Dynamics, Degradation, and Chemical Modification of Non-Coding RNA

by

John A. H. Hoerter

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2008

Doctoral Committee:

Associate Professor Nils G. Walter, Chair
Professor Michael D. Morris
Professor Vincent L. Pecoraro
Associate Professor Janine R. Maddock
Associate Professor Anna K. Mapp

for the loves of my life, Perrine and Camille

ACKNOWLEDGEMENTS

It is impossible to overstate the debt of gratitude I owe to my scientific mentors, family, and friends for their support and guidance during my graduate studies. First and foremost, my advisor Nils Walter has been a generous and incredibly supportive mentor, whose scientific acumen I admire and have taken advantage of through discussion of various issues, whether or not they have been directly related to my thesis work. The other members of my thesis committee, Vince Pecoraro, Mike Morris, Janine Maddock, and Anna Mapp have been very supportive throughout my tenure. I asked them to serve on my committee because I found them to be excellent teachers and scientists. Other mentors have included Chuck Barnes, Bert Hornback, David Rueda, Pat Holt, Hans Tanner, and Tom Hopkins.

My wife Perrine has tirelessly supported me and my work, and I look forward to our future as we grow and develop together. Our daughter Camille is the light of our lives, and we are proudly encouraging her to make her place in the world. My parents Nancy and Joseph have constantly promoted me and my brothers. Their support, intellectual stimulation, and love have always encouraged me to challenge myself. My brothers Christopher and Patrick are the few kindred-spirits I have found, and we experience and share emotional, personal, and intellectual development similarly, and often together. The rest of my family has also been very supportive.

My friends during graduate school have been true comrades, and I thank them for the pleasure of their company and the various extracurricular activities in which we have engaged. I also thank the members of the Walter Lab, current and past, for an excellent environment in which to study RNA.

John Hoerter
San Diego, California
30 November 2007

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

RNA	ribonucleic acid
ncRNA	non-coding RNA
EF-G	elongation factor G
mRNA	messenger RNA
tRNA	transfer RNA
rRNA	ribosomal RNA
snRNA	small nucleolar RNA
EF-Tu	elongation factor Tu
ATP	adenosine triphosphate
GTP	guanosine triphosphate
CTP	cytosine triphosphate
UTP	uridine triphosphate
UTR	untranslated region
HDV	hepatitis delta virus
VS	Varkud satellite
siRNA	small interfering RNA
miRNA	micro RNA
RNAi	RNA interference
NRON	non-coding repressor of NFAT
NFAT	nuclear factor of activated T-cells
RISC	RNA induced silencing complex
smFRET	single molecule FRET
FRET	fluorescence resonance energy transfer
RNP	ribonuclear protein
cryo-EM	cryo-electron microscopy
Ago2	Argonaute2
FWHM	full-width at half maximum
Sm	streptomycin
DMS	dimethylsulfate
dsRNA	double-stranded RNA
kb	kilo-base
A	adenosine
G	guanosine
C	cytosine
T	thymidine
U	uracil

GlcN6P	glucosamine-6-phosphate
RLC	RISC loading complex
HIV	human immunodeficiency virus
TAR	trans-activation response element
NMR	nuclear magnetic resonance (spectroscopy)
DTT	dithiothreitol
NTP	nucleotide triphosphate
EDTA	ethylenediaminetetraacetic
SDS	sodium dodecyl sulfate
RP-HPLC	reversed phase high performance liquid chromatography
DMSO	dimethylsulfoxide
FID	free induction decay
FL	fluorescein
TMR	tetramethylrhodamine
NOE(SY)	nuclear Overhauser effect (spectroscopy)
RNase	ribonuclease
MD	molecular dynamics
SL	spliced leader
tr-FRET	time-resolved FRET
HSQC	heteronuclear single quantum correlation
MES	2-(N-morpholino)ethanesulfonic acid
PNK	polynucleotide kinase
hpRNase	human pancreatic RNase
QDot	quantum dot
EDC	1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
MWCO	molecular weight cut-off
PCR	polymerase chain reaction
TBE	tris-borate-EDTA buffer
APES	(3-aminopropyl)triethoxy silane
DST	disuccinimidyl tartarate
PEG	poly(ethylene glycol)
TIRF	total internal reflection fluorescence
CCD	charge-coupled device
PAGE	poly-acrylamide gel electrophoresis
CAT	chloramphenicol acetyltransferase
WT	wild type
SD	standard deviation
RMS	root mean square
kDa	kilo-Dalton
GITC	2,3,4,6-tetra-O-acetyl-β -D-glucopyranosyl isothiocyanate

ABSTRACT

Our understanding of the function of RNA has grown significantly since the central dogma of molecular biology described RNA as a rudimentary conveyor of the genetic message. In the genomic era, we are beginning to learn the true depth and breadth of the non-coding (nc)RNA repertoire. A key attribute of RNA molecules is the conformational dynamics that they explore. Here, we have applied studies of structure, dynamics, and metal binding to models of helix 27 from 16S rRNA to elucidate the function of this important component of the ribosome. Our studies reveal the kinetic and thermodynamic framework within which this isolated helix undergoes secondary structure rearrangement. Both NMR and fluorescence techniques demonstrate millisecond exchange between the 885 and 888 conformations, defining an equilibrium constant close to one. Fluorescence studies also show that the antibiotic tetracycline interferes with this conformational exchange. Metal binding studies of helix 27 have confirmed data from ribosomal crystal structures and further correlated local and global metal binding features of this RNA.

Another general attribute of RNA is its inherent lability. The great promise of a new class of therapeutics based on small interfering (si)RNA molecules and the RNA interference pathway has forced researchers to overcome the nucleolytic vulnerability of RNA molecules, primarily by introducing chemical modifications. Work in this thesis demonstrates that siRNA degradation in blood serum is asymmetric, where the guide strand is predisposed to efficient degradation due to differential stability of the terminal base pairs. We further show that a simple pattern of chemical modifications greatly stabilizes siRNAs in regions particularly susceptible to nuclease cleavage. We have shown that 21 and 24-nucleotide siRNA-like double-stranded RNAs are specifically protected in cell extract, a result which demonstrates intracellular siRNA stability and may help explain recent results suggesting that, once inside the cell, chemical

modifications to siRNAs do not significantly increase the potency of the silencing effect. Taken together, the work presented in this thesis has helped to illuminate the function of important ncRNA molecules, furthering our understanding of how ncRNA contributes to the structural, catalytic, and regulatory landscape that defines the cellular lifecycle.

CHAPTER 1:

NON-CODING RNA IN THE GENOMIC ERA

1.1 Introduction

The scientific era subsequent to the publication of the human genome sequence (*1, 2*) has been—and continues to be—a revolutionary time in biomedicine. It was thought that knowledge of the sequence of human genes would greatly advance functional and therapeutic studies and finally provide definitive evidence for the origins of human complexity. While the human genome sequence has enabled rapid progress in the field of functional genomics, the simple knowledge of DNA sequence has proven insufficient to explain human complexity relative to lower organisms. Expectations for the number of protein coding genes in the human genome were greatly overestimated; the sheer size of the human genome (3×10^9 bases) suggested that the number of coding genes in humans would be much higher than the 21,858 noted in the Ensembl human genome annotation from September 2007, a total number of protein coding genes that is approximately equal to that of the worm *Caenorhabditis elegans*. It is known that >98.5% of the human genome does not code for protein (*3, 4*). The precise fraction of the genome that is actually transcribed into RNA is unclear. Published data suggests that 10-60% of the mamallian genome is transcribed (*5-8*), a large range that is reflective of the different techniques used to probe for transcription. However, it has been speculated that all of the non-repetitive elements of the genome are transcribed (*8*). This pervasive transcription of the human genome and new discoveries describing an expanded regulatory scope for RNA make it clear that the roles non-coding (nc)RNAs play in the cell are significant, and may harbor the keys to understanding the developmental and regulatory networks that underlie human complexity.

Phenomenologically, a number of disease states have been identified as originating from mutations in ncRNA, including a host of metabolic disorders associated with anticodon-loop mutations in a mitochondrial tRNA (*9*) and Type II diabetes derived from mutations in intronic and intergenic regions (*10*). Other ncRNAs that affect fundamental cellular processes have been identified, including six ncRNAs essential for cell viability, one repressor of Hedgehog signaling, and one repressor of the transcription factor nuclear-factor of activated T-cells (NFAT) (*11*).

ncRNAs are a diverse group of RNA molecules with a combination of structural, catalytic, or regulatory functions that range in size from alternatively spliced transcripts many kilo-bases (kb) in length to 22-nucleotide small microRNAs (*4*). Structural ncRNAs include small nuclear (sn)RNAs, ribosomal (r)RNAs, and transfer (t)RNAs. These molecules primarily provide scaffolds and elements for the sequence-specific recognition of RNA substrates. snRNAs assemble with a number of protein factors to form the spliceosome, the eukaryotic complex that is responsible for splicing introns out of pre-mRNAs before mRNAs are exported from the nucleus for translation. The snRNAs provide critical recognition of elements of the pre-mRNA including the branch site adenosine (*12*). In addition to the essential recognition and structural roles played by the snRNAs, evidence is emerging that snRNAs may also be the catalytic moiety of the splicesome (*13, 14*). Another group of structural RNAs, the rRNAs, co-assemble with more than 50 proteins to form the two ribosomal subunits. The 30S subunit and 16S rRNA binds the messenger (m)RNA and monitors the interaction of the tRNA anticodon and the mRNA codon. The universally conserved 16S rRNA bases A1492, A1493, and G530 intimately interact with the minor groove of the first two codon:anticodon base pairs, monitoring the presence of cognate tRNA in part via A-minor motifs (*15-17*). Once the presence of a cognate tRNA has been verified, a series of long-range interactions that still remain elusive transmit the correct codon:anticodon base pairing from the 30S subunit to the GTPase factor binding region of the 50S subunit, where EF-

Tu is induced to hydrolyze GTP while the ribosome moves toward full tRNA accommodation (*17*). Once the tRNA has been accommodated into the peptidyl-transferase site of the 50S subunit, the CCA ends of the A and P site tRNAs are aligned by conserved bases in 23S rRNA. The precise alignment of the tRNA, and by extension, alignment of the peptidyl-tRNA ester and the nucleophilic primary amine of the aminoacyl tRNA, provide for a lowering of the entropy of activation, accelerating the reaction by a factor of 2×10^7 over the uncatalyzed reaction (*18*).

Catalytic ncRNAs include the *glmS* riboswitch and the class of small ribozymes from sub-viral particles. The *glmS* riboswitch is an RNA structure located in the 5'-untranslated region (UTR) of the *glmS* mRNA from *B. subtilis*. The *glmS* protein product synthesizes the small molecule metabolite glucosamine-6-phosphate (GlcN6P), and recent results demonstrate that the 5'-UTR of the *glmS* mRNA harbors a ribozyme that activates upon the binding of GlcN6P, finally resulting in reduced expression of the *glmS* gene (*19, 20*). This example of a negative feedback loop regulated by the action of a metabolite responsive ribozyme is a novel mechanism for regulation of gene expression by a catalytic RNA. The small ribozymes are a structurally diverse class of RNAs named the hammerhead, hairpin, hepatitis delta virus (HDV), and *Neurospora* Varkud satellite (VS) ribozymes. Each of these ribozymes undergo a complex series of conformational changes associated with the various intermediates along the reaction pathway, finally catalyzing the site specific cleavage of its own RNA backbone (*21*). The cleavage events generate monomeric copies from rolling circle transcribed multimeric copies of their parent sub-viral genomes. These ribozymes have served as proto-typical systems for understanding the interplay of dynamics with the catalytic cycle in RNA systems (*22-24*).

Regulatory RNAs include small interfering (si)RNAs, micro (mi)RNAs, and the NRON RNA which acts as a repressor of the transcription factor NFAT (*11*). RNA interference (RNAi) is a cellular pathway that utilizes siRNAs and miRNAs to attenuate gene expression through different, yet related mechanisms. RNAi and siRNAs are discussed in more detail in subsequent sections of Chapter 1 of this thesis. miRNAs are expressed and processed into 21-23 nucleotide RNAs that go on to be incorporated into the RISC complex. Whereas siRNAs have perfect complementarity to the target and guide the cleavage of mRNA, miRNAs imperfectly base pair inside the 3'-UTR of mRNAs and lead to the inhibition of translation and sequestration of mRNA. miRNAs have been implicated in regulating developmental timing and, by consequence, in genetic mis-regulation associated with cancer (*25, 26*). Yet another example of a regulatory RNA is the non-coding repressor of NFAT (NRON) RNA (*11*). The NRON RNA is a large alternatively spliced ncRNA that may represent a new class of regulatory RNAs. The transcription factor NFAT is imported into the nucleus when phosphorylated in response to a calcium signal. The NRON RNA has been shown to bind NFAT, the kinase responsible for phosphorylating NFAT, the protein that activates the kinase in response to the calcium signal, and an importin protein that transports NFAT to the nucleus (*11*). This implies that NRON functions to localize these related factors to the same physical region of the cell, overcoming slow rates of activation due to dilution effects. If this is a general trend, perhaps common to many transcriptional activators, it will highlight both the structural and regulatory aspects of ncRNAs while providing a rich new paradigm for studying ncRNA function.

The work presented in this thesis will outline how studies of the dynamics, degradation, and chemical modification of ncRNAs provide important insight into the function and applications of this important class of molecules. The ncRNAs discussed in subsequent chapters of this thesis are representative RNAs from two of these groups, helix 27 from the structural rRNA 16S RNA and a regulatory siRNA.

1.2 RNA Dynamics

The dynamics of an RNA are intimately related to its folding. Although RNA folding is a hierarchical process, with secondary structure preceding tertiary structure formation, it is disordered. Strong secondary structure interactions and untimely formation of tertiary contacts, either representing a native contact or an off-pathway state, can lead to kinetically trapped states that require resolution (*27*). This multitude of folded states may also serve the purpose of supplying a number of functional conformations and enhancing the functional characteristics of a given molecule (*28*). A number of systems illustrate the intimate relationship between RNA dynamics and function, bridging several classes of RNA molecules. For example, the small catalytic RNA called the hairpin ribozyme undergoes a large multi-domain rearrangement to carry out catalysis (*24, 29*). Dynamics of very large RNAs are also important, including the large-amplitude motions involved in the ribosomal translocation step (*30*). Dynamics also prove important in binding events, highlighted by the structural dynamics of 16S rRNA during ribosome assembly (*31, 32*) and the interactions of the HIV-Tar RNA with small molecule substrate mimics and potential drugs (*33, 34*). These systems have been studied in detail using a number of key biophysical techniques including single molecule fluorescence resonance energy transfer (FRET) and NMR spectroscopy. These studies, enumerated below, also reiterate the necessity of understanding RNA dynamics in order to help understand function.

The two-way junction form of the hairpin ribozyme is an ideal system to study RNA folding, as it is comprised of two domains that dock together to promote catalysis. The folding of the ribozyme towards docking involves a large conformational transition that is ideally studied using FRET (*29*). Single molecule (sm)FRET studies have revealed that heterogeneous folding kinetics of the hairpin ribozyme underlie the heterogeneous cleavage kinetics observed in ensemble assays of catalytic activity (*24, 35*). Further studies have identified mutations far removed from the cleavage site that affect catalysis, suggesting a network of coupled motions in the ribozyme (*36*).

In the case of the much larger ribonuclear protein (RNP) complex of the ribosome, very large amplitude motions have been uncovered to be important in translation. Both cryo-EM and FRET studies have revealed large intersubunit motions during the process of translocation, where the A and P site tRNAs are moved through the ribosome to make space for the next aminoacyl tRNA (*30, 37*). These dramatic and reversible motions highlight the dynamic nature of the ribosome, requiring large-scale motions such as this in order to utilize and coordinate all macromolecular components of the translational cycle. Interestingly, it has been shown that EF-G is not absolutely required to promote translocation and that the antibiotic sparsomycin, which binds in the region of the peptidyl-transferase center of the 50S subunit, can catalyze translocation (*38*). This finding further shows that the underlying rearrangements of translocation are inherent to the ribosome, and that EF-G or sparsomycin serve to unlock these motions at the appropriate (or inappropriate, in case of the antibiotic) time. In addition, there is a need to understand the local motions of the ribosome that trigger, promote, and/or constitute these global conformational changes. In spite of the fact that the ribosome is a profoundly complex macromolecular machine whose detailed inner workings still remain elusive, it is clear that conformational dynamics are key components to the function of the ribosome.

The assembly of the ribosome is also a very dynamic process whereby a well-defined series of proteins bind to the rRNAs. While studies on the subunits as a whole have revealed important alternate conformations during assembly (*31*), detailed studies on models of regions of 16S rRNA have revealed important dynamics that enable binding of a series of ribosomal proteins, including S15, S6, and S18 (*32, 39, 40*). This important rRNA three-way junction samples conformational states that enable S15 binding, while S15 binding induces rearrangements that enable subsequent binding of S6 and S18. This interplay of RNA dynamics and ligand binding seems to be a general trend, as the HIV-Tar RNA also undergoes large dynamics in the free state, sampling conformations that are recognized by various ligands, including a Tat protein mimic and various small molecule drugs (*33, 34, 41, 42*).

