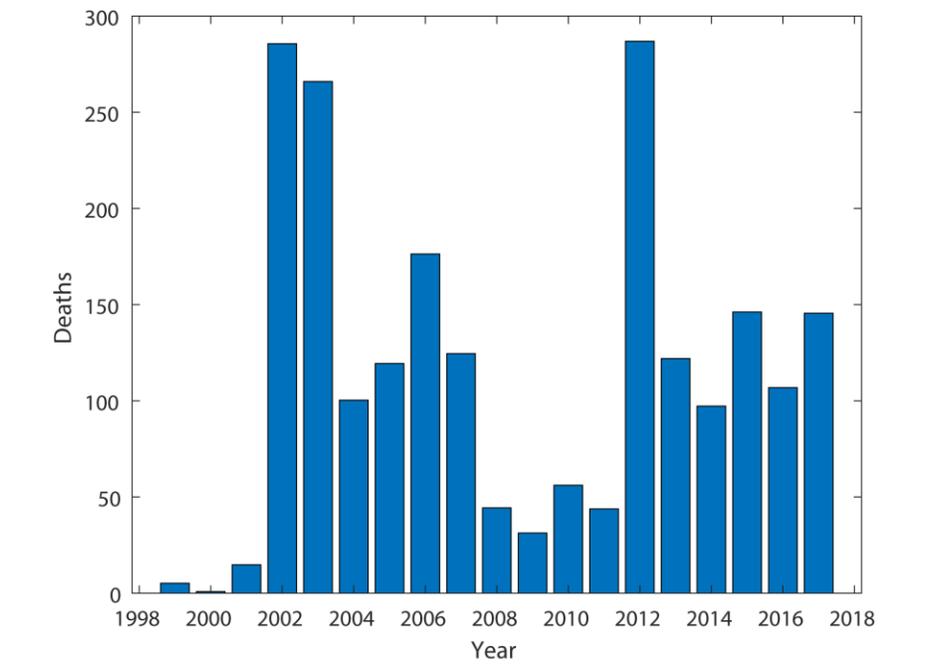


Figure 20: Deaths from West Nile Virus in the USA from 1999 to 2017. Source: Adapted from ⁴²⁷

The spread of WNV to Canada began in the summer of 2001, when WNV was detected in dead birds and mosquito pools in Ontario⁴²⁸⁻⁴³⁰. Clinical cases emerged in 2002 in Quebec and Ontario⁴³⁰, and spread to Manitoba, Saskatchewan and Alberta in the subsequent year, finally making landfall in British Columbia in 2009⁴³¹. There have been approximately 5700 human cases in Canada since 2002^{407,432}, and it has resulted in 48 deaths in Canada since 2013, including 26 deaths in 2018 (Figure 21)⁴³². West Nile virus currently covers a wide geographic area⁴⁰⁷, and is considered one of the most important causative agents of viral encephalitis worldwide⁴⁰⁷, along with Japanese encephalitic virus⁴³³⁻⁴³⁶.

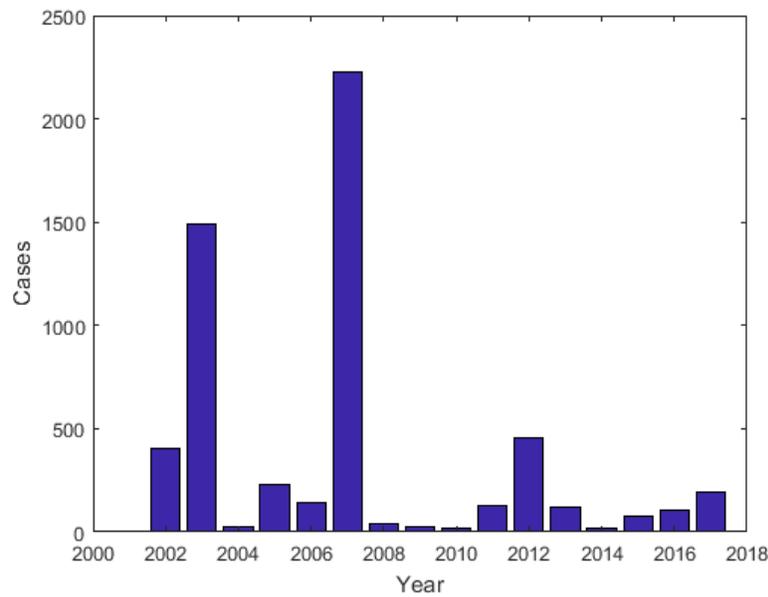


Figure 21: Cases of WNV in Canada from 2002 to 2017. Source: Adapted from ^{407,432}

West Nile Virus is likely to remain an endemic virus within Canada as birds, horses and other wildlife serve as a reservoir for the disease⁴³⁷⁻⁴⁴⁴. Human infection occurs most frequently via bites from infected mosquitos that fed on infected birds⁴⁴⁵⁻⁴⁴⁹. Birds and other animals are capable of infecting mosquitos because they produce sufficient viremia, whereas humans do not produce sufficient viremia to infect mosquitos^{450,451}. While rarely transmitting from human to human, it may be transmitted via blood transfusion⁴⁵²⁻⁴⁵⁵, organ transplant⁴⁵⁶, from mother-to-child during pregnancy^{457,458} and via breastfeeding⁴⁵⁹⁻⁴⁶³.

Thus far, a vaccine does not exist for WNV in humans, though four vaccines are available for horses⁴⁶⁴⁻⁴⁶⁸ which have been effective in previous outbreaks and trials⁴⁶⁹⁻⁴⁷³. So far, some therapeutics, such as human immunoglobulin⁴⁷⁴ and interferon-alpha⁴⁷⁵⁻⁴⁷⁹ have been applied with limited success. While the human cost of West Nile virus and other highly related flaviviruses provides suitable motivation for an understanding of its PRF mechanism, other viruses possess high relatedness with

WNV⁵⁵ and have a very similar PRF motif¹⁵³. It is possible that a study of this PRF motif may provide insights into means of disrupting other highly similar viruses of the JEV serogroup (e.g. Murray Valley Encephalitis virus, MVEV; Kunjin virus, KUNV; St. Louis Encephalitic virus, SLEV; Alfuy virus, ALFV; Usutu virus, USUV; Koutango virus, KOUV)⁵⁵. Furthermore, more general insights into frameshifting may be discovered, and allow for therapeutic targeting of PRF in viruses⁴⁸⁰.

4. Methods

This thesis uses optical tweezers to perform single molecule force spectroscopy (SMFS) on the West Nile Virus frameshift signal. This section describes the physical principles behind optical tweezers, how samples are prepared and measured, and how subsequent analysis of the data works.

4.1 Single-molecule force spectroscopy

Prior to the advent of experimental technology and methods allowing one to study single molecules, information about biomolecules typically came from measurements performed on large ensembles of biomolecules. While useful for determining some properties of nucleic acids, such as thermal stability, more information could be garnered by examining individual biomolecules. Single-molecule approaches allow one to examine transient states, rare conformations and dynamic behaviour of single molecules. Additionally, single-molecule approaches can examine sub-populations, which are lost in ensemble studies by averaging over many molecules.

Single-molecule force spectroscopy (SMFS) encompasses a variety of methods, including optical tweezers (OT, Figure 24), magnetic tweezers (MT, Figure 23), and atomic force microscopy (AFM, Figure 22). Molecular conformational changes may also be probed through other methods such as single-molecule Forster resonant energy transfer (smFRET), which are used standalone, or in conjunction with SMFS for additional detail.

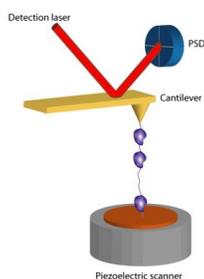


Figure 22: Atomic force microscopy of single biomolecules. Fluctuations in the cantilever with a known stiffness result in movements reflected by the reflection of a laser onto a position sensitive device (PSD).

Source:⁴⁸¹

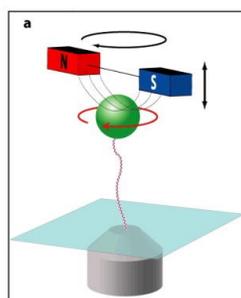


Figure 23: Force/torque spectroscopy by magnetic tweezers. Source:⁴⁸¹

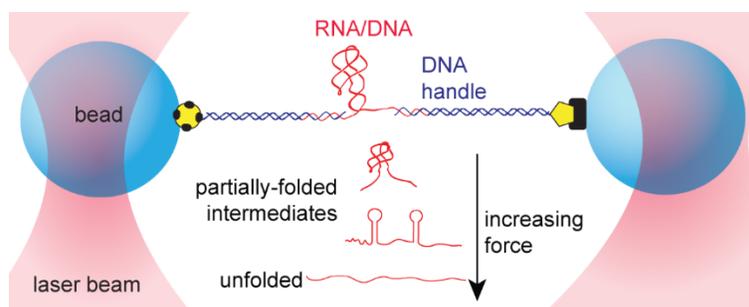


Figure 24: Single-molecule force spectroscopy of nucleic acid construct. Increasing force causes the construct to unfold, often through a series of intermediates. Source:⁴⁸²

4.1.1 Sample Preparation

SMFS measures the extension of a single biomolecule under an applied force. The nucleic acid construct, typically a biomolecule on the order of ~ 10 -100nm, is annealed to kb long handles which may then be attached to the probe. For optical and magnetic tweezers, the probe is a dielectric bead or superparamagnetic bead respectively, whereas for AFM, it is a cantilever. Attachments of handles to beads are made via molecular interactions; streptavidin, a protein purified from *Streptomyces avidinii*, binds with high affinity to biotin with a dissociation constant of roughly 10^{-15} M⁴⁸³. It is possible to synthesize biotin-labelled (d)UTP to create biotinylated nucleic acids⁴⁸⁴, which will attach to streptavidin coated beads. Digoxigenin, a protein found in the flowers and leaves of digitalis plants, binds with anti-digoxigenin with an affinity of approximately 10^{-9} M⁴⁸⁵. Additionally, digoxigenin-labelled (d) UTP can be synthesized into handles, binding to anti-digoxigenin labelled beads.

Before attachment to handles, beads are sonicated (agitated with sound waves) to break up any aggregates. Beads are attached to the full molecular construct via incubation at room temperature. The experimental solution consists of 50 mM 3-(N-morpholino) propanesulfonic acid (MOPS), pH 7.0, 130 mM KCl, 4 mM MgCl₂. The purpose of MOPS buffer is to prevent changes in pH over the course of the experiment. Solution conditions are chosen to be permissive to RNA folding in terms of pH⁴⁸⁶ and ionic strength⁴⁸⁷. While our experiment is not looking to replicate *in vivo* conditions, typical mammalian cells have comparable values. A mammalian cell has a monovalent salt concentration of 150mM and a divalent salt concentration of approximately 1mM^{488,489}, while our experiments take place at concentrations of 130mM for monovalent and 4mM for divalent. Additionally, the pH is similar, as human cells have a slightly basic pH of 7.4⁴⁹⁰, and experiments take place at pH 7.0.

One issue with optical trapping experiments is the gradual accumulation of reactive oxygen species in solution via photodissociation of O₂ by laser light^{491,492}. An oxygen scavenging system is added to reduce oxidative damage to the construct via removal of molecular O₂ by oxidizing glucose. The oxygen scavenging system consists of glucose oxidase, glucose and catalase (Figure 25). Two glucose

molecules react with two water and two O₂ (catalyzed by glucose oxidase) to form two molecules of gluconic acid and two molecules of hydrogen peroxide (H₂O₂). The two molecules of hydrogen peroxide are converted to two water molecules and an O₂ molecule via catalase. The net effect of these two processes is two molecules of glucose and one molecule of O₂ being converted into two molecules of gluconic acid (Figure 25).

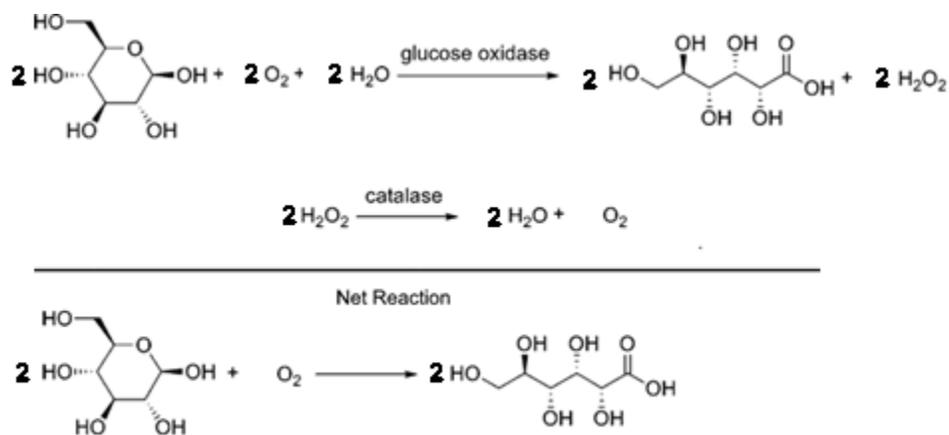


Figure 25: Oxygen scavenging reactions. Top. In the first reaction, catalyzed by glucose oxidase, two glucose molecules react with two O₂ molecules and two molecules of water to form two molecules of gluconic acid and two molecules of H₂O₂. The lower reaction converts two molecules of H₂O₂ to two molecules of water and one molecule of diatomic oxygen. Bottom. The net reaction is going from two molecules of glucose and one molecule O₂ to two molecules of gluconic acid.⁴⁹³

Additionally, ribonuclease (RNase) inhibitor is included, to prevent ribonucleases from degrading the construct⁴⁹⁴. RNases exist on skin, in dust suspended in the air, and on bench surfaces. As such, their ubiquity motivates anti-contamination measures such as storing pipette tips in a sealed container, cleaning work surfaces judiciously with RNase inactivating agents, and having dedicated pipettes for RNA work to avoid nuclease contamination⁴⁹⁵.

4.1.2 Slide Preparation

A sample slide is prepared, consisting of a narrow channel formed between two glass cover slides held together with two-sided tape. The glass is cleaned with dilute ethanol and plasma-cleaned to remove any contamination, prior to forming the channel. Sample solution is injected into the gap between cover slides and sealed with vacuum grease (Figure 26). At the optical tweezers apparatus, immersion oil is used on the objective lens, and the microscope is focused visually. Dumbbells are found by searching manually, usually in a raster fashion, over the microscope slide. Position is calibrated before every set of pulls. For measurements, position is changed linearly until a maximum force is reached, position is then linearly decreased until force is zero, producing the associated force-extension curve. This cycle may be repeated up to hundreds of times before a given molecule breaks.

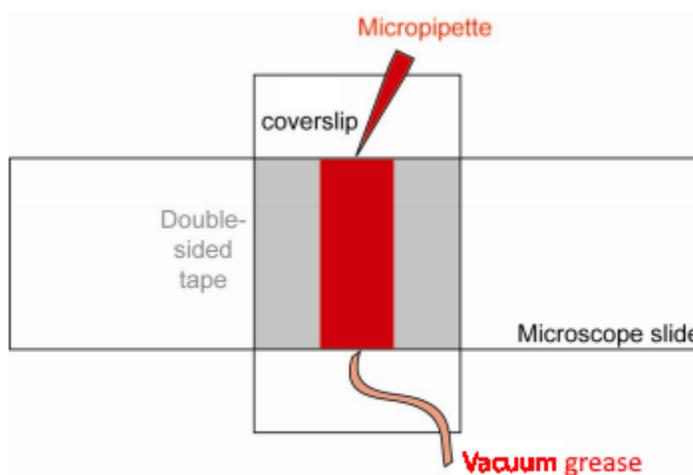


Figure 26: Flow cell schematic. Source⁴⁹⁶

4.1.3 Optical Tweezers Measurements

Optical tweezers, developed in the late 1980s⁴⁹⁷, allow for the trapping of small particles in a beam of laser light. The trapped particle is held in an approximately quadratic potential, experiencing a Hookean restoring force when perturbed. Additionally, the bead experiences a scattering force in the direction of light propagation.⁴⁹⁸

The scattering force on a spherical, isotropic particle of comparable wavelength to the incident light was first derived by Gustav Mie in 1908⁴⁹⁹. Rayleigh had previously developed equations for scattering by small particles of radius $a \ll \lambda$ ⁵⁰⁰. In the Rayleigh regime, scattering forces are given by Equation 1 below, where I_0 is the intensity of the incident light, σ the scattering cross section, n_m the index of refraction of the medium (water $n_m=1.33$), c the speed of light in a vacuum, λ the wavelength of the trapping beam and m the ratio of the index of refraction of the particle to that of the medium ($m=n_p/n_m$)⁴⁹⁸.

Equation 1: Scattering force of a particle in Rayleigh regime

$$\vec{F}_{scatt} = \frac{I_0 \sigma n_m}{c} \hat{z}$$

$$\sigma = \frac{128\pi^5 a^6}{3\lambda^4} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2$$

The trapping beam used is in the infrared regime, with a wavelength of 1064nm. The radii of the polystyrene beads used in the experiment were 300 and 410nm for the avidin DN (Vector Labs) and anti-digoxigenin (Roche) labelled beads respectively. The beads' radii are of the same order of magnitude as the trapping beam, so the Rayleigh approximation breaks down, though it is still informative.

The other important force is the gradient force. For a particle within an electric field $\vec{E}(\vec{r}, t)$ with intensity profile $I(\vec{r})$, the gradient force for a dipole with polarizability $\vec{p}(\vec{r}, t)$ is given by:

Equation 2: Gradient force on a particle

$$\vec{F}_{grad}(\vec{r}) = [\vec{p}(\vec{r}, t) \cdot \nabla] \vec{E}(\vec{r}, t) = \left(\frac{2\pi n_m a^3}{c} \right) \left(\frac{m^2 - 1}{m^2 + 2} \right) \nabla I(\vec{r}).$$

The total force is given by $\vec{F}(\vec{r}) = \vec{F}_{scatt}(\vec{r}) + \vec{F}_{grad}(\vec{r})$, and in order for trapping to occur, the minimum force along the z-axis must be negative⁵⁰¹. Another condition is that the trapping potential, given by $V(\vec{r}) = -\frac{2\pi n_m a^3}{c} \left(\frac{m^2 - 1}{m^2 + 2} \right) \nabla I(\vec{r})$, must have a much larger ($\sim 10x$) well depth than the average

kinetic energy of the nanosphere, which is $\frac{1}{2}k_B T$ in each direction, where k_B is the Boltzmann constant and T is the temperature⁵⁰¹.

4.1.3.1 Position calibration

Bead position is measured via the deflection of the trapping beam onto a position-sensitive detector (PSD). The PSD consists of two layers of laminar semiconductor. When a photon of light encounters a PIN photodiode, it creates an electron hole pair and a concomitant current (Figure 27). The resistance of the photodiode is constant throughout the layer, so light incident closer to a given side will have a larger current on that side. The position is proportional to the difference signal divided by the sum signal. The proportionality constants $k_{x,y}$ are determined via corroborating a known displacement of the piezoelectric stage to the ratio of the difference signal to the sum signal. The location of the light's incidence on the PIN photodiode is calculated via

Equation 3: Position of incidence of light on a PSD.

$$x = k_x * \frac{I_{x1} - I_{x2}}{I_{x1} + I_{x2}}$$

$$y = k_y * \frac{I_{y1} - I_{y2}}{I_{y1} + I_{y2}}$$

where I is the current or voltage measured.

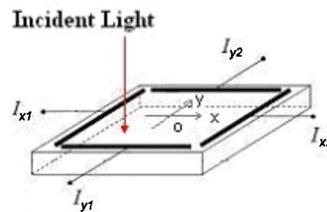


Figure 27: Schematic of position sensitive device (PSD). Source: Revised from Wikipedia, author Andycui1 (CC BY-SA 3.0)

Furthermore, one also needs to align the trapping beam with the PSD center, and be able to determine the proportionality between given deflections of the trapping beam using an acoustic-optical device (AOD). AODs respond within microseconds of a signal change, offering better response times than other deflection methods such as tilt-capable mirrors⁵⁰².

4.1.3.2 AODs

Acoustic-Optical Devices (AODs) act by transducing an AC electrical signal into a sound wave via a piezo-electric transducer. The sound wave propagates through a crystal, tellurium dioxide (TeO₂) in our experiment, creating evenly-spaced regions of high and low density⁵⁰³. The index of refraction n is related to the density by the following relations:

Gladstone-Dale Law

$$\frac{n - 1}{\rho} = CM^{-u}$$

Lorentz-Lorenz Law

$$\frac{n^2 - 1}{(n^2 + 2)\rho} = KM^{-v}$$

Drude Law

$$\frac{n^2 - 1}{\rho} = C_{drude}$$

where ρ is the density, C and K are material constants, u and v are empirical constants approximately equal to 0.4, and M is the average atomic mass⁵⁰⁴. In the relevant regimes for TeO₂ crystal, refractive index increases monotonically with ρ . Importantly, an incident sound wave on a TeO₂ crystal creates periodic regions of low and high density, behaving similar to a diffraction grating for incident light.

For light incident normal (Debye-Sears regime) to an AOD surface, the deflection angles are given by $\theta_D = n\lambda/\Lambda$, where n is the deflection order, λ is the wavelength of the incident light, and Λ is the acoustic wavelength. Deflection efficiencies increase with greater sound wave amplitude, which makes the differentials in density larger. This subsequently increases the magnitude of refractive index differences. Modification of the sound wave frequency changes the spacing between regions of higher density, allowing for modification of the deflection angle. For a diffraction grating, the deflection angle for normally incident light is $\sin\left(\frac{\theta}{2}\right) = \lambda/2\Lambda$, where Λ is the period of the diffraction grating and λ is the wavelength of the incident light. This reduces to $\theta \approx \lambda/\Lambda$ in the small angle regime. One can then modulate the deflection angle by changing the driving frequency of the piezoelectric device (Figure 28).

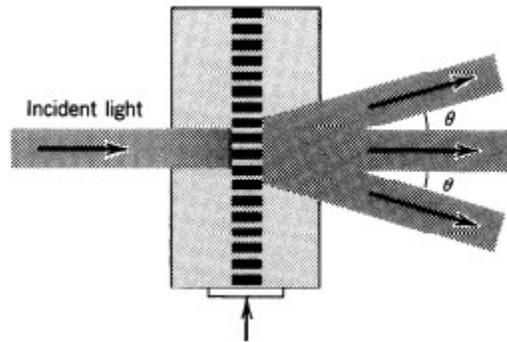


Figure 28: Deflection of incident light by a sound wave. Source: ⁵⁰³

In order to find a proportion between the frequency change of an AOD and the actual position of a trap, we reposition the trap a known distance away by changing the input frequency to the AOD, from this we may determine the scaling between change in AOD driving frequency and distance in the sample plane. Furthermore, with a calibrated position detector, one can trap a bead, move it by varying the AOD input frequency, and measure its subsequent position using the detection laser and PSD.

4.1.3.3 Force Calibration

In order to calculate the force, it is necessary to calculate the stiffness constant α of the system.

Stiffness may be calculated in several different ways; these are by power spectrum, Stokes drag, and variance⁵⁰⁵.

4.1.3.4 Power spectrum

The thermal motion of beads within an optical trap can be used to calculate the trap stiffness. The power spectral density for Brownian motion of a spherical particle in a harmonic potential is

Equation 4: Power spectrum

$$S_{xx}(f) = \frac{k_B T}{\pi^2 \beta (f_0^2 + f^2)}$$

$$\beta = 6\pi\eta a,$$

where k_B is Boltzmann's constant, T is the absolute temperature, f_0 is the roll off frequency, β is the hydrodynamic drag coefficient, η is the viscosity of the medium, and a is the bead radius. This is shown in Figure 29. For cases where the bead's center is a distance z from a surface, the value of β must be corrected with Faxen's law.

Equation 5: Faxen's law correction for motion near a surface

$$\beta = \frac{6\pi\eta a}{\left[1 - \frac{9}{16} \left(\frac{a}{z}\right) + \frac{1}{8} \left(\frac{a}{z}\right)^3 - \frac{45}{256} \left(\frac{a}{z}\right)^4 - \frac{1}{16} \left(\frac{a}{z}\right)^5\right]}$$

The trap stiffness is related to these parameters by $\alpha = 2\pi\beta f_0$.

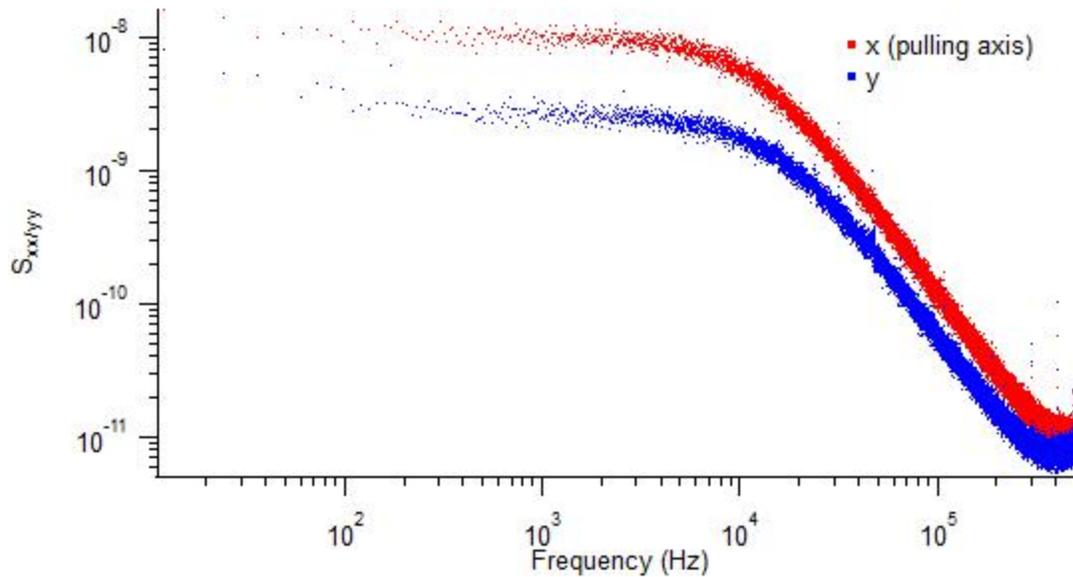


Figure 29: Power spectral density plot for bead trapped in x (pulling axis) and y (perpendicular to pulling axis).

Stokes Drag

Another complementary method for determining the trap stiffness is via observing the motions of a bead being driven by a triangular driving force of amplitude A_0 and frequency f . This results in a square wave of force being applied to the bead, with bead motion given by

Equation 6: Bead motion under triangular driving force in viscous medium

$$x(t) = \frac{\beta A_0 f}{2\alpha} [1 - e^{-\left(\frac{\alpha}{\beta}\right)t}],$$

where β is the drag coefficient of the bead and α is the trap stiffness. The value for β requires Faxen's correction, given by Equation 5. Typically, only the asymptotic value $\frac{\beta A_0 f}{2\alpha}$ is used to estimate the trap stiffness, as this is easiest to fit.

Variance

The variance method uses the equipartition theorem for an object in a harmonic potential to determine the stiffness. For any object in a harmonic potential with stiffness α , the equipartition theorem states that

Equation 7: Equipartition theorem

$$\frac{1}{2}k_B T = \frac{1}{2}\alpha\langle x^2 \rangle,$$

where x is the displacement from equilibrium. By measuring the variance in position, one can determine the stiffness without any prior knowledge of the bead's drag coefficient.

4.2 Analysis

The basic data structure for force-ramp experiments is the force-extension curve (FEC), a plot which yields much information. This section will explore what information may be garnered from a collection of FECs. First we will have a brief overview of the interpretation of FECs, showing how states and transitions can be determined. Secondly we will examine properties of conformational states that may be determined from FECs; these will demonstrate assignment of structural conformation and illustrate how to determine the properties of individual structures (contour length, unfolding force and free energy). Thirdly we will examine the properties of individual transitions, examining kinetic parameters.

4.2.1 Properties of biopolymers

We will begin, prior to the discussion of given structural states, by first defining the mechanical properties of biomolecules. As previously described, the assay consists of a molecular construct annealed to handles, which are attached to polystyrene beads via protein linkages. A common model used to describe the force-response of biomolecular chains is the Extensible Worm-Like Chain (EWLC hereby referred to as WLC) model, which consists of a continuously flexible rod^{506,507} (Figure 30).

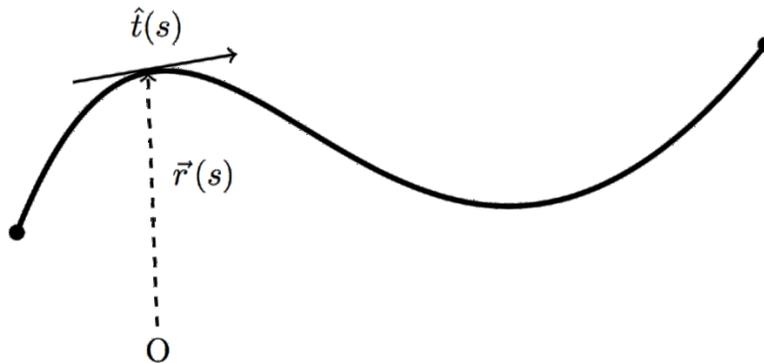


Figure 30: Worm-like chain model

This rod is defined by several key parameters:

- The contour length L_c : the total length along the backbone of the rod
- The persistence length L_p : the characteristic decay length for correlations between tangent segments $\langle \hat{t}(s) \cdot \hat{t}(0) \rangle = \langle \cos(\theta(s)) \rangle = e^{-s/P}$
- The stretch modulus K

The Hamiltonian of the system can be decomposed into three distinct terms: the entropic, enthalpic and external (from the applied force). The entropic term favors a short extension, owing to the greater number of conformations where segments are unaligned, additionally, the enthalpic term also favors a shorter extension, owing to the increase in energy concomitant with stretching. These terms compete with the external term, which favours extension. The Hamiltonian of the system for a given extension x is:

Equation 8: Hamiltonian for biomolecule under constant pulling force

$$H = H_{entropic} + H_{enthalpic} + H_{external} = \frac{1}{2} k_B T \int_0^{L_c} L_p \left(\frac{\partial^2 \vec{r}(s)}{\partial s^2} \right)^2 ds + \frac{1}{2} \frac{K}{L_c} x^2 - xF$$

Computing a partition function and finding the most probable force produces the following interpolation formula at low force ($<10\text{pN}$)⁵⁰⁶:

Equation 9: Worm-like chain fitting function

$$\frac{FL_p}{k_B T} = \frac{1}{4} \left(1 - \frac{x}{L_c} + \frac{F}{K} \right)^{-2} - \frac{1}{4} + \frac{x}{L_c} - \frac{F}{K}.$$

Additionally, at higher forces, the relation is well approximated by the following formula:

Equation 10: WLC in high force limit

$$x = L_c \left(1 - \frac{1}{2} \left(\frac{k_B T}{FL_p} \right)^{\frac{1}{2}} + \frac{F}{K} \right)$$

At low forces (<10 pN), the greatest contributions to the increase in length are due to the increasing alignment in the polymer backbone (entropic component), this is shown in Figure 31. Between 10 pN and 40-50pN, the greatest contribution to the Hamiltonian becomes the enthalpic component, the stretching of the individual bonds. At approximately 60pN, the overstretching transition occurs, where extension changes by a factor of 1.7 without a significant increase in force⁵⁰⁸. This regime is characterized by strand unpeeling, melting bubble formation, and the conversion of the NA from B-form (helical) to S-form (ladder-like), i.e. the double helix untwists^{509,510}. This turning over limits the range of useful data to below 50pN in our experiment.

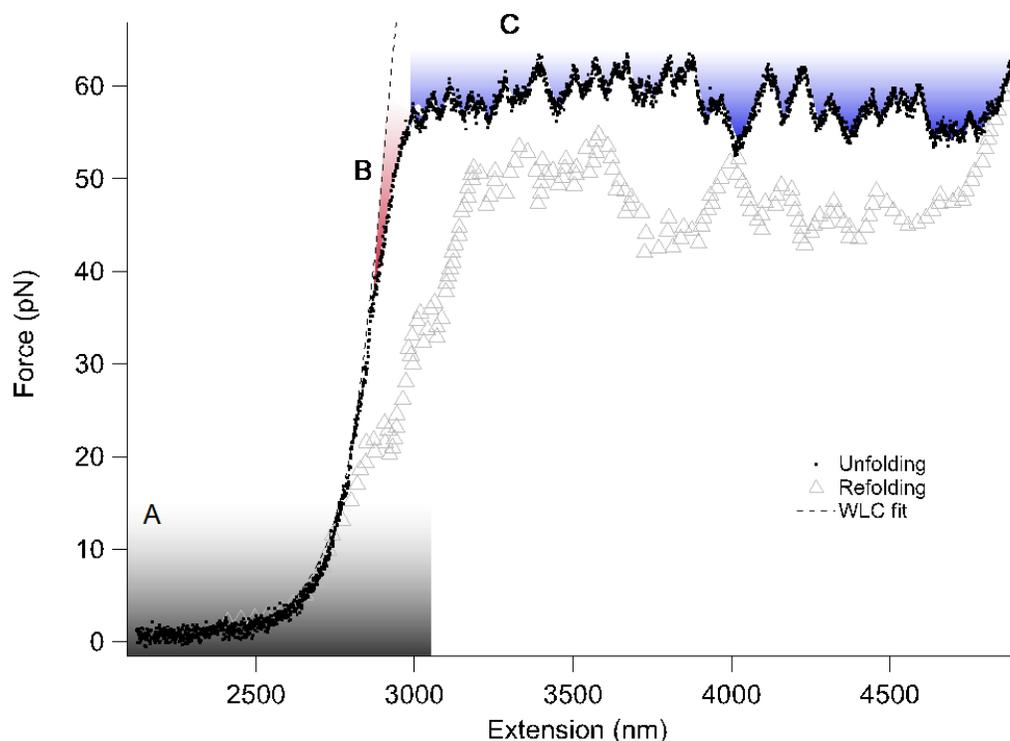


Figure 31: Unfolding of a single molecule⁵¹¹. The unfolding proceeds via (A), the entropic regime, where alignment of the polymer backbone causes an increase in extension. (B) The enthalpic regime, where increases in extension are largely due to bond stretching. (C) The overstretching regime, whereby B-DNA within the double helix converts to ladder like S-DNA. Additionally, this regime is characterized by strand melting and bubble formation, which are represented by the multiple rips observed in this regime.

Source:⁴⁸²

These values are dependent on solution conditions (temperature⁵¹², pH⁴⁸⁶, ionic concentration^{487,513–518} and presence of neutral ‘crowding’ cosolutes⁵¹⁹). However, our experiments are all taken under constant conditions, and values are close to the literature values for ssRNA ($L_p \sim 1\text{nm}$, $K \sim 1600\text{pN/nm}$)^{487,520} and dsDNA ($L_p \sim 50\text{nm}$, $K \sim 1100\text{pN/nm}$)^{511,521–524}.

4.2.2 Properties of states

Nucleic acids fold into secondary structure from the formation of intra-molecular base pairs. One typical component of secondary structure is the stem-loop or hairpin, which refers to a double helix of

Watson Crick (canonical base pairing A: T/G: C) base pairs, topped by a loop, typically of four nucleotides or more. An example of a stem is shown in Figure 32, consisting of a 13 bp stem with a bulge and a 5nt loop.

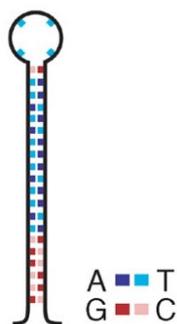


Figure 32: DNA hairpin consisting of a 20 bp stem with 4 nt loop. Source:⁵²⁵

Secondary structures form prior to tertiary structures, such as pseudoknots (example shown in a) or other structures such as kissing stem loops, which consist of two stem loops interacting by inter-loop contacts. This hierarchical folding is necessary, as tertiary structures contain the secondary contacts, as in a, where the pseudoknot consists of two stem loops.

When a structure is formed, nucleotides are sequestered away from contributing to the contour length. Single molecule studies apply force to these structures, thereby rupturing them and causing a concomitant increase in contour length (Figure 33). The difference in contour length will be equivalent to the length of the sequestered nucleotides (0.59nm/nt for ssRNA, 0.62 nm/nt for ssDNA)^{526,527} minus the width of the structure d_T (typically 2nm for a hairpin, $2\text{-}6\text{ nm}$ for a pseudoknot, though our data suggests larger d_T values)^{21,528}. For pseudoknots, one can occasionally find the crystal structure of a homologous pseudoknot structure, and calculate the d_T directly using molecular visualization software such as Visual Molecular Dynamics (VMD)⁵²⁹.

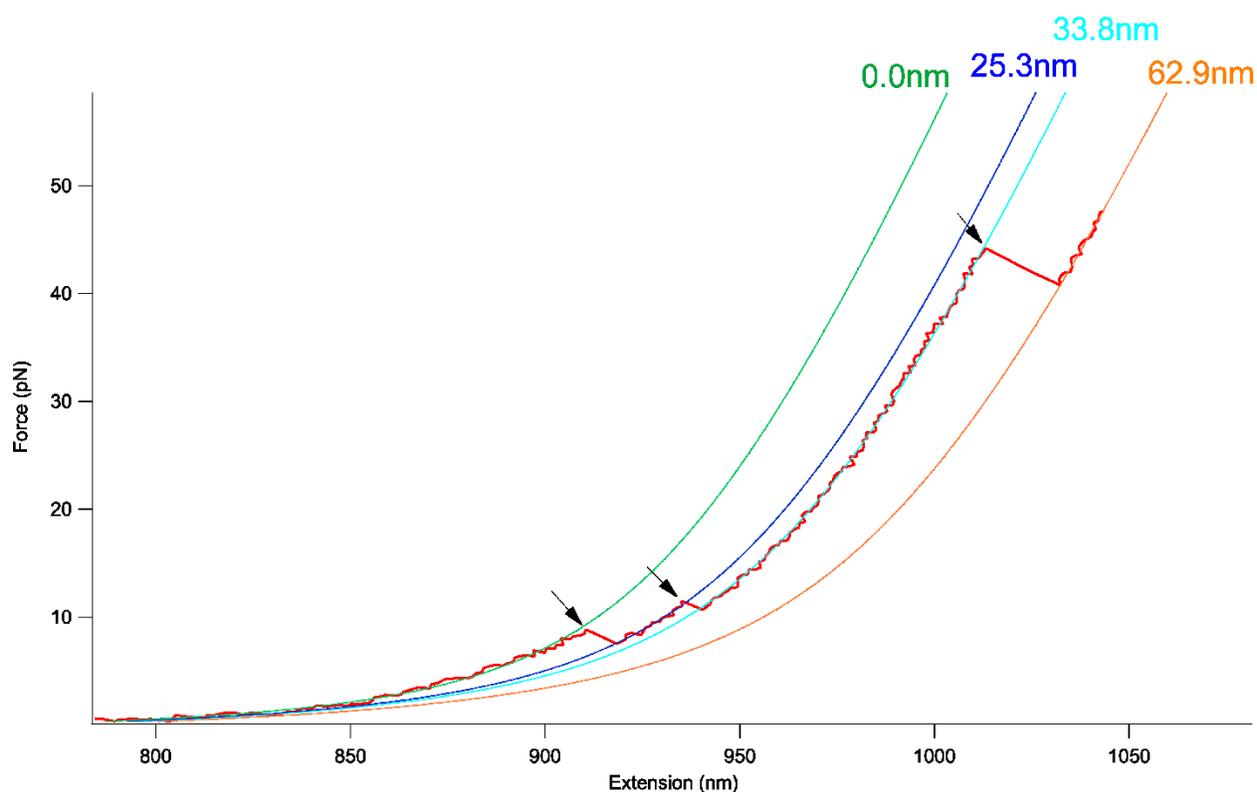


Figure 33: Fit to multiple branches of the WNV frameshift signal using the worm-like chain (WLC) model proposed by Marko and Siggia⁵⁰⁶. Unfolding forces are shown with arrows. The uppermost arrow corresponds to the unfolding of a pseudoknot, and as such has a much higher unfolding force than the other states.

For example, the closest crystal structure for the WNV NC_00942 pseudoknot (b) is the murine leukemia virus (MLV), which has a width of 5.6nm ⁵³⁰, and has 54nt sequestered in structure (Figure 34), as opposed to the WNV pseudoknot with 61nt (36.0 nm unfolded). Assuming a similar d_T , this gives a contour length change prediction of 30.4nm.

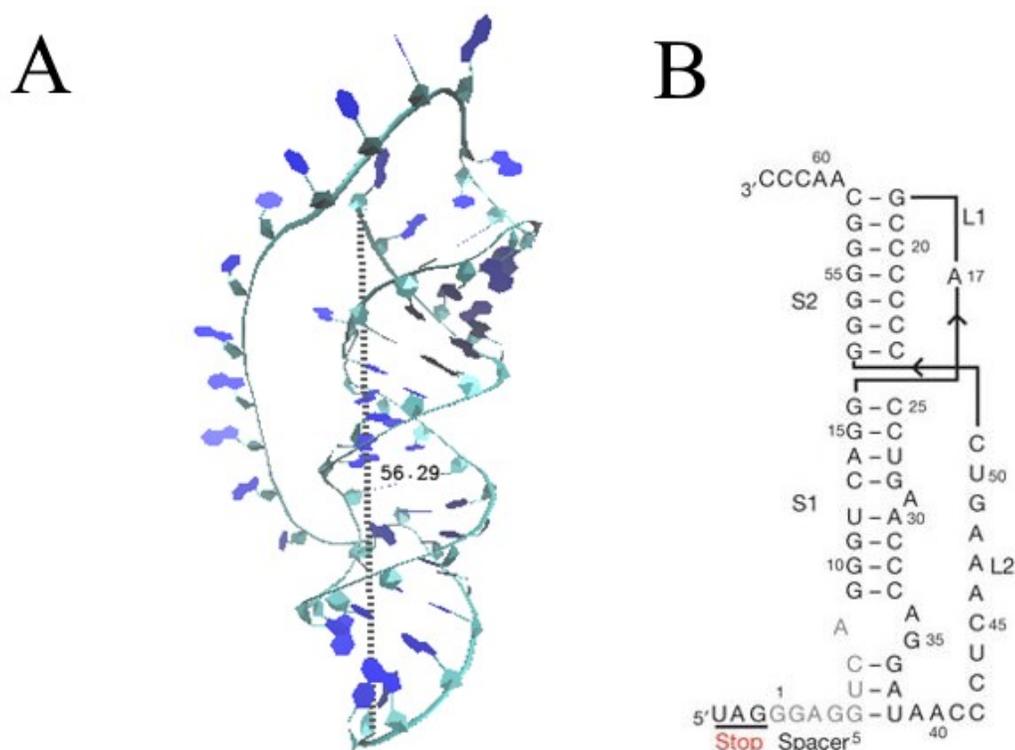


Figure 34: The Murine Leukemia Virus (MLV) read-through pseudoknot. Panel A. A visualization in VMD⁵²⁹ of the MLV read-through pseudoknot (RCSB ID: 2LC8)⁵³⁰, with a d_T of 56 angstroms. Panel B. A diagram of the MLV pseudoknot showing base pairings. Source: ⁵³⁰

Other pseudoknots lacking homologous structures are more difficult to calculate d_T , though the problem may still be approached. A DNA double helix rises approximately 0.33-0.34 nm/bp⁵³¹⁻⁵³³, whereas an RNA double helix rises 0.27-0.28 nm/bp^{531,534,535}. For the MLV virus, it consists of two stacked stems, with 11 and 7 bp respectively, for a total contribution to height of 5.0 nm. Additionally, Stem 1 contains 4 unpaired nts (0.59 nm/nt for ssRNA)⁵²⁶, which contribute to the height. It is unclear how an asymmetric unpaired nucleotide contributes to height, though we may discount them for now. There exists at least one set of unpaired nucleotides in the bulge at residues 8 and 34-35, contributing 0.6 nm to d_T , for a total of 5.6 nm, close to the calculated value.

This technique may be used to assign states to certain branches of the force extension curve with a computed contour length. Additionally, force determinations can be made by finding the point at which the rip occurs. This can present another data point in assigning a state. Tertiary structures, such as pseudoknots, tend to have a higher rupture force (20-50 pN)^{21,151} than mere secondary structures (5-15 pN)⁵²⁸, owing to the stabilizing influences of the extra contacts (Figure 35), in addition to increases in energetic barrier rigidity^{536,537}. As such, the unfolding force of a state may provide an additional means of classifying it.

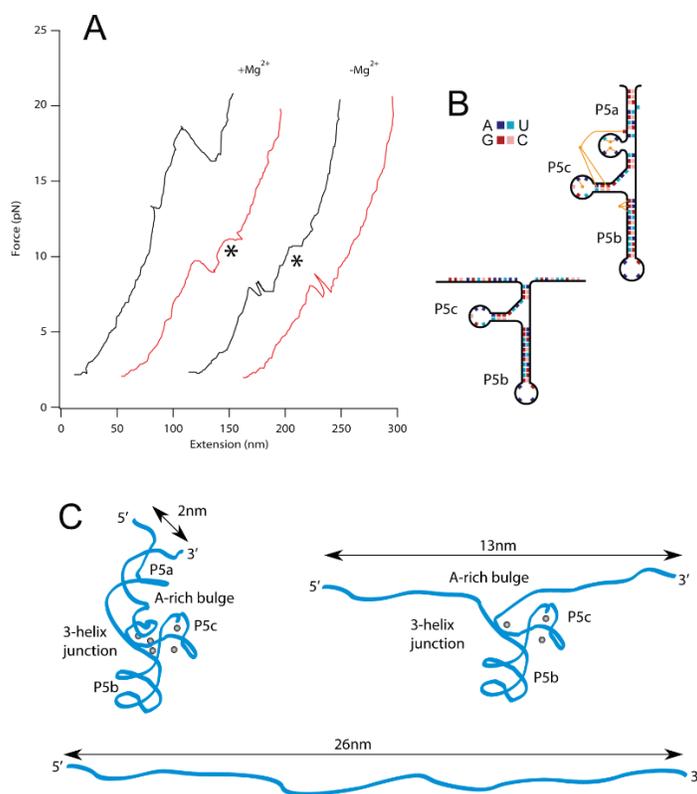


Figure 35: A classic experiment showcasing the higher unfolding forces of tertiary structure as opposed to secondary structure⁵³⁶. Panel A: Stems P5abc of the tetrahymena ribozyme have a higher unfolding force in the presence of Mg²⁺, which facilitates the formation of tertiary contacts (black, unfolding; red, refolding). Panel B: Sequence of P5abc with Mg²⁺ and without (lower). Panel C: Unfolding process starting at tertiary contacts in presence of Mg²⁺. Source:⁴⁸²

5. Novel Analysis Methods

5.1 State Assignment

Assigning states is a challenge in interpreting single-molecule force spectroscopy data. Here we present some techniques making this challenge easier. Examining a multi-state FEC, it consists of several branches separated by rips. Structurally, branches correspond to states whereas rips correspond to transitions between states. Detecting rips allows one to calculate unfolding force distributions, as well as determine the boundaries of branches.

5.1.1 Rip detection

One technique involves fitting branches in a semi-automated fashion; a program suggests the cursor location based on a rip detection schema. The rip detection schema finds outliers from the distribution of normalized $\bar{r}_i = \sqrt{(\Delta\bar{x}_i)^2 + (\Delta\bar{F}_i)^2}$, where $\Delta\bar{z}_i = \frac{\Delta z_i}{\max(z_i) - \min(z_i)}$ and Δz_i is the forward difference of z at i . The additional criterion of the line segment angle φ_i may be used for additional clarification, defined as $\varphi_i = \text{atan}\left(\frac{\Delta F_i}{\Delta x_i}\right)$ (Figure 36). One issue that arises is that rips can often travel through multiple points, resulting in multiple shorter-length values of \bar{r}_i . To obviate this issue, subsequent line segments \vec{r}_i and \vec{r}_{i+1} are joined if they are sufficiently aligned, such that $|\cos\theta_{i,i+1}| > \alpha$, where α is a threshold (usually 0.99), where $\cos\theta_{\{i,i+1\}} = \frac{\vec{r}_i \cdot \vec{r}_{i+1}}{\|\vec{r}_i\| * \|\vec{r}_{i+1}\|}$. For each criterion (\bar{r}_i, φ_i) , there is an associated distribution which may be thresholded to determine whether or not it is sufficiently anomalous to be classified as a rip. For \bar{r}_i , the threshold is a positive z-score of at least 3, for φ_i , the upper threshold is the first percentile (Figure 37).

With rips found, branches can be found between rips. The first branch is fit from the user-set start point of the FEC to the beginning of the first rip. Subsequent branches are fit from the end of the previous rip (point after start of previous rip) to the beginning of the next rip. The final branch is fit between the end of the last rip to the last point of the FEC, as specified by the user. With branches found spanning

point i to j , relevant information is recorded in a table by an automatic procedure: $i, j, x_i, x_j, F_i, F_j, W_i = \int_{x_0}^{x_i} F(x)dx, W_j = \int_{x_0}^{x_j} F(x)dx$, as well as all of the fitting parameters from a WLC fit to the branch. With state properties calculated, assignments can be made via finding the most likely state that this branch corresponds to. Classification may occur by a simple means, if we have N Gaussian populations $P_i(x) = A_i \exp[-\frac{(x - \mu_i)^2}{2\sigma_i^2}]$, with parameters A_i, μ_i , and σ_i representing the amplitude, mean and standard deviation of state i respectively, then the point x_j belongs to the population $k_j = \operatorname{argmax}_{i \in \{1, N\}} P_i(x_j)$. This technique may be used classify multiple observables, though we use it for contour length.

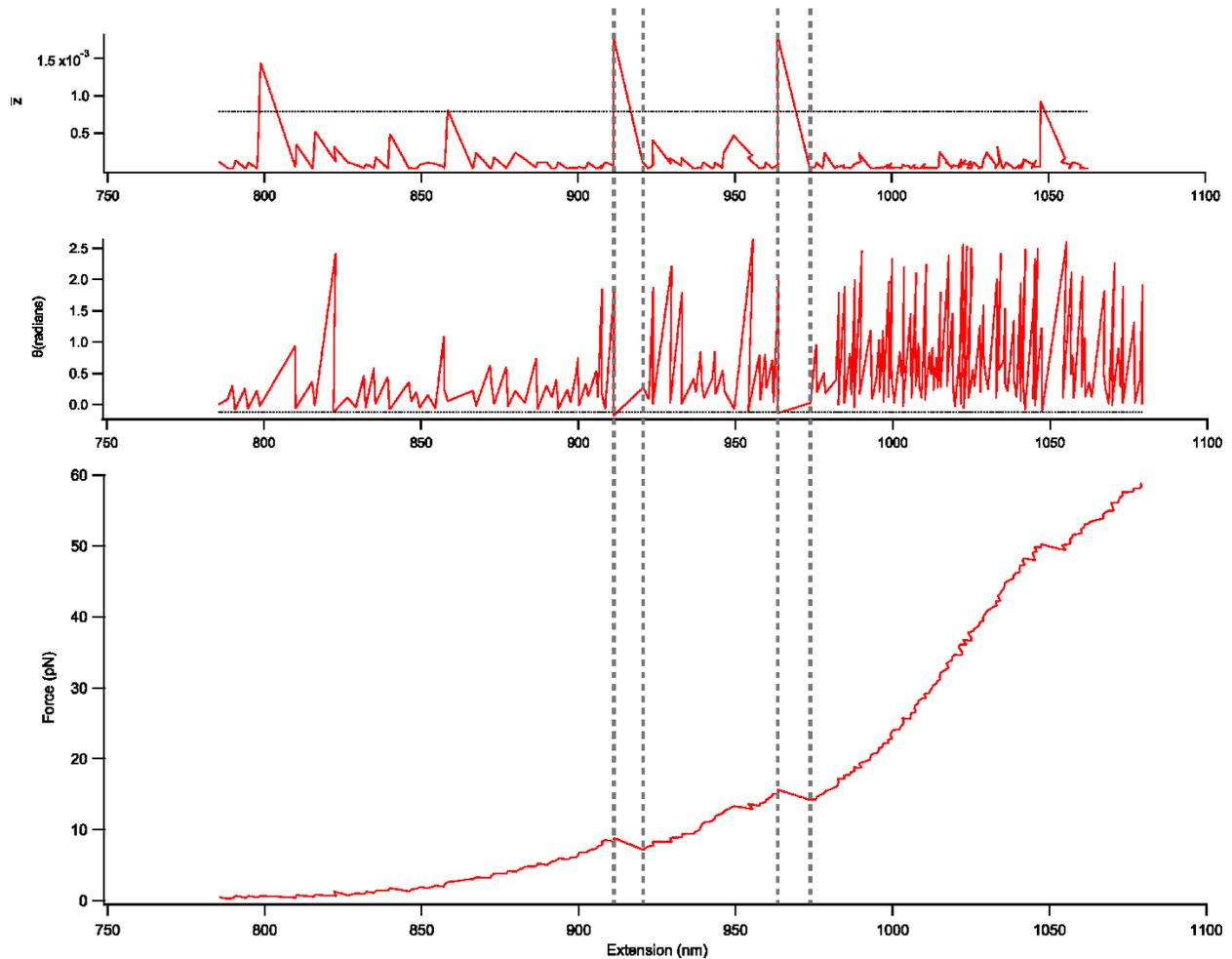


Figure 36: Rip detection schema. Rips are detected by finding anomalous values for feature scaled Euclidean distance of a rip (top) using a z-score cut-off of 3. An additional criterion is that the angle is

downward sufficiently, the upper cut-off being the 1st percentile for the distribution of angles $\varphi_i = \text{atan}\left(\frac{dy_i}{dx_i}\right)$ (middle). Shown below is the corresponding FEC, with dotted lines covering the rips. Care is taken to include the entire rip; for vectors $\vec{r}_i = (dx_i)\hat{x} + (dy_i)\hat{y}$, subsequent segments \vec{r}_i and \vec{r}_{i+1} are joined if $|\cos\theta_{i,i+1}| > \alpha$, where α is a threshold (usually $\alpha=0.99$) and $\cos\theta_{i,i+1} = \frac{\vec{r}_i \cdot \vec{r}_{i+1}}{\|\vec{r}_i\| \|\vec{r}_{i+1}\|}$.

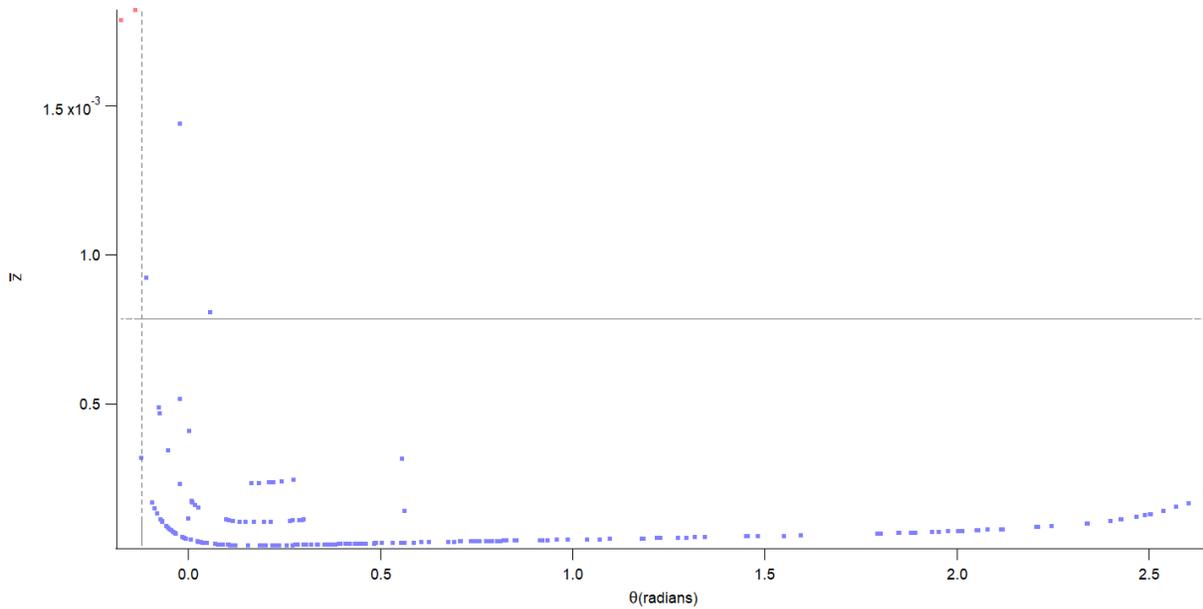


Figure 37: Decision boundary for rip detection. Angle values are shown on the x axis and \bar{z} values are shown on the y axis. The red points at the top left represent the rips detected using a z-score cut-off of 3 for the value of \bar{z} and a first percentile cutoff for the value of θ .

5.1.2 Moving window contour length fits

Another method involves taking a sliding window over the curve and computing the contour length in every window. This creates a subsequent curve of the contour length, which one can plot against force (Figure 39) or extension (Figure 38). For either, one sees a similar structure, plateaus with sudden steps. The plateaus correspond to motions *within* a state, whereas the steps, or, motions in contour length, correspond to transitions. This provides a picture of the transitions between states, and the collection of possible trajectories through distinct sequences of states. The contour length vs force graph provides

additional value in determining force ranges of high structural plasticity, by looking for forces where many distinct states (i.e. contour lengths) are possible.

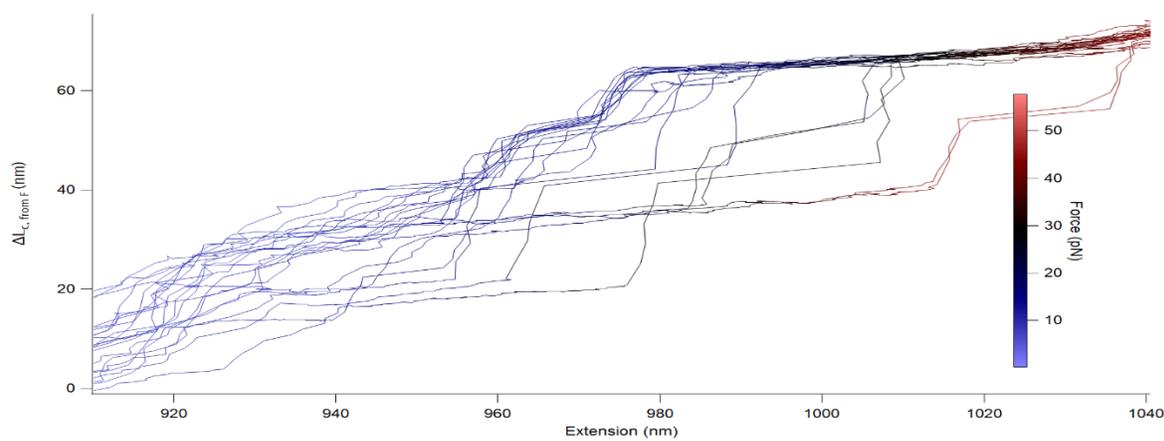


Figure 38: Trace of contour length vs extension for a collection of 20 FECs. The color scale shows the force at a given point. Multiple starting states are observed (note error in contour length increases precipitously at low force). Plateaus correspond to states. The kink character of the trajectories results from the sliding window fit to contour length responding slowly to a sudden force rip, whereas the extension changes immediately.

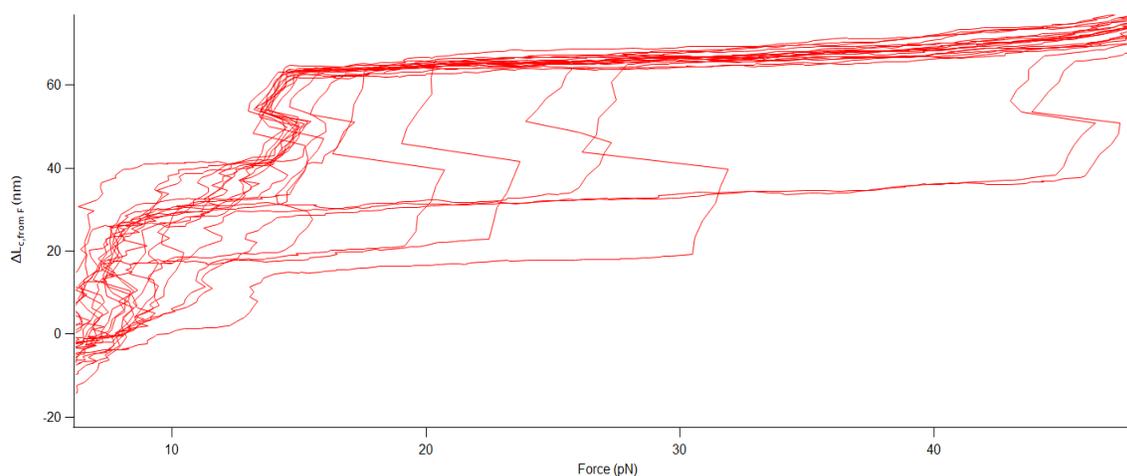


Figure 39: Trace of contour length vs force for a collection of 20 FECs. Horizontal motions correspond to motions within a state, whereas vertical motions correspond to transitions between states. Again, kinks

are due to the force changing more abruptly than the contour length, as the contour length is taken as a fit over multiple ($N=7$) points.