1.3 The Role of 16S rRNA Helix 27 in Translation

Translation is the process through which the ribosome, the cellular protein polymerase, decodes mRNA in a rapid, dynamic process with high fidelity between the mRNA codon and the amino acid finally incorporated into the growing protein chain. The rRNAs in prokaryotic ribosomes perform a number of functions in this process, the most important of which are verifying the correct codon:anticodon base pairing (*16*), catalyzing peptide bond formation in the active site of the large ribosomal subunit (*18, 43*), and engaging in the dynamic rearrangement of the ribosome to move the mRNA-tRNA complex by precisely one codon through the ribosome (*30*). A schematic of the elongation cycle of translation is pictured in Figure 1-1. The recent X-ray (*44-47*) and cryo-EM studies (*30, 48, 49*) of the ribosome have revealed and provided a structural

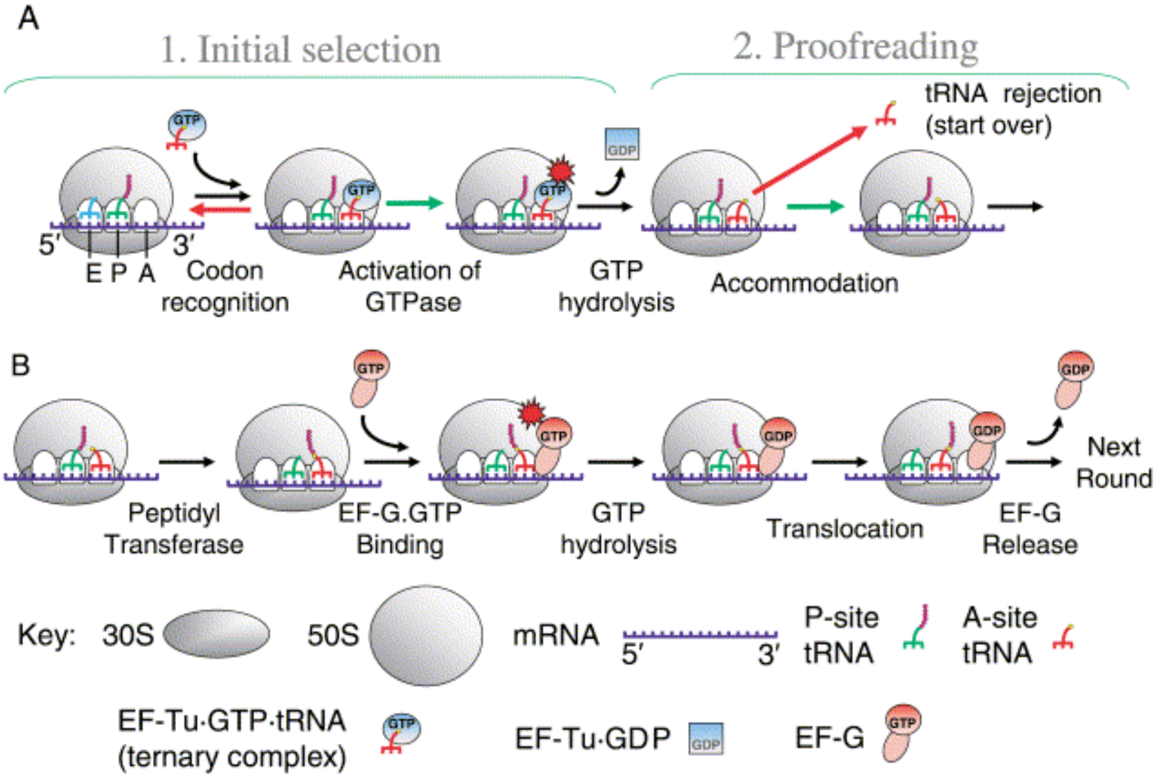


Figure 1-1. Schematic Description of translation elongation. A) Detailed cartoon indicating the orientation and interaction of the factors required for binding of aminoacyl tRNA to an open A site and the subsequent proofreading steps. B) Description of the final steps of the elongation cycle including translocation. Reprinted from reference (*17*) with permission from Elsevier.

basis for some rearrangements that constitute the translational cycle. However, it has proven difficult to identify the intrasubunit motions that contribute to the large (30-60 Å) motions inferred from low resolution cryo-EM structures. A central question in the field concerns how the individual elements of the ribosome rearrange during the large-scale dynamic motions that are revealed by the cryo-EM studies.

The most dramatic of these motions is associated with the process of translocation, where the translation factor EF-G catalyzes the ratchet-like rearrangement of the 30S subunit with respect to the 50S subunit, moves the mRNA:tRNA complex precisely one codon through the ribosome, and readies the A-site for the next incoming aminoacyl-tRNA (*17, 30*). Detailed models combining the X-ray and cryo-EM data have revealed that Helix 27 (H27) from 16S rRNA lies on the axis of the ratchet like intersubunit motion that describes translocation (*48, 49*). H27 also makes an important intersubunit bridge (*46*), and has been implicated in sensitivity to the antibiotics streptomycin and tetracycline (*50, 51*). Early studies on H27 suggested that it cycled between two alternate secondary structures and impacted the fidelity of the decoding process (*52, 53*). Recent work has called this hypothesis into question (*54*), but the work presented in Chapter 2 of this thesis demonstrates that an isolated H27 model system is dynamic, interchanges between two relevant conformations on the millisecond time scale, and that the dynamics are sensitive to the antibiotic tetracycline (*55*). It is unclear what functional role H27 may play in translation and how these H27 dynamics may be suppressed in the context of the ribosome. Chapters 2, 3 and 6 discuss the larger context of our studies on H27.

1.4 RNAi and Chemically Modified siRNAs

An attempt to make purple petunia flowers by overexpression of corresponding genes instead yielded white flowers (*56*). This unexplained phenomenon was an early indication of the RNAi pathway (*56*). Only when Fire and Mello demonstrated in 1998 that double-stranded (ds), but not single-stranded RNA mediates specific knockdown of genes in the nematode worm *C. elegans* (*57*), a discovery that was awarded with the

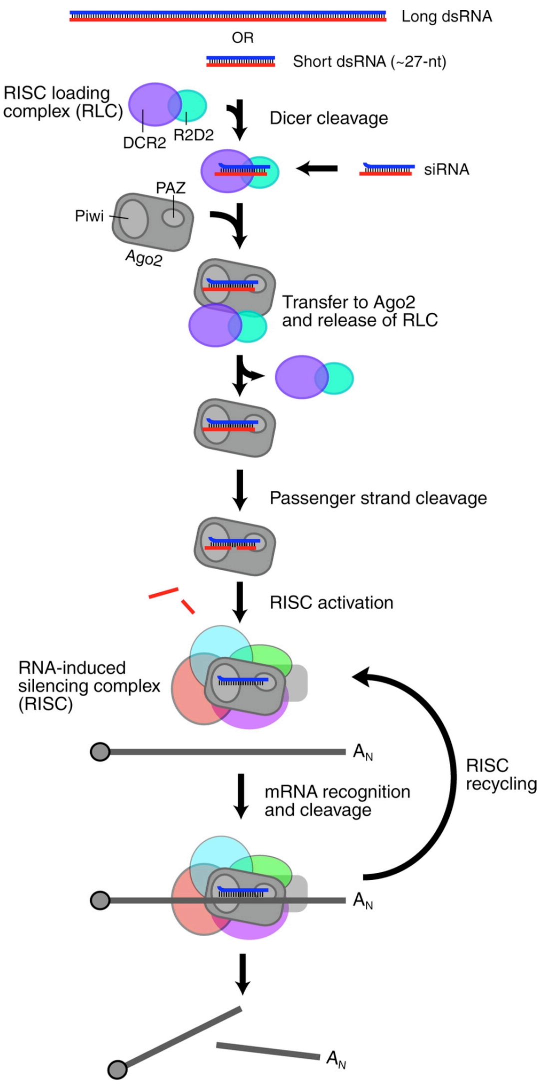


Figure 1-2. The RNA interference pathway. A cartoon depiction of the RNAi pathway, including Dicer dependent generation of siRNAs, RLC binding of siRNAs, and target cleavage. Reprinted, with permission, from the *Annual Review of Biomedical Engineering*, Volume 8, Copyright 2006 by Annual Reviews www.annualreviews.org (*58*).

Nobel Prize in 2006, did scientists become aware of the RNAi pathway (Figure 1-2). The RNAi pathway is evolutionarily ancient. Found in fungi, plants, and animals, the RNAi pathway most likely evolved to protect against viruses and selfish genetic elements, but adapted to regulate cellular development through the miRNA pathway (*25*). Early studies that demonstrated the utility of RNAi in mammalian cells not only introduced an indispensable reverse genetics tool, but also laid the groundwork for a new class of therapeutics (*59*).

In the presence of long, "naked" dsRNA molecules, a cellular RNase III-type enzyme called Dicer is called to action and processes the dsRNA into ~21-23 nucleotide pieces, called siRNAs (*60, 61*). These siRNAs are bound by Dicer and another protein called R2D2, which sense the thermodynamic asymmetry of the siRNA termini as R2D2 binds the 5' phosphate at the most stably base-paired end of the siRNA duplex and selects the strand to be excluded from the RNA-induced silencing complex (RISC) (*62*). The Dicer-R2D2-siRNA complex recruits Argonaute 2 (Ago2), the catalytic engine of RISC, and other proteins to form the RISC complex, after which the passenger strand of the siRNA is discarded and the RISC complex is ready to catalyze the sequence specific recognition and site specific cleavage of mRNA molecules and concomitant reduction in gene expression (*63*).

The exquisite specificity of target cleavage is somewhat compromised by the similarity of the siRNA and miRNA pathways (*64*). An siRNA can function as a miRNA, whose requirement for binding to the mRNA is considerably less stringent, involving complementarity of only an 8-nucleotide "seed sequence" compared to the 19-nucleotide perfect complementarity required for target cleavage with an siRNA guide (*64, 65*). While these issues can be overcome to some degree by bioinformatics analysis and careful siRNA design, recent work has shown that position specific 2'-O-Me modification can alleviate these off-target effects (*66*).

From a therapeutic standpoint, scientists can design dsRNA molecules that when introduced into the cell, function as siRNAs and degrade any RNA molecule identified by the sequence of the dsRNA drug. This pathway makes it possible to target essentially any intracellular RNA, including previously "undruggable" genetic targets, macromolecules that are not receptors or enzymes. This strategy is not without its limitations, which include delivery and stability of the dsRNA drug (*67, 68*). The inherent nuclease sensitivity of RNA molecules forced an early focus on the incorporation of chemical modifications into the siRNA strands, providing stability towards intracellular and extracellular ribonuclease (RNase) enzymes. The earliest such modification was the incorporation of deoxy nucleotides into the 3' overhangs characteristic of siRNAs (*59*), but have recently included a more diverse series of modifications. Specific correlations between modifications and function are listed in Table 1-1, but the general rules that have emerged are as follows: the 5'-end of the guide strand must have a free hydroxyl or phosphate; 2' ribose modifications including 2'-fluoro and 2'-O-Me are tolerated on both the guide and passenger strands; the RNA backbone can be stabilized by phosphorothioate modification; pyrimidine bases can be modified on the guide strand (*67*).

Recent work has suggested that chemical modification of siRNA is required for very different reasons, depending on the environment of the siRNA (*66, 69*). Chemical modifications clearly stabilize siRNAs in extracellular fluids like blood serum, as is demonstrated in Chapter 5 of this thesis and elsewhere (*70-79*). However, two recent reports in the literature and work presented in Chapter 4 of this thesis suggest that siRNAs are relatively protected inside the cell (*69, 80*). The origin of this protection is unclear, but it may arise from stable siRNA binding by Dicer and R2D2, restricting the siRNA at stage prior to RISC assembly, through an as yet unknown mechanism, until a target for the siRNA is present. These findings do not imply that chemical modifications to siRNA are obsolete in the intracellular milieu. In fact, it has been shown that 2'-O-Me chemical modifications at the one and two positions of the guide strand significantly reduce harmful "off-target" effects of siRNAs (*66*). Work in Chapter 5 has utilized this important pattern of chemical modifications to selectively stabilize particularly nuclease labile regions of an siRNA, thus satisfying independent extracellular and intracellular requirements by using the same modification.

siRNA therapeutics are being developed in a clinical setting, with early toxicity and efficacy studies underway with an inhaled siRNA targeting respiratory syncytial virus (*81*). siRNAs have been shown to decrease low-density lipoprotein cholesterol levels in both mouse and primate models of hypercholestorelemia (*82, 83*). These data are certainly promising, but there remains much to learn about siRNA function both in and outside the cell in order to support development, application, and future discovery of RNAi related therapies.

Passenger strand	*Guide strand*	*RNAi function*
5’-end capping	3’-end capping	Not affected
3’-end capping	Unmodified	Not affected
Unmodified	5’-end capping	Severely affected
2’-fluoro pyrimidines	2’-fluoro pyrimidines	Not affected
2’-deoxy	Unmodified	Moderately affected
Unmodified	2’-deoxy	Severely affected
2’-O-methyl	Unmodified	Moderately affected
Unmodified	2’-O-methyl	Severely affected
Unmodified	2’-O-methyl (position 1&2)	Reduced off-target
2’-O-(2-methoxyethyl)	Unmodified	Not affected
Unmodified	2’-O-(2-methoxyethyl)	Moderate to Severe
Phosphorothioate	Unmodified	Not affected
Unmodified	Phosphorothioate	Moderately affected
Phosphorothioate	Phosphorothioate	Moderately affected
3-methyl-U	3-methyl-U	Severely affected
Unmodified	C5-halogenated pyrimidine	Not affected

Table 1-1. siRNA chemical modifications and RNAi function. Strand specific impact of chemical modifications on RNAi potency. Reproduced from reference (*67*).

CHAPTER 2:

DYNAMICS INHERENT IN HELIX 27 FROM *ESCHERICHIA COLI* 16S RIBOSOMAL RNA[1]

2.1 Introduction

The ribosome is the universal protein biosynthesis machine. As such, it exhibits a high degree of structural organization which gives rise to the organized dynamic motions involved in the process of translation. Numerous large-scale structural rearrangements have been visualized by trapping and subsequent cryo-electron microscopy of intermediate functional states (*84, 85*). For example, upon binding of elongation factor G (EF-G), a large conformational reorganization is observed in which the small and large ribosomal subunits undergo a relative ratchet-like motion in conjunction with many local conformational changes, facilitating the translocation of the messenger (m)RNA and A- and P-site transfer (t)RNAs (*30, 48, 49, 86, 87*). The limited resolution of the cryo-EM maps (up to ~10 Å), however, prevents elucidation of the link between the observed global movement of the subunits and their local molecular reorganization. By contrast, recent advances in X-ray crystallographic analysis of ribosomal subunits have provided a wealth of atomic-resolution, yet largely static, structural information (*44, 88, 89*).

---

[1] Adapted with permission from Hoerter, J.A., Lambert, M.N., Pereira, M.J., and Walter, N.G., *Biochemistry* 43, 14624-36. Copyright 2004, American Chemical Society. FRET, UV crosslinking, and antibiotic assays were carried out by John Hoerter, NMR experiments were carried out by Meredith Lambert, and assignment of the UV crosslink was carried out by Miguel Pereira.

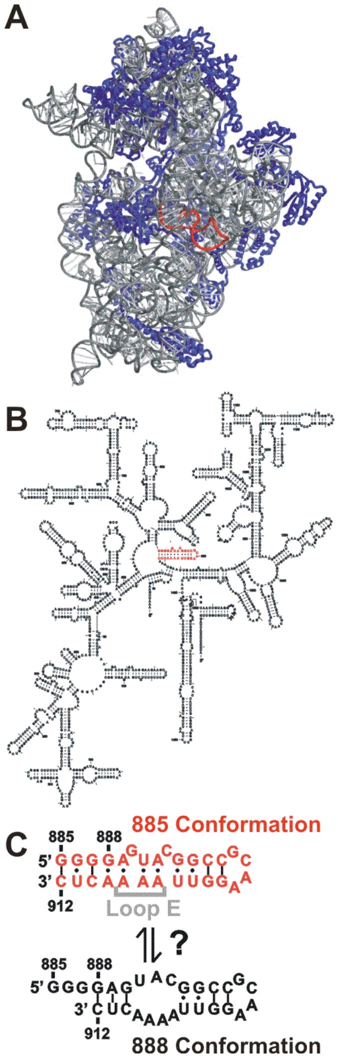


Figure 2-1. Helix 27 in the central domain of *E. coli* 16S rRNA. A) Location of H27 in the crystal structure of the 30S ribosomal subunit (PDB ID 1FJF) (*88*), as viewed from the interface with the 50S subunit. H27 is colored in red, the remainder of the 16S rRNA is in gray, and the small subunit proteins are in blue. This representation of the 30S was constructed with Nuccyl and rendered with MacPyMOL, DeLano Scientific, www.pymol.org. B) Secondary structure of the 30 S subunit, with H27 highlighted in red, adapted from (*90*). C) The 885 conformation, observed in all crystal structures of the 30S ribosomal subunit, is characterized by an 885-912 base pair and the 7-nucleotide loop E motif, whereas a proposed alternative base pairing scheme, the 888 conformation, manifests in an 888-912 base pair.