5.1.3 Whole trajectory categorization

Additionally, one may take a more holistic approach and consider the entire trajectory. For a given curve occupying n_s states, it undergoes $n_t=n_s-1$ transitions. As such, it is possible to calculate the probability of a given trajectory by multiplying the branch probabilities by the transition probabilities for a given trajectory through states. This approach allows one to use information about allowed transitions in order to make classifications, information that is otherwise discounted in other classification schemes. Furthermore, this classifies the entire trajectory, and is more robust than classifying individual states. The assigned sequence of states is given as the highest probability trajectory through n_s states. Probabilities for an individual state may be calculated in a model independent manner by examining the density of a particular state. This relies on previous state assignments.

For a state i consisting of N concatenated FEC coordinates (x_j, f_j) , $j=1, 2, \dots, N$, pixel densities can be calculated by determining the number of points within a given pixel, thereby creating a density map of state i . Doing this for all N_s states creates a corresponding density matrix $D_i(x, f)$ for each state. When classifying a new curve, it is pre-divided into branches using the semi-automated method mentioned above. For a branch k with N_b coordinates (x_l, f_l) , its density score for a given state $h_{i,k}$ is calculated as $h_{(i,k)} = \sum_{j=1}^{N_b} D(x_l, f_l)$. From this information alone, the probability P_s that a branch k is in a given state i is $P_s(i, k) = h_{i,k} / \sum_{m=1}^{N_s} h_{m,k}$. Transition probabilities $P_t(i_1, i_2)$ are taken as normalized transition frequencies. For a transition matrix T_{i_1, i_2} , where T_{i_1, i_2} represents the number of transitions from state i_1 to i_2 , the transition probability $P_t(i_1, i_2)$ is given by $P_t(i_1, i_2) = T_{i_1, i_2} / \sum_{m=1}^{N_s} T_{i_1, m}$. For a trajectory $B = \{k_1, k_2, \dots, k_t, \dots, k_{n_b}\}$ consisting of n_b branches total, the probability of a given assignment of states $A = \{i_1, i_2, \dots, i_t, \dots, i_{n_t}\}$ is

Equation 11: Pathway likelihood

$$P(A|B) = \left[\prod_{t=1}^{n_b} P_s(i_t, k_t) \right] \left[\prod_{t=1}^{n_b-1} P_t(i_t, i_{t+1}) \right] .$$

This can be used to find the maximum-likelihood pathway; an example is shown below (Figure 40).

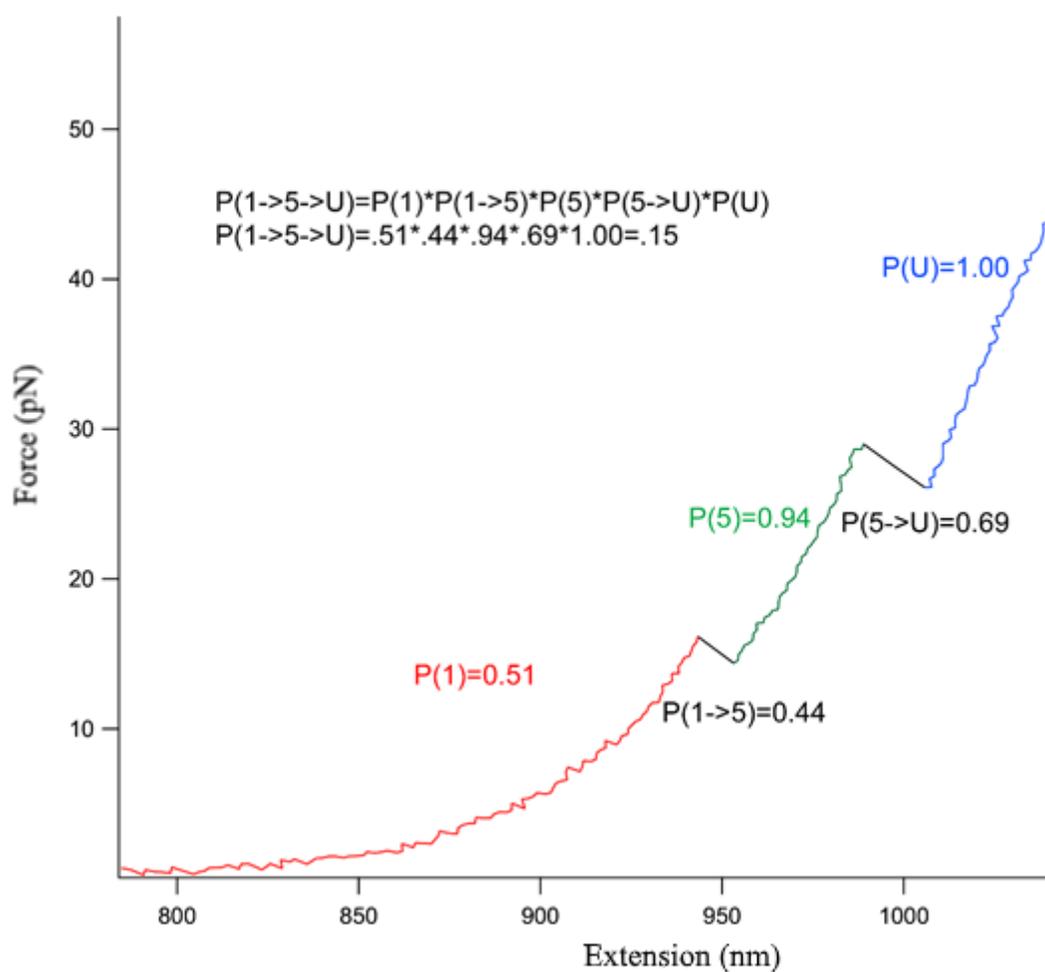


Figure 40: A maximum likelihood trajectory assignment of a force extension curve with three branches.

This technique works when adding data to an already populated dataset; it is especially useful in cases where there are disallowed transitions such that certain transition probabilities are zero.

6. Single Molecule Force Spectroscopy measurements of a high-efficiency frameshift stimulatory structure from West Nile virus

6.1 Introduction

In this chapter I describe measurements of the frameshift signal found in West Nile Virus (WNV) using optical tweezers. The results of this experiment are a structural assignment of the WNV frameshift signal, revealing high conformational heterogeneity and two distinct unfolding pathways.

6.2 Results

RNA containing the 111 nt downstream of the WNV slippery sequence, flanked on each side by kilobase-long “handle” sequences, was transcribed in vitro, annealed to ssDNA complementary to the handle regions, and attached to beads held in optical traps (Figure 41) as described previously²¹. The RNA was held near zero force for 3 s to permit folding, then the traps were separated at constant velocity to ramp up the force while measuring molecular extension, thereby generating force-extension curves (FECs). Unfolding FECs (Figure 42) typically displayed regions where the force rose nonlinearly with extension, representing the parts of the unfolding trajectory where the structure remained constant and the handles were stretched under tension, separated by “rips” in the curve where the extension increased abruptly concomitant with a drop in the force, indicating the unfolding of some part of the RNA structure. Notably, repeated unfolding of the same molecule revealed several different patterns of rips of different length in the FECs (Table 1), indicating the presence of a heterogeneous mixture of different conformational states in the mRNA.

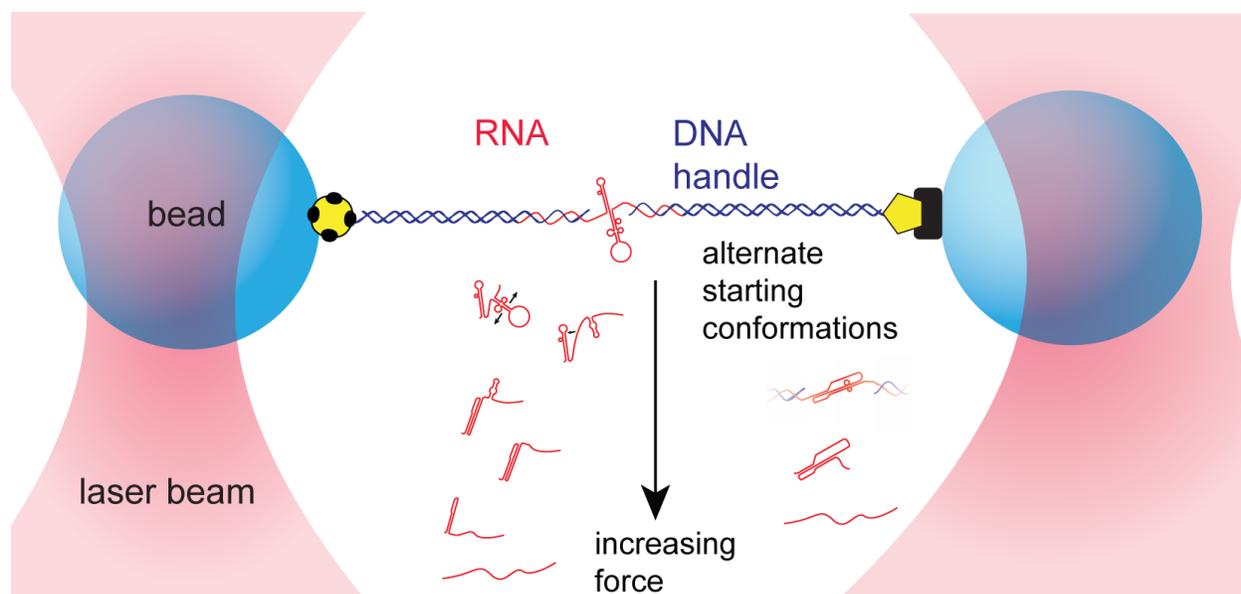


Figure 41: Single molecule force spectroscopy of the West Nile Virus frameshift signal. RNA construct is annealed to kb long DNA handles and attached to polystyrene beads via digoxigenin/anti-digoxigenin and streptavidin/biotin linkages. The construct adopts distinct starting conformations and goes through multiple intermediates while unfolding.

We characterized the structural transitions occurring in these FECs by fitting the curves to worm-like chain (WLC) polymer elasticity models before and after each rip (Equation 9 in Methods), using one WLC for the handles in series with a second WLC for the variable amount of unfolded RNA present in each conformation, in order to determine the contour length of unfolded RNA, L_c , before and after each transition (Figure 42

Figure 42). From this analysis, we identified at 7 conformational states having different lengths of unfolded RNA (Table 1), which would necessarily correspond to a minimum of 7 different structures in the frameshift signal. Some of these states (e.g. Figure 42

Figure 42, orange, dark blue) unfolded at the high forces characteristic of tertiary structures like pseudoknots¹⁴⁵, whereas the others unfolded in the range 10–20 pN more characteristic of simple

duplexes^{525,528}, suggesting the presence of two distinct kinds of pseudoknots as well as various helix and/or hairpin structures.

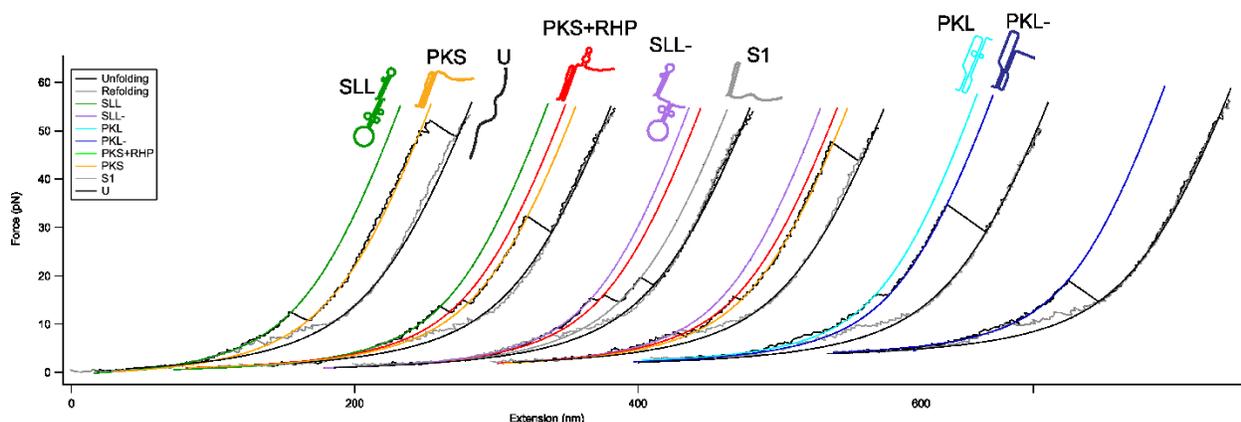


Figure 42: Representative force-extension curves (FECs) of the full-length West Nile Virus (WNV) NS1' frameshift signal (unfolding, black; refolding, grey). Worm-like chain fits show assigned structures^{139,153} (insets). Green. Stem-loop long (SLL/ state 1). Violet. Stem-Loop long partially unfolded (SLL-/ state 2). Cyan. Pseudoknot long (PKL/ state 3). Dark blue. Pseudoknot long, partially unfolded (PKL-/ state 4). Light Green. Pseudoknot short and residual hairpin (PKS+RHP/ state 5). Orange. Pseudoknot short (PKS/ state 6). Grey. Stem 1 (S1/ state 7). Black. Unfolded (U).

Observed state	Observed Lc (nm)	Proposed state	Expected Lc (nm)
1	60.7 ± 0.4	SLL	61.3
2	52 ± 1	SLL-	51.3
3	51 ± 1	PKL	51.5
4	48.6 ± 0.8	PKL-	48.9
5	40.3 ± 0.6	PKS+RHP	40.8
6	30 ± 1	PKS	29.8
7	21 ± 1	S1	20.4

Table 1: Observed and proposed states in wild-type data

To identify the structures corresponding to each state in the FECs, we matched the observed L_c values to the expectations for the pseudoknot structures proposed in the literature as well as the stem-loop structures predicted by mfold⁵³⁸. Seven structures were found to match the observed L_c values within experimental error (Table 1): the pseudoknots proposed by both Atkins et al.¹⁵³ and Dinman et al.¹³⁹ (respectively PKS and PKL), a partially folded version of PKL (PKL-), fully and partially folded versions of the double-hairpin proposed by Dinman et al. (respectively SLL and SLL-), a combination of short hairpin and PKS (PKS+RHP), and an extended version of stem 1 from PKS (S1) that is the same as one of the hairpins in SLL. With these preliminary assessments made, it is possible to define the transition map of the system, the set of transitions which occur between states, shown in Figure 43.

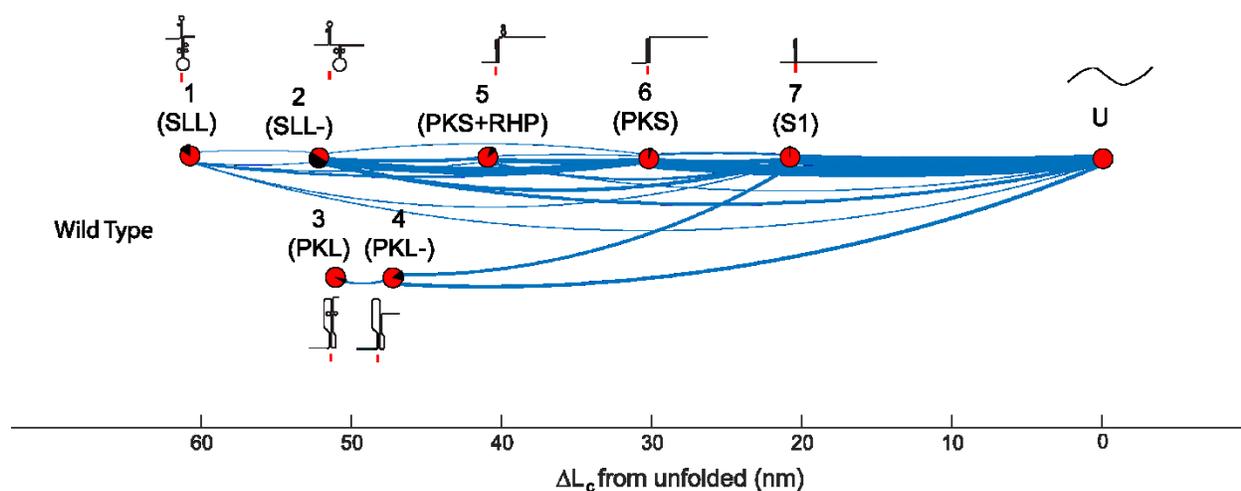


Figure 43: Transition map of WNV states during force ramp. Line thickness is proportional to number of transitions. Convex lines represent unfolding whereas concave lines are refolding events. Pie charts show proportion of curves starting in that state. Structure position shows predicted contour length whereas pie chart position shows observed contour length.

To confirm these assignments, we re-measured the FECs in the presence of anti-sense oligomers that bound to specific regions of the mRNA sequence (Figure 44). Binding of the anti-sense oligos to the mRNA would be expected to block the formation of the base pairs indicated by the shaded regions in Figure 44. Considering first the oligo that blocked formation of stem 2 in both PKS and PKL (Figure 44, green), we found that all unfolding events had forces in the range 10–20 pN (Figure 46), consistent with secondary structure only, confirming that the oligomer blocked formation of the pseudoknots. Indeed, not only were the pseudoknot structures PKS, PKL, and PKL- prevented from forming, but based on the L_c values from WLC fits to the different states in these FECs (Table 4), so were all the other states with the exception of S1. In their stead, various combinations of S1 with non-native helices that are energetically disfavored in the absence of anti-sense oligomer binding were observed, matching predictions from secondary-structure prediction tools^{538–540} (Table 2, Figure 45). A second anti-sense oligomer was used to block base-pairing at the 3' end of the RNA sequence, disrupting the SLL, SLL-, and PKL structures while permitting the formation of PKS, PKS+RHP, and PKL- structures (Figure 42, Figure 47).

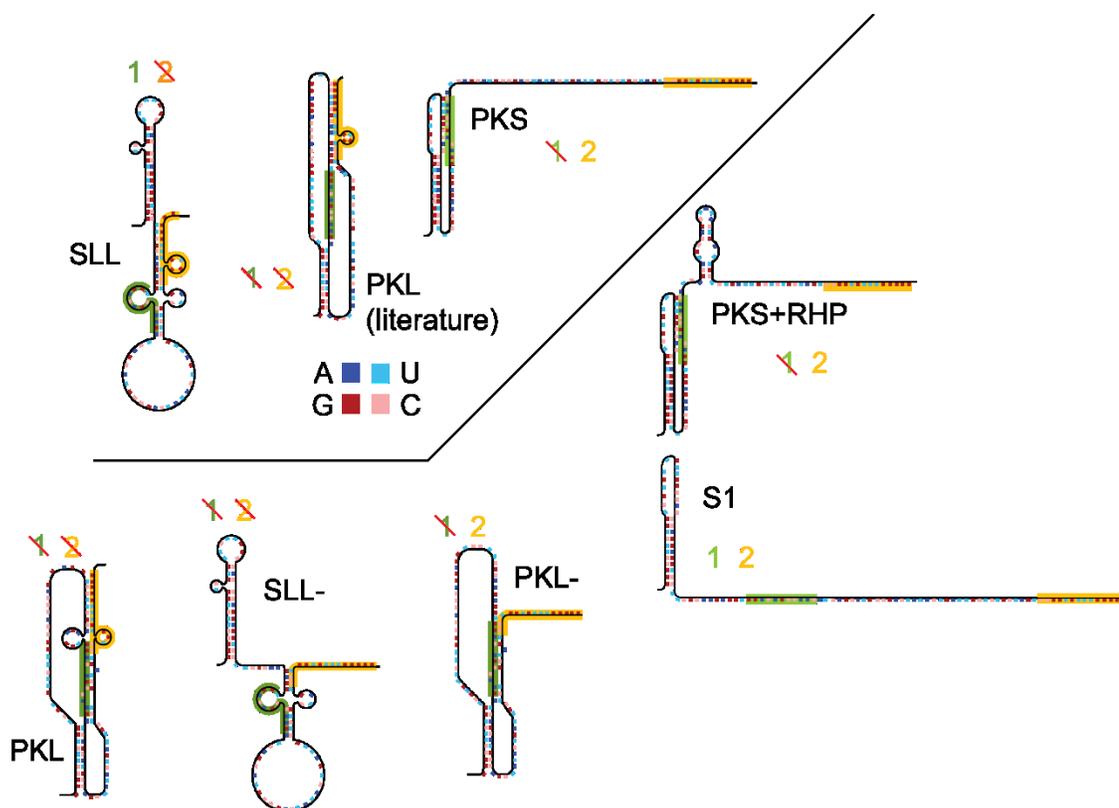


Figure 44: Disruption of structures by oligonucleotides. Shown are the proposed structures with the oligonucleotide binding locations shown in green (oligo 1) and orange (oligo 2). Structures disrupted by a given oligo have that oligo number crossed out with a red line. Oligo 1 was designed to disrupt the formation of the short pseudoknot¹⁵³, whereas oligo 2 was designed to interfere with the formation of the long pseudoknot¹³⁹. SLL is not disrupted by oligo 1, whereas SLL- is disrupted by both oligonucleotides. PKS is only disrupted by oligo 1. PKL is disrupted by both oligos, while PKL- is maintained in the presence of oligo 2, albeit in a slightly modified form, owing to the binding of the oligonucleotide to the 5' end. Structures in the top left correspond to predicted structures from literature (PKL is modified); while other structures are predicted from states in our experiment.

	Oligo Type								
		WT		O1		O2		Predicted	
Structure	SLL	60.7	0.4	61.3	0.2			61.3	
	SLL-	52	1					51.3	
	PKL	51	1					51.5	
	PKL-	48.6	0.8			47.1	0.3	48.9	47.6
	S1-4			49	1			49.5	
	S1-3			42	1			43.2	
	PKS+RHP	40.3	0.6			40.3	0.7	40.8	
	PKS	30	1			29.7	0.6	29.8	
	S1S2			32	1			31.4	
S1	21	1	19.5	0.5	20	1	20.4		
U	0		0		0		0		

Table 2: Table of all calculated contour lengths. Note: two values exist for predicted PKL-, corresponding to the WT and oligo 2 predictions respectively.

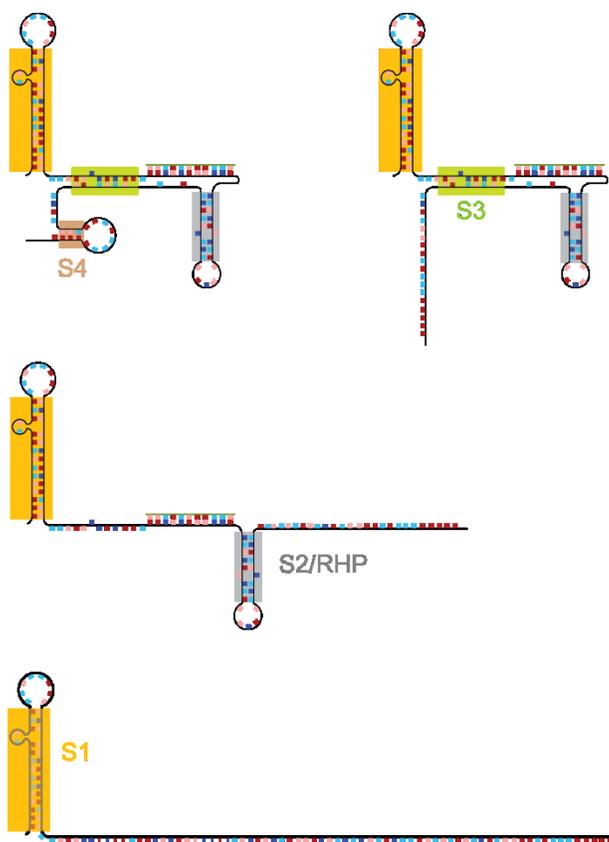


Figure 45: Novel Structures formed in presence of oligo 1

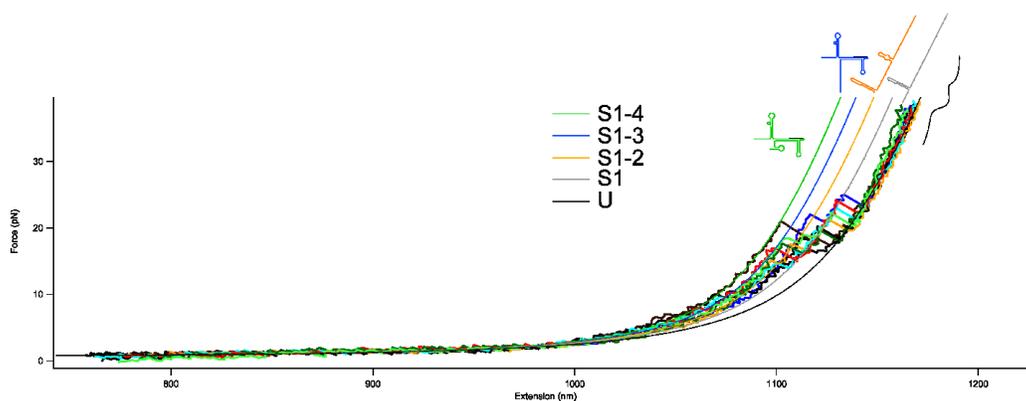


Figure 46: Force-extension curves in the presence of oligo 1. State 7/ S1 from the WT data is preserved.

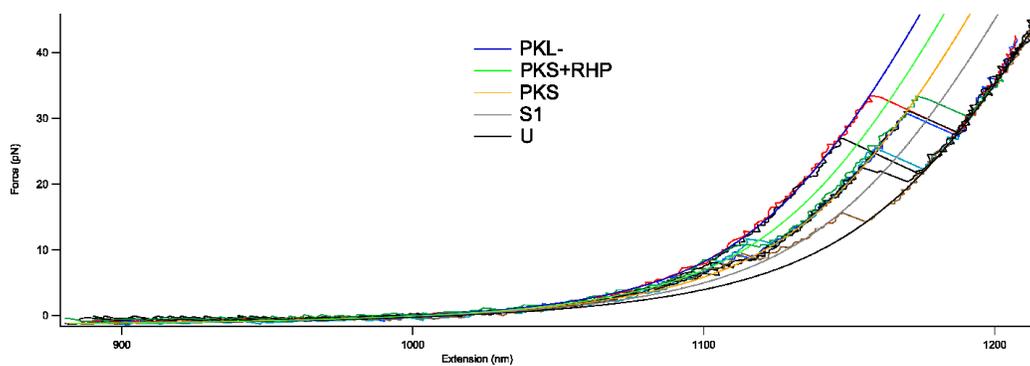


Figure 47: Force-extension curves in the presence of oligo 2. Oligo 2 maintains most of the structures in the wild-type sample, eliminating the three most folded WT states (SLL, SLL-, and PKL). States 4,5,6 and 7 (PKL-,PKS+RHP,PKS,S1 respectively) are preserved.

6.2.1 The WNV PRF signal forms the short pseudoknot predicted by Atkins and Firth

Examining the contour length data (Table 1), we see a strong correspondence between the predicted values for the short pseudoknot structure (PKS) to state 6, as well as the isolated 3' stem (S1) of that structure to state 7. Additionally, the unfolding force of PKS is high with large variance (mean \pm standard deviation: 24 ± 10 pN), as is expected for a pseudoknot, whereas the unfolding force for stem 1 is significantly lower with lower variance (17 ± 3 pN) (Table 4).

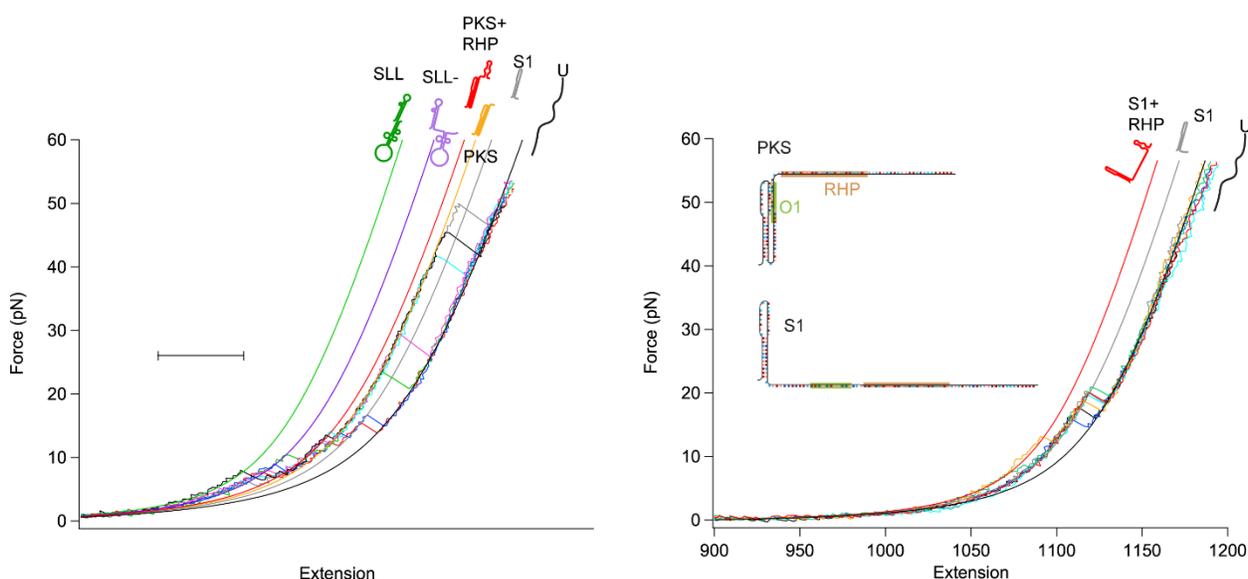


Figure 48: Oligo 1 abolishes PKS and maintains S1. Left. WT data, showing states 1,2,5,6 and 7 (SLL,SLL-,PKS+RHP,PKS and S1 respectively). Right. Oligo 1 abolishes PKS, maintaining S1/ state 7.

This pseudoknot also shows up in minimal energy predictions as the third and fourth optimal structures (-16.84 kcal/mol and -16.30 kcal/mol vs -18.05 kcal/mol minimum free energy; MFE) using Dirks and Pierce (DP09) parameters⁵⁴¹. Using Cao and Chen parameters⁵⁴¹ (CC06), this pseudoknot represents the two most optimal structures, suggesting it is energetically favored to form. Stem 1 is also robust in predictions and shows up in the majority of predicted structures for the whole sequence and becomes even more common when one restricts the sequence to the 5' end containing S1.