Clearly, 16S ribosomal (r)RNA is a central player in these structural rearrangements. It is the primary interaction partner for messenger RNA, it is responsible for message decoding and proofreading which enable subsequent peptidyl transfer, and it appears to undergo substantial shape changes including an opening and closing motion upon cognate tRNA selection (*91, 92*). While recent normal mode analyses on simplified models suggest that shape-dependent dynamic properties of the global architecture may provide the framework for large-scale conformational changes in the ribosome (*49*), details of the temporal and spatial coordination between local and global rearrangements are not well understood.

Helix 27 (H27) of 16S rRNA (comprising nucleotides 885 to 912 in *Escherichia coli* 16S rRNA) is a highly conserved RNA motif in the central domain of small subunit RNA. It lies on the axis of the inter-subunit ratchet motion in a location of potentially large conformational dynamics (*49*) (Figure 2-1), it is packed against the decoding center in helix 44 that senses and reports on correct tRNA selection, and its tetraloop acts as an inter-subunit bridge (*46, 48, 49, 86, 88, 89*). All of these observations are suggestive of a potential key role of H27 in ribosome dynamics and function, and more specifically in the linkage between correct codon recognition and the resulting cascade of steps that ultimately result in translocation. Evolutionary conservation was originally used to predict a base pair between G888 and C912 (*93*), but is also consistent with base pairing between G885 and C912 given the fact that the 3' most CUC triplet can pair with either the 5' most GGG or the subsequent GAG triplet by a simple 3-nucleotide slippage (Figure 2-1) (*94*). Dahlberg and co-workers, inspired by the finding from Stutz and co-workers that a C912U mutation conferred streptomycin resistance, isolated mutants that suppress the effects of a deleterious C912G mutant in *E. coli*. They demonstrated that a compensatory mutation, G885C, restored growth rates to nearly those of wild-type, experimentally supporting the 912-885 base-pairing scheme (*53*). Subsequently, the same group introduced mutations into 16S rRNA to selectively stabilize either the 912-885 or the 912-888 base-pairing scheme (henceforth referred to as the 885 and 888 conformations, respectively) and tested for a resulting phenotype in *E. coli* (*52*). In conjunction with cryo-electron microscopy studies of the mutated ribosomes, this work supported the notion that H27 cycles between the 885 and 888 conformations during translation and that this local conformational switch impacts the accuracy of A site tRNA decoding by triggering numerous changes in global ribosome architecture (*95*). However, only the 885 conformation has been observed in all functional ribosome states investigated by X-ray crystallography to date (*54, 89, 91, 92, 96, 97*), where it appears to be stabilized by the common loop E motif (*98*).

This discrepancy prompted Dahlberg and co-workers to reexamine their genetic approach. They found that two secondary mutations, C1192U in 16S RNA and A2058G in 23S RNA, introduced as genetic markers and initially thought to be silent, had "synergistic effects" with mutations in H27. In the absence of these secondary mutations, an *E. coli* strain that produces mutant ribosomes incapable of adopting the 888 conformation exhibits essentially the same growth rate as wild-type, calling the original switch helix hypothesis into question (*54*).

With this recent change in the interpretation of the existing genetic, biochemical, and structural data a number of questions arise – Is a slippery sequence such as that of H27 indeed trapped in only one conformational state, the 885 conformation? Or is conformational exchange still occurring, and with what rate? Given that base pairs in RNA can be as short-lived as sub-milliseconds (*99, 100*), it is possible that conformational exchange of a slippery sequence may actually be quite rapid. Insight into the kinetics and thermodynamics of such conformational exchange will not only help understand local structural dynamics in the ribosome, but will also shed light on conformational changes in slippery sequences involved in biological processes such as ribosomal frameshifting (*101*) or substrate recognition in the *Neurospora* VS ribozyme (*102*).

Here, we have used the isolated H27 sequence to define a kinetic and thermodynamic framework of conformational exchange between the 885 and 888 conformations. Similar studies of isolated structural motifs from the ribosome have previously led to fundamental insights into the hierarchical assembly of the central domain of the 30S subunit (*40*), the mechanism of decoding and its modulation by antibiotics (*103*), and the thermodynamic stability of novel RNA motifs discovered in the ribosomal crystal structures (*104*). Utilizing a complementary set of techniques, including ^{1}H NMR spectroscopy, UV-induced photo-cross-linking, and steady-state fluorescence resonance energy transfer (FRET) spectroscopy, we find that the isolated H27, under a variety of ionic conditions, exists in a dynamic equilibrium between the 885 and 888 conformers with millisecond exchange times and an equilibrium constant close to 1. FRET assays also show that the antibiotic tetracycline appears to specifically interfere with an induced shift towards the 888 conformation, a finding that is consistent with crystallographic localization of tetracycline bound to the 885 conformation of H27 in the 30S ribosomal subunit and that may pertain to tetracycline's mode of action. Taken together, our results invoke the notion of a structurally very dynamic helix 27 and beg the question of how these inherent structural dynamics may be suppressed within the ribosome.

2.2 Materials and Methods

RNA and DNA Preparation. Three different RNA constructs derived from H27 of *E. coli* 16S rRNA were generated for our NMR studies. The *Native* construct (GGG GAG UAC GGC CGC AAG GUU AAA ACU C) represents the unmodified *E. coli* H27 sequence, comprising nucleotides 885-912. RNA constructs *885* (5'-GGG GAG UAC GGC CGC AAG GUU AAA ACC C-3') and *888* (5'-GGG GAG UAC GGC CGC AAG GUU AAA ACU CCC C-3') were designed to specifically represent the H27 912-885 and 912-888 base-pairing schemes, respectively (Figure 2-1). These constructs suppress slippage of the base pairs at the helix terminus through a U911C mutation in the *885* construct, thus affording three stable G-C pairs in the 885 conformation versus a destabilizing A:C mismatch in the 888 conformation, and through extending the terminal helix of the *888* construct by three G:C base pairs that only can form in the 888 conformation. The constructs were run-off transcribed from double-stranded DNA templates containing an upstream T7 RNA polymerase promoter. Transcription reactions contained 40 mM Tris-HCl, pH 7.5, 15 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 2 mM spermidine, 4 mM of each rNTP, 5 units/ml inorganic pyrophosphatase, and 0.1 mg/ml T7 RNA polymerase (purified in His-tagged form from an over-expressing strain as described (*105*)) and were incubated at 37 ^{o}C overnight (~16 h). Full-length transcript was isolated after denaturing, 8 M urea, 20 % (w/v) polyacrylamide gel electrophoresis by UV shadowing, diffusion elution of small gel slices into 5 mM EDTA, and ethanol precipitation. For NMR, the RNA was further purified by anion exchange chromatography on Sephadex A-25 (Sigma), and dialysis into NMR buffer (10 mM Na_iPO_4, pH 6.4, 0.1 mM EDTA, 50 mM NaCl) by ultrafiltration using Centricon-3 concentrators (Amicon). Samples were concentrated to approximately 200 mL, and NMR buffer was added to a final volume of 225-250 mL, including the addition of 5% final volume of 99.9% D_2O (Aldrich or Cambridge Isotope Labs) for lock. Final RNA concentrations ranged from 0.5-1.2 mM as measured by UV absorption, where 1 A_{260} unit equals 37 μg/mL of RNA. Microvolume NMR tubes (Shigemi) were used for NMR data collection.

RNA construct *EH27* (5'-CCG CCU GGG GAG UAC GGC CGC AAG GUU AAA ACU C-3') for UV-cross-linking experiments contains an extended H27 sequence (nucleotides 879-912) in which a 5' overhang facilitates the binding of DNA oligonucleotides. It was synthesized by Dharmacon, Inc. (Lafayette, CO), deprotected as recommended by the manufacturer, purified on a denaturing 8 M urea, 20 % (w/v) polyacrylamide gel, eluted into crush and soak buffer (500 mM NH_4OAc, 0.1% sodium dodecyl sulfate (SDS) 0.1 mM EDTA), chloroform extracted, precipitated as described above, and further purified by C_8-reverse-phase HPLC with a linear acetonitrile gradient in triethylammonium acetate as described (*28, 106*). For determining the binding rate of *9mer* DNA to *EH27*, this construct was 5' fluorescein labeled to generate construct *5'FEH27*. To this end, *EH27* RNA (final concentration of 0.9-1 μM) was mixed with 5 mM [γ-S]ATP (Sigma), and 1 unit/μl of T4 polynucleotide kinase (Takara) in 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, and 5 mM DTT and incubated at 37 °C for 3 h (*107*). The phosphorothioated RNA was phenol-chloroform extracted, ethanol precipitated, and desalted using a Centrispin Column (Princeton Separations). It was then diluted to 0.3 μM and mixed with 5 mM fluorescein-5-maleimide (Molecular Probes) in 7.5 mM $Na_2B_4O_7$ and 14% DMSO, incubated at 70 °C for 30 minutes, ethanol precipitated, and purified by C_8-reverse-phase HPLC (*28, 106*). Appropriate fractions were pooled and passed over a NAP-5 gel filtration column (Amersham Phamacia Biotech) to yield the final stock solution of *5'FEH27*.

RNA construct *FEH27T* for FRET experiments was synthesized in a singly labeled form by Howard Hughes Medical Institute Biopolymer/Keck Foundation Biotechnology Resource Laboratory at the Yale University School of Medicine (New Haven, CT) with the same sequence as construct *EH27* (5'-CCG CCU GGG GAG UAC GGC CGX AAG GUU AAA ACY C-3'), except the cytidine in the variable position in the GNRA tetraloop and uracil 911, indicated as X and Y respectively, were replaced with the uracil analogs Fluorescein dT and Amino-Modifier C6 dT (Glen Research), respectively. This synthetic RNA was deprotected by treatment with triethylammonium trihydrofluoride, and purified by denaturing gel electrophoresis, and C_8-reverse-phase HPLC as described above (*28, 106*). It was chloroform extracted and then reacted with tetramethylrhodamine succinimidyl ester (Molecular Probes) in 35 mM Tris-HCl, pH 7.7, and 14% DMSO, incubated at room temperature overnight (~16 h) re-purified by C_8-reverse-phase HPLC, and passed over a NAP-5 column, thus yielding the doubly labeled FRET construct *FEH27T* (*28, 106*).

The DNA sequences used in this study are the following: *9mer* (5'-CCC AGG CGG-3'); *F2* (5'-CCC AGG CGG TCG ACT TA-3'); *cF2* (5'-TAA GTC GAC CGC CTG GG-3') and *Disrupter* (5'-TAC TCC-3'). Oligodeoxynucleotides were synthesized and desalted by Invitrogen (Carslbad, CA). Synthetic DNA was used as supplied provided that it gave a clear solution upon addition of water and appeared homogeneous as judged by denaturing polyacrylamide gel electrophoresis. Otherwise, the material was gel purified as above. Stock concentrations were determined by UV absorbance as described above for RNA.

NMR Spectroscopy. All NMR spectra were acquired in NMR buffer (10 mM Na_iPO_4, pH 6.4, 0.1 mM EDTA, 50 mM NaCl) at 8 ˚C, unless otherwise noted, on a Varian Inova 800 MHz spectrometer equipped with a triple resonance $^1H/^{13}C/^{15}N$ probe with Z-gradients. Each spectrum was collected using a water flip-back solvent suppression scheme followed by a WATERGATE pulse scheme (*108, 109*). Quadrature detection was achieved by implementation of the States-TPPI method in all spectra (*110*). Data for the *888* construct were collected over 54 scans. 2,816 complex points were acquired for a sweep width of 20,000 Hz, and a recycle delay of 1.3 seconds was used. Data for constructs *885* and *Native* were collected in 512 scans with 8,000 complex points and a spectral width of 20,000 Hz. A 1.5 second recycle delay was used. Spectra were processed with nmrPipe and visualized with Igor Pro (WaveMetrics) (*111*). Spectra were referenced to the solvent resonance position, 4.97 ppm, at 8 °C. All spectra were processed using a solvent filter and a cosine-bell apodization function. Each FID was zero-filled once before Fourier transforming.

Spectra of the *885*, *888*, and *Native* constructs were baseline corrected and normalized to the Lorentzian fit area of the imino proton resonance arising from U904 (~11.8 ppm). A calculated spectrum was constructed as a linear combination of the construct *885* and *888* spectra and minimized against the *Native* spectra using Levenberg-Marquart least-squares regression as implemented in the program Igor Pro, using only the relative contribution of the 888 conformation, f_{888} (with $f_{885} = 1-f_{888}$), as fitting parameter. The spectral region containing the exchange-broadened hump centered at approximately 11 ppm in the *Native* spectrum was omitted for the fit.

Thermal Denaturation. Melting temperatures (T_m) for the *885*, *888*, and *Native* NMR constructs were determined from thermal denaturation profiles recorded on a Beckman DU640B spectrophotometer equipped with High Performance Temperature Controller and Micro Auto 6 T_m cell holder. RNA samples (1 μM, 300 μL) were prepared in 10 mM sodium phosphate, pH 6.4, 0.1 mM EDTA, and 1.0 M sodium chloride and degassed for 5 minutes prior to obtaining UV melting curves. The temperature was increased from 15 ºC to 85 ºC at a rate of 1 °C/min, and the absorbance at 260 nm was recorded every 0.5 ºC. The first derivatives of the resulting melting profiles were fit to Gaussian distributions using the spectrophotometer as well as MicroCal Origin 7.0 software, yielding Tm's as the maxima of the Gaussians. Three independent melting profiles were recorded for each construct to estimate the experimental error in T_m to approximately ±2 °C.

UV-Induced Photo-Cross-linking. EH27 RNA was 5' end labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. In case of the *in vitro* transcribed constructs *885* and *888*, a 5-fold excess of *Disrupter* DNA was added to interfere with RNA secondary structure, the mixture heated to 70 °C and cooled to room temperature, the RNA dephosphorylated with calf intestine alkaline phosphatase, phenol-chloroform extracted, and ethanol precipitated prior to 5'-^{32}P labeling. The radiolabeled *885* and *888* constructs were further purified by denaturing, 8 M urea, 20 % (w/v) polyacrylamide gel electrophoresis, elution into 10 mM Tris-HCl, pH 7.6, 1 mM EDTA overnight, and ethanol precipitation.

5'-^{32}P labeled *EH27*, *885*, and *888* constructs (<2.5 nM) were annealed in standard buffer as described above. The solutions were incubated in wells of a microplate floating in a circulating water bath at 25 °C. A hand held UV lamp (λ = 254 nm, Spectroline EF-180C) was positioned ~1.5 cm above the microplate. Appropriate time points were taken over a 120-minute time interval, mixed with 10 μL of 80% formamide, 0.025% xylene cyanol, 0.025% bromophenol blue, and 50 mM EDTA and separated from uncross-linked one by denaturing, 8 M urea, 20 % (w/v) polyacrylamide gel electrophoresis and quantified using a Molecular Dynamics Storm 840 phosphorimager. The intensity of the upper of the two UV induced cross-links was calculated as the fraction of all three bands in the lane, plotted as a function of time and fit to a single-exponential single-exponential association function $y = y_o + A_1(1 - e^{(t/\tau_1)})$ to extract a cross-linking rate constant $k_{obs} = \tau_1^{-1}$. For assignment of the location of the cross-link in *EH27*, unlabeled *EH27* was exposed to UV light, the upper cross-linked RNA band (*cEH27*) was gel purified, ethidium bromide stained and eluted as described above for radiolabeled RNA. It was then 5' end radiolabeled as described above, or 3' end labeled using [^{32}P]pCp and T4 RNA ligase. The RNA was analyzed on an 8 M urea, 15 % wedged polyacrylamide sequencing gel, alongside ladders from partial digestion with G-specific RNase T_1 and alkaline hydrolysis of both cross-linked and uncross-linked *EH27* (*112*).

Steady-state fluorescence spectroscopy. Doubly labeled FRET construct *FEH27T* was annealed at 20 nM in standard buffer (50 mM Tris-HCl, pH 7.5, 50 mM NH_4Cl, 20 mM $MgCl_2$) by heating to 70 °C for 2 min, followed by cooling to room temperature over 10 min. Steady-state fluorescence measurements were performed at 25 °C on an Aminco-Bowman Series 2 (AB2) spectrofluorometer (Thermo Electron Corp.) as described previously (*113, 114*). The sample (156 μL) was placed in the fluorometer cuvette, and fluorescein was excited at 490 nm (4 nm bandwidth), and fluorescence emission was recorded simultaneously at the fluorescein (520 nm, 8 nm bandwidth) and tetramethylrhodamine (585 nm, 8 nm bandwidth) wavelengths via an oscillating monochromator. For the fuel strand experiment, DNA *F2* (1 μL) was manually added to a final concentration of 100 nM. After the signals reached a new equilibrium position, DNA *cF2* (1 μL) was added to a final concentration of 140 nM. The FRET ratio Q = (F_{585}/F_{520}) was calculated and its time trace recorded along with the single fluorophore data. To test the impact of antibiotics on the DNA induced shift from the 885 to the 888 conformation, tetracycline hydrochloride, streptomycin sulfate (both ICN Biomedical Inc.), and kanamycin sulfate (Fisher Scientific), were prepared at 10 mM stock concentration (used within 24 h; the tetracycline was protected from light). 40 nM *FEH27T* was annealed as described above in buffer consisting of 600 mM Tris-HCl, pH 7.5, 100 mM NH_4Cl, and 40 mM $MgCl_2$. After cooling to room temperature, water and antibiotic stock solution were added to reach a final concentration of 20 nM *FEH27T* in a buffer similar to standard buffer, except that it contained a variable concentration of antibiotic and 300 mM instead of 50 mM Tris-HCl to stabilize the pH at 7.5. This solution was preincubated at 25°C for 1-2 minutes, fluorescence data acquisition initiated as described above, and *9mer* DNA manually added to a final concentration of 400 nM. (Addition and mixing of the DNA took ~10 s in all cases.) The FRET ratio, Q, of time traces upon *9mer* addition was normalized to its average value before *9mer* addition, Q_0, by calculating $(Q-Q_0)/Q_0$, to yield a relative FRET efficiency, which was fit to the single exponential increase function as described above to extract a rate constant $k_{obs} = 1/\tau_1$ (*28, 106*). None of the antibiotics absorb at 490 nm or emit in the donor to acceptor wavelength range, ruling out any spectroscopic interference.