Further support for the assignment of state 6 to the short pseudoknot comes from examining transitions from state 6 (putative PKS state). The fact that state 6/PKS only transitions into state 7/S1 (F_{unfold} : 13 pN) or U (F_{unfold} : 29 pN) is what we would expect given this assignment (Figure 43, Table 3).

Additionally, oligo 1 data shows that oligo 1 abolishes state 6/PKS and maintains state 7/S1 (Table 4, Figure 48). This is consistent with our structural predictions whereby oligo 1 binds to the 3' stem of PKS, thereby disrupting it without disrupting S1 (Figure 44).

6.2.2 A short hairpin forms at the 3' end of the short pseudoknot

The residual hairpin structure (RHP) was postulated to solve a problem of a slightly more folded state (5 in Figure 43) than PKS often transitioning to PKS (6 in Figure 43) at low force (mean \pm SEM: 13.0 ± 0.4 pN). Structural predictions were made based on the contour length difference of 10 nm between state 6 (PKS) and state 5 (PKS+RHP) (Table 2). In addition, since this state persists in the presence of oligo 2, it cannot include the nucleotides bound by that oligo. MFOLD predictions produce it when the sequence is limited to between the 3' end of PKS and the 5' end of the oligo 2 binding site. Additionally, observance of back and forth transitions between state 5 and state 6 (PKS+RHP and PKS, respectively) in slow unfolding data (Figure 49, middle) suggests this. Furthermore, when pulled to low force (~ 20 pN) in order to maintain state 6/PKS for many pulls, the state 5 (PKS+RHP) forms consistently (Figure 49, right). This supports the notion that state 5 includes PKS, in addition to a short hairpin (RHP).

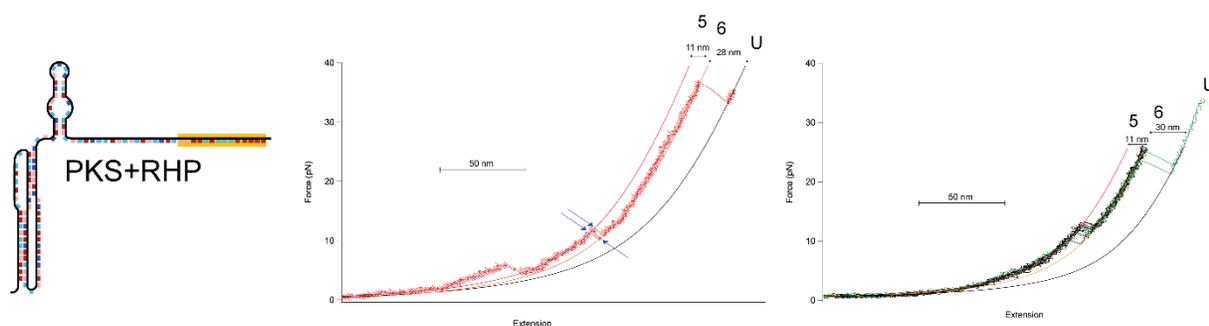


Figure 49: Support for the existence of a hairpin at the 3' end of PKS. Left. Proposed structure with oligo 2 shown in orange. State exists in presence of oligo 2 and consequently must be formed from bases at 3' end of PKS and 5' of oligo 2. Middle. Slow pulling data taken at 20nm/s reveals hairpin switching when PKS is formed. Arrows show switching events. Right. Pulling to lower force maintains PKS over many

pulls, and RHP forms consistently as other more folded structures along that pathway (SLL, SLL-) are disrupted.

Table 3: Average unfolding force for specific transitions in WT data.

		Unfolding force (pN)						
		To state						
		SLL-	PKL	PKL-	PKS+RHP	PKS	S1	U
From state	SLL	7±1			8±1	9±1	12±1	15±1
	SLL-				11±1	11±1	12±1	17±2
	PKL			12±1				
	PKL-						12±1	21±1
	PKS+RHP					11±1	15±2	19±3
	PKS						13±1	29±3
	S1							17±1

6.2.3 A set of double hairpins forms

Initial contour length measurements suggest that the set of dual stem-loops (stem-loop long; SLL) forms, as experimental measurements for state 1 are close (60.7 ± 0.4 nm) to the predicted 61.3 nm. This set of stem-loops consists of stem 1 (S1) as well as a weaker 3' stem loop. Stem 1 is slightly different in SLL vs in PKS, owing to Stem 2 of PKS base pairing with the loop of stem 1 to form the pseudoknot (Figure 44).

State 2 is predicted to be a slightly unfolded version of SLL, termed SLL-. This structure was predicted using MFOLD by using the measured contour length change (9 nm between SLL and SLL-) to determine the number of unfolded bases (~15-17 nt). SLL and SLL- both form in the WT data and are unfold at low force (SLL, 10.8 ± 0.3 pN; SLL-, 11.4 ± 0.2 pN) owing to the low stability of the 3' hairpin. Furthermore, when Oligo 2 is bound, states 1 and 2 (SLL and SLL-, respectively) are disrupted, as predicted, as oligo 2 binds the 3' end of the RNA construct disrupting the base pairs forming the 3' stem (Figure 44).

There exists transitions from state 1 and 2 (SLL and SLL-) to either state 5 or state 6 (PKS+RHP or PKS), a phenomenon that requires bases unpairing and repairing quickly (Figure 41 left inset, Figure

44). Stem 1 is preserved between the stem-loop and short pseudoknot structures, so all that is required is a pairing of the bases formerly sequestered in the 3' stem of SLL to pair with the loop of S1, thereby forming PKS (Figure 41). Other research has found a precedent for this phenomenon; a pseudoknot capable of spontaneous interchange with a set of double hairpins⁵⁴². We propose that RHP is a residual hairpin left from the conversion of SLL to PKS. While the literature structure for SLL does not include RHP, it is still quite likely to form within the structure and be maintained during conversion from SLL or SLL- to the Atkin's pseudoknot. Forcing RHP to form in the SLL structure entails a 4% sub optimality using MFOLD predictions (-39.1 vs -40.4 kcal/mol), compared with unconstrained; these predictions are shown in Figure 50. Taking the SLL structure with RHP and forcing bases at the stem origin to be unfolded, one gets the identity of an SLL- state (contour length difference of 8.9 nm from SLL) which is both disrupted by oligo 1 and oligo 2 (Figure 50).

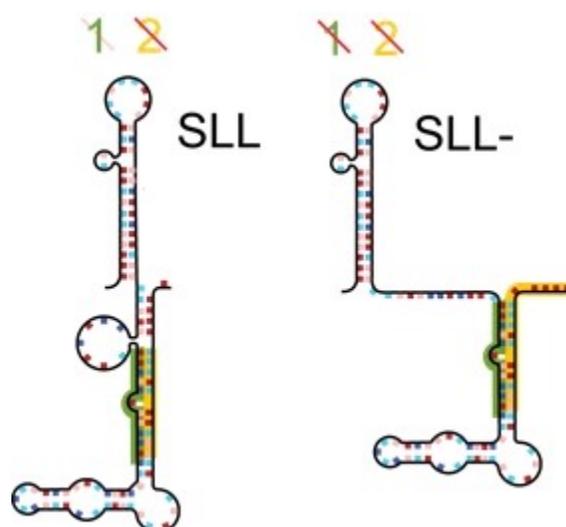


Figure 50: Potential modifications of SLL structures such that RHP is formed.

6.2.4 Evidence for the long pseudoknot (PKL) assignment

Going by contour length data, it appears that state 3 corresponds to the long pseudoknot (PKL). State 3 has an obligate transition into state 4 (putatively PKL-), a structure similar to PKL but with some 3' bases unfolded (Figure 44 PKL-). State 4 unfolds at high force into the unfolded state, or sometimes through S1 at low force, where S1 can form (Figure 5/E). The presence of state 4/PKL- is unexpected in oligo 2 data (Figure 5/F) given the literature structure (Figure 5/A), though modifications to the 3' stem of the long pseudoknot (Figure 5/B, predicted with MFOLD⁵³⁸) allow for a slightly 3' unfolded long pseudoknot state (PKL-) to form (Figure 5/C). The binding of oligo 2 is expected to maintain the PKL- state, with the slight change that some 3' bases of PKL- are unpaired (Figure 5/D). The unfolding of state 3 into state 4 (PKL into PKL-) occurs at low force (12 ± 1 pN, Table 3). The low stability of these contacts is consistent with the raw SHAPE data from Moomau et al. 2016^{139,543}. In this SHAPE data, the base pairs at the 3' end of the long pseudoknot are quite transient, as the bases remain relatively flexible, i.e. unpaired. Additionally, state 3 (PKL-) in the presence of oligo 2 is slightly (1.5nm) shorter than the PKL-WT state (Table 4). This is likely due to the binding of additional bases at the 3' end of PKL- (Figure 44D), thereby resulting in a lesser contour length change to unfolded (1.3 nm predicted).

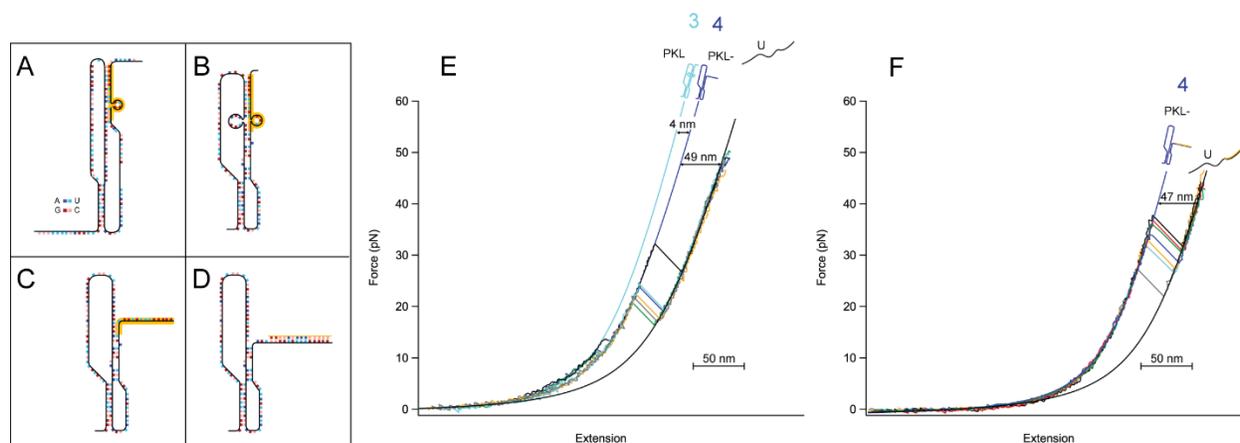


Figure 51: Some modifications to originally proposed structure. A. Original structure proposed by Moomau with oligo 2 binding site shown in orange. The binding of oligo 2 should disrupt the formation of the long pseudoknot, but a modified form (PKL-) is maintained (panel F). (B) Modified structure maintaining 5' stem of PKL, but with additional base pairing predicted via MFOLD. (C) Diagram of PKL- state, which forms as the 3' bases unpeel from the pseudoknot. (D). Structural prediction of PKL- state in the presence of oligo 2. (E) Force extension curves of the PKL pathway in WT data. (F). Force extension curves of the PKL- pathway in the presence of oligo 2.

Contour Length		Oligo Type							
		WT		O1		O2		MFOLD	
Structure	SLL	60.7	0.4	61.3	0.2			61.3	
	SLL-	52	1					51.3	
	PKL	51	1					51.5	
	PKL-	48.6	0.8			47.1	0.3	48.9	47.6
	PKS+RHP	40.3	0.6			40.3	0.7	40.8	
	PKS	30	1			29.7	0.6	29.8	
	S1	21	1	19.5	0.5	19.5	0.6	20.4	
	U	0		0		0		0	
Foces		WT		O1		O2		MFOLD(kcal/mol)	
Structure	SLL	10.8	0.3	10	1			-39.8	
	SLL-	11.4	0.2					-32.1	
	PKL	12	0.3						
	PKL-	21	1			19	2		
	PKS+RHP	13	0.4			12	1	-26.5	
	PKS	24	1			23	1	-22.2	
	S1	16.7	0.2	16.3	0.3	17	1	-19.9	
	U								

Table 4: Comparison of structure predictions and experimental measurements. Top. Experimental and predicted contour length changes. Bottom. Experimental average unfolding forces with energies calculated using MFOLD. Note, these should not be interpreted as correlated with unfolding force.

To conclude, we include in Table 4 the assigned structures, their experimental measurements and the corresponding predictions. WNV appears to form multiple distinct structures and adopt multiple distinct pathways. The stem-loop pathway (starts in SLL,SLL-,PKS+RHP,PKS,S1; Figure 52) is adopted 80±2% of the time, whereas the long pseudoknot pathway (starts in PKL,PKL-; Figure 53) is adopted the other 20±2% of the time.

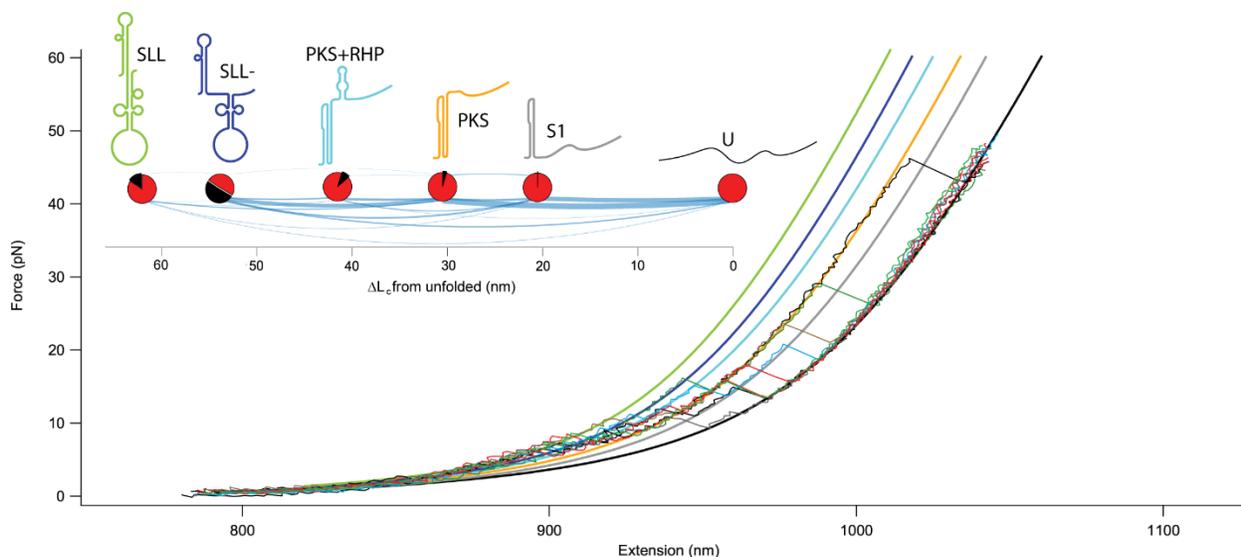


Figure 52: The stem loop pathway. Unfolding FECs for the SLL pathway of the WT construct with WLC fits. Inset. Unfolding pathway for the SLL pathway, with contour lengths shown as positions on the x-axis. States included: 1,2,5,6 and 7, corresponding to SLL, SLL-,PKS+RHP,PKS and S1 respectively.

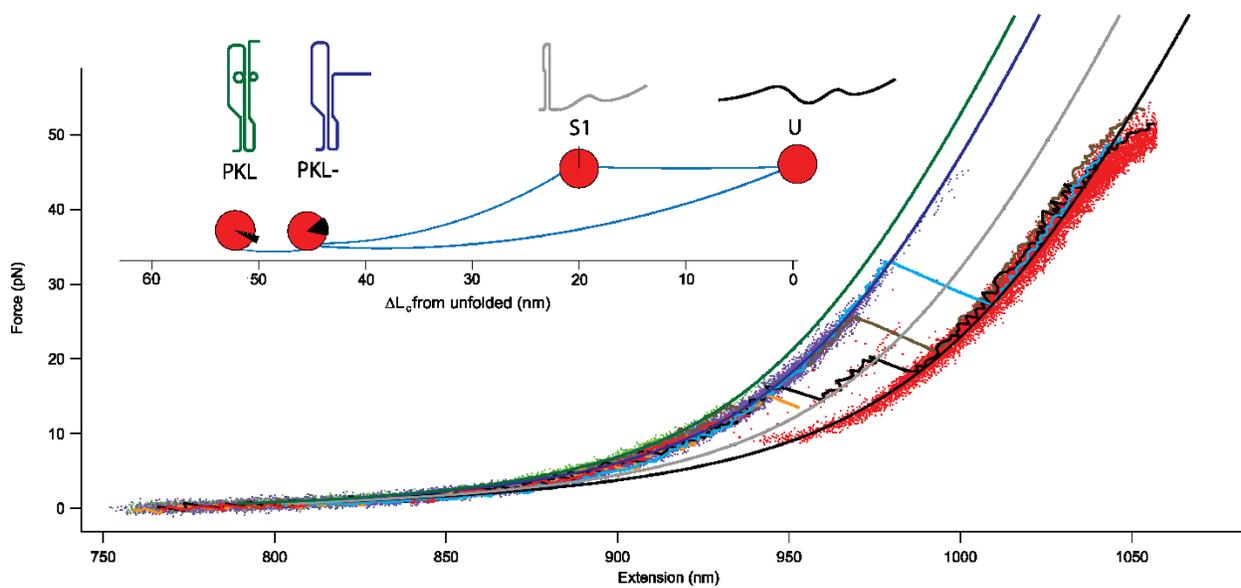


Figure 53: The long pseudoknot pathway. Unfolding FECs for the PKL pathway of the WT construct with WLC fits included. Inset shows unfolding pathway, diagrams show structures and match colour of corresponding WLC fits. States included: 3,4 and 7, corresponding to PKL, PKL- and S1 respectively.

6.2.5 The West Nile Virus frameshifting pseudoknot exhibits high heterogeneity in the force range of ribosomal stalling.

The WNV PRF signal exhibits multiple starting conformations (Figure 54) and exhibits high diversity of occupied states below 15 pN (Figure 55). Figure 55 shows the proportion of FECs in a given state at a given force, showing high state diversity below 20 pN. For comparison, the stalling force for the prokaryotic ribosome is 13 pN ¹⁶⁵, a potentially relevant force, given that PRF requires ribosomal pausing²⁶⁹.

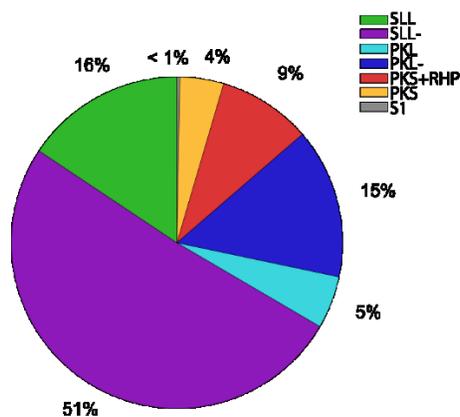


Figure 54: The West Nile virus frameshift signal is highly structurally plastic, exhibiting multiple starting conformations.

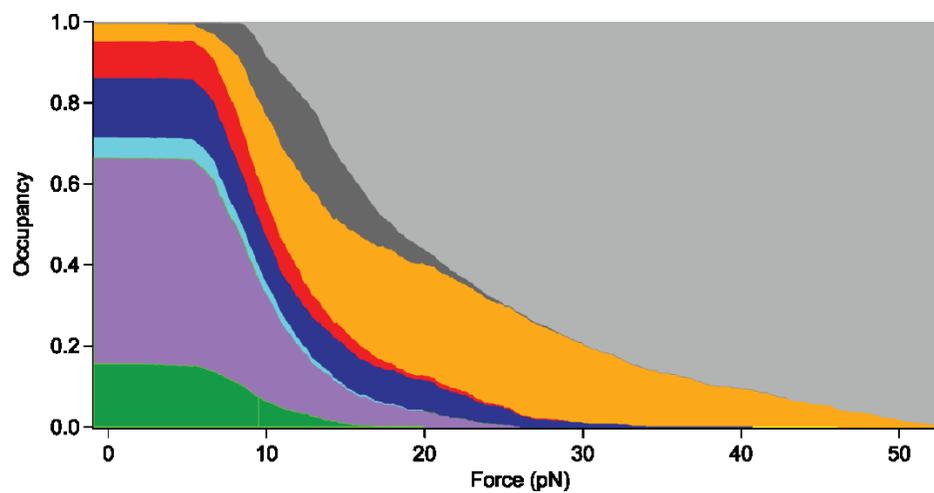


Figure 55: *State occupancy for the WNV PRF signal. Shown is the state occupancy as a function of force.*

State colours are the same as in Figure 54..

6.3 Discussion

We provide a structural characterization of the West Nile Virus frameshift signal. The West-Nile virus frameshifting pseudoknot adopts multiple starting conformations and multiple different pathways, presenting a strong example of functional heterogeneity. This study represents the first full characterization of the unfolding pathways of a very high efficiency frameshifting pseudoknot, exhibiting two distinct pathways and multiple distinct structures. The high conformational heterogeneity of the structure is likely related to its ability to induce frameshifting; it has been hypothesized that frameshifting emerges from variance in mRNA tension during ribosomal stalling²¹. Differences in mRNA tension, which can cause the ribosome to switch reading frame can emerge during ribosomal stalling by the stimulatory structure switching between different conformations. If there are multiple distinct structures that may form in the force range of ribosomal stalling (~13 pN, Figure 55)¹⁶⁵, and these structures are demonstrated to transition into one another (Figure 43), it is highly possible that conformational switching occurs under applied tension by the ribosome. This switching may cause fluctuations in mRNA tension which are great enough to switch the ribosomal reading frame. One strong consideration would be if ribosomal unfolding differs from unfolding by optical tweezers. Ribosomal unfolding would involve interactions between the 5' end of the stimulatory structure and the ribosomal entry tunnel²⁷⁵, and this may bias the formation of certain states.

A natural avenue for future work would be to look at other high efficiency frameshifting pseudoknots; if they too exhibit high conformational heterogeneity, it would provide support for the hypothesis that frameshifting is driven by conformational heterogeneity via fluctuations in mRNA tension during attempted translocation by the ribosome.

7. An improved metric for correlating structural heterogeneity of frameshift signals with PRF efficiency

7.1 Introduction

Few attempts have been made to operationalize a quantitative measure of the conformational plasticity of a molecular system. One way, from Ritchie's 2012 study of frameshifting pseudoknots, put forth the percent alternately folded as a measure²¹. This *ad hoc* definition, while capturing the relationship between conformational plasticity and frameshift efficiency well for the data²¹, still presents challenges. The first main challenge is the fact that the basis for choosing one structure as the native structure may not be well founded; one gets radically different values depending on one's choice of native state. In the study mentioned, the native state was defined as the state with the greatest contour length change from unfolded, i.e. the most folded state. This is somewhat arbitrary, and this *ad hoc* definition does not correspond well to factors explaining frameshifting. One proposed explanation for the link between conformational plasticity and frameshifting is that the stimulatory structure is able to adopt multiple different conformations when partially unfolded during ribosomal stalling, thereby causing fluctuations in the tension of the mRNA being translated. These fluctuations may cause the ribosome to switch reading frame, such that subsequent translation proceeds in the -1 frame.

The other main challenge emerges when one studies more complex systems, where more than two starting states are present. The current definition does not capture the extra conformational plasticity reflected in having multiple alternate starting states, nor does it capture the competition between multiple structures, it merely reports the percent alternately folded.

Another issue arises from the use of the percent alternatively folded as a metric of conformational plasticity. Predicting frameshift efficiency using this metric brings one to a maximum predicted frameshift efficiency of approximately 30% (Figure 13), at which the percent alternatively folded nears 100% and cannot go any higher. There is not a physical basis to this maximum value, and some pseudoknots frameshift at levels beyond that¹³⁹.

7.2 Results

Borrowing the concept of Shannon entropy⁵⁴⁴ from information theory, we use this to create a quantitative metric for the diversity of states of a system. The Shannon entropy is defined as

Equation 13: Shannon entropy

$$H'(F) = - \sum_{i=1}^N p_i(F) \ln p_i(F),$$

where $p_i(F)$ is the proportion of the total population in subpopulation i at force F . This definition obviates the need to assign a native state and is also robust to the presence of multiple states. Additionally, metrics of conformational plasticity have not previously looked at force dependence. With the state proportions known for each force, we may calculate the Shannon entropy at each force level. This definition is applied to West Nile virus data in Figure 56.

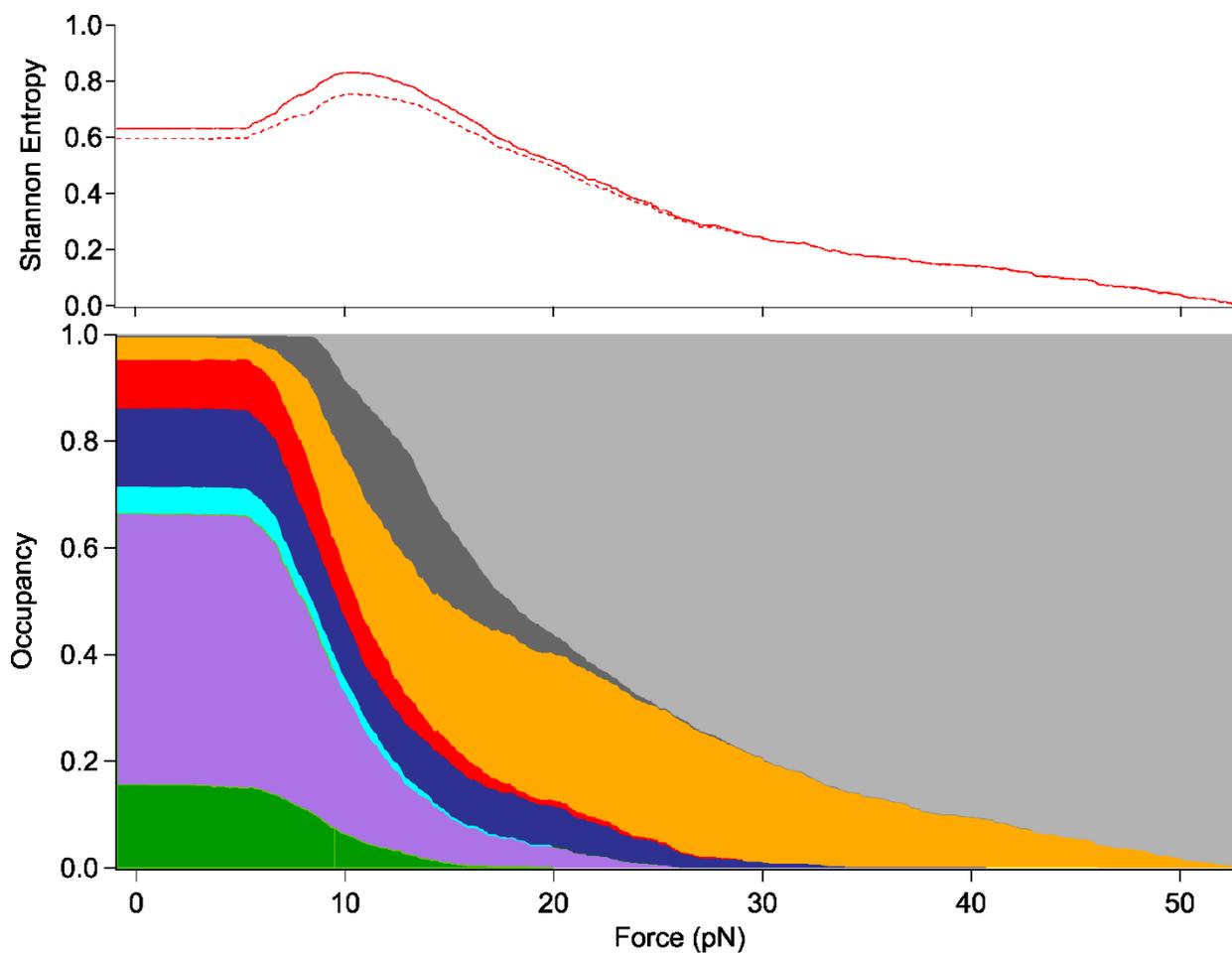


Figure 56: Force-dependent conformational plasticity for WNV PRF signal. (Top) Shannon entropy as a function of force derived from state occupancy as a function of force (Bottom). The dotted line represents the Shannon entropy in the case that PKS and PKS+RHP are joined, as both states may be functionally similar as their 5' ends (end closest to the ribosome) are identical.

This study includes PRF stimulating pseudoknots from West-Nile virus, with measurements taken in the previous chapter. Additionally, data from other pseudoknots taken from Ritchie *et al.*²¹ is used: bacteriophage T2 gene 32 (PT2G32)⁵⁴⁵, sugar cane yellow leaf (Sc YLV) and a mutated version (C27A)^{146,147}, pea enation mosaic virus (PEMV1)⁵⁴⁶, simian retrovirus (SRV1)⁵⁴⁷, severe acute respiratory syndrome coronavirus (SARS CoV)²⁵⁰, human endogenous retrovirus (HERV)⁵⁴⁵, mouse mammary tumor virus (MMTV)^{548,549}, and Visna–Maedi retrovirus (VMV)⁵⁵⁰.