The initial FRET ratio Q_0 for all samples was the same within 15%. The amplitude from the fit, or relative FRET increase A, was averaged for a minimum of two independent assays and plotted against the tetracycline concentration. These data were fit to a modified form of the Hill equation:

$$y = A_{obs,0} - A_{\infty}\frac{[tetracycline]}{K_{D,app} + [tetracycline]} \quad \textbf{Equation 2-1}$$

where $A_{obs,0}$ is the average amplitude change observed in the absence of tetracycline, A_∞ is the amplitude change for an infinite tetracycline concentration, and $K_{D, app}$ is the apparent dissociation constant for tetracycline.

For measurement of fast kinetics upon addition of high concentrations of *9mer* DNA (no antibiotics), a stopped flow mixing technique was employed based on the Milli-Flow Reactor of the AB2 spectrofluorometer (Thermo Electron Corp.). *FEH27T* RNA (800 μL) was prepared and annealed at 80 nM (diluted 1:2 in the experiment) in standard buffer as described above. *9mer* DNA was prepared at twice its final concentration in standard buffer, and equal volumes of the two solutions were rapidly mixed in the Milli-Flow Reactor at 25 °C. Excitation was set at 490 nm (8 nm bandwidth) while the tetramethylrhodamine fluorescence emission increase was monitored at 585 nm (16 nm bandwidth). (Monitoring the fluorescein emission decrease at 520 nm yielded rate constants within 20% of those for the tetramethylrhodamine, which is consistent with FRET data where both wavelengths could be monitored simultaneously). Data at *9mer* concentrations lower than 350 μM were collected with 1-ms time resolution, while data at concentrations greater than 350 μM were collected at 300-μs resolution. 5-25 time traces were averaged and fit to a single exponential increase function as described above, yielding rate constants $k_{obs} = 1/\tau$. Error bars (dk) stem from the calculated error in τ_1 ($d\tau_1$) derived from the fit ($dk = \frac{d\tau_1}{\tau_1^2}$). The *9mer* concentration dependence of k_{obs} was fit to eq 2-2 (see below) using Levenberg-Marquart least-squares regression implemented in Origin 7.0 (Microcal) to extract the elementary rate constants.

For an independent determination of the *9mer* binding rate constant (k_{on}), *5'FEH27* was prepared at a concentration of 10 nM and annealed in standard buffer as described above for *FEH27T* and additionally degassed prior to measurement by placing the sample in a Speed-Vac vacufuge for 15 minutes. The sample (150 μL) was placed in the fluorometer cuvette with excitation at 490 nm (4 nm bandwidth) while emission was monitored at 520 nm (8 nm bandwidth). Given a stable equilibrium fluorescence signal from *5'FEH27*, *9mer* DNA was added to final concentrations of 100 nM to 1 μM, and the resulting kinetic fluorescence increase was fit to a single-exponential increase function as described above. The resulting rate constant, $k_{obs} = \tau_1^{-1}$, was plotted against the *9mer* concentration, and these data fit by linear regression, where the slope represents the second order rate constant associated with *9mer* binding to *5'FEH27*.

2.3 Results

G890 as a key indicator for the 885 and 888 conformations

Since the base-pairing patterns in the proposed H27 conformations are distinct, we determined one-dimensional (1D) imino proton (^{1}H) NMR spectra of three different constructs derived from nucleotides 885 to 912 in *Escherichia coli* 16S rRNA (Figure 2-2): Construct *Native* is unmodified; construct *885* carries a U911C mutation to selectively stabilize the 885 conformation; and construct *888* has an extended terminal helix to stabilize the 888 conformation. The secondary structures of the "locked" constructs *885* and *888* are identical between bases 895-904, and contain a characteristic structural element, a GNRA tetraloop (where N can be any nucleotide, here C, and R is a purine, here A). The imino ^{1}H-NMR spectra of the *885* and *888* constructs therefore exhibit several common features (Figure 2-2A), assigned based on imino-imino and imino-amino nuclear Overhauser effects (NOEs) as well as ^{1}H-^{15}N HSQC spectra. The resonance of the first G of the tetraloop appears around 10.6-10.7 ppm, an upfield position consistent with previous GNRA tetraloop studies (*115, 116*); evidence is observed for three base pairs immediately adjacent to the tetraloop, as indicated by the presence of imino proton resonances corresponding to the G895, G902, and G903 NH1 protons. The G902 and G903 NH1 resonances are observed between 12 and 14 ppm, consistent with Watson-Crick base pairing of these imino protons, while an expected upfield chemical shift to 11.57 ppm is observed for the G895 NH1 resonance involved in the G895:U904 wobble pair. In addition, U904 of this wobble pair shows an expected resonance at 11.77 ppm (Figure 2-2A).

Notably, the G890 NH1 chemical shift is very different in the *885* and *888* spectra (Figure 2-2A). In the *888* construct, G890 is predicted to pair with C910, which leads to a G890 NH1 resonance at 12.19 ppm, corresponding to a Watson-Crick base paired imino

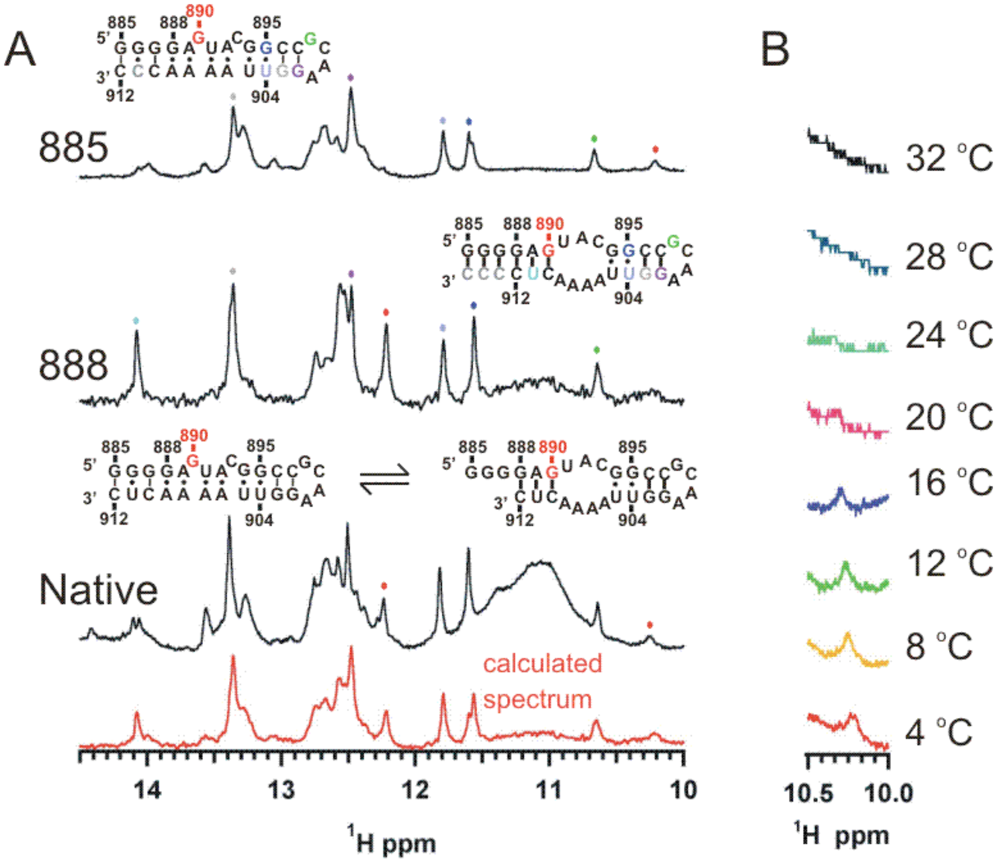


Figure 2-2. 1D imino ^{1}H-NMR spectra of three RNA constructs based on *E. coli* H27. A) Spectra of the *885*, *888*, and *Native* constructs collected at at 8 ˚C. The spectrum in red is a linear combination of the *885* and *888* constructs, fitted to the spectrum of the *Native* construct. The relative fractions of the 885 and 888 conformations are $f_{885} = 0.573 \pm 0.04$ and $f_{888} = 0.427 \pm 0.04$, respectively. B) A temperature series on the *Native* construct shows that the resonance at ~10.2 ppm, which arises from the G890 imino proton in the 885 conformation, disappears above 20 °C through line broadening, much below the RNA melting temperature (63 °C) and the temperatures at which other resonances are lost (30-50 °C).

proton. In the *885* construct, where G890 is predicted to be looped out to form two hydrogen bonds, one each to the adjacent U891 and to a phosphate oxygen on the opposing strand (*88*). The *885* G890 NH1 resonance is found as a relatively broad peak of diminished intensity at 10.21 ppm. This is consistent with NMR spectra of the loop E motif of the sarcin-ricin loop of the large-subunit ribosomal RNA, where the corresponding G (G2655 in *E. coli* 23S rRNA, G4319 in rat 28S rRNA) has a nearly identical chemical shift (10.18 ppm) (*117, 118*). Our findings therefore identify G890 as a key indicator of whether H27 resides in the 885 or 888 conformation.

In the spectrum of the *Native* construct, as expected, ^{1}H-NMR resonances are found that correspond to the GNRA tetraloop and the three adjacent base pairs, which are shared among all NMR constructs (Figure 2-2A). However, we also observe a broad featureless hump, centered around 11 ppm, that is unique to the *Native* construct. In long (~350 ms) mixing time NOESY spectra, this hump reveals NOEs to Watson-Crick base-paired G and C amino protons, suggesting that it contains G imino proton resonances at the helix terminus partially protected from chemical exchange. In addition, we found two different resonances for the G890 NH1 proton, at 10.21 ppm and 12.19 ppm, that are characteristic of the 885 and 888 conformers, respectively (Figure 2-2A). This observation shows that the *Native* construct partially occupies both conformations, and that the 885 and 888 conformations are either in slow exchange compared to the NMR time scale or reside in different sub-populations of the RNA. We also performed a temperature series that shows that specifically the resonance at 10.21 ppm, which arises from G890 in the 885 conformation, disappears above 20 ^{o}C due to line broadening (Figure 2-2B), much below the melting temperature of the *Native* construct under these ionic conditions (63 ºC) and the temperatures at which other resonances are lost (30-50 ^{o}C). This is consistent with the previous observation of enhanced conformational dynamics of the bulged G in the loop E motif of 28S rRNA compared to Watson-Crick base pairs (*118*).

To quantify the relative contributions of the 885 and 888 conformations to the *Native* structure, we fit a linear combination of the spectra of the *885* and *888* constructs to the spectrum of the *Native* construct, using the relative contribution of the 888

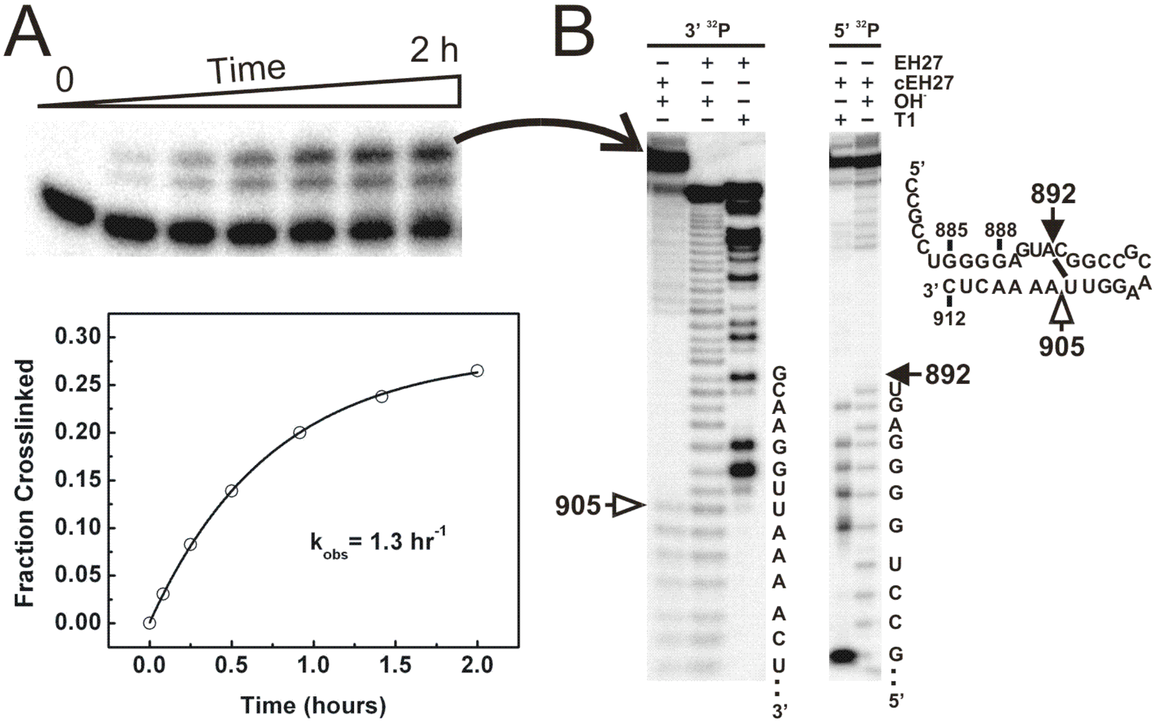


Figure 2-3. UV-induced photo-crosslinking of construct EH27. A) Time course of cross-linking. After only 5 min exposure to UV light, two bands of cross-linked RNA are apparent after separation on a denaturing 20% polyacrylamide gel, the uppermost of which grows to ~28% of the total RNA over a 2-h time period with an observed rate constant of 1.3 hr^{-1}. B) Mapping the major cross-linking site. The major (upper) cross-link (termed *cEH27*) was prepared in large scale, radiolabeled either on the 3' or 5' ends for comparative analysis on sequencing gels by alkaline hydrolysis (OH$^-$) and RNase T1 digestion together with the uncross-linked *EH27*. The cross-link leads to a gap in the alkaline ladder since OH$^-$-induced backbone cleavage 3' of a nucleotide anywhere between the cross-linked bases will not yield a fragment. Please note that cleavage 3' of U905 still yields a band in the OH$^-$ lane of the 3' radiolabeled *cEH27*, while cleavage 3' of A892 is associated with the first missing fragment in the OH$^-$ lane of the 5' radiolabeled *cEH27*, thus indicating that the cross-link (dash in the H27 depiction) occurs between nucleotides A892 and U905. Our sequence assignments are indicated alongside the sequencing ladders.

conformation, f_{888} (with $f_{885} = 1-f_{888}$), as the only fitting parameter (Experimental Section) (*119*). An optimized calculated spectrum is found for $f_{888} = 0.427 \pm 0.04$, suggesting that the *Native* construct significantly populates both the 885 and 888 conformations under NMR conditions (Figure 2-2A).

A novel UV-induced photo-crosslink in H27

The 885 conformation of H27 observed in the 30S subunit crystal structures is stabilized by the common loop E motif (*88, 89*), which often yields a specific intrahelical photo-cross-link upon UV irradiation (*98*). However, in multiple UV-irradiation studies of the ribosome only long-range cross-links involving G894 of H27 and the 5' domain of 16S rRNA were observed (*120-122*). Irradiation of construct *EH27* in standard buffer (chosen to resemble buffer conditions used for kinetic assays of ribosome function (*123*): 50 mM Tris-HCl, pH 7.5, 50 mM NH_4Cl, 20 mM $MgCl_2$) at 25 ^{o}C with UV light at 254 nm resulted in a major and a minor RNA-RNA photo-cross-link (Figure 2-3). The major cross-link has a lower gel mobility, accumulates with a single-exponential rate constant of 1.3 hr^{-1}, and saturates at ~28% after two hours. (A minor, slightly higher-mobility cross-link also accumulates single-exponentially, is enhanced by the addition of magnesium, and appears to be located near the 3' terminus.) We purified the major cross-linked species, ^{32}P-radiolabeled it either on the 5' or 3' end, and analyzed it using alkali and RNase T1 (G-specific) degradation. We thus found that the cross-link covalently couples nucleotides A892 and U905 (Figure 2-3), which is distinct from the loop E-specific cross-link that is observed between the nucleotides equivalent to U891 and A906 (*98*). Our findings therefore suggest that a structure different from a loop E motif causes the efficient cross-link in H27. Interestingly, neither of the two conformationally "locked" constructs *885* and *888* formed a cross-link under the same conditions.