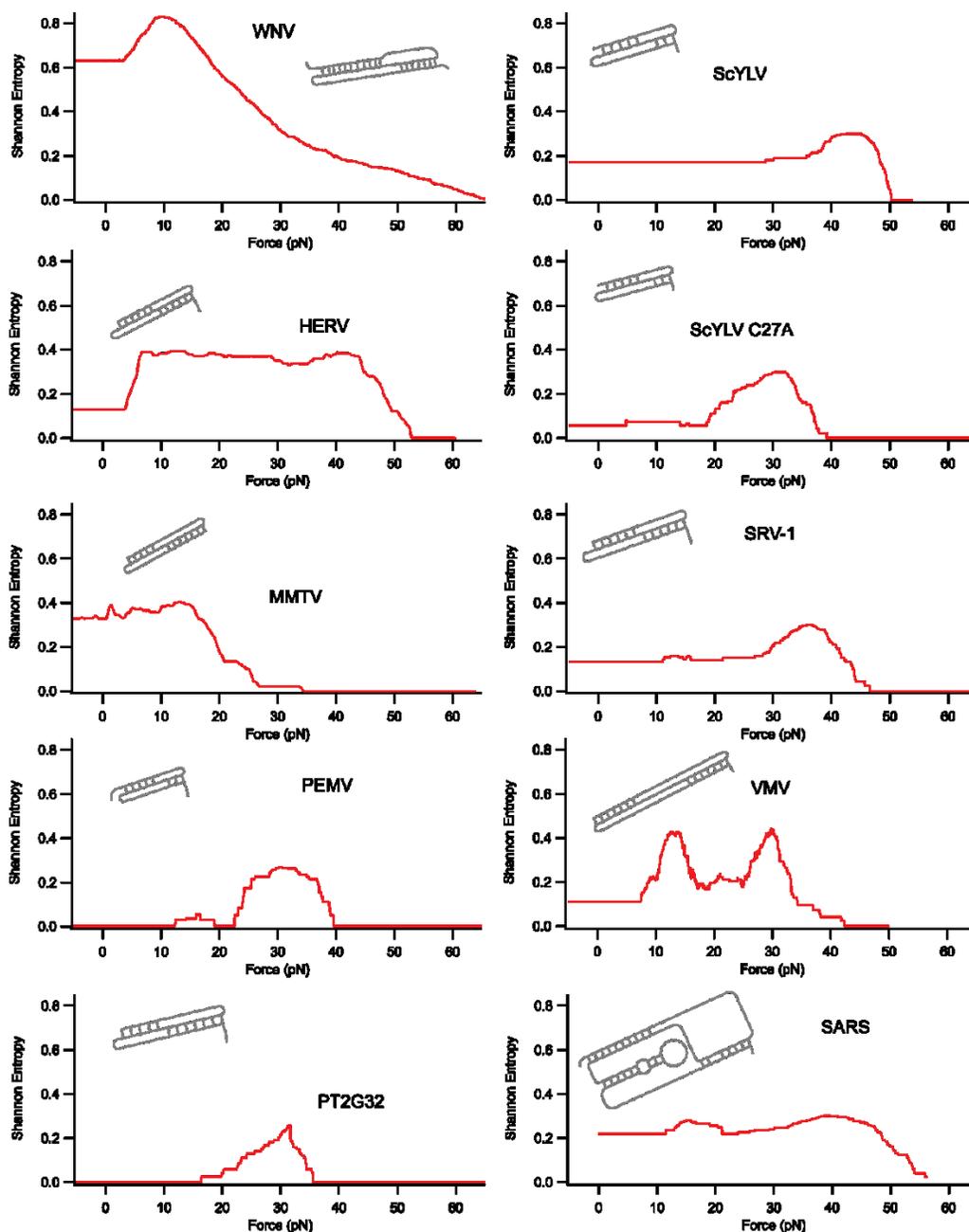


Figure 57: Force dependent entropy for 10 frameshifting pseudoknots.

The force-dependent Shannon entropy for 10 frameshifting pseudoknots is calculated and shown in Figure 57. We demonstrate moderate correlation ($R^2=0.81$) between the Shannon entropy observed at zero force and frameshift efficiency (Figure 58). This value is slightly better than the R^2 value of 0.77 calculated from the Ritchie 2012 paper using their method of quantifying conformational plasticity (Figure 13), which also uses the proportion of starting states (i.e. zero force). The more

physically relevant force range is the average between 11 and 15 pN, as the ribosomal stalling force is 13 pN¹⁶⁵. The Shannon entropy value averaged in this region produces a better correlation ($R^2=0.86$, Figure 59).

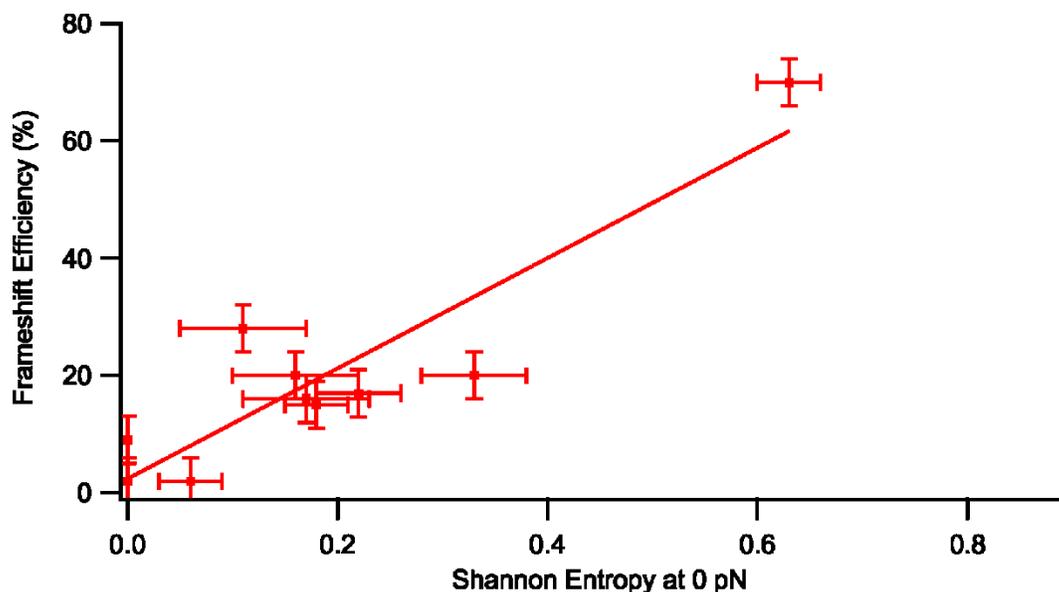


Figure 58: Relationship between zero force SE and FSE. Pearson's $R^2=0.81$. Errors in in SE are calculated as $\delta H = -\sum_{i=1}^N \delta p_i (\ln(p_i) + 1)$, where $\delta p_i = \sqrt{p_i(1-p_i)}/\sqrt{N_c}$, where N_c is the number of FECs. Error in frameshift efficiency (FSE) are put as 4%, as in ²¹

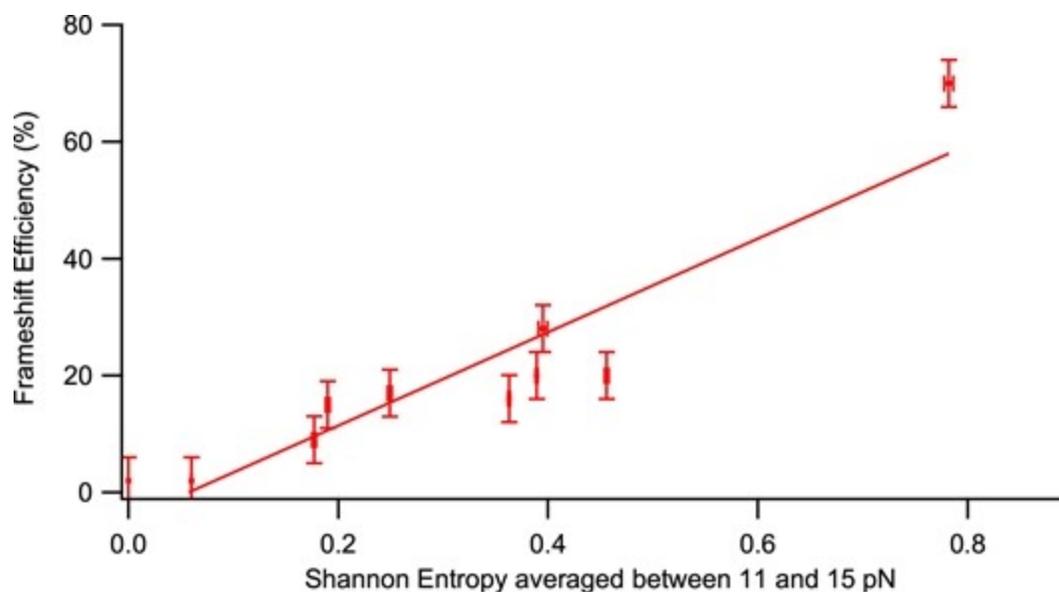


Figure 59: Relationship between SE averaged in the range of 11-15 pN and FSE. Errors in SE show standard error of mean for values for SE in force range 11-15 pN, the stalling force range of the ribosome¹⁶⁵. Pearson's $R^2=0.86$

7.3 Discussion

We propose a novel measure of conformational plasticity which may obviate some of the problems associated with the prior metric of percent alternately folded. These problems are the dependence on the choice of native state and the inability to account for multiple possible starting states. The use of the Shannon entropy as a metric for conformational plasticity obviates these problems, with the additional benefit of allowing a force-dependent metric of conformational plasticity, allowing one to look at force ranges relevant to models of frameshifting, such as the ribosomal stalling force of 13 pN¹⁶⁵. Hypotheses suggest that PRF emerges from fluctuations in mRNA tension during ribosomal stalling due to the stimulatory structure adopting multiple different structures²¹. The entropy-based metric may capture competition between states better; it reflects the diversity of states available (for the molecule to form) at a given force range.

We observe a correlation between the value of Shannon entropy in the force range of 11-15pN and the frameshifting efficiency, accounting for 86% of the variance in FS efficiency. The zero-force value for Shannon entropy correlates with frameshift efficiency with an R^2 value of 0.81. Together, these results support the hypothesis that frameshifting efficiency is a phenomenon associated with conformational plasticity, a hypothesis with growing empirical support.

A possible concern is that these state assignments are based off of single molecule measurements in solution, and not unfolding by the ribosome. It is possible that unfolding by the ribosome biases the adoption of certain states and thus changes both the state occupancies and the Shannon entropy in that force range. One possible approach to this would be to simulate ribosomal unfolding using nanopores, which can mimic the tension on the mRNA in the ribosomal entry tunnel. This approach of using nanopores may better approximate the actual diversity of structures encountered during ribosomal stalling.

It may be possible to use structural assignments to predict the distribution of tension in the mRNA during ribosomal stalling. The level of tension for a given structure may possibly be predicted by the distance from the bound tRNA in the slippery sequence to the part of the stimulatory structure blocking the ribosomal entry tunnel (Figure 12, ribosome not shown). This, along with the known state proportions, can produce a distribution of tension in the mRNA. This may be a better predictor of frameshift efficiency if the cause of PRF is tension fluctuations in mRNA. With this analysis, it is possible that the few bases at the 5' end of the stimulatory structure are most important for frameshifting. One issue with the analysis presented in this chapter is that all states are treated as equal, and some states may not be functionally different from others. For example, in the WNV PRF signal, the PKS state differs from another state (PKS+RHP) only by the formation of a hairpin at the 3' end of the pseudoknot (Figure 44). While these are counted as distinct states, they likely do not have largely different effects on mRNA tension. Regardless, the effect of merging these two states is minor, shown as a dotted line in Figure 56.

8. Conclusion

We present the mechanical characterization of a high conformational plasticity pseudoknot within West Nile virus (WNV). Structures observed from single molecule force spectroscopy measurements match structures previously predicted using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) analysis as well as bioinformatic predictions. The pseudoknot exhibits high structural plasticity, unfolding along distinct pathways, and occupying many distinct states.

Furthermore, we present a new measure of conformational plasticity, which corrects some of the previously held weaknesses of prior definitions. West Nile virus is highly structurally plastic by this definition, and the new definition allows its structural plasticity to be easily quantified, skirting the earlier issues of the prior definition. Furthermore, there exists a correlation between the conformational plasticity of frameshifting pseudoknots at this peak and the pseudoknot frameshift efficiency.

The analysis of WNV presents a structural assignment for a high efficiency frameshifting pseudoknot, which can be useful for determining the structural basis of PRF. Several frameshifting pseudoknots have been previously been characterized in the literature, but in WNV we see an even greater variety of states. Merely from a structural perspective, WNV presents an interesting case study, highlighting potential transitions and characterizing another highly plastic pseudoknot.

The analysis clarifying the link between conformational plasticity and frameshift efficiency presents a new method of quantifying structural plasticity, borrowing concepts from information theory. Furthermore, it produces a metric which can possibly be used to predict frameshifting efficiency (though this would be difficult from sequence information alone). Additionally, it suggests a mechanism relating the conformational plasticity at the ribosomal stalling force (13pN) to the frameshift efficiency, suggesting a role for structural transitions within this force range.

8.1 Future Work

Further study would include the characterization of more high efficiency frameshifting pseudoknots, seeing if they exhibit structural heterogeneity. If this is in fact a commonality between high efficiency frameshift stimulatory structures, this knowledge could potentially feed back into bioinformatics techniques to predict frameshifting given sequence or structural data. This could potentially expand the repertoire of known frameshift stimulators.

Examining the motions of ribosomes during PRF is a promising approach^{152,315}. It is highly likely that the dynamics of the single molecule are inadequate at describing the mechanism of PRF, as unfolding via optical tweezers is likely different than unfolding by the ribosome. Unfolding the RNA construct through a nanopore, may provide a better simulation of ribosomal unfolding, while providing a simpler assay than ribosomal translocation work, as in³¹⁵.

Phylogenetic analysis may be a novel approach to studying frameshifting. It may be possible to reconstruct the time evolution of a given PRF signal via phylogenetic tree reconstruction⁵⁵¹⁻⁵⁵³ and ancestral sequence reconstruction (ASR)^{554,555}. Additionally, for the case of insect-borne flaviviruses of the JEV serogroup (of which WNV is a member), there appears to be distinct divergence events between functional PRF and non-functional PRF⁵⁵. Such events can be studied using phylogenetic software⁵⁵⁶ to determine a minimal change of sequence which deactivates a given PRF site and yield insights into the structural and sequence signature of PRF.

Furthermore, one may examine the impact of binding ligands on the conformational plasticity of the frameshift stimulatory structure. This has been applied in the case of the SARS coronavirus, where an anti-frameshifting ligand⁸¹ lowers the conformational plasticity and subsequently the frameshift efficiency⁸³. An example highly related to WNV is the application of flavonol and isoflavonoid compounds to treat Japanese encephalitic virus in cells⁷⁷, which achieve their action through binding the frameshift site, a sequence highly conserved between JEV and WNV¹⁵³. These compounds may potentially produce similar anti-viral effects for the highly related members of the JEV serogroup, of

which WNV is a member⁵⁵. The WNV PRF signal possesses high sequence similarity ($56/61=92\%$ ⁵⁵⁷) to JEV. This similarity may be useful as there exist two natural compounds, Kaempferol and Daidzin which have been shown to inhibit JEV pathology, likely via binding to the frameshift site⁷⁷. Kaempferol is a flavonol present in high levels in black and green tea, and low-to-moderate levels in broccoli, apples, leeks and kale⁵⁵⁸⁻⁵⁶⁰. Some preliminary experiments have shown anti-viral activity in cell culture against influenza⁵⁶¹ and hepatitis B viruses⁵⁶². Daidzin is an isoflavonoid found in soy foods and kudzu vines⁵⁶³, commonly used for its effect at blunting blood alcohol spikes⁵⁶⁴. In addition, both have antioxidant effects^{560,565-567}. Both compounds have been observed to inhibit JEV infectivity *in vitro*⁷⁷. Similar compounds baicalein and quercetin also display anti-JEV activity⁵⁶⁸ (Figure 60), though their mechanism has not been studied, unlike daidzin and kaempferol⁷⁷.

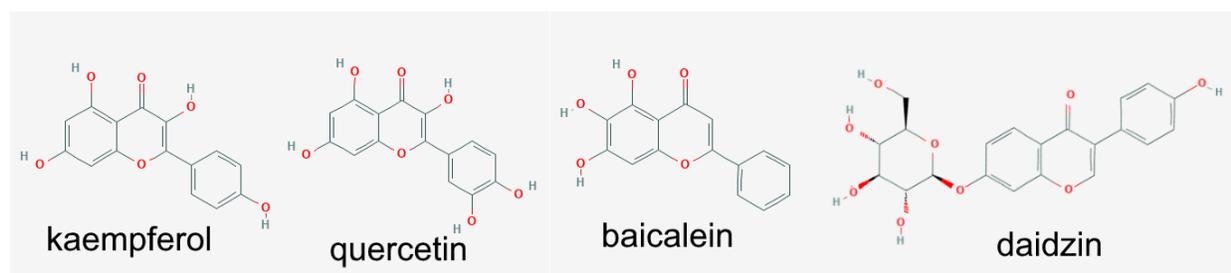


Figure 60: Flavonoid compounds exhibiting anti-Japanese encephalitic virus effects. Source: PubChem (<https://pubchem.ncbi.nlm.nih.gov>)

Additionally, both compounds were shown to bind to the frameshift site using electrospray ionization mass spectrometry (ESI-MS)⁵⁶⁹, isothermal titration calorimetry (ITC)⁵⁷⁰, and docking simulations⁵⁷¹. Due to the similarity of the frameshift signals of JEV and WNV, it is possible that the same compounds bind to the WNV frameshift signal and disrupts frameshifting, much as a novel ligand binds to the SARS pseudoknot⁸¹, reducing the conformational plasticity of the pseudoknot⁸³. In fact, the computationally predicted binding site for JEV (Figure 61) is completely conserved between JEV and WNV, along with the 11 nt 5' of the site and the 12 nt 3' of the site (Figure 62). Therefore, given the anti-JEV properties of Kaempferol and Daidzin, and the target being similar between JEV and WNV, it is

possible that these compounds could potentially work as anti-viral drugs for WNV. Additionally, they may have effects on the wider related JEV serogroup, consisting of flaviviruses closely related to JEV (including WNV) possessing homologous PRF sites^{55,153}.

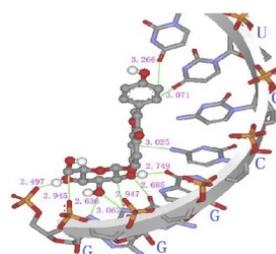


Figure 61: Binding site of JEV shown with daidzin bound. Source:⁷⁷ (CC BY)

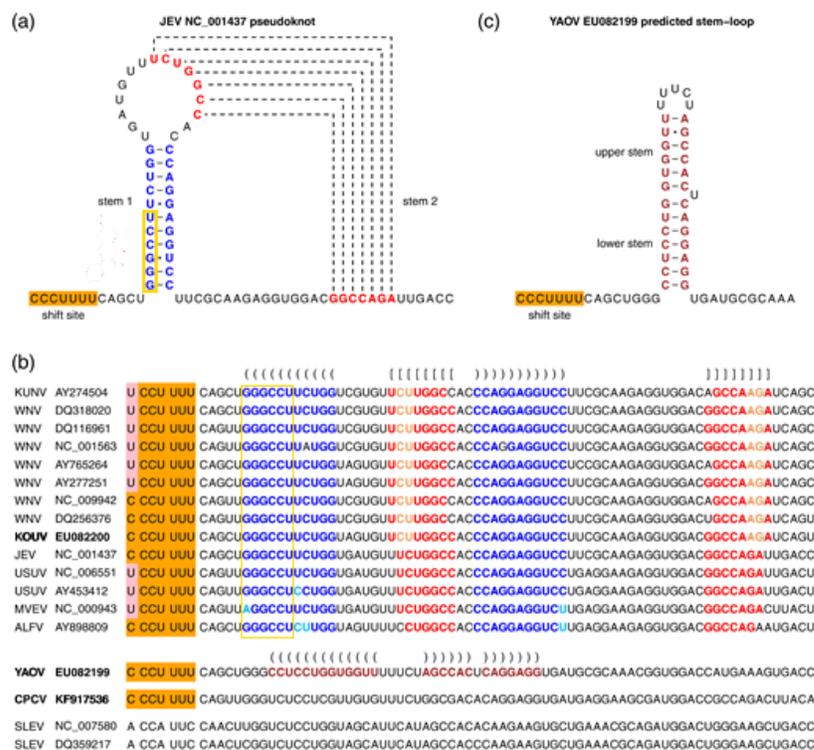


Figure 62: Frameshift stimulatory sequence in the Japanese Encephalitic virus (JEV) serogroup, which includes West Nile virus (WNV). Note that the sequences shown consist of the slippery site (orange), a five nucleotide spacer region (sequence CAGYU, note $Y=\{C,U\}$ in IUPAC notation⁵⁷²), and a 61nt stimulatory structure. Stem 1 is highlighted in blue and Stem 2 is highlighted in red. Light blue and orange bases show substitutions that preserve base pairing. The orange rectangle shows the JEV binding site for Kaempferol and Daidzin. Source: ⁵⁵(CC BY)

9. Literature Cited

1. Beadle, G. W. & Tatum, E. L. Genetic Control of Biochemical Reactions in Neurospora. *Proc. Natl. Acad. Sci.* **27**, 499–506 (1941).
2. WANG, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomed. Rep.* **3**, 152–158 (2015).
3. Dinman, J. D. Mechanisms and Implications of Programmed Translational Frameshifting. *Wiley Interdiscip. Rev. RNA* **3**, 661–673 (2012).
4. Dinman, J. D. Control of gene expression by translational recoding. *Adv. Protein Chem. Struct. Biol.* **86**, 129–149 (2012).
5. Gesteland, R. F. & Atkins, J. F. Recoding: Dynamic Reprogramming of Translation. *Annu. Rev. Biochem.* **65**, 741–768 (1996).
6. Brierley, I., Gilbert, R. J. C. & Pennell, S. Pseudoknot-Dependent Programmed —1 Ribosomal Frameshifting: Structures, Mechanisms and Models. in *Recoding: Expansion of Decoding Rules Enriches Gene Expression* (eds. Atkins, J. F. & Gesteland, R. F.) 149–174 (Springer New York, 2010).
7. Mahmoudabadi, G., Milo, R. & Phillips, R. Energetic cost of building a virus. *Proc. Natl. Acad. Sci.* **114**, E4324–E4333 (2017).
8. Maynard, N. D., Gutschow, M. V., Birch, E. W. & Covert, M. W. The Virus as Metabolic Engineer. *Biotechnol. J.* **5**, 686 (2010).
9. Aller, S., Scott, A., Sarkar-Tyson, M. & Soyer, O. S. Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses. *J. R. Soc. Interface* **15**, 20180125 (2018).

10. Cui, J., Schlub, T. E. & Holmes, E. C. An Allometric Relationship between the Genome Length and Virion Volume of Viruses. *J. Virol.* **88**, 6403–6410 (2014).
11. Tzlil, S., Kindt, J. T., Gelbart, W. M. & Ben-Shaul, A. Forces and Pressures in DNA Packaging and Release from Viral Capsids. *Biophys. J.* **84**, 1616–1627 (2003).
12. Perlmutter, J. D., Qiao, C. & Hagan, M. F. Viral genome structures are optimal for capsid assembly. *eLife* **2**,
13. Hu, Y., Zandi, R., Anavitarte, A., Knobler, C. M. & Gelbart, W. M. Packaging of a Polymer by a Viral Capsid: The Interplay between Polymer Length and Capsid Size. *Biophys. J.* **94**, 1428–1436 (2008).
14. Atkins, J. F., Elseviers, D. & Gorini, L. Low Activity of β -Galactosidase in Frameshift Mutants of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **69**, 1192–1195 (1972).
15. Celis, J. E. & Smith, J. D. Nonsense mutations and tRNA suppressors. in (Academic Press, 1979).
16. Moghal, A., Mohler, K. & Ibba, M. Mistranslation of the genetic code. *FEBS Lett.* **588**, 4305–4310 (2014).
17. Kramer, E. B. & Farabaugh, P. J. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* **13**, 87–96 (2007).
18. Dinman, J. D. Programmed Ribosomal Frameshifting Goes Beyond Viruses. *Microbe Wash. DC* **1**, 521–527 (2006).
19. Huang, X., Cheng, Q. & Du, Z. A genome-wide analysis of RNA pseudoknots that stimulate efficient-1 ribosomal frameshifting or readthrough in animal viruses. *BioMed Res. Int.* **2013**, 984028 (2013).

20. Cao, S. & Chen, S.-J. Predicting ribosomal frameshifting efficiency. *Phys. Biol.* **5**, 016002 (2008).
21. Ritchie, D. B., Foster, D. a N. & Woodside, M. T. Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16167–72 (2012).
22. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**, 9–17 (1985).
23. Bell, N. M. & Lever, A. M. L. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol.* **21**, 136–144 (2013).
24. Anderson, E. C. & Lever, A. M. L. Human Immunodeficiency Virus Type 1 Gag Polyprotein Modulates Its Own Translation. *J. Virol.* **80**, 10478–10486 (2006).
25. Vaishnav, Y. N. & Wong-Staal, F. The biochemistry of AIDS. *Annu. Rev. Biochem.* **60**, 577–630 (1991).
26. Biswas, P., Jiang, X., Pacchia, A. L., Dougherty, J. P. & Peltz, S. W. The Human Immunodeficiency Virus Type 1 Ribosomal Frameshifting Site Is an Invariant Sequence Determinant and an Important Target for Antiviral Therapy. *J. Virol.* **78**, 2082–2087 (2004).
27. Nikolaitchik, O. A. & Hu, W.-S. Deciphering the Role of the Gag-Pol Ribosomal Frameshift Signal in HIV-1 RNA Genome Packaging. *J. Virol.* **88**, 4040–4046 (2014).
28. Chen, C. & Montelaro, R. C. Characterization of RNA Elements That Regulate Gag-Pol Ribosomal Frameshifting in Equine Infectious Anemia Virus. *J. Virol.* **77**, 10280–10287 (2003).

29. Jacks, T. *et al.* Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* **331**, 280–283 (1988).
30. Baril, M., Dulude, D., Gendron, K., Lemay, G. & Brakier-Gingras, L. Efficiency of a programmed -1 ribosomal frameshift in the different subtypes of the human immunodeficiency virus type 1 group M. *RNA N. Y. N* **9**, 1246–1253 (2003).
31. Li, Y. *et al.* Transactivation of programmed ribosomal frameshifting by a viral protein. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2172-2181 (2014).
32. Tsai, T. W., Yang, H., Yin, H., Xu, S. & Wang, Y. High-Efficiency ‘-1’ and ‘-2’ Ribosomal Frameshiftings Revealed by Force Spectroscopy. *ACS Chem. Biol.* **12**, 1629–1635 (2017).
33. Fang, Y. *et al.* Efficient -2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein. *Proc. Natl. Acad. Sci.* 201211145 (2012).
doi:10.1073/pnas.1211145109
34. Jacobs, J. L., Belew, A. T., Rakauskaite, R. & Dinman, J. D. Identification of functional, endogenous programmed -1 ribosomal frameshift signals in the genome of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **35**, 165–174 (2007).
35. Wang, R., Xiong, J., Wang, W., Miao, W. & Liang, A. High frequency of +1 programmed ribosomal frameshifting in *Euplotes octocarinatus*. *Sci. Rep.* **6**, 21139 (2016).
36. Ivanov, I. P. & Atkins, J. F. Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: close to 300 cases reveal remarkable diversity despite underlying conservation. *Nucleic Acids Res.* **35**, 1842–1858 (2007).
37. Kurian, L., Palanimurugan, R., Gödderz, D. & Dohmen, R. J. Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nature* **477**, 490–494 (2011).

38. Craigen, W. J. & Caskey, C. T. Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* **322**, 273–275 (1986).
39. Bekaert, M., Atkins, J. F. & Baranov, P. V. ARFA: a program for annotating bacterial release factor genes, including prediction of programmed ribosomal frameshifting. *Bioinforma. Oxf. Engl.* **22**, 2463–2465 (2006).
40. Jäger, G., Nilsson, K. & Björk, G. R. The Phenotype of Many Independently Isolated +1 Frameshift Suppressor Mutants Supports a Pivotal Role of the P-Site in Reading Frame Maintenance. *PLOS ONE* **8**, e60246 (2013).
41. Brégeon, D., Colot, V., Radman, M. & Taddei, F. Translational misreading: a tRNA modification counteracts a +2 ribosomal frameshift. *Genes Dev.* **15**, 2295–2306 (2001).
42. Weiss, R. B., Dunn, D. M., Atkins, J. F. & Gesteland, R. F. Slippery Runs, Shifty Stops, Backward Steps, and Forward Hops: -2, -1, +1, +2, +5, and +6 Ribosomal Frameshifting. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 687–693 (1987).
43. Huang, W. M. *et al.* A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science* **239**, 1005–1012 (1988).
44. Weiss, R. B., Dunn, D. M., Atkins, J. F. & Gesteland, R. F. Ribosomal Frameshifting from -2 to +50 Nucleotides. in *Progress in Nucleic Acid Research and Molecular Biology* (eds. Cohn, W. E. & Moldave, K.) **39**, 159–183 (Academic Press, 1990).
45. Plant, E. P. Ribosomal Frameshift Signals in Viral Genomes. *Viral Genomes - Mol. Struct. Divers. Gene Expr. Mech. Host-Virus Interact.* (2012). doi:10.5772/26550
46. Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E. & Baranov, P. V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **44**, gkw530 (2016).

47. Antonov, I., Coakley, A., Atkins, J. F., Baranov, P. V. & Borodovsky, M. Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic Acids Res.* **41**, 6514–6530 (2013).
48. Naphthine, S. Prokaryotic-style frameshifting in a plant translation system: conservation of an unusual single-tRNA slippage event. *EMBO J.* **22**, 3941–3950 (2003).
49. Antonov, I., Baranov, P. & Borodovsky, M. GeneTack database: genes with frameshifts in prokaryotic genomes and eukaryotic mRNA sequences. *Nucleic Acids Res.* **41**, D152-156 (2013).
50. Rodríguez, I., García, P. & Suárez, J. E. A Second Case of -1 Ribosomal Frameshifting Affecting a Major Virion Protein of the Lactobacillus Bacteriophage A2. *J. Bacteriol.* **187**, 8201–8204 (2005).
51. Ajon, M. *et al.* UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili: UV-inducible DNA exchange in hyperthermophilic archaea. *Mol. Microbiol.* **82**, 807–817 (2011).
52. Xu, J., Hendrix, R. W. & Duda, R. L. Conserved Translational Frameshift in dsDNA Bacteriophage Tail Assembly Genes. *Mol. Cell* **16**, 11–21 (2004).
53. Cobucci-Ponzano, B. *et al.* The gene of an archaeal α -l-fucosidase is expressed by translational frameshifting. *Nucleic Acids Res.* **34**, 4258–4268 (2006).
54. Raes, J. & Van de Peer, Y. Functional divergence of proteins through frameshift mutations. *Trends Genet.* **21**, 428–431 (2005).
55. Moureau, G. *et al.* New Insights into Flavivirus Evolution, Taxonomy and Biogeographic History, Extended by Analysis of Canonical and Alternative Coding Sequences. *PLoS ONE* **10**, (2015).