FRET demonstrates reversibility of conformational switching

To more directly test whether the 885 and 888 conformations of H27 interconvert, we designed a FRET labeled construct *FEH27T*, in which the 5' end of H27 is extended by 6 nucleotides of 16S rRNA sequence. This design allows for an assay inspired by a previously described DNA-fuelled molecular machine (*124*), in which the partially complementary DNA oligonucleotide *F2* binds to the 5' overhang as well as the adjacent

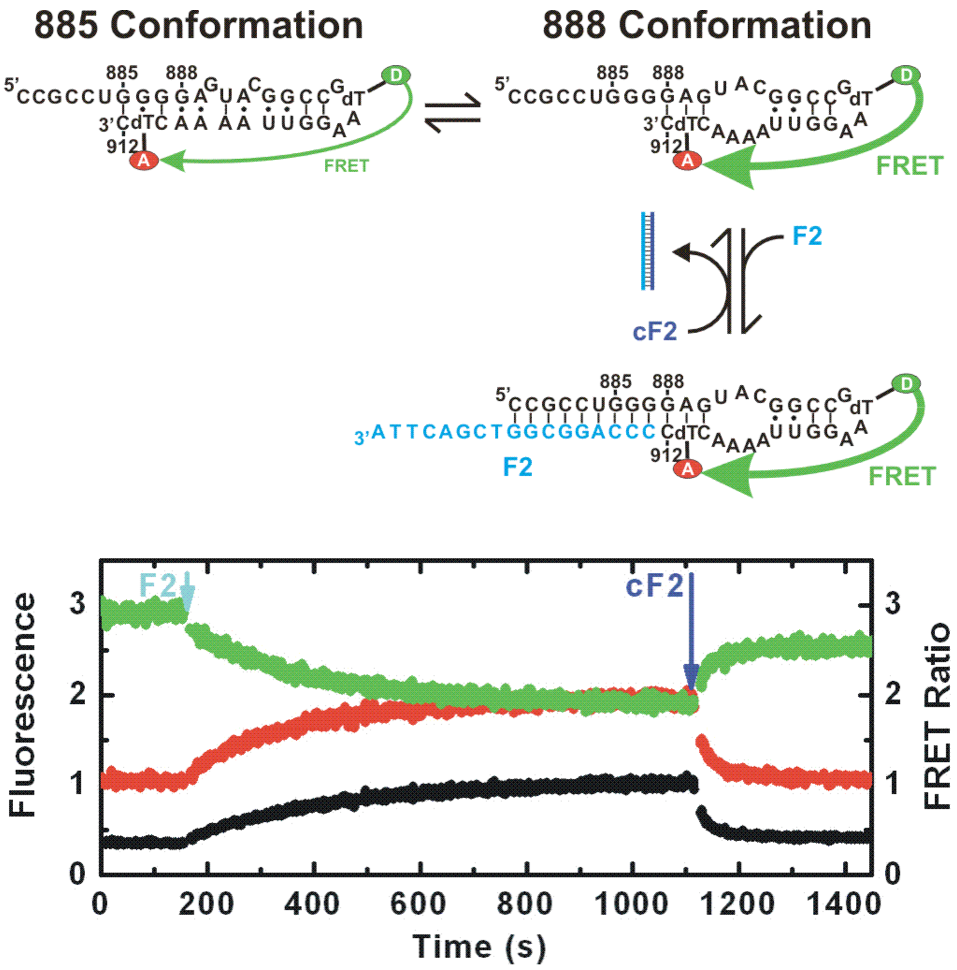


Figure 2-4. Scheme and fluorescence time course of a FRET monitored experiment demonstrating reversibility of conformational switching of H27. Construct *FEH27T* is doubly labeled with donor (fluorescein, green) and acceptor (tetramethylrhodamine, red) fluorophores. Upon addition of the partially complementary DNA oligonucleotide *F2* (indicated in cyan) the donor emission decreases, while the acceptor emission and the resulting FRET ratio (black) increase, as expected from a shift from the 885 to the 888 conformation. Upon addition of DNA oligonucleotide *cF2* (indicated in blue), which forms a DNA duplex with *F2* and thus removes it from *FEH27T*, FRET decreases back to its initial value.

nucleotides 885 to 887, eliminating the possibility of the RNA assuming the 885 conformation thereby shifting the population toward the 888 conformation. Such a conformational shift is expected to result in a FRET increase between the donor fluorophore (fluorescein) attached to the N of the GNRA tetraloop and the acceptor fluorophore (tetramethylrhodamine) coupled to the base in position 911 (Figure 2-4). The fluorophore pair is predicted to have a distance of ~42 Å in the 885 conformation of the 30S ribosomal subunit (*88, 125*), which is close to the pair's Förster distance of ~55 Å, making it very sensitive to changes in that distance (*28*). Preliminary time resolved FRET analysis of *FESH2T* at saturating Mg^{2+} shows a decrease in the measured intrafluorophore distance of more than 5 Å upon addition of *9mer* DNA. DNA oligonucleotide *F2* itself contains a 3' overhang, which allows its removal by addition of the fully complementary DNA oligonucleotide *cF2* to form a long DNA duplex (Figure 2-4). If conformational switching of *FEH27T* is reversible, this should lead to a FRET decrease.

Indeed, when we incubated 20 nM *FEH27T* in standard buffer (50 mM Tris-HCl, pH 7.5, 50 mM NH_4Cl, 20 mM $MgCl_2$), at 25 °C, we found a relatively low FRET ratio of ~0.36. Upon addition of 100 nM of *F2* DNA, we observed an exponential decrease in donor signal, accompanied by a synchronous increase in the acceptor signal, consistent with an increase in FRET (Figure 2-4). After ~800 s, the donor and acceptor signals leveled off at a FRET ratio of ~1.0. Next, we added 140 nM of the fully complementary *cF2*, which led to an exponential decrease over ~300 s to a FRET ratio of ~0.42, signifying an essentially complete reversal to the initial FRET ratio (Figure 2-4). In a control experiment, addition of *cF2* to *FEH27T* in the absence of *F2* resulted in no changes in the FRET ratio, as expected. These results provide direct evidence for full reversibility of conformational switching between the 885 and 888 conformations in H27. Furthermore, we noticed a dependence of the kinetics of the FRET increase on the *F2* concentration, suggesting that the observed rates report on a bimolecular reaction.

Extracting fast rate constants of conformational switching

Our initial observation that the FRET increase as induced by the addition of partially complementary DNA oligonucleotide *F2* exhibits a dependence on the DNA

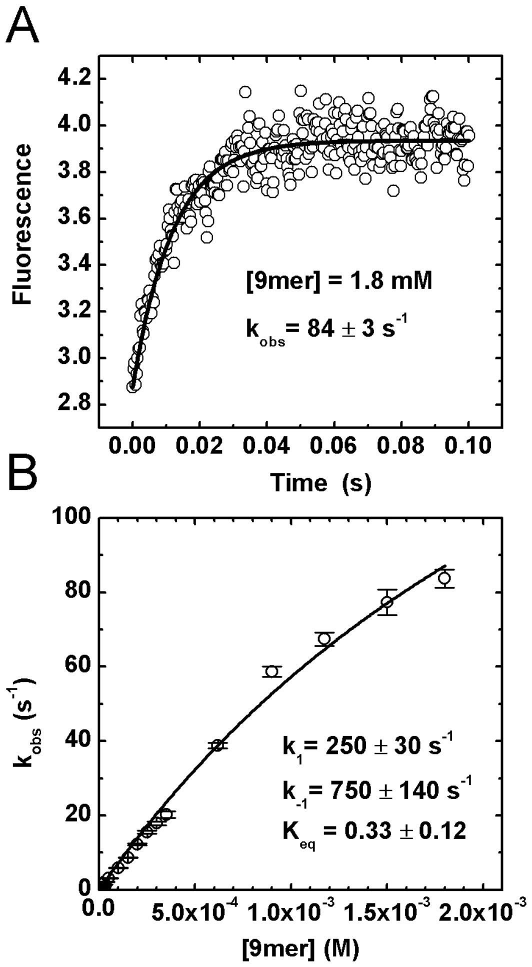
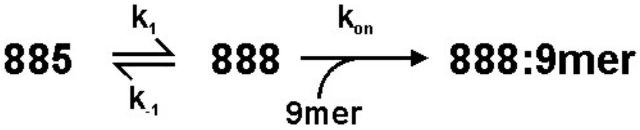


Figure 2-5. Stopped-flow kinetics monitoring the acceptor fluorescence increase upon 9mer addition at different concentrations to shift the FRET labeled FEH27T construct into the 888 conformation. A) Circles, fluorescence emission at 585 nm with excitation at 490 nm, averaged over 14 individual stopped-flow time traces collected upon addition of 1.8 mM *9mer*; black line, single-exponential fit yielding $k_{obs} = 84 \pm 3\ s^{-1}$. B) Plot of k_{obs} derived from data sets as in (A), collected over a range of *9mer* concentrations. Black line, fit to eq 2-2 based on kinetic Scheme 2-1, yielding values for the forward and reverse rate constants, k_1 and k_{-1}, respectively, which characterize the rapid exchange between the 885 and 888 conformers.

concentration suggests that a bimolecular reaction, rather than unimolecular conformational switching, is rate limiting. To further test this notion and to ask whether a rate constant for conformational switching can be extracted at saturating DNA concentration, we performed a DNA titration experiment. We utilized a DNA oligonucleotide, referred to as *9mer*, that binds to the same 879-887 region of *FEH27T* as *F2*, but lacks its 3' overhang. A representative stopped-flow time course of the acceptor fluorescence signal (representative of the increase in FRET; the donor fluorescence decreased with similar kinetics), taken at 1.8 mM *9mer*, is shown in Figure 2-5A. The signal increase was >30% and was complete in well under 100 ms after DNA addition. The data were fit with a single-exponential increase function, yielding a rate constant of $84 \pm 3\ s^{-1}$. By performing this experiment over a broad range (2 µM to 1.8 mM) of *9mer* concentrations, we obtained the titration curve shown in Figure 2-5B.

The rate constants observed for the induced change in acceptor fluorescence are linearly dependent on the *9mer* DNA concentration up to ~900 µM. At *9mer* concentrations of >1 mM, a deviation from this linear dependence was observed, indicative of the beginning of *9mer* saturation as unimolecular conformational switching becomes rate limiting (and thus *9mer* concentration independent). To extract rate constants for conformational switching, we considered the following simplified reaction mechanism:



Scheme 2-1

This minimal reaction scheme is based on the fact that we have evidence for reversible switching between the 885 and 888 conformations through our "fuel-strand" experiment of Figure 2-4 and assumes that *9mer* binding captures and traps H27 in the high-FRET 888 conformation, thus generating the observed FRET increase. This model does not account for *9mer* first binding to the 885 conformation and subsequently trapping the 888 conformation. This mechanism does allow us to derive the following analytical expression for k_{obs}:

$$k_{obs} = \frac{k_1 k_{on}[9mer]}{k_1 + k_{-1} + k_{on}[9mer]} \qquad \textbf{Equation 2-2}$$

To reduce the number of variables in our fit, an independent determination of the *9mer* binding rate constant k_{on} was necessary. To this end, ,the initially unlabeled RNA termed *EH27*,(identical in sequence to *FEH27T*, excepting modification) was enzymatically 5' phosphorothioated, then labeled with fluorescein-5-maleimide to yield *5'FEH27* (Materials and Methods Section). The 5' fluorescein label of this construct responds to the adjacent binding of *9mer* through nucleotide-specific fluorescence dequenching (*126*) independent of conformational switching. As expected, the observed rate constant of the fluorescence decrease upon addition of *9mer* DNA was linearly dependent on *9mer* concentration; the slope of this linear dependence yielded the binding rate constant of *9mer* to *FEH27* as previously described (*126*), with $k_{on} = 2.96 \pm 0.45 \times 10^5$ M^{-1} s^{-1}. The resulting fit of eq 2-2 to the titration data in Figure 2-5B then gave $k_1 = 250 \pm 30$ s^{-1} and $k_{-1} = 750 \pm 150$ s^{-1} for conformational switching under standard conditions (50 mM Tris-HCl, pH 7.5, 50 mM NH_4Cl, 20 mM $MgCl_2$, at 25 ^{o}C). The equilibrium constant between the 888 and 885 conformations ($[888]/[885] = k_1/k_{-1}$) is therefore 0.33 ± 0.12.

Tetracycline inhibits conformational switching

Tetracycline is known to bind to the 30S ribosomal subunit close to A892 of H27 (*50, 127*). To ask whether tetracycline may specifically interfere with conformational switching of H27, we studied the effects of tetracycline and, as controls, the aminoglycoside antibiotics streptomycin and kanamycin on the *9mer*-induced conformational shift of the FRET labeled construct *FEH27T* into the 888 conformation. As shown in Figure 2-6A, neither addition of 1 mM streptomycin nor kanamycin has any impact on the observed FRET change upon *9mer* addition. By contrast, in the presence of 1 mM tetracycline an ~50% decrease in the relative change in FRET efficiency is observed. We performed similar experiments in the presence of increasing concentrations of tetracycline, leading to a hyperbolic decrease in the relative FRET change, which yielded an apparent dissociation constant for tetracycline of 120 ± 40 μM (Figure 2-6B).

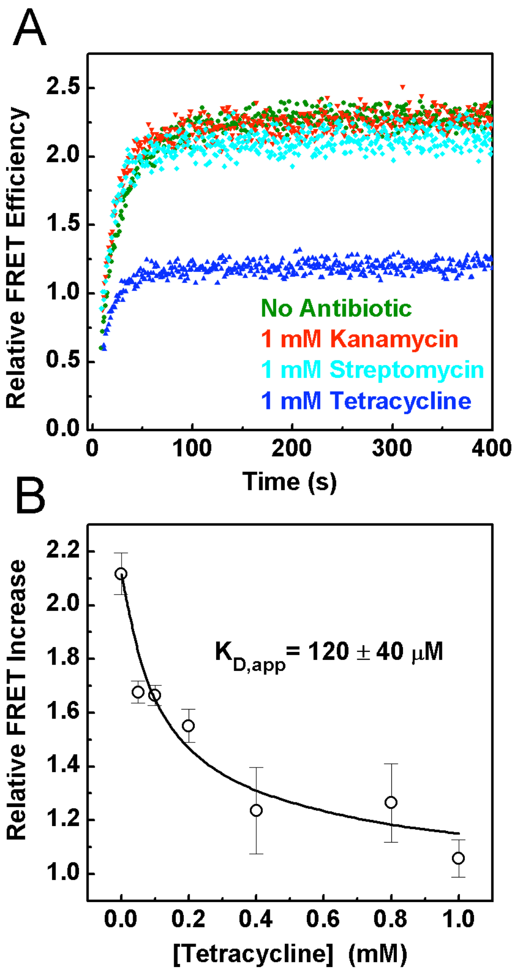


Figure 2-6. Effects of streptomycin, kanamycin, and tetracycline on the 9mer-induced shift of the FRET labeled construct FEH27T to the 888 conformation. A) Plot of the relative FRET efficiency recorded after the addition of 400 nM *9mer* DNA to *FEH27T* in the presence of 1 mM tetracycline, streptomycin, or kanamycin as indicated. B) Plot of the change in the relative FRET efficiency upon addition of 400 nM *9mer* in the presence of increasing concentrations of tetracycline; black line, fit of the data to a modified form of the Hill equation (eq 2-1) to derive an apparent dissociation constant for tetracycline, $K_{D,app}$.

2.4 Discussion

The diverse functions of RNA in storage, processing, and regulation of genetic information derive from its structural diversity. RNA can form long straight helices when storing genetic information of a virus, or assume a fold as intricate as a ribosomal particle when performing a tightly controlled catalytic task. The limited four-letter alphabet of RNA allows for a multitude of (nearly) isoenergetic alternative structures, which can interconvert as required for a specific task such as the translational cycle. This is perhaps the main reason why the ribosome was conserved up to the present day as an RNA-based enzyme after it enabled the template directed synthesis of peptides and protein enzymes in a prebiotic RNA World (*44, 128*).

The kinetics and thermodynamics of interconversion of alternative folds of an RNA determine its functional performance. NMR studies and MD simulations suggest that individual base pair lifetimes are in the millisecond time range (*99, 100*), while a full ribosomal cycle probably takes several tens to hundreds of milliseconds (*43, 129*), suggesting that individual structural rearrangements in the ribosome must occur on the intermediate time scale. Here we have used a combination of 1D ^{1}H-NMR spectroscopy, UV-induced photo-cross-linking, and fluorescence spectroscopy to acquire a kinetic and thermodynamic framework for a possible conformational rearrangement in the slippery sequence of H27 from *E. coli* 16S rRNA using model systems of identical sequence. We find a specific NMR resonance — that of the imino proton of G890, which in the 885 conformation is looped out in a loop E motif but in the 888 conformation is involved in a Watson-Crick base pair — to be a key indicator for the presence of the 885 and 888 conformations. Both the 885 and 888 conformations are highly populated in the isolated H27 under NMR conditions, with a slight, 57%-over-43% excess of the 885 conformation (Figure 2-2). UV irradiation produces an efficient intramolecular photo-cross-link between nucleotides 892 and 905 that is distinct from an 891/906 cross-link expected for a loop E motif (Figures 2-3 and 2-7). Taken together with the fact that neither of the two “locked” conformers cross-links, these findings further support the notion that the isolated H27 coexists in multiple conformations. While these observations do not provide direct evidence for an exchange between conformations, fluorescence resonance energy transfer (FRET) shows that addition of a partially complementary DNA oligonucleotide rapidly shifts the 885 to the 888 conformation, revealing an elementary rate constant for the forward reaction of 250 s^{-1} under conditions adapted from assays of ribosomal function (Figures 2-4 and 2-5). The reversal is three-fold as fast, thus thermodynamically favoring the 885 over the 888 conformation, possibly due to the stabilizing effect of the loop E motif. Finally, the antibiotic tetracycline, known to bind to the 885 conformation of H27 in the ribosomal crystal structures (*50, 127*) (Figure 2-7), suppresses the induced shift to the 888 conformation with a concentration dependence yielding an apparent dissociation constant of 120 μM. This is consistent with a stabilizing thermodynamic effect of tetracycline on the 885 conformation.

Surprisingly Rapid Exchange Kinetics of a Slippery Sequence. The rate constants we estimate for the reversible interconversion of the 885 and 888 conformers of H27, 250 s^{-1} and 750 s^{-1}, are only somewhat slower than the opening rates of individual base pairs, depending on their nature and position in an RNA duplex. For example, non-terminal G:C base pairs were found to open to an extent sufficient for imino proton exchange with solvent at rate constants of ~20 s^{-1} to ~400 s^{-1}, whereas non-terminal A:U base pairs have estimated opening rate constants faster than 10,000 s^{-1} (*99*). End fraying often leads to additional acceleration of the opening of terminal base pairs (*130*). The fact that the reversible 885-to-888 conformational change, entailing a three-base-pair slippage (Figure 2-1C), can occur surprisingly rapidly compared to opening of a single base pair suggests that slippage is mechanistically not much more challenging than a simple base pair opening. Interestingly, little is known about the underlying molecular mechanism of slippage, despite the fact that it can contribute significantly to biologically important processes such as DNA mutagenesis and evolution of genomes (*131*) as well as ribosomal frameshifting (*101*). [Our results suggest that the opening of (or at least partial loss of hydrogen bonding in) a single base pair such as G887:C910 in the 885 conformer (Figure 2-1C) may be the starting point for slippage. The adjacent G888:A909 noncanonical pair may then open to accommodate the first new pairing, G888:C910, followed by rapid adjustments forming base pairs G887:U911 and G886:C912. Similar sliding has to happen twice more until the 888 conformation with its G888:C912 base pair is reached.] One of the multiple conformational intermediates of such a sequential sliding mechanism may then lead to efficient cross-linking of nucleotides A892 and U905, which are nearly 8 Å apart and separated by an unpaired C893 in the 885 conformation of the ribosomal crystal structures (Figure 2-7) (*88*). The latter notion is further corroborated by the fact that we do not find evidence in the NMR spectra of any of our three H27 constructs for involvement of U905 or G894 (G's also have iminos) in a G894:U905 base pair (Figure 2-2), which is consistent with the cross-linking region being dynamic.

The proposed mechanism for H27 conformational exchange is likely analogous to the essential rearrangement observed in the substrate domain of the VS ribozyme, where a series of adjacent guanines also facilitates slippage (*102, 132*). A related exchange between frayed base pairs has been proposed as a mechanism for branch migration in DNA recombination events (*133, 134*). As may be expected, the topologically more challenging branch migration between four DNA strands proceeds with a rate constant of ~35 s^{-1} (*133*) roughly an order of magnitude slower than our conformational exchange in H27 from 16S rRNA. Crothers and co-workers previously studied a large structural rearrangement involving branch migration in the spliced leader (SL) RNA from *Leptomonas collosoma* (*135*). In a similar approach as employed here, the authors used partially complementary DNA oligonucleotides to specifically shift the secondary structure of SL RNA towards one or the other of its two nearly isoenergetic conformations. Consistent with the larger scale of rearrangements in SL RNA compared to H27, the authors estimated an interconversion rate constant of ~7 s^{-1} (*135*), about two orders of magnitude slower than H27 conformational exchange.

Thermodynamics Slightly Favors the 885 Conformation. Using the nearest-neighbor free energy parameters of Zuker and Turner (*136, 137*) the 888 conformer of H27 is computationally predicted to be 2.5 kcal/mole more stable than the 885 conformer (-8.9 versus -6.4 kcal/mol). By contrast, the 888:885 equilibrium constant derived from our FRET experiments is 0.33, that is, it slightly favors the 885 conformation by ~0.7 kcal/mol. Certainly, differences in conditions (1 M NaCl, 37 ^{o}C for the computational prediction, 50 mM NH_4Cl, 20 mM $MgCl_2$, 25 ^{o}C in our experimental studies) may contribute to this discrepancy, especially since under NMR conditions (50 mM NaCl, 8 ˚C) we find an intermediate 888:885 distribution of 0.427/0.573 = 0.75. However, the computational algorithm does not account for the extended noncanonical base pairs in the loop E motif of the 885 conformation and in fact adds a penalty to the energy calculation for the larger "internal loop" represented by this motif (*136*). Of course, the 888 conformation will also likely contain (unaccounted for) noncanonical base pairs, which will partially counter the imbalance in the calculation. But it is also tempting to speculate that perhaps the 885 conformation's loop E motif is a particularly stable structure that slightly stabilizes the 885 over the 888 conformation, as we observe experimentally. There is reason to believe that the loop E motif may be a particularly stable structural element. Loop E has been recognized as a widespread organizing motif in the ribosomal RNAs as well as the hairpin ribozyme (*98*). The loop E motif, when isolated from the ribosomal matrix, is conducive to NMR spectroscopy (*117*),X-ray crystallographic structure determination (*138*) , thermodynamic measurements (*139*), and is reported to remain stable in extended molecular dynamics simulations (*140, 141*). Interestingly, we find it to be thermodynamically stabilized even further by tetracycline binding, which may be explained by the crystallographically observed interaction with A892 and C893 in the 885 conformation of H27 (Figure 2-7) (*50, 127*). Nevertheless, the 885 and 888 conformations are still close to isoenergetic, suggesting that yet unknown noncanonical base pairs must be stabilizing the 888 conformation as well.

The Structural and Functional Context of H27 in the Ribosome. In all crystal structures of the 30S ribosomal subunit, the minor groove of the H27 loop E motif closely packs against H44 of the 3' terminal domain, which harbors the decoding center in 16S rRNA. Additionally, two bases adjacent to the H27 loop E, C893 and G894, make hydrogen bonding and stacking interactions, respectively, with U244 from H11 of the 5' domain (*88*). It is plausible that these interactions within the ribosomal matrix further stabilize the 885 conformation over the 888 conformation, both thermodynamically and kinetically. A particular boost to the original switch helix hypothesis that proposed cycling of H27 between the 885 and 888 conformations during translation (*52, 95*) came from the fact that in all crystal structures H27 knits together distant parts of 16S rRNA,

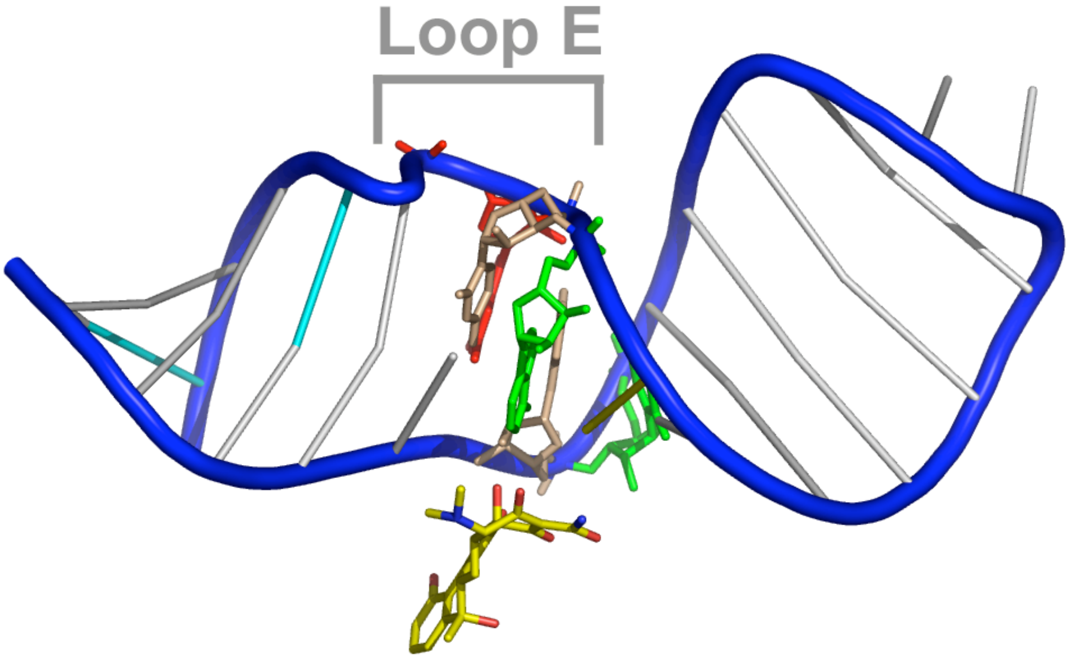


Figure 2-7.H27 from *E. coli* 16 rRNA with bound tetracycline. (PDB ID 1HNW) (*50*) Nucleotides 885 and 888 are colored in cyan and the looped out G890, part of loop E and producing a key resonance in our NMR studies, is colored red. A cross-link characteristic for loop E involves U891 and A906 (tan), whereas the cross-link observed in our studies of the isolated H27 occurs between A892 and U905 (both green). In the crystal structure, tetracycline (yellow) interacts with nucleotides A892 and C893 (yellow stick). This figure was constructed with Nuccyl and rendered with MacPyMOL.

particularly the decoding center in H44 with H11 of the 5' domain. Upon message decoding, H27 then could plausibly transduce local dynamics into the global structural rearrangements observed by cryo-electron microscopy (*30, 48, 49, 86, 87*). If indeed no H27 conformational cycling occurs in present day ribosomes, as is the current assertion (*54, 92*), it is intriguing to ask how tertiary structure contacts in the ribosomal matrix may so completely suppress the inherent dynamics we observe for the two nearly isoenergetic conformations of an isolated H27. Given the high evolutionary conservation of the H27 sequence (*93*) one may speculate that perhaps a prebiotic RNA-only ribosome relied on conformational switching. Yet when translocation became catalyzed by an EF-G-like protein cofactor, more subtle conformational adjustments around H27 began to suffice, especially in a region that overall appears to be very dynamic (*49, 142*). Such subtle adjustments may involve, for example, only partial slippage by one or two, rather than all three nucleotides of the full 885-to-888 conformational switch. However, there would seem to remain at least the formal possibility that the catalytic action of EF-G may mask the deleterious effects in a cell line only producing ribosomes incapable of adopting the 888 conformation, and that in wild type ribosomes, the H27 switch may still be active.

Suppression of conformational adjustments around H27 may also present another mode of action for the broad-spectrum antibiotic tetracycline that primarily seems to inhibit translation by blocking binding of aminoacylated tRNA to the ribosomal A-site. Two independent atomic resolution crystal structures of *T. thermophilus* 30S subunits in complex with tetracycline revealed multiple binding sites for the antibiotic (*50, 127*). A primary binding site is located near H34 in the 30S A-site, which is in agreement with previous biochemical studies as well as cryo-EM localization of the Tet(O) protection protein (*143, 144*). Tetracycline also inserts between H27 and H11 to bind to A892 and C893, becoming the secondary binding site that shows a slightly lower occupancy relative to the primary binding site when 80 μM tetracycline is soaked into 30S crystals (*50*) (Figure 2-7). This secondary site is also observed in the ribosome upon addition of 120 μM tetracycline through the protection of A892 from DMS modification (*145*) and a direct UV-cross-link between G890 and the antibiotic (*146*). Strikingly, our apparent tetracycline dissociation constant of 120 μM based on suppression of H27 conformational dynamics matches the concentration range where it appears to bind to H27 in the ribosome. Clearly, tetracycline is promiscuous in its binding to RNA since four additional binding sites are partially occupied in one of the two 30S crystal structures (*127*) and it has been observed to inhibit catalysis in both the hammerhead and hepatitis delta virus ribozymes (*147, 148*). However, the apparent dissociation constants are markedly reduced (to ~500 μM) in the latter cases, suggesting that the tetracycline effect we observe for H27 is quite specific. Thus, while the ribosome may no longer fully exploit the rapid conformational dynamics inherent to a slippery sequence such as that of H27 from *E. coli* 16S rRNA, a broad-spectrum antibiotic such as tetracycline still specifically binds H27, presumably to inhibit translation. We are only starting to glimpse the complex linkage between local and global structural rearrangements in the ribosomal machinery during protein biosynthesis, yet it is clear that model systems such as H27 will help bridge our current gap in understanding.

CHAPTER 3:

SOLUTION PROBING OF METAL ION BINDING BY HELIX 27 FROM *ESCHERICHIA COLI* 16S RIBOSOMAL RNA[2]

3.1 Introduction

As the most abundant cellular divalent metal ion, Mg^{2+} is thought to be crucial to RNA function by promoting folding through a reduction of the electrostatic repulsion between the negatively charged backbone phosphates. RNA electrostatic potential and ion distribution calculations in combination with spectroscopic and thermodynamic measurements have led to our current understanding that a majority of Mg^{2+} ions at physiologic (millimolar) concentrations readily traverse the solvent accessible surface of an RNA and move from one negative potential sink to the next on the millisecond timescale (*149-152*). Any RNA dynamics will be linked to rearrangements of this diffusely bound, fully hydrated cation cloud around the RNA and *vice versa*. Occasionally a Mg^{2+} ion may site-specifically reside in a region of particularly high electronegative surface potential, such as close to phosphates partially buried within a tightly folded RNA, however, the energetic penalty for partial dehydration must be overcome to form an inner-sphere coordinated metal-RNA chelate (*150, 152*). Recent

---

[2] Adapted with permission from Lambert, M.N., Hoerter, J.A., Pereira, M.J., and Walter, N.G., *RNA* 11, 1688-700. Copyright 2005, Cold Spring Harbor Press. Meredith Lambert carried out the NMR experiments, John Hoerter carried out the time-resolved FRET experiments, and Miguel Pereira carried out the Terbium(III)-footprinting.

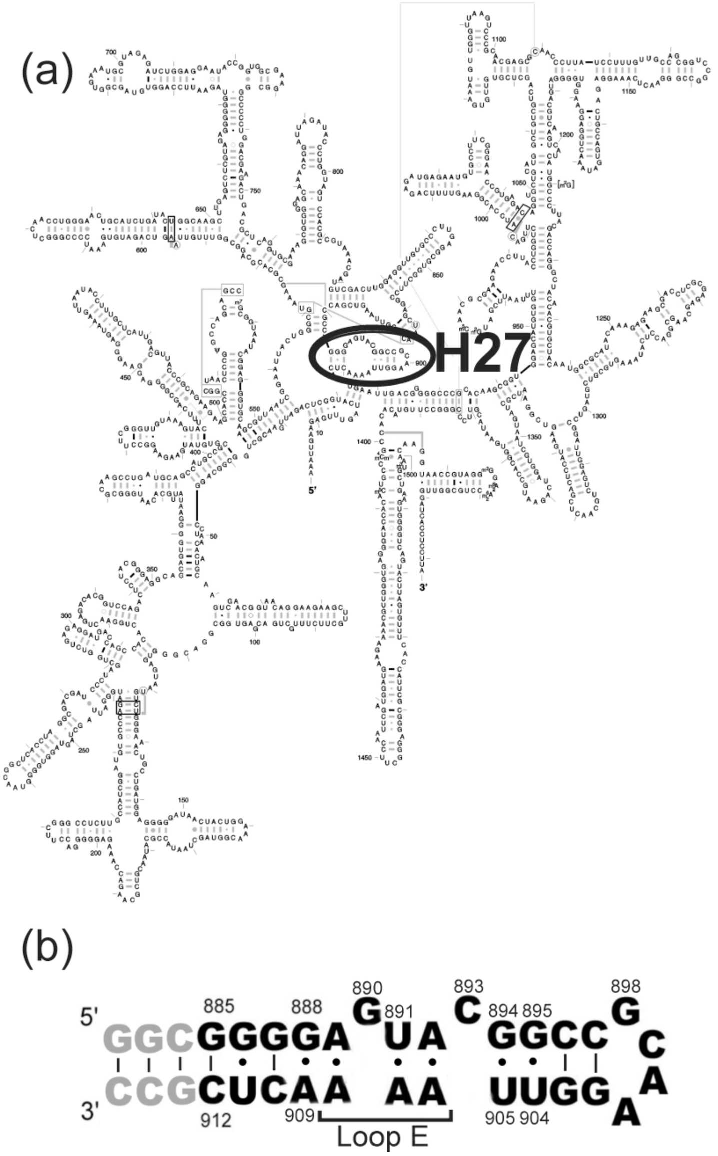


Figure 3-1. Helix 27 is centrally located within 16S rRNA. a) Secondary structure map of 16S rRNA from *E. coli*, (*90*) with helix 27 circled. b) Helix 27 NMR construct. The proposed 912-885 secondary structure of the helix 27 sequence is shown in black. Gray nucleotides were added to preferentially stabilize the 885 conformation for NMR study. Dashes represent canonical Watson-Crick base pairs, dots represent known non-Watson-Crick base pair interactions.

advances in x-ray crystallography have yielded many insightful RNA crystal structures, which often depict metal ions bound in specific sites. Relatively little is known, however, about how such seemingly stable metal ion binding sites in the frozen solid state of a crystal compare with metal ion binding properties in solution (*150-154*).