56. Möllenbeck, M., Gavin, M. C. & Klobutcher, L. A. Evolution of programmed ribosomal frameshifting in the TERT genes of Euplotes. *J. Mol. Evol.* **58**, 701–711 (2004).
57. Farabaugh, P. J., Kramer, E., Vallabhaneni, H. & Raman, A. Evolution of +1 programmed frameshifting signals and frameshift-regulating tRNAs in the order Saccharomycetales. *J. Mol. Evol.* **63**, 545–561 (2006).
58. IVANOV, I. P., GESTELAND, R. F. & ATKINS, J. F. Evolutionary specialization of recoding: Frameshifting in the expression of *S. cerevisiae* antizyme mRNA is via an atypical antizyme shift site but is still +1. *RNA* **12**, 332–337 (2006).
59. Kramer, E. M., Su, H.-J., Wu, C.-C. & Hu, J.-M. A simplified explanation for the frameshift mutation that created a novel C-terminal motif in the APETALA3 gene lineage. *BMC Evol. Biol.* **6**, 30 (2006).
60. Okamura, K., Feuk, L., Marquès-Bonet, T., Navarro, A. & Scherer, S. W. Frequent appearance of novel protein-coding sequences by frameshift translation. *Genomics* **88**, 690–697 (2006).
61. Brierley, I., Jenner, A. J. & Inglis, S. C. Mutational analysis of the “slippery-sequence” component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **227**, 463–479 (1992).
62. Johnson, A. D. An extended IUPAC nomenclature code for polymorphic nucleic acids. *Bioinformatics* **26**, 1386–1389 (2010).
63. Ketteler, R. On programmed ribosomal frameshifting: The alternative proteomes. *Front. Genet.* **3**, 1–10 (2012).
64. Ten Dam, E. B., Pleij, C. W. A. & Bosch, L. RNA pseudoknots: Translational frameshifting and readthrough on viral RNAs. *Virus Genes* **4**, 121–136 (1990).

65. Chen, X. *et al.* Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting. *EMBO J.* **14**, 842–52 (1995).
66. Ritchie, D. B. *et al.* Conformational dynamics of the frameshift stimulatory structure in HIV-1. *RNA N. Y. N* **23**, 1376–1384 (2017).
67. Yu, C. H., Noteborn, M. H., Pleij, C. W. A. & Olsthoorn, R. C. L. Stem-loop structures can effectively substitute for an RNA pseudoknot in -1 ribosomal frameshifting. *Nucleic Acids Res.* **39**, 8952–8959 (2011).
68. Yang, L. *et al.* Single-Molecule Mechanical Folding and Unfolding of RNA Hairpins: Effects of Single A-U to A·C Pair Substitutions and Single Proton Binding and Implications for mRNA Structure-Induced -1 Ribosomal Frameshifting. *J. Am. Chem. Soc.* **140**, 8172–8184 (2018).
69. Roca, J. *et al.* Monovalent ions modulate the flux through multiple folding pathways of an RNA pseudoknot. *Proc. Natl. Acad. Sci.* 201717582 (2018). doi:10.1073/pnas.1717582115
70. Hori, N., Denesyuk, N. A. & Thirumalai, D. Salt Effects on the Thermodynamics of a Frameshifting RNA Pseudoknot under Tension. *J. Mol. Biol.* **428**, 2847–2859 (2016).
71. Ivanov, I. P., Matsufuji, S., Murakami, Y., Gesteland, R. F. & Atkins, J. F. Conservation of polyamine regulation by translational frameshifting from yeast to mammals. *EMBO J.* **19**, 1907–1917 (2000).
72. Rom, E. & Kahana, C. Polyamines regulate the expression of ornithine decarboxylase antizyme in vitro by inducing ribosomal frame-shifting. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 3959–3963 (1994).

73. Petros, L. M., Howard, M. T., Gesteland, R. F. & Atkins, J. F. Polyamine sensing during antizyme mRNA programmed frameshifting. *Biochem. Biophys. Res. Commun.* **338**, 1478–1489 (2005).
74. Higashi, K. *et al.* Enhancement of +1 Frameshift by Polyamines during Translation of Polypeptide Release Factor 2 in *Escherichia coli*. *J. Biol. Chem.* **281**, 9527–9537 (2006).
75. Rato, C., Amirova, S. R., Bates, D. G., Stansfield, I. & Wallace, H. M. Translational recoding as a feedback controller: systems approaches reveal polyamine-specific effects on the antizyme ribosomal frameshift. *Nucleic Acids Res.* **39**, 4587–4597 (2011).
76. Yordanova, M. M., Wu, C., Andreev, D. E., Sachs, M. S. & Atkins, J. F. A Nascent Peptide Signal Responsive to Endogenous Levels of Polyamines Acts to Stimulate Regulatory Frameshifting on Antizyme mRNA. *J. Biol. Chem.* **290**, 17863–17878 (2015).
77. Zhang, T. *et al.* Anti-Japanese-encephalitis-viral effects of kaempferol and daidzin and their RNA-binding characteristics. *PloS One* **7**, e30259 (2012).
78. Brierley, I. Macrolide-induced ribosomal frameshifting: a new route to antibiotic resistance. *Mol. Cell* **52**, 613–615 (2013).
79. Gupta, P., Kannan, K., Mankin, A. S. & Vázquez-Laslop, N. Regulation of gene expression by macrolide-induced ribosomal frameshifting. *Mol. Cell* **52**, 629–642 (2013).
80. Brakier-Gingras, L., Charbonneau, J. & Butcher, S. E. Targeting frameshifting in the human immunodeficiency virus. *Expert Opin. Ther. Targets* **16**, 249–258 (2012).
81. Park, S.-J., Kim, Y.-G. & Park, H.-J. Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J. Am. Chem. Soc.* **133**, 10094–10100 (2011).

82. Anzalone, A. V., Lin, A. J., Zairis, S., Rabadan, R. & Cornish, V. W. Reprogramming eukaryotic translation with ligand-responsive synthetic RNA switches. *Nat. Methods* **13**, 453–458 (2016).
83. Ritchie, D. B., Soong, J., Sikkema, W. K. A. & Woodside, M. T. Anti-frameshifting Ligand Reduces the Conformational Plasticity of the SARS Virus Pseudoknot. *J. Am. Chem. Soc.* **136**, 2196–2199 (2014).
84. Matsumoto, S., Caliskan, N., Rodnina, M. V., Murata, A. & Nakatani, K. Small synthetic molecule-stabilized RNA pseudoknot as an activator for –1 ribosomal frameshifting. *Nucleic Acids Res.* **46**, 8079–8089 (2018).
85. Ofori, L. O. *et al.* High-Affinity Recognition of HIV-1 Frameshift-Stimulating RNA Alters Frameshifting in Vitro and Interferes with HIV-1 Infectivity. *J. Med. Chem.* **57**, 723–732 (2014).
86. Marcheschi, R. J., Mouzakis, K. D. & Butcher, S. E. Selection and characterization of small molecules that bind the HIV-1 frameshift site RNA. *ACS Chem. Biol.* **4**, 844–854 (2009).
87. Cardno, T. S. *et al.* HIV-1 and Human PEG10 Frameshift Elements Are Functionally Distinct and Distinguished by Novel Small Molecule Modulators. *PLOS ONE* **10**, e0139036 (2015).
88. Lin, Y. H. & Chang, K. Y. Rational design of a synthetic mammalian riboswitch as a ligand-responsive -1 ribosomal frame-shifting stimulator. *Nucleic Acids Res.* **44**, 9005–9015 (2016).
89. Napthine, S. *et al.* Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat. Commun.* **8**, (2017).

90. Naphthine, S. *et al.* A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. *Nucleic Acids Res.* **44**, 5491–5503 (2016).
91. Gurvich, O. L., Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Expression Levels Influence Ribosomal Frameshifting at the Tandem Rare Arginine Codons AGG_AGG and AGA_AGA in *Escherichia coli*. *J. Bacteriol.* **187**, 4023–4032 (2005).
92. Kobayashi, Y., Zhuang, J., Peltz, S. & Dougherty, J. Identification of a cellular factor that modulates HIV-1 programmed ribosomal frameshifting. *J. Biol. Chem.* **285**, 19776–84 (2010).
93. Meskauskas, A. & Dinman, J. D. Ribosomal protein L5 helps anchor peptidyl-tRNA to the P-site in *Saccharomyces cerevisiae*. *RNA N. Y. N* **7**, 1084–1096 (2001).
94. Barry, J. K. & Miller, W. A. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11133–11138 (2002).
95. Tajima, Y., Iwakawa, H., Kaido, M., Mise, K. & Okuno, T. A long-distance RNA–RNA interaction plays an important role in programmed -1 ribosomal frameshifting in the translation of p88 replicase protein of Red clover necrotic mosaic virus. *Virology* **417**, 169–178 (2011).
96. Loughran, G., Firth, A. E. & Atkins, J. F. Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome. *Proc. Natl. Acad. Sci.* **108**, E1111–E1119 (2011).
97. Ahn, D.-G. *et al.* Interference of ribosomal frameshifting by antisense peptide nucleic acids suppresses SARS coronavirus replication. *Antiviral Res.* **91**, 1–10 (2011).

98. Neuman, B. W. *et al.* Inhibition, escape, and attenuated growth of severe acute respiratory syndrome coronavirus treated with antisense morpholino oligomers. *J. Virol.* **79**, 9665–9676 (2005).
99. Karthikeyan, S., Dinman, J. D., Mankin, A. S., Thomas, P. & Jones, J. E. Programmed Ribosomal Frameshifting Generates a Copper Transporter and a Copper Chaperone from the Same Gene Article Programmed Ribosomal Frameshifting Generates a Copper Transporter and a Copper Chaperone from the Same Gene. 207–219 (2017).
doi:10.1016/j.molcel.2016.12.008
100. Caliskan, N. *et al.* Conditional Switch between Frameshifting Regimes upon Translation of dnaX mRNA. *Mol. Cell* **66**, 558-567.e4 (2017).
101. BARANOV, P. V. P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA* **10**, 221–230 (2004).
102. Vimaladithan, A., Pande, S., Zhao, H. & Farabaugh, P. J. Peptidyl-tRNAs promote translational frameshifting. *Nucleic Acids Symp. Ser.* 190–193 (1995).
103. Hong, S. *et al.* Mechanism of tRNA-mediated +1 ribosomal frameshifting. *Proc. Natl. Acad. Sci.* **115**, 11226–11231 (2018).
104. Temperley, R., Richter, R., Dennerlein, S., Lightowers, R. N. & Chrzanowska-Lightowers, Z. M. Hungry codons promote frameshifting in human mitochondrial ribosomes. *Science* **327**, 301 (2010).
105. Gallant, J. A. & Lindsley, D. Ribosomes can slide over and beyond “hungry” codons, resuming protein chain elongation many nucleotides downstream. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13771–13776 (1998).

106. Gallant, J. *et al.* On the role of the starved codon and the takeoff site in ribosome bypassing in *Escherichia coli*. *J. Mol. Biol.* **342**, 713–724 (2004).
107. Lindsley, D., Gallant, J. & Guarneros, G. Ribosome bypassing elicited by tRNA depletion. *Mol. Microbiol.* **48**, 1267–1274 (2003).
108. Barak, Z., Lindsley, D. & Gallant, J. On the Mechanism of Leftward Frameshifting at Several Hungry Codons. *J. Mol. Biol.* **256**, 676–684 (1996).
109. Weiss, R., Lindsley, D., Falahee, B. & Gallant, J. On the mechanism of ribosomal frameshifting at hungry codons. *J. Mol. Biol.* **203**, 403–410 (1988).
110. Lindsley, D. & Gallant, J. On the directional specificity of ribosome frameshifting at a ‘hungry’ codon. *Proc. Natl. Acad. Sci.* **90**, 5469–5473 (1993).
111. Farabaugh, P. J. *Programmed Alternative Reading of the Genetic Code: Molecular Biology Intelligence Unit.* (Springer Science & Business Media, 2012).
112. Liao, P.-Y., Gupta, P., Petrov, A. N., Dinman, J. D. & Lee, K. H. A new kinetic model reveals the synergistic effect of E-, P- and A-sites on +1 ribosomal frameshifting. *Nucleic Acids Res.* **36**, 2619–2629 (2008).
113. Pande, S., Vimaladithan, A., Zhao, H. & Farabaugh, P. J. Pulling the ribosome out of frame by +1 at a programmed frameshift site by cognate binding of aminoacyl-tRNA. *Mol. Cell. Biol.* **15**, 298–304 (1995).
114. Atkins, J. F. & Björk, G. R. A gripping tale of ribosomal frameshifting: extragenic suppressors of frameshift mutations spotlight P-site realignment. **73**, (2009).
115. Waas, W. F., Druzina, Z., Hanan, M. & Schimmel, P. Role of a tRNA Base Modification and Its Precursors in Frameshifting in Eukaryotes. *J. Biol. Chem.* **282**, 26026–26034 (2007).

116. Vimaladithan, A. & Farabaugh, P. J. Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. *Mol. Cell. Biol.* **14**, 8107–8116 (1994).
117. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463 (2004).
118. Sherwood, A. V. & Henkin, T. M. Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses. *Annu. Rev. Microbiol.* **70**, 361–374 (2016).
119. Namy, O., Rousset, J.-P., Naphine, S. & Brierley, I. Reprogrammed Genetic Decoding in Cellular Gene Expression. *Mol. Cell* **13**, 157–168 (2004).
120. Spall, V. E., Shanks, M. & Lomonosoff, G. P. Polyprotein Processing as a Strategy for Gene Expression in RNA Viruses. *Semin. Virol.* **8**, 15–23 (1997).
121. Koch, G. *Biosynthesis, Modification, and Processing of Cellular and Viral Polyproteins*. (Elsevier, 2012).
122. Hellen, C. U. T., Kraeusslich, H. G. & Wimmer, E. Proteolytic processing of polyproteins in the replication of RNA viruses. *Biochemistry* **28**, 9881–9890 (1989).
123. Yost, S. A. & Marcotrigiano, J. Viral Precursor Polyproteins: Keys of Regulation from Replication to Maturation. *Curr. Opin. Virol.* **3**, 137–142 (2013).
124. Dulude, D., Berchiche, Y. A., Gendron, K., Brakier-Gingras, L. & Heveker, N. Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. *Virology* **345**, 127–136 (2006).
125. Shehu-Xhilaga, M., Crowe, S. M. & Mak, J. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol.* **75**, 1834–1841 (2001).

126. Finch, L. K. *et al.* Characterization of Ribosomal Frameshifting in Theiler's Murine Encephalomyelitis Virus. *J. Virol.* **89**, 8580–8589 (2015).
127. Hung, M., Patel, P., Davis, S. & Green, S. R. Importance of Ribosomal Frameshifting for Human Immunodeficiency Virus Type 1 Particle Assembly and Replication. *J. Virol.* **72**, 4819–4824 (1998).
128. Gendron, K., Dulude, D., Lemay, G., Ferbeyre, G. & Brakier-Gingras, L. The virion-associated Gag-Pol is decreased in chimeric Moloney murine leukemia viruses in which the readthrough region is replaced by the frameshift region of the human immunodeficiency virus type 1. *Virology* **334**, 342–352 (2005).
129. Karacostas, V., Wolffe, E. J., Nagashima, K., Gonda, M. A. & Moss, B. Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* **193**, 661–671 (1993).
130. Plant, E. P., Rakauskaite, R., Taylor, D. R. & Dinman, J. D. Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *J. Virol.* **84**, 4330–4340 (2010).
131. McDonagh, P., Sheehy, P. A. & Norris, J. M. In vitro inhibition of feline coronavirus replication by small interfering RNAs. *Vet. Microbiol.* **150**, 220–229 (2011).
132. Dinman, J. D., Ruiz-Echevarria, M. J., Czaplinski, K. & Peltz, S. W. Peptidyl-transferase inhibitors have antiviral properties by altering programmed -1 ribosomal frameshifting efficiencies: development of model systems. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6606–6611 (1997).

133. Dinman, J. D., Ruiz-Echevarria, M. J. & Peltz, S. W. Translating old drugs into new treatments: ribosomal frameshifting as a target for antiviral agents. *Trends Biotechnol.* **16**, 190–196 (1998).
134. Goss Kinzy, T. *et al.* New targets for antivirals: the ribosomal A-site and the factors that interact with it. *Virology* **300**, 60–70 (2002).
135. Ye, Q. *et al.* A single nucleotide mutation in NS2A of Japanese encephalitis-live vaccine virus (SA14-14-2) ablates NS1' formation and contributes to attenuation. *J. Gen. Virol.* **93**, 1959–1964 (2012).
136. Kendra, J. A. *et al.* Ablation of Programmed -1 Ribosomal Frameshifting in Venezuelan Equine Encephalitis Virus Results in Attenuated Neuropathogenicity. *J. Virol.* **91**, (2017).
137. Melian, E. B. *et al.* Programmed ribosomal frameshift alters expression of west Nile virus genes and facilitates virus replication in birds and mosquitoes. *PLoS Pathog.* **10**, e1004447 (2014).
138. Melian, E. B. *et al.* NS1' of flaviviruses in the Japanese encephalitis virus serogroup is a product of ribosomal frameshifting and plays a role in viral neuroinvasiveness. *J. Virol.* **84**, 1641–1647 (2010).
139. Moomau, C., Musalgaonkar, S., Khan, Y. A., Jones, J. E. & Dinman, J. D. Structural and Functional Characterization of Programmed Ribosomal Frameshift Signals in West Nile Virus Strains Reveals High Structural Plasticity Among cis-Acting RNA Elements. *J. Biol. Chem.* **1937**, (2016).
140. Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *J. Gen. Virol.* **93**, 1385–1409 (2012).

141. Belew, A. T. & Dinman, J. D. Frameshifting : Implications for Disease and Therapeutics. **14**, 172–178 (2015).
142. Chen, G., Chang, K.-Y., Chou, M.-Y., Bustamante, C. & Tinoco, I. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of -1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12706–11 (2009).
143. Hansen, T. M., Reihani, S. N. S., Oddershede, L. B. & Sørensen, M. A. Correlation between mechanical strength of messenger RNA pseudoknots and ribosomal frameshifting. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5830–5 (2007).
144. Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. C. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244–247 (2006).
145. Green, L., Kim, C. H., Bustamante, C. & Tinoco, I. Characterization of the Mechanical Unfolding of RNA Pseudoknots. *J. Mol. Biol.* **375**, 511–528 (2008).
146. Cornish, P. V., Stammli, S. N. & Giedroc, D. P. The global structures of a wild-type and poorly functional plant luteoviral mRNA pseudoknot are essentially identical. *RNA N. Y. N* **12**, 1959–1969 (2006).
147. Cornish, P. V., Hennig, M., Giedroc, D. P. & Tinoco, I. A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting. *Pnas* **102**, 12694–12699 (2005).
148. Cornish, P. V. & Giedroc, D. P. Pairwise coupling analysis of helical junction hydrogen bonding interactions in luteoviral RNA pseudoknots. *Biochemistry* **45**, 11162–11171 (2006).

149. Grentzmann, G., Ingram, J. A., Kelly, P. J., Gesteland, R. F. & Atkins, J. F. A dual-luciferase reporter system for studying recoding signals. *RNA* **4**, 479–486 (1998).
150. Naphthine, S., Liphardt, J., Bloys, A., Routledge, S. & Brierley, I. The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *J. Mol. Biol.* **288**, 305–320 (1999).
151. Zhong, Z. *et al.* Mechanical unfolding kinetics of the SRV-1 gag-pro mRNA pseudoknot: possible implications for -1 ribosomal frameshifting stimulation. *Sci. Rep.* **6**, 39549 (2016).
152. Wu, B. *et al.* Translocation kinetics and structural dynamics of ribosomes are modulated by the conformational plasticity of downstream pseudoknots. *Nucleic Acids Res.* (2018).
doi:10.1093/nar/gky636
153. Firth, A. E. & Atkins, J. F. A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. *Viol. J.* **6**, 14 (2009).
154. Nelson, D. L., Cox, M. M. & Lehninger, A. L. *Lehninger principles of biochemistry.* (W.H. Freeman, 2013).
155. Berg, J. M., Tymoczko, J. L. & Stryer, L. Eukaryotic Protein Synthesis Differs from Prokaryotic Protein Synthesis Primarily in Translation Initiation. *Biochem. 5th Ed.* (2002).
156. Jackson, R. J. Alternative mechanisms of initiating translation of mammalian mRNAs. *Biochem. Soc. Trans.* **33**, 1231–1241 (2005).
157. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
158. Wilson, D. N. & Cate, J. H. D. The Structure and Function of the Eukaryotic Ribosome. *Cold Spring Harb. Perspect. Biol.* **4**, a011536 (2012).

159. Colussi, T. M. *et al.* Initiation of translation in bacteria by a structured eukaryotic IRES RNA. *Nature* **519**, 110–113 (2015).
160. Shine, J. & Dalgarno, L. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**, 34–38 (1975).
161. Filipowicz, W. & Haenni, A. L. Binding of ribosomes to 5'-terminal leader sequences of eukaryotic messenger RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 3111–3115 (1979).
162. Hellen, C. U. T. & Sarnow, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* **15**, 1593–1612 (2001).
163. Pisarev, A. V., Shirokikh, N. E. & Hellen, C. U. T. Translation initiation by factor-independent binding of eukaryotic ribosomes to internal ribosomal entry sites. *C. R. Biol.* **328**, 589–605 (2005).
164. Lobanov, A. V., Turanov, A. A., Hatfield, D. L. & Gladyshev, V. N. Dual functions of codons in the genetic code. *Crit. Rev. Biochem. Mol. Biol.* **45**, 257–265 (2010).
165. Kaiser, C. M. & Tinoco, I. Probing the Mechanisms of Translation with Force. *Chem. Rev.* **114**, 3266–3280 (2014).
166. Low, J. T. *et al.* Structure and Dynamics of the HIV-1 Frameshift Element RNA. *Biochemistry* **53**, 4282–4291 (2014).
167. Dustin Ritchie. Programmed Ribosomal Frameshifting.
168. Drummond, D. A. & Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nat. Rev. Genet.* **10**, 715–724 (2009).
169. Curran, J. F. & Yarus, M. Base substitutions in the tRNA anticodon arm do not degrade the accuracy of reading frame maintenance. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6538–6542 (1986).

170. Manickam, N., Nag, N., Abbasi, A., Patel, K. & Farabaugh, P. J. Studies of translational misreading in vivo show that the ribosome very efficiently discriminates against most potential errors. *RNA* **20**, 9–15 (2014).
171. Parker, J. Errors and alternatives in reading the universal genetic code. *Microbiol. Rev.* **53**, 273–298 (1989).
172. Kramer, E. B. & Farabaugh, P. J. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA* **13**, 87–96 (2007).
173. Kurland, C. G. Translational Accuracy and the Fitness of Bacteria. *Annu. Rev. Genet.* **26**, 29–50 (1992).
174. Loewe, L. & Hill, W. G. The population genetics of mutations: good, bad and indifferent. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1153–1167 (2010).
175. Ohno, S. Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2421–2425 (1984).
176. Ranz, J. M., Ponce, A. R., Hartl, D. L. & Nurminsky, D. Origin and evolution of a new gene expressed in the *Drosophila* sperm axoneme. *Genetica* **118**, 233–244 (2003).
177. Michel, A. M. *et al.* Observation of dually decoded regions of the human genome using ribosome profiling data. *Genome Res.* **22**, 2219–2229 (2012).
178. Belew, A. T., Advani, V. M. & Dinman, J. D. Endogenous ribosomal frameshift signals operate as mRNA destabilizing elements through at least two molecular pathways in yeast. *Nucleic Acids Res.* **39**, 2799–2808 (2011).

179. Harger, J. W. & Dinman, J. D. An in vivo dual-luciferase assay system for studying translational recoding in the yeast *Saccharomyces cerevisiae*. *RNA N. Y. N* **9**, 1019–1024 (2003).
180. McNabb, D. S., Reed, R. & Marciniak, R. A. Dual Luciferase Assay System for Rapid Assessment of Gene Expression in *Saccharomyces cerevisiae*. *Eukaryot. Cell* **4**, 1539–1549 (2005).
181. Bekaert, M. *et al.* Recode-2: New design, new search tools, and many more genes. *Nucleic Acids Res.* **38**, 69–74 (2009).
182. Belew, A. T., Hepler, N. L., Jacobs, J. L. & Dinman, J. D. PRFdb: a database of computationally predicted eukaryotic programmed -1 ribosomal frameshift signals. *BMC Genomics* **9**, 339 (2008).
183. Beremand, M. N. & Blumenthal, T. Overlapping genes in rna phage: a new protein implicated in lysis. *Cell* **18**, 257–266 (1979).
184. Atkins, J. F., Gesteland, R. F., Reid, B. R. & Anderson, C. W. Normal tRNAs promote ribosomal frameshifting. *Cell* **18**, 1119–1131 (1979).
185. Dunn, J. J. & Studier, F. W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166**, 477–535 (1983).
186. Condron, B. G., Atkins, J. F. & Gesteland, R. F. Frameshifting in gene 10 of bacteriophage T7. *J. Bacteriol.* **173**, 6998–7003 (1991).
187. Coleman, J., Inouye, M. & Atkins, J. Bacteriophage MS2 Lysis Protein Does Not Require Coat Protein to Mediate Cell Lysis. *J. Bacteriol.* **153**, 1098–1100 (1983).

188. Wilson, W., Malim, M. H., Mellor, J., Kingsman, A. J. & Kingsman, S. M. Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. *Nucleic Acids Res.* **14**, 7001–7016 (1986).
189. Clare, J. & Farabaugh, P. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 2829–2833 (1985).
190. Bekaert, M., Richard, H., Prum, B. & Rousset, J.-P. Identification of programmed translational -1 frameshifting sites in the genome of *Saccharomyces cerevisiae*. *Genome Res.* **15**, 1411–1420 (2005).
191. Byun, Y., Moon, S. & Han, K. A general computational model for predicting ribosomal frameshifts in genome sequences. *Comput. Biol. Med.* **37**, 1796–1801 (2007).
192. Moon, S., Byun, Y., Kim, H.-J., Jeong, S. & Han, K. Predicting genes expressed via -1 and +1 frameshifts. *Nucleic Acids Res.* **32**, 4884–4892 (2004).
193. Theis, C., Reeder, J. & Giegerich, R. KnotInFrame: prediction of -1 ribosomal frameshift events. *Nucleic Acids Res.* **36**, 6013–6020 (2008).
194. Ingolia, N. T. Ribosome Footprint Profiling of Translation throughout the Genome. *Cell* **165**, 22–33 (2016).
195. Atkins, J. F., Loughran, G. & Baranov, P. V. A [Cu]rious Ribosomal Profiling Pattern Leads to the Discovery of Ribosomal Frameshifting in the Synthesis of a Copper Chaperone. *Mol. Cell* **65**, 203–204 (2017).
196. Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Recoding: translational bifurcations in gene expression. *Gene* **286**, 187–201 (2002).

197. Cobucci-Ponzano, B., Rossi, M. & Moracci, M. Recoding in Archaea. *Mol. Microbiol.* **55**, 339–348 (2005).
198. Caliskan, N., Peske, F. & Rodnina, M. V. Changed in translation: mRNA recoding by –1 programmed ribosomal frameshifting. *Trends Biochem. Sci.* **40**, 265–274 (2015).
199. Belew, A. T., Meskauskas, A. & Musalgaonkar, S. Ribosomal frameshifting in the CCR5 mRNA is regulated by miRNAs and the NMD pathway. *Nature* **512**, (2014).
200. Advani, V. M., Belew, A. T. & Dinman, J. D. Yeast telomere maintenance is globally controlled by programmed ribosomal frameshifting and the nonsense-mediated mRNA decay pathway. *Translation* **1**, e24418 (2013).
201. Brierley, I. & Dos Ramos, F. J. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.* **119**, 29–42 (2006).
202. Plant, E. P. & Dinman, J. D. The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front. Biosci. J. Virtual Libr.* **13**, 4873–4881 (2008).
203. Gareiss, P. C. & Miller, B. L. Ribosomal frameshifting: an emerging drug target for HIV. *Curr. Opin. Investig. Drugs Lond. Engl.* **10**, 121–128 (2009).
204. Griffiths, A. Slipping and Sliding: frameshift mutations in herpes simplex virus thymidine kinase and drug-resistance. *Drug Resist. Updat. Rev. Comment. Antimicrob. Anticancer Chemother.* **14**, 251–259 (2011).
205. Park, S. B., Seronello, S., Mayer, W. & Ojcius, D. M. Hepatitis C Virus Frameshift/Alternate Reading Frame Protein Suppresses Interferon Responses Mediated by Pattern Recognition Receptor Retinoic-Acid-Inducible Gene-I. *PLOS ONE* **11**, e0158419 (2016).

206. Howard, M. T., Gesteland, R. F. & Atkins, J. F. Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. *RNA N. Y. N* **10**, 1653–1661 (2004).
207. Henderson, C. M., Anderson, C. B. & Howard, M. T. Antisense-induced ribosomal frameshifting. *Nucleic Acids Res.* **34**, 4302–4310 (2006).
208. Shigemoto, K. *et al.* Identification and characterisation of a developmentally regulated mammalian gene that utilises -1 programmed ribosomal frameshifting. *Nucleic Acids Res.* **29**, 4079–4088 (2001).
209. Morales, S., Monzo, M. & Navarro, A. Epigenetic regulation mechanisms of microRNA expression. *Biomol. Concepts* **8**, 203–212 (2017).
210. Cai, Y., Yu, X., Hu, S. & Yu, J. A Brief Review on the Mechanisms of miRNA Regulation. *Genomics Proteomics Bioinformatics* **7**, 147–154 (2009).
211. Vidigal, J. A. & Ventura, A. The biological functions of miRNAs: lessons from in vivo studies. *Trends Cell Biol.* **25**, 137–147 (2015).
212. Nickless, A., Bailis, J. M. & You, Z. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell Biosci.* **7**, 26 (2017).
213. Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* **16**, 665–677 (2015).
214. Peccarelli, M. & Kebaara, B. W. Regulation of Natural mRNAs by the Nonsense-Mediated mRNA Decay Pathway. *Eukaryot. Cell* **13**, 1126–1135 (2014).
215. Plant, E. P., Wang, P., Jacobs, J. L. & Dinman, J. D. A programmed -1 ribosomal frameshift signal can function as a cis-acting mRNA destabilizing element. *Nucleic Acids Res.* **32**, 784–90 (2004).

216. Hug, N., Longman, D. & Cáceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* **44**, 1483–1495 (2016).
217. Bhuvanagiri, M., Schlitter, A. M., Hentze, M. W. & Kulozik, A. E. NMD: RNA biology meets human genetic medicine. *Biochem. J.* **430**, 365–377 (2010).
218. Miller, J. N. & Pearce, D. A. Nonsense-mediated decay in genetic disease: friend or foe? *Mutat. Res. Rev. Mutat. Res.* **762**, 52–64 (2014).
219. Belew, A. T. & Dinman, J. D. Cell cycle control (and more) by programmed -1 ribosomal frameshifting: implications for disease and therapeutics. *Cell Cycle Georget. Tex* **14**, 172–8 (2015).
220. Seligmann, H. & Pollock, D. D. The ambush hypothesis: hidden stop codons prevent off-frame gene reading. *DNA Cell Biol.* **23**, 701–705 (2004).
221. Singh, T. R. & Pardasani, K. R. Ambush hypothesis revisited: Evidences for phylogenetic trends. *Comput. Biol. Chem.* **33**, 239–244 (2009).
222. Morgens, D. W., Chang, C. H. & Cavalcanti, A. R. Ambushing the ambush hypothesis: predicting and evaluating off-frame codon frequencies in Prokaryotic Genomes. *BMC Genomics* **14**, 418 (2013).
223. Abrahams, L. & Hurst, L. D. Refining the Ambush Hypothesis: Evidence That GC- and AT-Rich Bacteria Employ Different Frameshift Defence Strategies. *Genome Biol. Evol.* **10**, 1153–1173 (2018).
224. Zarraga, I. G. *et al.* Nonsense-Mediated mRNA Decay Caused by a Frameshift Mutation in a Large Kindred of Type 2 Long QT Syndrome. *Heart Rhythm Off. J. Heart Rhythm Soc.* **8**, 1200–1206 (2011).

225. Matsufuji, S. *et al.* Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* **80**, 51–60 (1995).
226. Topp, S. & Gallivan, J. P. Guiding bacteria with small molecules and RNA. *J. Am. Chem. Soc.* **129**, 6807–6811 (2007).
227. Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.* **32**, 1610–1614 (2004).
228. Dixon, N. *et al.* Reengineering orthogonally selective riboswitches. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2830–2835 (2010).
229. Ceres, P., Garst, A. D., Marcano-Velázquez, J. G. & Batey, R. T. Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. *ACS Synth. Biol.* **2**, 463–472 (2013).
230. Ceres, P., Trausch, J. J. & Batey, R. T. Engineering modular ‘ON’ RNA switches using biological components. *Nucleic Acids Res.* **41**, 10449–10461 (2013).
231. Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P. F. & Mörl, M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Res.* **41**, 2541–2551 (2013).
232. Robinson, C. J. *et al.* Modular riboswitch toolsets for synthetic genetic control in diverse bacterial species. *J. Am. Chem. Soc.* **136**, 10615–10624 (2014).
233. Grate, D. & Wilson, C. Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer-ligand complex. *Bioorg. Med. Chem.* **9**, 2565–2570 (2001).
234. Suess, B. *et al.* Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Res.* **31**, 1853–1858 (2003).

235. Weigand, J. E. *et al.* Screening for engineered neomycin riboswitches that control translation initiation. *RNA N. Y. N* **14**, 89–97 (2008).
236. Wittmann, A. & Suess, B. Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Mol. Biosyst.* **7**, 2419–2427 (2011).
237. Klauser, B., Atanasov, J., Siewert, L. K. & Hartig, J. S. Ribozyme-based aminoglycoside switches of gene expression engineered by genetic selection in *S. cerevisiae*. *ACS Synth. Biol.* **4**, 516–525 (2015).
238. Werstuck, G. & Green, M. R. Controlling gene expression in living cells through small molecule-RNA interactions. *Science* **282**, 296–298 (1998).
239. Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14283–14288 (2007).
240. Beisel, C. L., Bayer, T. S., Hoff, K. G. & Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol. Syst. Biol.* **4**, 224 (2008).
241. Kumar, D., An, C.-I. & Yokobayashi, Y. Conditional RNA interference mediated by allosteric ribozyme. *J. Am. Chem. Soc.* **131**, 13906–13907 (2009).
242. Ausländer, S., Ketzer, P. & Hartig, J. S. A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Mol. Biosyst.* **6**, 807–814 (2010).
243. Chen, Y. Y., Jensen, M. C. & Smolke, C. D. Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8531–8536 (2010).

244. Nomura, Y., Zhou, L., Miu, A. & Yokobayashi, Y. Controlling mammalian gene expression by allosteric hepatitis delta virus ribozymes. *ACS Synth. Biol.* **2**, 684–689 (2013).
245. Hsu, H. T., Lin, Y. H. & Chang, K. Y. Synergetic regulation of translational reading-frame switch by ligand-responsive RNAs in mammalian cells. *Nucleic Acids Res.* **42**, 14070–14082 (2014).
246. Beilstein, K., Wittmann, A., Grez, M. & Suess, B. Conditional control of mammalian gene expression by tetracycline-dependent hammerhead ribozymes. *ACS Synth. Biol.* **4**, 526–534 (2015).
247. Jacks, T. & Varmus, H. E. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* **230**, 1237–1242 (1985).
248. Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **55**, 447–458 (1988).
249. Nikolić, E. I. C., King, L. M., Vidakovic, M., Irigoyen, N. & Brierley, I. Modulation of Ribosomal Frameshifting Frequency and Its Effect on the Replication of Rous Sarcoma Virus. *J. Virol.* **86**, 11581–11594 (2012).
250. Baranov, P. V. *et al.* Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* **332**, 498–510 (2005).
251. Dinman, J. D., Icho, T. & Wickner, R. B. A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 174–178 (1991).
252. Dinman, J. D. & Wickner, R. B. Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J. Virol.* **66**, 3669–3676 (1992).

253. Harger, J. W., Meskauskas, A. & Dinman, J. D. An 'integrated model' of programmed ribosomal frameshifting. *Trends Biochem. Sci.* **27**, 448–454 (2002).
254. Plant, E. P. *et al.* The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. *RNA N. Y. N* **9**, 168–174 (2003).
255. Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. Signals for ribosomal frameshifting in the rous sarcoma virus gag-pol region. *Cell* **55**, 447–458 (1988).
256. Farabaugh, P. J. Programmed translational frameshifting. *Microbiol. Rev.* **60**, 103–134 (1996).
257. Weiss, R. B., Dunn, D. M., Shuh, M., Atkins, J. F. & Gesteland, R. F. E. coli ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biol.* **1**, 159–169 (1989).
258. Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. C. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244–247 (2006).
259. Moran, S. J. *et al.* The mechanics of translocation: a molecular 'spring-and-ratchet' system. *Struct. Lond. Engl. 1993* **16**, 664–672 (2008).
260. Léger, M., Dulude, D., Steinberg, S. V. & Brakier-Gingras, L. The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed -1 ribosomal frameshift. *Nucleic Acids Res.* **35**, 5581–5592 (2007).
261. Licznar, P. *et al.* Programmed translational -1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs. *EMBO J.* **22**, 4770–4778 (2003).
262. Chen, G., Wen, J.-D. & Tinoco, I. Single-molecule mechanical unfolding and folding of a pseudoknot in human telomerase RNA. *Rna* **13**, 2175–88 (2007).

263. Lin, Z., Gilbert, R. J. C. & Brierley, I. Spacer-length dependence of programmed -1 or -2 ribosomal frameshifting on a U6A heptamer supports a role for messenger RNA (mRNA) tension in frameshifting. *Nucleic Acids Res.* **40**, 8674–8689 (2012).
264. Yusupova, G. Z., Yusupov, M. M., Cate, J. H. D. & Noller, H. F. The Path of Messenger RNA through the Ribosome. *Cell* **106**, 233–241 (2001).
265. Kurkcuoglu, O., Doruker, P., Sen, T. Z., Kloczkowski, A. & Jernigan, R. L. The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. *Phys. Biol.* **5**, 046005 (2008).
266. Yusupova, G., Jenner, L., Rees, B., Moras, D. & Yusupov, M. Structural basis for messenger RNA movement on the ribosome. *Nature* **444**, 391–394 (2006).
267. Demeshkina, N., Jenner, L., Yusupova, G. & Yusupov, M. Interactions of the ribosome with mRNA and tRNA. *Curr. Opin. Struct. Biol.* **20**, 325–332 (2010).
268. Tholstrup, J., Oddershede, L. B. & Sørensen, M. A. mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res.* **40**, 303–313 (2012).
269. Kontos, H., Napthine, S. & Brierley, I. Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol. Cell. Biol.* **21**, 8657–70 (2001).
270. Yang, J.-R., Chen, X. & Zhang, J. Codon-by-Codon Modulation of Translational Speed and Accuracy Via mRNA Folding. *PLOS Biol.* **12**, e1001910 (2014).
271. Endoh, T. & Sugimoto, N. Mechanical insights into ribosomal progression overcoming RNA G-quadruplex from periodical translation suppression in cells. *Sci. Rep.* **6**, 22719 (2016).

272. Brierley, I., Rolley, N. J., Jenner, A. J. & Inglis, S. C. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **220**, 889–902 (1991).
273. Rodnina, M. V. The ribosome in action: Tuning of translational efficiency and protein folding. *Protein Sci. Publ. Protein Soc.* **25**, 1390–1406 (2016).
274. Schuller, A. P. & Green, R. Roadblocks and resolutions in eukaryotic translation. *Nat. Rev. Mol. Cell Biol.* **19**, 526–541 (2018).
275. Qu, X. *et al.* The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* **475**, 118–121 (2011).
276. Qu, X., Lancaster, L., Noller, H. F., Bustamante, C. & Tinoco, I. Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14458–14463 (2012).
277. Tinoco, I., Kim, H.-K. & Yan, S. Frameshifting dynamics. *Biopolymers* **99**, n/a-n/a (2013).
278. Chen, C. *et al.* Dynamics of translation by single ribosomes through mRNA secondary structures. *Nat. Struct. Mol. Biol.* **20**, 582–588 (2013).
279. Xie, P. Dynamics of forward and backward translocation of mRNA in the ribosome. *PLoS One* **8**, e70789 (2013).
280. Mouzakis, K. D., Lang, A. L., Vander Meulen, K. A., Easterday, P. D. & Butcher, S. E. HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Res.* **41**, (2013).
281. Qin, P., Yu, D., Zuo, X. & Cornish, P. V. Structured mRNA induces the ribosome into a hyper-rotated state. *EMBO Rep.* **15**, 185–90 (2014).

282. Kim, H.-K. *et al.* A frameshifting stimulatory stem loop destabilizes the hybrid state and impedes ribosomal translocation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5538–5543 (2014).
283. Jia, H., Wang, Y. & Xu, S. Super-resolution force spectroscopy reveals ribosomal motion at sub-nucleotide steps. *Chem. Commun. Camb. Engl.* **54**, 5883–5886 (2018).
284. Xie, P. Model of ribosomal translocation coupled with intra- and inter-subunit rotations. *Biochem. Biophys. Rep.* **2**, 87–93 (2015).
285. Xie, P. Model of the pathway of -1 frameshifting: Long pausing. *Biochem. Biophys. Rep.* **5**, 408–424 (2016).
286. Xie, P. & Chen, H. Mechanism of ribosome translation through mRNA secondary structures. *Int. J. Biol. Sci.* **13**, 712–722 (2017).
287. Kim, H.-K. & Tinoco, I. EF-G catalyzed translocation dynamics in the presence of ribosomal frameshifting stimulatory signals. *Nucleic Acids Res.* **45**, 2865–2874 (2017).
288. Caliskan, N., Katunin, V. I., Belardinelli, R., Peske, F. & Rodnina, M. V. Programmed -1 frameshifting by kinetic partitioning during impeded translocation. *Cell* **157**, 1619–1631 (2014).
289. Ling, C. & Ermolenko, D. N. Structural insights into ribosome translocation. *Wiley Interdiscip. Rev. RNA* **7**, 620–636 (2016).
290. Belardinelli, R., Sharma, H., Peske, F., Wintermeyer, W. & Rodnina, M. V. Translocation as continuous movement through the ribosome. *RNA Biol.* **13**, 1197–1203 (2016).
291. Peske, F., Savelsbergh, A., Katunin, V. I., Rodnina, M. V. & Wintermeyer, W. Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation. *J. Mol. Biol.* **343**, 1183–1194 (2004).

292. Rodnina, M. V. & Wintermeyer, W. The ribosome as a molecular machine: the mechanism of tRNA-mRNA movement in translocation. *Biochem. Soc. Trans.* **39**, 658–662 (2011).
293. Holtkamp, W., Wintermeyer, W. & Rodnina, M. V. Synchronous tRNA movements during translocation on the ribosome are orchestrated by elongation factor G and GTP hydrolysis. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **36**, 908–918 (2014).
294. Petrov, A., Chen, J., O’Leary, S., Tsai, A. & Puglisi, J. D. Single-molecule analysis of translational dynamics. *Cold Spring Harb. Perspect. Biol.* **4**, a011551 (2012).
295. Fischer, N., Konevega, A. L., Wintermeyer, W., Rodnina, M. V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333 (2010).
296. Córdova, N. J., Ermentrout, B. & Oster, G. F. Dynamics of single-motor molecules: the thermal ratchet model. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 339–343 (1992).
297. Peskin, C. S., Odell, G. M. & Oster, G. F. Cellular motions and thermal fluctuations: the Brownian ratchet. *Biophys. J.* **65**, 316–324 (1993).
298. Vale, R. D. & Oosawa, F. Protein motors and Maxwell’s demons: Does mechanochemical transduction involve a thermal ratchet? *Adv. Biophys.* **26**, 97–134 (1990).
299. Liu, T. *et al.* Direct measurement of the mechanical work during translocation by the ribosome. *eLife* **3**, (2014).
300. Jenner, L. B., Demeshkina, N., Yusupova, G. & Yusupov, M. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.* **17**, 555–560 (2010).

301. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
302. Andreev, D. E. *et al.* Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res.* **45**, 513–526 (2017).
303. Sinha, D. K., Bhalla, U. S. & Shivashankar, G. V. Kinetic measurement of ribosome motor stalling force. *Appl. Phys. Lett.* **85**, 4789–4791 (2004).
304. Wang, M. D. *et al.* Force and Velocity Measured for Single Molecules of RNA Polymerase. *Science* **282**, 902–907 (1998).
305. Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**, 437–447 (2003).
306. John, R., Davenport, Wuite, G. J. L., Landick, R. & Bustamante, C. Single-Molecule Study of Transcriptional Pausing and Arrest by *E. coli* RNA Polymerase. *Science* **287**, 2497–2500 (2000).
307. Yin, H. *et al.* Transcription Against an Applied Force. *Science* **270**, 1653–1657 (1995).
308. Caliskan, N., Katunin, V. I., Belardinelli, R., Peske, F. & Rodnina, M. V. Programmed –1 Frameshifting by Kinetic Partitioning during Impeded Translocation. *Cell* **157**, 1619–1631 (2014).
309. Chen, J. *et al.* Dynamic pathways of –1 translational frameshifting. *Nature* **512**, 328–332 (2014).
310. Roy, R., Hohng, S. & Ha, T. A Practical Guide to Single Molecule FRET. *Nat. Methods* **5**, 507–516 (2008).

311. Hellenkamp, B. *et al.* Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nat. Methods* **15**, 669–676 (2018).
312. Schuler, B. Single-molecule FRET of protein structure and dynamics - a primer. *J. Nanobiotechnology* **11**, S2 (2013).
313. Chen, J. *et al.* Single-Molecule FRET Studies of HIV TAR–DNA Hairpin Unfolding Dynamics. *J. Phys. Chem. B* **118**, 12130–12139 (2014).
314. Kang, H., Hines, J. V. & Tinoco, I. Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus. *J. Mol. Biol.* **259**, 135–147 (1996).
315. Yan, S., Wen, J. D., Bustamante, C. & Tinoco, I. Ribosome Excursions during mRNA Translocation Mediate Broad Branching of Frameshift Pathways. *Cell* **160**, (2015).
316. West Nile virus lineage 1, complete genome. (2018).
317. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616 (2006).
318. Behm-Ansmant, I., Helm, M. & Motorin, Y. Use of Specific Chemical Reagents for Detection of Modified Nucleotides in RNA. *Journal of Nucleic Acids* (2011).
doi:10.4061/2011/408053
319. Turner, R., Shefer, K. & Ares, M. Safer one-pot synthesis of the 'SHAPE' reagent 1-methyl-7-nitroisatoic anhydride (1m7). *RNA N. Y. N* **19**, 1857–1863 (2013).
320. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).

321. Patterson, J. T., Nickens, D. G. & Burke, D. H. HIV-1 reverse transcriptase pausing at bulky 2' adducts is relieved by deletion of the RNase H domain. *RNA Biol.* **3**, 163–169 (2006).
322. Loughrey, D., Watters, K. E., Settle, A. H. & Lucks, J. B. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.* **42**, e165–e165 (2014).
323. Lucks, J. B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci.* **108**, 11063–11068 (2011).
324. Mortimer, S. A., Trapnell, C., Aviran, S., Pachter, L. & Lucks, J. B. SHAPE-Seq: High-Throughput RNA Structure Analysis. *Curr. Protoc. Chem. Biol.* **4**, 275–297 (2012).
325. Watters, K. E., Yu, A. M., Strobel, E. J., Settle, A. H. & Lucks, J. B. Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods* **103**, 34–48 (2016).
326. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7287–7292 (2004).
327. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
328. Rivas, E. & Eddy, S. R. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* **285**, 2053–2068 (1999).
329. Firth, A. E., Chung, B. Y., Fleeton, M. N. & Atkins, J. F. Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. *Virol. J.* **5**, 108 (2008).

330. Hofacker, I. L. *et al.* Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucleic Acids Res.* **26**, 3825–3836 (1998).
331. Hofacker, I. L. & Stadler, P. F. Automatic detection of conserved base pairing patterns in RNA virus genomes. *Comput. Chem.* **23**, 401–414 (1999).
332. Mason, P. W., McAda, P. C., Dalrymple, J. M., Fournier, M. J. & Mason, T. L. Expression of Japanese encephalitis virus antigens in *Escherichia coli*. *Virology* **158**, 361–372 (1987).
333. Mason, P. W. Maturation of Japanese encephalitis virus glycoproteins produced by infected mammalian and mosquito cells. *Virology* **169**, 354–364 (1989).
334. Blitvich, B. J., Mackenzie, J. S., Coelen, R. J., Howard, M. J. & Hall, R. A. A novel complex formed between the flavivirus E and NS1 proteins: analysis of its structure and function. *Arch. Virol.* **140**, 145–156 (1995).
335. Chen, L.-K. *et al.* Persistence of Japanese Encephalitis Virus Is Associated with Abnormal Expression of the Nonstructural Protein NS1 in Host Cells. *Virology* **217**, 220–229 (1996).
336. Blitvich, B. J., Scanlon, D., Shiell, B. J., Mackenzie, J. S. & Hall, R. A. Identification and analysis of truncated and elongated species of the flavivirus NS1 protein. *Virus Res.* **60**, 67–79 (1999).
337. Hall, R. A., Kay, B. H., Burgess, G. W., Clancy, P. & Fanning, I. D. Epitope Analysis of the Envelope and Non-structural Glycoproteins of Murray Valley Encephalitis Virus. *J. Gen. Virol.* **71**, 2923–2930 (1990).
338. Ahlquist, P. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nat. Rev. Microbiol.* **4**, 371–382 (2006).
339. Nicholson, B. L. & White, K. A. Functional long-range RNA–RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.* **12**, 493–504 (2014).

340. Brinton, M. A. The Molecular Biology of West Nile Virus: A New Invader of the Western Hemisphere. *Annu. Rev. Microbiol.* **56**, 371–402 (2002).
341. Faggioni, G. *et al.* West Nile alternative open reading frame (N-NS4B/WARF4) is produced in infected West Nile Virus (WNV) cells and induces humoral response in WNV infected individuals. *Viol. J.* **9**, 283 (2012).
342. Mentoor, J. L. D. *et al.* Full-Genome Sequence of a Neuroinvasive West Nile Virus Lineage 2 Strain from a Fatal Horse Infection in South Africa. *Genome Announc* **4**, e00740-16 (2016).
343. Barzon, L. *et al.* Whole genome sequencing and phylogenetic analysis of West Nile virus lineage 1 and lineage 2 from human cases of infection, Italy, August 2013. *Eurosurveillance* **18**, 20591 (2013).
344. Lwande, O. W. *et al.* Whole genome phylogenetic investigation of a West Nile virus strain isolated from a tick sampled from livestock in north eastern Kenya. *Parasit. Vectors* **7**, 542 (2014).
345. Wengler, G., Czaya, G., Farber, P. M. & Hegemann, J. H. In vitro synthesis of West Nile virus proteins indicates that the amino-terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. *J. Gen. Virol.* **72**, 851–858 (1991).
346. VanBlargan, L. A. *et al.* Context-Dependent Cleavage of the Capsid Protein by the West Nile Virus Protease Modulates the Efficiency of Virus Assembly. *J. Virol.* **89**, 8632–8642 (2015).
347. Monaghan, S., Ryan, M. D. & Flint, M. Virus-encoded proteinases of the Flaviviridae. *J. Gen. Virol.* **79**, 947–959 (1998).

348. Stocks, C. E. & Lobigs, M. Signal Peptidase Cleavage at the Flavivirus C-prM Junction: Dependence on the Viral NS2B-3 Protease for Efficient Processing Requires Determinants in C, the Signal Peptide, and prM. *J. Virol.* **72**, 2141–2149 (1998).
349. Bera, A. K., Kuhn, R. J. & Smith, J. L. Functional Characterization of cis and trans Activity of the Flavivirus NS2B-NS3 Protease. *J. Biol. Chem.* **282**, 12883–12892 (2007).
350. Castle, E., Nowak, T., Leidner, U., Wengler, G. & Wengler, G. Sequence analysis of the viral core protein and the membrane-associated proteins V1 and NV2 of the flavivirus West Nile virus and of the genome sequence for these proteins. *Virology* **145**, 227–236 (1985).
351. Wengler, G., Castle, E., Leidner, U., Nowak, T. & Wengler, G. Sequence analysis of the membrane protein V3 of the flavivirus West Nile virus and of its gene. *Virology* **147**, 264–274 (1985).
352. Kanai, R. *et al.* Crystal Structure of West Nile Virus Envelope Glycoprotein Reveals Viral Surface Epitopes. *J. Virol.* **80**, 11000–11008 (2006).
353. Castle, E., Leidner, U., Nowak, T., Wengler, G. & Wengler, G. Primary structure of the West Nile flavivirus genome region coding for all nonstructural proteins. *Virology* **149**, 10–26 (1986).
354. Muller, D. A. & Young, P. R. The flavivirus NS1 protein: Molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Res.* **98**, 192–208 (2013).
355. Leung, J. Y. *et al.* Role of Nonstructural Protein NS2A in Flavivirus Assembly. *J. Virol.* **82**, 4731–4741 (2008).

356. Bessaud, M. *et al.* Functional characterization of the NS2B/NS3 protease complex from seven viruses belonging to different groups inside the genus Flavivirus. *Virus Res.* **120**, 79–90 (2006).
357. Ambrose, R. L. & Mackenzie, J. M. A Conserved Peptide in West Nile Virus NS4A Protein Contributes to Proteolytic Processing and Is Essential for Replication ∇ . *J. Virol.* **85**, 11274–11282 (2011).
358. Zmurko, J., Neyts, J. & Dallmeier, K. Flaviviral NS4b, chameleon and jack-in-the-box roles in viral replication and pathogenesis, and a molecular target for antiviral intervention. *Rev. Med. Virol.* **25**, 205–223 (2015).
359. Laurent-Rolle, M. *et al.* The NS5 Protein of the Virulent West Nile Virus NY99 Strain Is a Potent Antagonist of Type I Interferon-Mediated JAK-STAT Signaling. *J. Virol.* **84**, 3503–3515 (2010).
360. Koo, Q. Y. *et al.* Conservation and Variability of West Nile Virus Proteins. *PLOS ONE* **4**, e5352 (2009).
361. Roehrig, L. R. P. and J. T. West Nile Virus: A Reemerging Global Pathogen - Volume 7, Number 4—August 2001 - Emerging Infectious Diseases journal - CDC.
doi:10.3201/eid0704.017401
362. Brinton, M. A. Replication Cycle and Molecular Biology of the West Nile Virus. *Viruses* **6**, 13–53 (2013).
363. Frese, M. *et al.* Internal Ribosome Entry Site-Based Attenuation of a Flavivirus Candidate Vaccine and Evaluation of the Effect of Beta Interferon Coexpression on Vaccine Properties. *J. Virol.* **88**, 2056–2070 (2014).

364. Plante, K. *et al.* Novel chikungunya vaccine candidate with an IRES-based attenuation and host range alteration mechanism. *PLoS Pathog.* **7**, e1002142 (2011).
365. Guerbois, M. *et al.* IRES-driven expression of the capsid protein of the Venezuelan equine encephalitis virus TC-83 vaccine strain increases its attenuation and safety. *PLoS Negl. Trop. Dis.* **7**, e2197 (2013).
366. Rossi, S. L. *et al.* IRES-based Venezuelan equine encephalitis vaccine candidate elicits protective immunity in mice. *Virology* **437**, 81–88 (2013).
367. Volkova, E. *et al.* IRES-dependent replication of Venezuelan equine encephalitis virus makes it highly attenuated and incapable of replicating in mosquito cells. *Virology* **377**, 160–169 (2008).
368. Dinman, J. D. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip. Rev. RNA* **3**, 661–673 (2012).
369. Hussmann, K. L., Vandergaast, R., Zheng, K., Hoover, L. I. & Fredericksen, B. L. Structural proteins of West Nile virus are a major determinant of infectious particle production and fitness in astrocytes. *J. Gen. Virol.* **95**, 1991–2003 (2014).
370. Takamatsu, Y. *et al.* NS1' protein expression facilitates production of Japanese encephalitis virus in avian cells and embryonated chicken eggs. *J. Gen. Virol.* **95**, 373–383 (2014).
371. Akey, D. L. *et al.* Flavivirus NS1 Structures Reveal Surfaces for Associations with Membranes and the Immune System. *Science* **343**, 881–885 (2014).
372. Rastogi, M., Sharma, N. & Singh, S. K. Flavivirus NS1: a multifaceted enigmatic viral protein. *Virol. J.* **13**, (2016).
373. Liu, J. *et al.* Flavivirus NS1 protein in infected host sera enhances viral acquisition by mosquitoes. *Nat. Microbiol.* **1**, 16087 (2016).

374. Youn, S., Ambrose, R. L., Mackenzie, J. M. & Diamond, M. S. Non-structural protein-1 is required for West Nile virus replication complex formation and viral RNA synthesis. *Viol. J.* **10**, 339 (2013).
375. Strauss, J. H. & Strauss, E. G. With a Little Help from the Host. *Science* **283**, 802–804 (1999).
376. Tao, Y. J. & Ye, Q. RNA Virus Replication Complexes. *PLoS Pathog.* **6**, (2010).
377. Young, L. B., Melian, E. B., Setoh, Y. X., Young, P. R. & Khromykh, A. A. Last 20 aa of the West Nile virus NS1' protein are responsible for its retention in cells and the formation of unique heat-stable dimers. *J. Gen. Virol.* **96**, 1042–1054 (2015).
378. Fan, W. F. & Mason, P. W. Membrane association and secretion of the Japanese encephalitis virus NS1 protein from cells expressing NS1 cDNA. *Virology* **177**, 470–476 (1990).
379. Lin, Y. L. *et al.* DNA immunization with Japanese encephalitis virus nonstructural protein NS1 elicits protective immunity in mice. *J. Virol.* **72**, 191–200 (1998).
380. Falconar, A. K. & Young, P. R. Immunoaffinity purification of native dimer forms of the flavivirus non-structural glycoprotein, NS1. *J. Virol. Methods* **30**, 323–332 (1990).
381. Takamatsu, Y., Raekiansyah, M., Morita, K. & Hayasaka, D. NS1' Protein Expression in the JaOArS982 Strain of Japanese Encephalitis Virus Does Not Enhance Virulence in Mice. *Trop. Med. Health* **43**, 233–237 (2015).
382. Kuno, G., Chang, G.-J. J., Tsuchiya, K. R., Karabatsos, N. & Cropp, C. B. Phylogeny of the Genus Flavivirus. *J. Virol.* **72**, 73–83 (1998).

383. Blitvich, B. J. & Firth, A. E. Insect-Specific Flaviviruses: A Systematic Review of Their Discovery, Host Range, Mode of Transmission, Superinfection Exclusion Potential and Genomic Organization. *Viruses* **7**, 1927–1959 (2015).
384. Cook, S. *et al.* Molecular evolution of the insect-specific flaviviruses. *J. Gen. Virol.* **93**, 223–234 (2012).
385. Blitvich, B. J. & Firth, A. E. A Review of Flaviviruses that Have No Known Arthropod Vector. *Viruses* **9**, (2017).
386. Gould, E. A., Moss, S. R. & Turner, S. L. Evolution and dispersal of encephalitic flaviviruses. *Arch. Virol. Suppl.* 65–84 (2004).
387. St. John, K. Review Paper: The Shape of Phylogenetic Treespace. *Syst. Biol.* **66**, e83–e94 (2017).
388. Nei, M. Phylogenetic Analysis in Molecular Evolutionary Genetics. *Annu. Rev. Genet.* **30**, 371–403 (1996).
389. Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* **13**, 303–314 (2012).
390. Reis, M. dos, Donoghue, P. C. J. & Yang, Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* **17**, 71–80 (2016).
391. Bromham, L. & Penny, D. The modern molecular clock. *Nat. Rev. Genet.* **4**, 216–224 (2003).
392. Drake, J. W. & Holland, J. J. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13910–13913 (1999).
393. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* **9**, 267–276 (2008).

394. Drake, J. W. Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 4171–4175 (1993).
395. Lipson, M. *et al.* Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92–95 (2018).
396. Mathieson, I. *et al.* The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
397. Holmes, E. C., Dudas, G., Rambaut, A. & Andersen, K. G. The Evolution of Ebola virus: Insights from the 2013–2016 Epidemic. *Nature* **538**, 193–200 (2016).
398. Nelson, M. I. *et al.* Global migration of influenza A viruses in swine. *Nat. Commun.* **6**, 6696 (2015).
399. Thézé, J. *et al.* Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico. *Cell Host Microbe* **23**, 855-864.e7 (2018).
400. Bryant, J. E., Holmes, E. C. & Barrett, A. D. T. Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. *PLoS Pathog.* **3**, (2007).
401. Gould, E. A., de Lamballerie, X., Zanotto, P. M. de A. & Holmes, E. C. Origins, evolution, and vector/host coadaptations within the Genus Flavivirus. in *Advances in Virus Research* **59**, 277–314 (Academic Press, 2003).
402. Allicock, O. M. *et al.* Phylogeography and Population Dynamics of Dengue Viruses in the Americas. *Mol. Biol. Evol.* **29**, 1533–1543 (2012).
403. Briese, T., Jia, X. Y., Huang, C., Grady, L. J. & Lipkin, W. I. Identification of a Kunjin/West Nile-like flavivirus in brains of patients with New York encephalitis. *Lancet Lond. Engl.* **354**, 1261–1262 (1999).