Helix (H)27 is located within the decoding region of 16S ribosomal (r)RNA from *Escherichia coli* (Fig. 3-1a) (*45*). Cryo-electron microscopy studies have indicated that relative ratcheting motions of the 30S and 50S ribosomal subunits during translational translocation are centered around H27 (*49*), which serves as an inter-subunit bridge in a very dynamic region of the ribosome (*155*). In 1997 Lodmell and Dahlberg proposed that the H27 stem-loop is dynamic and alternates between two plausible base-pairing patterns (Fig. 3-1b), the 912-885 (885) and 912-888 (888) conformations, early in the translational cycle (*52*). Subsequent crystal structures of “open” and “closed” forms of the 30S subunit in complex with aminoglycoside antibiotics consistently depict H27 in the 885 conformation (*91*), as do crystal structures of hyperaccurate ribosomes containing S12 mutations (*97*). Follow-up studies by the Dahlberg group revealed that their previous findings originated from a synergistic effect between H27 and selective marker mutations (*54*), suggesting that H27 indeed does not need to switch between conformations during translation. In contrast, an isolated H27 exists in a rapid dynamic equilibrium between the 885 and 888 conformations, indicative of a low energy barrier for conformational switching (*55*). A simple extension of the terminal helix of this isolated H27 stabilizes the 885 conformation (*55*), making it a suitable system to probe the metal ion binding of this ribosomal structural element in solution. This is of particular interest since the 885 conformation contains a loop E motif (Fig. 3-1b), a common structural motif in RNA (*98*). Bacterial 30S ribosomal subunit crystal structures depicting magnesium ion binding sites consistently place two such Mg^{2+} binding sites at opposite ends of H27 (*44-46, 156, 157*) ; one (Mg73) adjacent to the G886-U911 pair, and another one (Mg126) within the GNRA tetraloop (Fig. 3-1b), a common metal ion binding site (*154, 158-160*). The metal-RNA distances are all >2.8 Å, consistent with the notion that these metal ions are outer-sphere (diffusely) bound. However, it is unclear whether in solution similar and/or additional sites are occupied.

To begin to probe the relationship between metal ion binding sites inferred from crystal data of H27 and those in solution, we have applied NMR chemical shift assays, Tb^{3+} footprinting, and time-resolved (tr)-FRET based distance measurements to an extended construct representing the 885 conformation of H27 from *E. coli* 16S rRNA (Fig. 3-1b). Our NMR and Tb^{3+} footprinting data verify the crystallographically observed base-pairing pattern and Mg^{2+} binding sites in solution (*44-46, 156, 157*). In addition, NMR analyses point to metal ion interactions in the vicinity of the tandem G894-U905 and G895-U904 wobble pairs that are not predicted by the crystal structures, while time-resolved (tr)-FRET measurements reveal an end-to-end distance increase in H27 upon association with Mg^{2+}. We propose that only a subset of H27-metal ion interactions are described in crystal structures of the small ribosomal subunit where magnesium ion binding sites are observed, and that local dynamics on the NMR timescale afforded by solution conditions may contribute to these differences. Our studies of H27 highlight potential discrepancies between RNA-metal ion interactions in solution and in crystals.

3.2 Materials and Methods

Design and synthesis of the helix 27 NMR construct. Helix 27 RNA construct 885 (5'-GGC GGG GAG UAC GGC CGC AAG GUU AAA ACU CGC C-3') (Fig. 3-1b) was designed with the addition of extra G and C nucleotides to the helix terminus of the natural H27 sequence to stabilize the 912-885 conformation over the 912-888 and prevent switching between the two H27 base-pairing patterns. The molecule was transcribed *in vitro* from a double-stranded DNA template containing a T7 RNA polymerase promoter region, as described previously (*55*). For the uniformly $^{13}C/^{15}N$-labeled sample, isotopically enriched nucleotides (Silantes) were substituted for unlabeled nucleotide triphosphates. Following transcription, EDTA was added to the reactions to a final concentration of 60 mM, and reaction mixtures were extracted with an equal volume of phenol followed by two extractions with equal volumes of a 24:1 $CHCl_3$:isoamyl alcohol mixture. Centricon-3 centrifugal filter devices (Amicon) were used to concentrate the protein-free extracts and remove excess nucleoside triphosphates. H27 885 samples were then separated from abortive transcripts and template DNA by electrophoresis on denaturing, 8 M urea, 20% polyacrylamide gels. RNA was detected by UV shadowing and appropriate bands were excised and eluted by crushing and soaking in 5 mM EDTA. The RNAs were further purified by anion exchange with Sephadex A-25 resin, and exchanged into NMR buffer (10 mM Na_iPO_4, pH 6.4, 0.1 mM EDTA, 50 mM NaCl) using Centricon-3 concentrators. Samples were concentrated to approximately 200 µL, and 99.9% D_2O (Aldrich) and NMR buffer were each added to the samples in predetermined amounts so that the final volume for each sample was 225 µL, and the final concentration of D_2O (for lock) was 5%. Final concentrations, determined by absorbances at 260 nm, of the 885 construct were ~0.35 mM and ~0.6 mM, for the homonuclear and $^{13}C/^{15}N$ isotopically labeled samples, respectively. D_2O-matched microvolume NMR tubes (Shigemi) were used for data collection.

NMR data acquisition and processing. One-dimensional Mg^{2+} titration experiments with the helix 27 885 construct were performed at 4 ºC on a 600 MHz Bruker Avance spectrometer. Spectra were collected with 128 transients, and a jump-and-return water suppression scheme coupled with a WATERGATE pulse train was utilized (*108, 109*). Sterile-filtered 1 M and 100 mM stock solutions of magnesium chloride hexahydrate, 99.995% (Aldrich), were carefully added to each sample to stepwise increase their Mg^{2+} concentrations. Magnesium ion concentrations noted in Figure 3-3 and referred to throughout the text reflect the final concentration of the added metal ions in molar equivalents, relative to the RNA concentration, and are not adjusted to reflect the free metal ion concentrations according to uptake of metal ions by the RNA or the 0.1 mM EDTA background present in the NMR sample buffer. The sample was allowed to re-equilibrate at 5 ºC and was re-locked and re-shimmed, the probe was re-tuned, and the $\pi/2$ pulse recalibrated for each Mg^{2+} concentration. Spectra were processed with NMRPipe using a solvent filter and a cosine-bell apodization function, and zero-filled once before Fourier transforming (*111*). NMRDraw was used to visualize each spectrum and detect peaks (*111*). Data were then reprocessed using Bruker software to create the stacked plot shown in Figure 3-3. Plots of chemical shift (accurate to an error of about ±0.004 ppm due to limited digital resolution of the spectrum (sweep width (ppm) divided by number of acquisition points)) vs. added Mg^{2+} equivalents were fit with the following binding isotherm for a system in fast exchange, where the RNA concentration is of the same magnitude as the metal ion concentration (*161, 162*).

$$\delta_{obs} = \delta_f + \left(\left(\delta_b - \delta_f\right) \frac{\left\{ \left[[ion]_t + [RNA]_t + Mg_{1/2} \right] - \sqrt{\left([ion]_t + [RNA]_t + Mg_{1/2}\right)^2 - \left(4[RNA]_t[ion]_t\right)} \right\}}{2[RNA]_t} \right)$$

Equation 3-1

δ_{obs} is the observed chemical shift, δ_f is the chemical shift of the unbound imino proton, δ_b is the chemical shift of the fully bound proton, $[\mathrm{ion}]_t$ is the total added magnesium ion concentration in mM, $[RNA]_t$ is the RNA concentration, and $Mg_{1/2}$ is the magnesium half-titration point. Chemical shifts reported are accurate within the digital resolution of the spectrum, or sweep width (ppm) divided by the number of acquisition points. The digital resolution in the Mg^{2+} titration data is therefore 0.004 ppm. It should be noted that the Mg^{2+} concentrations used to compute the binding isotherms are added metal ion concentrations, and do not reflect the true free metal ion concentration since the polyanionic RNA uptakes an unknown number of the metal ions in solution. Therefore, the extracted magnesium half-titration points, $Mg_{1/2}$, can only be compared relative to those of other resonance shifts within the same sample (*163*).

NOESY spectra of exchangeable protons with and without cobalt(III) hexammine, as well as the imino ^{1}H-^{15}N HSQC spectrum, were also acquired at 5 ºC on the Bruker Avance 600 MHz spectrometer. NOESY spectra were collected with 350 and 150 ms mixing times for samples with and without 2 mM hexamminecobalt(III) chloride, 99.999% (Aldrich), respectively. NOESYs were acquired with 2048 complex points, 256 increments, and a sweep width of 12,019.231 Hz in both the directly and indirectly detected dimensions. The carrier was placed at the water proton resonance position in all cases. In the ^{1}H-^{15}N HSQC experiment, ^{15}N decoupling parameters, delays, sweep width (1824.568 Hz) and carrier position (150.911 Hz) were optimized for observation of imino ^{1}H-^{15}N cross peaks. 2048 complex points and 512 increments were collected, with a ^{1}H sweep width of 17985.612 Hz. In these experiments, quadrature detection was achieved by implementation of the States-TPPI method (*110*), and spectra were processed and visualized with NMRPipe and NMRView software (*111*). Processing for these spectra included apodization with a cosine-bell function, application of a solvent filter, and zero-filling once prior to Fourier transforming.

Terbium(III) footprinting experiments. Helix 27 construct Tb885 (5'-CCG CCU GGG GAG UAC GGC CGC AAG GUU AAA ACU CAG GCG G-3') was obtained commercially (Dharmacon) and deprotected according to manufacturer's protocols. Tb885 was subsequently 5' end labeled using [γ-^{32}P]ATP and purified by denaturing, 8 M urea, 20% polyacrylamide gel electrophoresis. The band corresponding to the H27 construct was excised and eluted by diffusion into 1 mM EDTA overnight at 4 ºC, and then ethanol precipitated. For the footprinting reactions, radiolabeled Tb885 and ($\geq$ 4 nanomoles) was first dissolved in 5 μL 2× annealing buffer (100 mM Tris pH 7.5, 40 mM $MgCl_2$, and 100 mM NH_4Cl) and 4 μL distilled deionized water, heated at 70 ºC for two minutes, and allowed to cool at room temperature for an additional five minutes. 1 μL of 10× Tb^{3+} solution (20 mM $TbCl_3$, 5 mM MES pH 5.5) was added to the reaction mixture and incubated at 25 ºC for two hours. A control reaction containing 5 mM MES, pH 5.5 buffer, instead of 10× Tb^{3+} buffer, was also run in parallel. 50 mM EDTA was added to the 10 μL reactions to stop it, and the Tb885 RNA construct was subsequently ethanol precipitated overnight. Precipitated Tb885 RNA was analyzed on an 8M urea, wedged 20% polyacrylamide sequencing gel after resuspension in loading buffer containing 80% formamide and 0.025% each of xylene cyanol and bromophenol blue. The Tb885 RNA sample was run alongside a sequencing ladder of RNA partially digested by G-specific RNase T_1 and alkaline hydrolysis. Gels were exposed to phosphorimager screens overnight at room temperature, and bands were quantified using ImageQuant software of a Molecular Dynamics Storm 840 phosphorimager. The relative extent of excision at each nucleotide (Π) was calculated as described (*164*) from the following equation:

$$\Pi = \frac{\left[\dfrac{\text{band intensity at nucleotide } x}{\left(\sum_i \text{band intensity at nucleotide } i \right)} \right]_{\text{2 mM Tb(III)}}}{\left[\dfrac{\text{band intensity at nucleotide } x}{\left(\sum_i \text{band intensity at nucleotide } i \right)} \right]_{\text{0 mM Tb(III)}}} .. \quad \textbf{Equation 3-2}$$

where x is the analyzed nucleotide position in the RNA, and 0 mM Tb(III) is the control reaction, described above. Data from the Tb885 constructs were further normalized relative to the value calculated for Π_{C899}, which allowed for comparison of results from different gels. Results were within 5% error between experiments.

Time-resolved FRET measurements. RNA construct F885T for FRET experiments was synthesized in a singly fluorescein-labeled form by the Howard Hughes Medical Institute Biopolymer/Keck Foundation Biotechnology Resource Laboratory at the Yale University School of Medicine (New Haven, CT) with the same sequence as construct 885 (5'-GGC GGG GAG UAC GGC CGX AAG GUU AAA ACY CGC C-3'), except that the C899 within the GCAA tetraloop and U911, indicated as X and Y respectively, were replaced with Fluorescein dT and Amino-Modifier C6 dT (Glen Research), respectively. The synthetic RNA was deprotected, purified, and subsequently labeled with tetramethylrhodamine succinimidyl ester (Molecular Probes) as previously described (*55*).

The global structures of the doubly-labeled, conformationally locked H27 construct F885T was studied as a function of Mg^{2+} concentration by tr-FRET analysis. The H27 RNA was annealed by heating at 70 ºC for two minutes and cooling to room temperature over ten minutes. The sample was then incubated at 25 ºC for ≥ 5 min in 50 mM Tris-HCl, pH 7.5, 50 mM NH_4Cl with an appropriate $MgCl_2$ concentration, prior to collection of time-resolved emission profiles of the donor fluorescein using time-correlated single-photon counting similar to previously described procedures (*23, 29*). To measure donor-acceptor distance distributions, two time-resolved fluorescence decays were collected, with and without the acceptor fluorophore. The fluorescein emission decay in the donor-only complex was used to extract three intrinsic donor lifetimes with their fractional contributions by a sum-of exponentials fit. The data from the doubly labeled complex were then fit with the Förster model for distance distributions, as described (*165*). The full width at half maximum (FWHM) of the derived distance distributions did not increase by more than 10% over the whole range of Mg^{2+} concentrations. An additional adjustable fitting parameter that corrects for singly labeled (fluorescein only) RNA was included in these analyses; incomplete acceptor labeling and separation of doubly from singly labeled material combined to give a singly labeled fraction of up to 30%. A control was performed in which the fraction of singly labeled material was intentionally increased by 20%; the singly labeled fraction found in the resulting fit increased accordingly by 20%, as expected. The fits and inter-fluorophore distances at higher Mg^{2+} concentrations were not substantially different when the fraction of singly labeled material was fixed to the value found at zero Mg^{2+} or, alternatively, freely varied.

Mg^{2+} was titrated by incremental addition of 0.5 µL aliquots of appropriate $MgCl_2$ stock solutions, taking into account the volume change; the volume at the end of any given titration increased by not more than 8%. Single distance distributions fit well, as judged by the low reduced χ^2 values (<1.3) and by evenly distributed residuals. To extract absolute distances, a value of 55 Å was used for the Förster distance, R_0, of fluorescein and tetramethylrhodamine (*165*), based on a value of 2/3 for an isotropic orientation factor, which was supported by the high mobility of the fluorophores as evident from their relatively low fluorescence anisotropies (~0.12 for fluorescein and ~0.26 for tetramethylrhodamine) observed in the F885T FRET construct. Mean fluorophore distances were fit with a modified Hill equation as follows:

$$y = R_o + \left[\frac{\left(R_{\max} - R_o\right) x^n}{x^n + Mg_{1/2}^n} \right] \qquad \textbf{Equation 3-3}$$

where $Mg_{1/2}$ is the magnesium half-titration point, n is the Hill coefficient (which was close to unity in both fits indicating lack of cooperativity in metal ion binding), R_0 is the initial FRET distance at 0 mM Mg^{2+}, and $R_{\max}$ is the FRET distance at saturating magnesium ion concentrations. A similar titration was also performed in an NMR-like phosphate buffer (10 mM NaP_i, pH 7.0, 50 mM NaCl, 0.1 mM EDTA) . The pH had to be raised from 6.4 to 7.0 since fluorescein emission is pH dependent and the detected signal diminishes as the pH drops below neutral. Nonetheless, FRET data collected in phosphate buffer gave the same absolute distance changes as data collected in Tris buffer, although the values were offset by approximately 0.5 Å. Extracted $Mg_{1/2}$ values for the two data sets were within error of each other.