404. Lanciotti, R. S. Origin of the West Nile Virus Responsible for an Outbreak of Encephalitis in the Northeastern United States. *Science* **286**, 2333–2337 (1999).
405. May, F. J., Davis, C. T., Tesh, R. B. & Barrett, A. D. T. Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. *J. Virol.* **85**, 2964–2974 (2011).
406. Baillie, G. J. *et al.* Phylogenetic and evolutionary analyses of St. Louis encephalitis virus genomes. *Mol. Phylogenet. Evol.* **47**, 717–728 (2008).
407. Chancey, C., Grinev, A., Volkova, E. & Rios, M. The Global Ecology and Epidemiology of West Nile Virus. *BioMed Research International* (2015). doi:10.1155/2015/376230
408. Bernkopf, H., Levine, S. & Nerson, R. Isolation of West Nile Virus in Israel. *J. Infect. Dis.* **93**, 207–218 (1953).
409. Goldblum, N., Sterk, V. & Paderski, B. WEST NILE FEVER THE CLINICAL FEATURES OF THE DISEASE AND THE ISOLATION OF WEST NILE VIRUS FROM THE BLOOD OF NINE HUMAN CASES. *Am. J. Epidemiol.* **59**, 89–103 (1954).
410. Sejvar, J. J. Clinical Manifestations and Outcomes of West Nile Virus Infection. *Viruses* **6**, 606–623 (2014).
411. Drebot, M. A. & Artsob, H. West Nile virus. *Can. Fam. Physician* **51**, 1094–1099 (2005).
412. Goldblum, N., Jasinska-Klingberg, W., Klingberg, M., Marberg, K. & Sterk, V. The natural history of West Nile Fever. I. Clinical observations during an epidemic in Israel. *Am. J. Hyg.* **64**, 259–269 (1956).
413. Jupp, P. G. The Ecology of West Nile Virus in South Africa and the Occurrence of Outbreaks in Humans. *Ann. N. Y. Acad. Sci.* **951**, 143–152 (2001).

414. Murgue, B., Murri, S., Triki, H., Deubel, V. & Zeller, H. G. West Nile in the Mediterranean basin: 1950-2000. *Ann. N. Y. Acad. Sci.* **951**, 117–126 (2001).
415. Spigland, I., Jasinska-Klingberg, W., Hofshi, E. & Goldbltjm, N. Clinical and Laboratory Observations in an Outbreak of West Nile Fever in Israel in 1957. *Harefuah* **54**, 275–81 (1958).
416. Pruzanski, W. & Altman, R. Encephalitis due to West Nile fever virus. *World Neurol.* **3**, 524–528 (1962).
417. Carey, D. E., Rodrigues, F. M., Myers, R. M. & Webb, J. K. Arthropod-borne viral infections in children in Vellore, South India, with particular reference to dengue and West Nile viruses. *Indian Pediatr.* **5**, 285–296 (1968).
418. George, S., Gourie-Devi, M., Rao, J. A., Prasad, S. R. & Pavri, K. M. Isolation of West Nile virus from the brains of children who had died of encephalitis. *Bull. World Health Organ.* **62**, 879–882 (1984).
419. Flatau, E., Kohn, D., Daher, O. & Varsano, N. West Nile fever encephalitis. *Isr. J. Med. Sci.* **17**, 1057–1059 (1981).
420. Asnis, D. S., Conetta, R., Teixeira, A. A., Waldman, G. & Sampson, B. A. The West Nile Virus Outbreak of 1999 in New York: The Flushing Hospital Experience. *Clin. Infect. Dis.* **30**, 413–418 (2000).
421. Centers for Disease Control and Prevention (CDC). Outbreak of West Nile-like viral encephalitis--New York, 1999. *MMWR Morb. Mortal. Wkly. Rep.* **48**, 845–849 (1999).
422. Nash, D. *et al.* The outbreak of West Nile virus infection in the New York City area in 1999. *N. Engl. J. Med.* **344**, 1807–1814 (2001).

423. Lanciotti, R. S. *et al.* Rapid detection of west nile virus from human clinical specimens, field-collected mosquitoes, and avian samples by a TaqMan reverse transcriptase-PCR assay. *J. Clin. Microbiol.* **38**, 4066–4071 (2000).
424. Reisen, W. K. Ecology of West Nile Virus in North America. *Viruses* **5**, 2079–2105 (2013).
425. Brown, E. B. E. *et al.* Assessing the Risks of West Nile Virus–Infected Mosquitoes from Transatlantic Aircraft: Implications for Disease Emergence in the United Kingdom. *Vector Borne Zoonotic Dis.* **12**, 310–320 (2012).
426. Johnston, B. L. & Conly, J. M. West Nile virus - Where did it come from and where might it go? *Can. J. Infect. Dis.* **11**, 175–178 (2000).
427. Final Cumulative Maps and Data | West Nile Virus | CDC. (2018). Available at: <https://www.cdc.gov/westnile/statsmaps/cumMapsData.html>. (Accessed: 2nd November 2018)
428. Giordano, B. V., Kaur, S. & Hunter, F. F. West Nile virus in Ontario, Canada: A twelve-year analysis of human case prevalence, mosquito surveillance, and climate data. *PLOS ONE* **12**, e0183568 (2017).
429. West Nile virus - Mosquitoes no longer just an annoyance! *Can. J. Infect. Dis.* **14**, 150–153 (2003).
430. Drebot, M. A. *et al.* West Nile virus surveillance and diagnostics: A Canadian perspective. *Can. J. Infect. Dis.* **14**, 105–114 (2003).
431. David Roth *et al.* West Nile Virus Range Expansion into British Columbia. *Emerg. Infect. Dis. J.* **16**, 1251 (2010).
432. Canada, P. H. A. of & Canada, P. H. A. of. West Nile virus and other mosquito-borne disease national surveillance report. *aem* (2015). Available at:

- <https://www.canada.ca/en/public-health/services/diseases/west-nile-virus/west-nile-virus-other-mosquito-borne-disease.html#y2016>. (Accessed: 4th November 2018)
433. Chappell, J. D. & Dermody, T. S. 134 - Biology of Viruses and Viral Diseases. in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition)* (eds. Bennett, J. E., Dolin, R. & Blaser, M. J.) 1681-1693.e4 (Content Repository Only!, 2015). doi:10.1016/B978-1-4557-4801-3.00134-X
434. Chapter 2 - Mosquito-Borne Diseases. in *Zika Virus Disease* (ed. Qureshi, A. I.) 27–45 (Academic Press, 2018). doi:10.1016/B978-0-12-812365-2.00003-2
435. Barrett, A. D. T. & Weaver, S. C. 51 - Arboviruses: alphaviruses, flaviviruses and bunyaviruses: Encephalitis; yellow fever; dengue; haemorrhagic fever; miscellaneous tropical fevers; undifferentiated fever. in *Medical Microbiology (Eighteenth Edition)* (eds. Greenwood, D., Barer, M., Slack, R. & Irving, W.) 520–536 (Churchill Livingstone, 2012). doi:10.1016/B978-0-7020-4089-4.00066-4
436. Burrell, C. J., Howard, C. R. & Murphy, F. A. Chapter 36 - Flaviviruses. in *Fenner and White's Medical Virology (Fifth Edition)* (eds. Burrell, C. J., Howard, C. R. & Murphy, F. A.) 493–518 (Academic Press, 2017). doi:10.1016/B978-0-12-375156-0.00036-9
437. Colpitts, T. M., Conway, M. J., Montgomery, R. R. & Fikrig, E. West Nile Virus: Biology, Transmission, and Human Infection. *Clin. Microbiol. Rev.* **25**, 635–648 (2012).
438. Murray, K. O., Walker, C. & Gould, E. The virology, epidemiology, and clinical impact of West Nile virus: a decade of advancements in research since its introduction into the Western Hemisphere. *Epidemiol. Infect.* **139**, 807–817 (2011).
439. Gubler, D. J. The continuing spread of West Nile virus in the western hemisphere. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **45**, 1039–1046 (2007).

440. Molaei, G., Andreadis, T. G., Armstrong, P. M., Anderson, J. F. & Vossbrinck, C. R. Host feeding patterns of *Culex* mosquitoes and West Nile virus transmission, northeastern United States. *Emerg. Infect. Dis.* **12**, 468–474 (2006).
441. Rappole, J. H., Derrickson, S. R. & Hubálek, Z. Migratory birds and spread of West Nile virus in the Western Hemisphere. *Emerg. Infect. Dis.* **6**, 319–328 (2000).
442. Hamer, G. L. *et al.* Host Selection by *Culex pipiens* Mosquitoes and West Nile Virus Amplification. *Am. J. Trop. Med. Hyg.* **80**, 268–278 (2009).
443. Kent, R., Juliusson, L., Weissmann, M., Evans, S. & Komar, N. Seasonal blood-feeding behavior of *Culex tarsalis* (Diptera: Culicidae) in Weld County, Colorado, 2007. *J. Med. Entomol.* **46**, 380–390 (2009).
444. Kilpatrick, A. M., Daszak, P., Jones, M. J., Marra, P. P. & Kramer, L. D. Host heterogeneity dominates West Nile virus transmission. *Proc. R. Soc. Lond. B Biol. Sci.* **273**, 2327–2333 (2006).
445. Hayes, E. B. *et al.* Epidemiology and Transmission Dynamics of West Nile Virus Disease. *Emerg. Infect. Dis.* **11**, 1167–1173 (2005).
446. Troupin, A. & Colpitts, T. M. Overview of West Nile Virus Transmission and Epidemiology. in *West Nile Virus: Methods and Protocols* (ed. Colpitts, T. M.) 15–18 (Springer New York, 2016). doi:10.1007/978-1-4939-3670-0_2
447. Sampathkumar, P. West Nile Virus: Epidemiology, Clinical Presentation, Diagnosis, and Prevention. *Mayo Clin. Proc.* **78**, 1137–1144 (2003).
448. Turell, M. J., O’Guinn, M. L., Dohm, D. J. & Jones, J. W. Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus. *J. Med. Entomol.* **38**, 130–134 (2001).

449. Dohm, D. J., Sardelis, M. R. & Turell, M. J. Experimental vertical transmission of West Nile virus by *Culex pipiens* (Diptera: Culicidae). *J. Med. Entomol.* **39**, 640–644 (2002).
450. Bowen, R. A. & Nemeth, N. M. Experimental infections with West Nile virus. *Curr. Opin. Infect. Dis.* **20**, 293 (2007).
451. Diamond, M. S. Virus and Host Determinants of West Nile Virus Pathogenesis. *PLOS Pathog.* **5**, e1000452 (2009).
452. Rios, M., Daniel, S., Chancey, C., Hewlett, I. K. & Stramer, S. L. West Nile Virus Adheres to Human Red Blood Cells in Whole Blood. *Clin. Infect. Dis.* **45**, 181–186 (2007).
453. Meny, G. M., Santos-Zabala, L., Szallasi, A. & Stramer, S. L. West Nile virus infection transmitted by granulocyte transfusion. *Blood* **117**, 5778–5779 (2011).
454. Pisani, G., Cristiano, K., Pupella, S. & Liunbruno, G. M. West Nile Virus in Europe and Safety of Blood Transfusion. *Transfus. Med. Hemotherapy* **43**, 158–167 (2016).
455. Pealer, L. N. *et al.* Transmission of West Nile Virus through Blood Transfusion in the United States in 2002. *N. Engl. J. Med.* **349**, 1236–1245 (2003).
456. Iwamoto, M. *et al.* Transmission of West Nile Virus from an Organ Donor to Four Transplant Recipients. *N. Engl. J. Med.* **348**, 2196–2203 (2003).
457. Stewart, R. D., Bryant, S. N. & Sheffield, J. S. West Nile Virus Infection in Pregnancy. *Case Reports in Infectious Diseases* (2013). doi:10.1155/2013/351872
458. Centers for Disease Control and Prevention (CDC). Intrauterine West Nile virus infection--New York, 2002. *MMWR Morb. Mortal. Wkly. Rep.* **51**, 1135–1136 (2002).
459. Hinckley, A. F., O’Leary, D. R. & Hayes, E. B. Transmission of West Nile Virus Through Human Breast Milk Seems to Be Rare. *Pediatrics* **119**, e666–e671 (2007).

460. Centers for Disease Control and Prevention (CDC). Possible West Nile virus transmission to an infant through breast-feeding--Michigan, 2002. *MMWR Morb. Mortal. Wkly. Rep.* **51**, 877–878 (2002).
461. Montgomery, S. P. *et al.* Transfusion-associated transmission of West Nile virus, United States 2003 through 2005. *Transfusion (Paris)* **46**, 2038–2046 (2006).
462. Centers for Disease Control and Prevention (CDC). Fatal West Nile virus infection after probable transfusion-associated transmission--Colorado, 2012. *MMWR Morb. Mortal. Wkly. Rep.* **62**, 622–624 (2013).
463. Petersen, L. R. & Busch, M. P. Transfusion-transmitted arboviruses. *Vox Sang.* **98**, 495–503 (2010).
464. Seino, K. K. *et al.* Comparative Efficacies of Three Commercially Available Vaccines against West Nile Virus (WNV) in a Short-Duration Challenge Trial Involving an Equine WNV Encephalitis Model. *Clin. Vaccine Immunol. CVI* **14**, 1465–1471 (2007).
465. Ostlund, E. N. *et al.* Equine West Nile encephalitis, United States. *Emerg. Infect. Dis.* **7**, 665–669 (2001).
466. Turell, M. J. *et al.* DNA vaccine for West Nile virus infection in fish crows (*Corvus ossifragus*). *Emerg. Infect. Dis.* **9**, 1077–1081 (2003).
467. Yang, J. S. *et al.* Induction of potent Th1-type immune responses from a novel DNA vaccine for West Nile virus New York isolate (WNV-NY1999). *J. Infect. Dis.* **184**, 809–816 (2001).
468. Minke, J. M. *et al.* Recombinant canarypoxvirus vaccine carrying the prM/E genes of West Nile virus protects horses against a West Nile virus-mosquito challenge. *Arch. Virol. Suppl.* 221–230 (2004).

469. Davidson, A. H. *et al.* Immunologic responses to West Nile virus in vaccinated and clinically affected horses. *J. Am. Vet. Med. Assoc.* **226**, 240–245 (2005).
470. Gardner, I. A. *et al.* Incidence and effects of West Nile virus infection in vaccinated and unvaccinated horses in California. *Vet. Res.* **38**, 109–116 (2007).
471. Salazar, P. *et al.* Outcome of equids with clinical signs of West Nile virus infection and factors associated with death. *J. Am. Vet. Med. Assoc.* **225**, 267–274 (2004).
472. Schuler, L. A., Khaitsa, M. L., Dyer, N. W. & Stoltenow, C. L. Evaluation of an outbreak of West Nile virus infection in horses: 569 cases (2002). *J. Am. Vet. Med. Assoc.* **225**, 1084–1089 (2004).
473. Ward, M. P. *et al.* Investigation of an outbreak of encephalomyelitis caused by West Nile virus in 136 horses. *J. Am. Vet. Med. Assoc.* **225**, 84–89 (2004).
474. Agrawal, A. G. & Petersen, L. R. Human immunoglobulin as a treatment for West Nile virus infection. *J. Infect. Dis.* **188**, 1–4 (2003).
475. Kalil, A. C. *et al.* Use of interferon-alpha in patients with West Nile encephalitis: report of 2 cases. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **40**, 764–766 (2005).
476. Sayao, A.-L. *et al.* Calgary experience with West Nile virus neurological syndrome during the late summer of 2003. *Can. J. Neurol. Sci. J. Can. Sci. Neurol.* **31**, 194–203 (2004).
477. Anderson, J. F. & Rahal, J. J. Efficacy of interferon alpha-2b and ribavirin against West Nile virus in vitro. *Emerg. Infect. Dis.* **8**, 107–108 (2002).
478. Morrey, J. D. *et al.* Effect of interferon-alpha and interferon-inducers on West Nile virus in mouse and hamster animal models. *Antivir. Chem. Chemother.* **15**, 101–109 (2004).
479. Gea-Banacloche, J. *et al.* West Nile virus: pathogenesis and therapeutic options. *Ann. Intern. Med.* **140**, 545–553 (2004).

480. Hu, H.-T., Cho, C.-P., Lin, Y.-H. & Chang, K.-Y. A general strategy to inhibiting viral -1 frameshifting based on upstream attenuation duplex formation. *Nucleic Acids Res.* **44**, 256–266 (2016).
481. Neuman, K. C. & Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **5**, 491–505 (2008).
482. Ritchie, D. B., Woodside, M. T. & Halma, M. T. Studying the structural dynamics of nucleic acids with single-molecule force spectroscopy.
483. Green, N. M. Avidin. in *Advances in Protein Chemistry* (eds. Anfinsen, C. B., Edsall, J. T. & Richards, F. M.) **29**, 85–133 (Academic Press, 1975).
484. Chevalier, J., Yi, J., Michel, O. & Tang, X.-M. Biotin and Digoxigenin as Labels for Light and Electron Microscopy in Situ Hybridization Probes: Where Do We Stand? *J. Histochem. Cytochem.* **45**, 481–491 (1997).
485. Marko-Varga, G. & Oroszlan, P. *Emerging Technologies in Protein and Genomic Material Analysis*. (Elsevier, 2003).
486. Williams, M. C., Wenner, J. R., Rouzina, I. & Bloomfield, V. A. Effect of pH on the overstretching transition of double-stranded DNA: evidence of force-induced DNA melting. *Biophys. J.* **80**, 874–81 (2001).
487. Chen, H. *et al.* Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 799–804 (2012).
488. Lodish, H. *et al.* Intracellular Ion Environment and Membrane Electric Potential. *Mol. Cell Biol.* 4th Ed. (2000).
489. Bohr, editor of compilation.), Henri. *Handbook of molecular biophysics : methods and applications*. (Weinheim : Wiley-VCH, 2009).

490. Kruse, C. R. *et al.* The effect of pH on cell viability, cell migration, cell proliferation, wound closure, and wound reepithelialization: In vitro and in vivo study. *Wound Repair Regen.* **25**, 260–269 (2017).
491. Neuman, K., Chadd, E., Liou, G., Bergman, K. & Block, S. Characterization of photodamage to escherichia coli in optical traps. *Biophys. J.* **77**, 2856–2863 (1999).
492. Peterman, E. J. G., Gittes, F. & Schmidt, C. F. Laser-Induced Heating in Optical Traps. *Biophys. J.* **84**, 1308–1316 (2003).
493. Benesch, R. E. & Benesch, R. Enzymatic Removal of Oxygen for Polarography and Related Methods. *Science* **118**, 447–448 (1953).
494. K Dickson, M Haigis & R Raines. Ribonuclease Inhibitor: Structure and Function. *Prog Nucleic Acid Res Mol Biol* **80**, 349–374 (2005).
495. Working with RNA: the basics. (2012).
496. Shaevitz, J. W. A Practical Guide to Optical Trapping. 19
497. S Chu, JE Bjorkholm, A Ashkin & A Cable. Experimental Observation of Optically Trapped Atoms. *Phys. Rev. Lett.* **57**, 314 (1986).
498. Neuman, K. C. & Block, S. M. Optical trapping. *Rev. Sci. Instrum.* **75**, 2787–2809 (2004).
499. Mie, G. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Ann. Phys.* **330**, 377–445 (1908).
500. Rayleigh, Lord. XXXIV. *On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky.* Lond. Edinb. Dublin *Philos. Mag. J. Sci.* **47**, 375–384 (1899).

501. Li, T. Physical Principle of Optical Tweezers. in *Fundamental Tests of Physics with Optically Trapped Microspheres* (ed. Li, T.) 9–20 (Springer New York, 2013).
doi:10.1007/978-1-4614-6031-2_2
502. Le Gall, A. *et al.* Simultaneous calibration of optical tweezers spring constant and position detector response. *Opt. Express* **18**, 26469 (2010).
503. Acousto-Optics. in *Fundamentals of Photonics* 799–831 (John Wiley & Sons, Ltd, 2001).
doi:10.1002/0471213748.ch20
504. Sangwal, K. & Kucharczyk, W. Relationship between density and refractive index of inorganic solids. *J. Phys. Appl. Phys.* **20**, 522 (1987).
505. Neuman, K. C. & Block, S. M. Optical trapping. *Rev. Sci. Instrum.* **75**, 2787–2809 (2004).
506. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
507. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599–1600 (1994).
508. Lai, P.-Y. & Zhou, Z. B- to S-form transition in double-stranded DNA with basepair interactions. *Phys. Stat. Mech. Its Appl.* **321**, 170–180 (2003).
509. Williams, M. C., Rouzina, I. & McCauley, M. J. Peeling back the mystery of DNA overstretching. *Proc. Natl. Acad. Sci.* **106**, 18047–18048 (2009).
510. King, G. A. *et al.* Revealing the competition between peeled ssDNA, melting bubbles, and S-DNA during DNA overstretching using fluorescence microscopy. *Proc. Natl. Acad. Sci.* **110**, 3859–3864 (2013).
511. Gross, P. *et al.* Quantifying how DNA stretches, melts and changes twist under tension. *Nat. Phys.* **7**, 731–736 (2011).

512. de Lorenzo, S., Ribezzi-Crivellari, M., Ricardo Arias-Gonzalez, J., Smith, S. B. & Ritort, F. A Temperature-Jump Optical Trap for Single-Molecule Manipulation. *Biophys. J.* **108**, 2854–2864 (2015).
513. Baumann, C. G., Smith, S. B., Bloomfield, V. A. & Bustamante, C. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci.* **94**, 6185–6190 (1997).
514. Brunet, A. *et al.* Dependence of DNA Persistence Length on Ionic Strength of Solutions with Monovalent and Divalent Salts: A Joint Theory-Experiment Study. *Macromolecules* **48**, 3641–3652 (2015).
515. Zarudnaya, M. I. *et al.* Dependence of DNA persistence length on ionic conditions. *Biopolym. Cell* **33**, 81–91 (2017).
516. Tinland, B., Pluen, A., Sturm, J. & Weill, G. Persistence Length of Single-Stranded DNA. *Macromolecules* **30**, 5763–5765 (1997).
517. Tan, Z. & Chen, S.-J. Predicting Electrostatic Forces in RNA Folding. *Methods Enzymol.* **469**, 465–87 (2009).
518. Heilman-Miller, S. L., Thirumalai, D. & Woodson, S. A. Role of counterion condensation in folding of the Tetrahymena ribozyme. I. Equilibrium stabilization by cations. *J. Mol. Biol.* **306**, 1157–1166 (2001).
519. Lima, C. H. M., Rocha, M. S. & Ramos, E. B. Unfolding DNA condensates produced by DNA-like charged depletants: A force spectroscopy study. *J. Chem. Phys.* **146**, (2017).
520. Greenleaf, W. J., Frieda, K. L., Foster, D. A. N., Woodside, M. T. & Block, S. M. Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **319**, 630–633 (2008).

521. Herrero-Galán, E. *et al.* Mechanical Identities of RNA and DNA Double Helices Unveiled at the Single-Molecule Level. *J. Am. Chem. Soc.* **135**, 122–131 (2013).
522. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* **271**, 795–799 (1996).
523. Bryant, Z. *et al.* Structural transitions and elasticity from torque measurements on DNA. *Nature* **424**, 338–341 (2003).
524. Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. Stretching DNA with optical tweezers. *Biophys. J.* **72**, 1335–1346 (1997).
525. Woodside, M. T. *et al.* Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science* **314**, 1001–1004 (2006).
526. Saenger, W. *Principles of Nucleic Acid Structure*. (Springer New York, 1984).
527. Chi, Q., Wang, G. & Jiang, J. The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Phys. Stat. Mech. Its Appl.* **392**, 1072–1079 (2013).
528. Woodside, M. T. *et al.* Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6190–6195 (2006).
529. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
530. Houck-loomis, B. *et al.* An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* **480**, 561–564 (2011).

531. Taylor, P., Rixon, F. & Desselberger, U. Rise per base pair in helices of double-stranded rotavirus RNA determined by electron microscopy. *Virus Res.* **2**, 175–182 (1985).
532. Waters, J. T. *et al.* Transitions of Double-Stranded DNA Between the A- and B-Forms. *J. Phys. Chem. B* **120**, 8449–8456 (2016).
533. Sinden, R. R. CHAPTER 1 - Introduction to the Structure, Properties, and Reactions of DNA. in *DNA Structure and Function* (ed. Sinden, R. R.) 1–57 (Academic Press, 1994). doi:10.1016/B978-0-08-057173-7.50006-7
534. Lang, D., Steely, H. T., Kao, C.-Y. & Ktistakis, N. T. Length, mass, and denaturation of double-stranded RNA molecules compared with DNA. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* **910**, 271–281 (1987).
535. Lipfert, J. *et al.* Double-stranded RNA under force and torque: Similarities to and striking differences from double-stranded DNA. *Proc. Natl. Acad. Sci.* **111**, (2014).
536. Liphardt, J. Reversible Unfolding of Single RNA Molecules by Mechanical Force. *Science* **292**, 733–737 (2001).
537. Cao, S. & Chen, S.-J. Predicting structures and stabilities for H-type pseudoknots with interhelix loops. *RNA* **15**, 696–706 (2009).
538. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
539. Kato, Y., Sato, K., Asai, K. & Akutsu, T. Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming. *Nucleic Acids Res.* **40**, W29-34 (2012).
540. Bellaousov, S. & Mathews, D. H. ProbKnot: Fast prediction of RNA secondary structure including pseudoknots. *RNA* **16**, 1870–1880 (2010).

541. Andronescu, M. S., Pop, C. & Condon, A. E. Improved free energy parameters for RNA pseudoknotted secondary structure prediction. *RNA N. Y. N* **16**, 26–42 (2010).
542. Wu, Y.-J., Wu, C.-H., Yeh, A. Y.-C. & Wen, J.-D. Folding a stable RNA pseudoknot through rearrangement of two hairpin structures. *Nucleic Acids Res.* **42**, 4505–4515 (2014).
543. Dinman, J. D., University of Maryland. Personal Communication, Unpublished Data.
544. SHANNON, C. E. A MATHEMATICAL THEORY OF COMMUNICATION. *BELL Syst. Tech. J.* **27**, 379–423 (1948).
545. Wang, Y. *et al.* Comparative studies of frameshifting and nonframeshifting RNA pseudoknots: a mutational and NMR investigation of pseudoknots derived from the bacteriophage T2 gene 32 mRNA and the retroviral gag-pro frameshift site. *RNA N. Y. N* **8**, 981–96 (2002).
546. Nixon, P. L. *et al.* Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J. Mol. Biol.* **322**, 621–633 (2002).
547. Michiels, P. J. *et al.* Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* **310**, 1109–23 (2001).
548. Chamorro, M., Parkin, N. & Varmus, H. E. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc. Natl. Acad. Sci.* **89**, 713–717 (1992).
549. Shen, L. X. & Tinoco, I. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* **247**, 963–978 (1995).
550. Pennell, S. *et al.* The stimulatory RNA of the Visna-Maedi retrovirus ribosomal frameshifting signal is an unusual pseudoknot with an interstem element. *RNA* **14**, 1366–1377 (2008).

551. Jill Harrison, C. & Langdale, J. A. A step by step guide to phylogeny reconstruction. *Plant J.* **45**, 561–572 (2006).
552. De Bruyn, A., Martin, D. P. & Lefeuvre, P. Phylogenetic reconstruction methods: an overview. *Methods Mol. Biol. Clifton NJ* **1115**, 257–277 (2014).
553. Singh, G. B. Introduction to Phylogenetic Reconstruction. in *Fundamentals of Bioinformatics and Computational Biology: Methods and Exercises in MATLAB* (ed. Singh, G. B.) 235–252 (Springer International Publishing, 2015). doi:10.1007/978-3-319-11403-3_13
554. Joy, J. B., Liang, R. H., McCloskey, R. M., Nguyen, T. & Poon, A. F. Y. Ancestral Reconstruction. *PLOS Comput. Biol.* **12**, e1004763 (2016).
555. Matsumoto, T., Akashi, H. & Yang, Z. Evaluation of Ancestral Sequence Reconstruction Methods to Infer Nonstationary Patterns of Nucleotide Substitution. *Genetics* **200**, 873–890 (2015).
556. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
557. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
558. Somerset, S. M. & Johannot, L. Dietary flavonoid sources in Australian adults. *Nutr. Cancer* **60**, 442–449 (2008).
559. Chen, A. Y. & Chen, Y. C. A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. *Food Chem.* **138**, 2099–2107 (2013).
560. Manach, C. *et al.* Bioavailability, metabolism and physiological impact of 4-oxo-flavonoids. *Nutr. Res.* **16**, 517–544 (1996).

561. Jeong, H. J. *et al.* Neuraminidase inhibitory activities of flavonols isolated from *Rhodiola rosea* roots and their in vitro anti-influenza viral activities. *Bioorg. Med. Chem.* **17**, 6816–6823 (2009).
562. Li, J. *et al.* In vitro and in vivo anti-hepatitis B virus activities of a plant extract from *Geranium carolinianum* L. *Antiviral Res.* **79**, 114–120 (2008).
563. Nicotra, F. *Organic and Bio-molecular Chemistry - Volume II.* (EOLSS Publications, 2009).
564. Xie, C.-I. *et al.* Daidzin, an Antioxidant Isoflavonoid, Decreases Blood Alcohol Levels and Shortens Sleep Time Induced by Ethanol Intoxication. *Alcohol. Clin. Exp. Res.* **18**, 1443–1447 (1994).
565. Choi, E. J. & Kim, G.-H. The antioxidant activity of daidzein metabolites, O-desmethylangolensin and equol, in HepG2 cells. *Mol. Med. Rep.* **9**, 328–332 (2014).
566. Tatsimo, S. J. N. *et al.* Antimicrobial and antioxidant activity of kaempferol rhamnoside derivatives from *Bryophyllum pinnatum*. *BMC Res. Notes* **5**, 158 (2012).
567. Velloso, J. C. R. *et al.* Antioxidant and cytotoxic studies for kaempferol, quercetin and isoquercitrin. *Eclética Quím.* **36**, 07–20 (2011).
568. Johari, J., Kianmehr, A., Mustafa, M. R., Abubakar, S. & Zandi, K. Antiviral activity of baicalein and quercetin against the Japanese encephalitis virus. *Int. J. Mol. Sci.* **13**, 16785–16795 (2012).
569. Ho, C. *et al.* Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* **24**, 3–12 (2003).
570. Duff, M. R., Grubbs, J. & Howell, E. E. Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity. *J. Vis. Exp. JoVE* (2011). doi:10.3791/2796

571. Forli, S. *et al.* Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **11**, 905–919 (2016).
572. IUPAC codes. Available at: <http://www.bioinformatics.org/sms2/iupac.html>. (Accessed: 18th October 2018)