3.3 Results

Establishment of the base-pairing pattern of the 885 conformation

Previous studies by our group have indicated that an isolated H27 exists in a rapid dynamic equilibrium between the 885 and 888 conformations, suggesting a low energy barrier to conformational switching (*55*). To stabilize a single conformation and suppress any conformational switching, we here have designed an NMR construct with three G-C (C-G) pairs added to the helix terminus that "lock" it into the 885 conformation (Fig. 3-1b). To verify the predicted base-pairing pattern of the 885 conformation, NOESY spectra of exchangeable protons were collected and imino proton resonances assigned by standard strategies. In particular, we used imino-imino and imino-amino NOEs to identify sequential residues, and ^{1}H-^{15}N HSQC cross peaks to unambiguously distinguish G from U imino proton resonances (*166, 167*). Figure 3-2 shows relevant regions of exchangeable NOESY and ^{1}H-^{15}N HSQC spectra of 885 with sequential connectivities drawn and resonance assignments labeled. Assignments were initially made in 10 mM Na_iPO_4, pH 6.4, 0.1 mM EDTA, 50 mM NaCl, in the absence of Mg^{2+}, so that the effect of metal ions on specific resonances could subsequently be assessed.

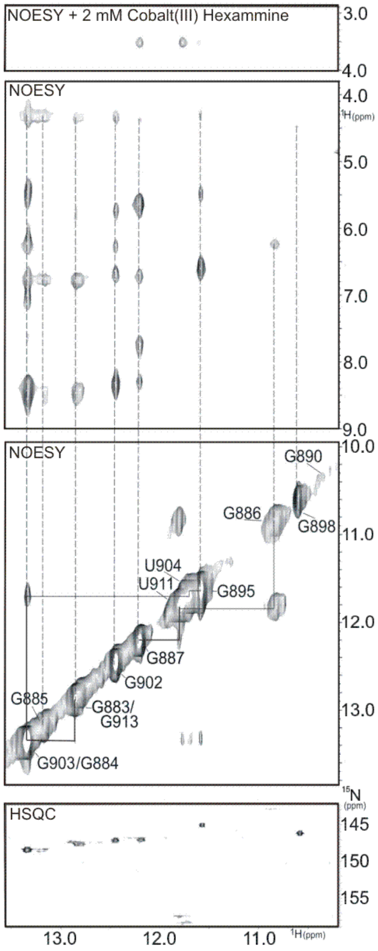


Figure 3-2. Two-dimensional NMR spectra used to assign imino proton resonances in the 885 construct. Imino-amino and imino-imino regions of 150-ms mixing time NOESY spectra of 885, center two panels. Connectivities are drawn with solid lines between nucleotides demonstrating imino-imino NOEs, and nucleotide assignments for imino protons are labeled along the diagonal. Dashed lines connect imino-amino NOEs with the diagonal. Bottom panel depicts portions of a ^{1}H-^{15}N HSQC spectrum that unambiguously distinguishes imino nitrogen resonances of U's (downfield, 156-162 ppm) from those of G's (upfield, 146-149 ppm). Top panel is a portion of a 350 ms mixing time NOESY spectrum acquired with the addition of 2 mM cobalt(III) hexamine. Strong cross peaks are observed between individual imino proton resonances and the averaged $Co(NH_3)_6^{3+}$ hexammine resonance at ~3.7 ppm. NMR data was processed with NMRPipe and spectra were visualized using NMRDraw software (*111*).

NMR spectra of the 885 construct do not show any U NH3 imino proton resonances within the 12-15 ppm Watson-Crick region (Fig. 3-2), consistent with the base-pairing pattern in crystal structures that is devoid of canonical A-U base pairs (Fig. 3-1b) (*44-46, 156, 157*). Instead, the imino proton NMR spectrum of our 885 construct displays all of the characteristics of the loop E motif, as first documented by Szewczak and Moore in the ribosomal sarcin-ricin loop (*117*). This includes a far upfield shifted NH1 imino proton resonance belonging to the looped-out G890 (Fig. 3-2). In addition, the spectrum shows a sharply upfield shifted G NH1 resonance at 10.6 ppm, a GNRA tetraloop hallmark, which we were able to assign to G898 (Fig. 3-2). Two upfield U imino proton resonances were also observed between 11.5 and 12.0 ppm that display strong NOEs to upfield G NH1 resonances, indicative of G-U pairing. These resonances were assigned to the U NH3 resonances within the G895-U904 and G886-U911 wobble pairs (Figs. 3-1b and 3-2). The line widths of a number of resonances throughout the molecule are broadened slightly, including G886 NH1, U911 NH3, U904 NH3, and G890 NH1, providing evidence for molecular motion near the G-U wobble pairs and the looped out base. Nevertheless, the assignable number of imino proton resonances in Figure 3-2, together with the high resolution of the spectrum, indicates that the 885 construct forms a single, well-defined hairpin. This hairpin contains key features of the loop E motif, namely a non Watson-Crick paired G within the helix and at least two G-U pairs as evidenced by NOESY spectra of exchangeable protons. Therefore, it most likely assumes the conformation observed for H27 in crystal structures of the 30S subunit and the intact ribosome, as expected (*44-46, 156, 157*).

NMR based detection of cobalt(III) hexammine and Mg^{2+} binding

To begin to pinpoint binding sites of multivalent metal ions in solution, we utilized cobalt(III) hexammine ($Co(NH_3)_6^{3+}$) as a probe for outer-sphere Mg^{2+} coordination in the 885 construct. $Co(NH_3)_6^{3+}$ has traditionally been employed as a fully coordinated magnesium hexahydrate mimic because it is strongly positively charged, and its octahedral geometry and the hydrogen-bonding capabilities of its exchange-inert ammonium ligands resemble those of the primary-shell water molecules of $Mg(H_2O)_6^{2+}$

(*158, 168, 169*). $Co(NH_3)_6^{3+}$ amino protons give rise to an averaged ^1H-NMR signal at ~3.7 ppm, and NOEs to this distinctive resonance accurately identify metal ion binding sites within an RNA (*158, 168, 169*). Therefore, we collected NOESY spectra of exchangeable protons for the 885 construct upon addition of hexamminecobalt(III) chloride to a final concentration of 2 mM. Most of the imino proton resonances in the H27 construct exhibit weak cross peaks to the hexammine resonance and thus association with the metal-ion complex, indicative of diffuse ion binding within the helix grooves. Only a select few resonances display stronger NOEs with the hexammine resonance, suggestive of more specific binding of $Co(NH_3)_6^{3+}$ (Fig. 3-2, top panel). Significantly, relevant crystal structures of the 30S ribosomal subunit depict a Mg^{2+} ion binding in the major groove of H27 near the G886-U911 pair (*45, 156*), consistent with the NOEs we observe from the G887 NH1 and U911 NH3 resonances to the averaged resonance of the ammonium ligands of the $Co(NH_3)_6^{3+}$ complex (Fig. 3-2). Observation of such NOEs may be expected, since $Co(NH_3)_6^{3+}$ was previously demonstrated to specifically bind to the unusual major groove geometry created by a tandem G-U wobble pair (*168*). In contrast, we do not observe a cross peak between the ammonium ligands of $Co(NH_3)_6^{3+}$ and the tetraloop G898 NH1 resonance.

Next, magnesium chloride titrations were performed to detect potential Mg^{2+} binding sites by chemical shift analysis for comparison with those from our $Co(NH_3)_6^{3+}$ experiments and those found in relevant 30S ribosomal subunit crystal structures (*44-46, 156, 157*) . Figure 3-3a shows one-dimensional ^1H-NMR spectra of the 885 construct acquired in the presence of increasing molar equivalents of Mg^{2+} (at an RNA concentration of 0.35 mM in 10 mM Na_iPO_4, pH 6.4, 50 mM NaCl). Almost every imino proton resonance in the molecule shifts slightly (typically < 0.05 ppm) in position (Fig. 3-3b), indicating that Mg^{2+} binds diffusely throughout the molecule, and/or metal binding induces structural changes that result in a slight shifting of imino proton resonances. As is typical in NMR spectra of RNA, imino resonances broaden with the addition of high concentrations of Mg^{2+} ions; however, addition of low (millimolar) concentrations of Mg^{2+} first results in the narrowing of several resonances in the molecule, including those belonging to G886 NH1, G895 NH1, G898 NH1, G903 NH1, and U904 NH3, consistent

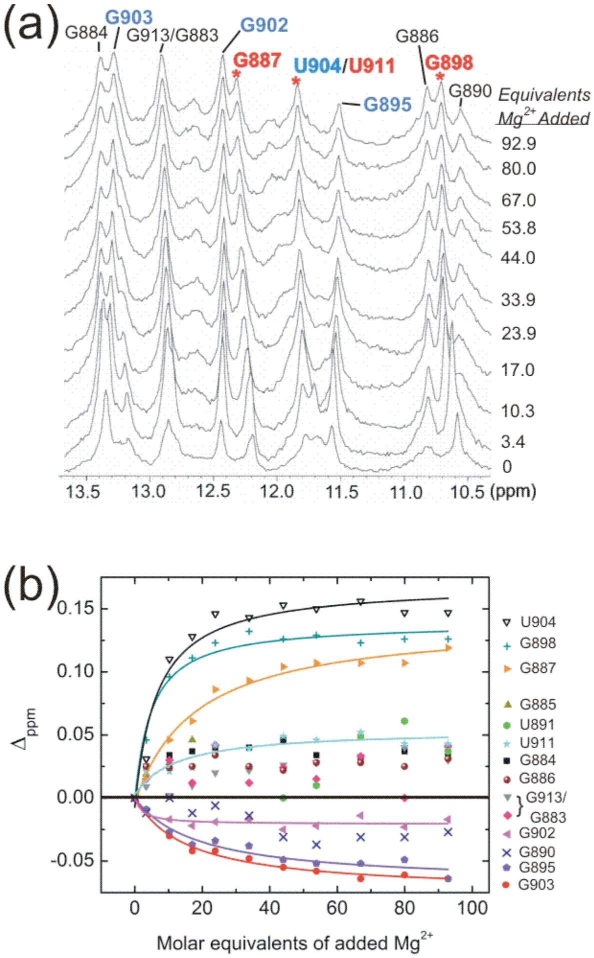


Figure 3-3. NMR-detected magnesium titrations of the H27 885 construct. a) Stacked 885 imino proton spectra acquired with increasing Mg^{2+} concentration, as indicated by added molar equivalents. Imino proton resonance assignments are labeled, and those belonging to G887, G898, U911, are marked with asterisks in red, and those belonging to U904, G903, G902 and G895 are labeled in blue. b) Plot of chemical shift changes for imino resonances of the 885 construct in dependence of the Mg^{2+} molar equivalents. Curves are shown for those resonances that fit to binding isotherms describing a one metal per site model (*161, 162*). Individual resonance positions were found at each metal ion concentration using the peak finder utility in nmrDraw and defining parameters for both a reasonable threshold for positive peak detection and a χ^2 probability threshold for noise peak rejection by χ^2 test.

with Mg^{2+}-induced stabilization of local and/or global architecture within the molecule (Fig. 3-3a).

Specific resonances show relatively significant chemical shift changes with increasing metal ion concentration, suggesting possible specific association with Mg^{2+} (Fig. 3-3 a and b), and/or more substantial rearrangements of local architecture upon metal binding. For these resonances, chemical shift changes as a function of Mg^{2+} concentration were fit well with binding isotherms based on stoichiometric, fast exchange binding between the RNA site and a Mg^{2+} ion (*161, 162*). The resulting relative Mg^{2+} half-titration points ($Mg_{1/2}$) are shown in Table 3-1. The tetraloop G898 NH1 resonance demonstrates a particularly large (0.13 ppm) downfield shift with a corresponding $Mg_{1/2}$ of 1.5 ± 0.3 mM. This observation is consistent with previous studies that have shown that GNRA tetraloops bind Mg^{2+} tightly (*154, 158-160*). Other resonances that display significant downfield shifts were those of U904 NH3 (0.16 ppm) and G887 NH1 (0.12 ppm), with relative Mg^{2+} half-titration points of $Mg_{1/2} = 2.0 \pm 0.6$ mM and $Mg_{1/2} = 6.1 \pm 1.1$ mM, respectively (Fig. 3-3b, Table 3-1). The resonance shift of U911 NH3 was less significant (0.05 ppm), but fit a binding isotherm, yielding $Mg_{1/2} = 5.4 \pm 1.2$ mM (Fig. 3-3b, Table 3-1). Other imino resonances display slight upfield shifts with increasing Mg^{2+} concentration that fit our stoichiometric binding model, namely those belonging to G895 (-0.06 ppm), G902 (-0.02 ppm), and G903 (-0.06 ppm); extracted relative Mg^{2+} half-titration points were $Mg_{1/2} = 6.5 \pm 2.3$ mM, $Mg_{1/2} = 0.7 \pm 0.6$ mM, and $Mg_{1/2} = 5.4 \pm 0.9$ mM, respectively (Fig. 3-3b, Table 3-1). Among the latter six nucleotides, G902 participates in the closing base pair of the GCAA tetraloop, while G895, U904, and U911 participate in G-U wobble pairs, and G887 and G903 are adjacent to G-U wobble pairs. Such chemical shift changes are consistent with metal ion binding near these locations, as reported in other RNAs featuring G-U pairs (*158, 168, 169*) and GNRA tetraloops (*154, 158-160*). These data, however, cannot distinguish between the possibilities that the chemical shifts are due to either a direct impact of the cation or structural rearrangements affected by nearby metal ion binding. The fact that G887 and U911 imino proton resonances display strong NOESY crosspeaks to the ammonium ligands of $Co(NH_3)_6^{3+}$ (Fig. 3-2) does provide such a distinction and supports the existence of a metal ion binding site in this region, next to the G886-U911 wobble pair. By contrast, we found no

Nucleotide	$Mg_{1/2}$ (mM)[*]	Δ_{max} (ppm)[†]	Strong NOE to $Co(NH_3)_6^{3+}$ amino protons?	Relative fraction cleaved in Tb^{3+} footprinting experiments (Normalized Π)
G887	6.1 ± 1.1	0.117 ± 0.004	Yes	0.19
G895	6.5 ± 2.3	-0.057 ± 0.004	No	0.42
G898	1.5 ± 0.3	0.132 ± 0.004	No	0.61
G902	0.7 ± 0.6	-0.020 ± 0.004	No	0.38
G903	5.4 ± 0.9	-0.064 ± 0.004	No	0.30
U904	2.0 ± 0.6	0.157 ± 0.004	No	0.41
U911	4.2 ± 2.4	0.048 ± 0.004	Yes	0.16

Table 3-1. Comparison of solution NMR spectroscopy and terbium footprinting data for potential metal ion binding sites within H27. [*] $Mg_{1/2}$ values were determined from the one metal per site model described in Materials and Methods, where the total Mg^{2+} ion concentration is used as input (*161, 162*). This concentration necessarily does not reflect the (unknown) free metal ion concentration so that the derived relative half-titration points $Mg_{1/2}$ should only be compared amongst the resonance shifts observed here (*163*).

[†] Reported Δ_{max}values are accurate within the digital resolution of the spectrum, which is 0.004 ppm.

evidence for a strong NOE between the tetraloop G898 NH1 resonance and the $Co(NH_3)_6^{3+}$ amino proton, nor was there evidence for such an NOE involving the U904 NH3 resonance, although these two resonances display the most significant shifts upon addition of Mg^{2+}; a possibility is that Mg^{2+} and $Co(NH_3)_6^{3+}$ may bind to (slightly) different sites within H27.

Terbium(III) footprinting confirms secondary structure and metal binding

The lanthanide metal ion terbium(III) has been used to footprint the secondary and tertiary structure of various RNAs and to identify potential regions of enhanced structural dynamics and Mg^{2+} binding (*112, 164, 170-172*). Tb^{3+} facilitates RNA backbone scission by first binding near the negatively charged backbone phosphate groups, and then abstracting the proton from a nearby 2’ hydroxyl group. The resulting 2’ oxyanion attacks the adjacent phosphodiester linkage, which results in scission of the nucleotide chain at that location (*170, 173*). Specific metal binding sites, single-stranded tracts, and non-Watson-Crick base-paired regions are particularly vulnerable to Tb^{3+}-induced hydrolysis (*174-176*). However, slow scission at all nucleotides will eventually occur at sufficiently high terbium(III) concentrations.

Here we have utilized Tb^{3+}-mediated footprinting as a complementary biochemical technique to map the secondary structure and Mg^{2+} binding sites in the 885 conformation of H27. Trace amounts (< 4 nmol) of ^{32}P radiolabeled footprinting construct Tb885 (with a further extended terminal helix) were incubated with 2 mM Tb^{3+} in 50 mM Tris pH 7.5, 20 mM $MgCl_2$, 50 mM NH_4Cl at 25 ºC for two hours and subsequently analyzed on a denaturing polyacrylamide sequencing gel. Normalized fractions of terbium(III)-mediated scission (expressed as ratio Π, see Materials and Methods), obtained from the quantified gel bands, are shown in Figure 3-4a. A schematic of the Tb885 secondary structure with quantified Tb^{3+} hits (occurring between the indicated nucleotides) is shown in Figure 3-4b. It reveals that, in general, nucleotides G888 through A907 of Tb885 are cut with significantly (up to 2-fold) greater probability than regions G881 through G887 and A908 through C912. These more labile backbone regions are components of the loop E motif, adjacent G-U tandem wobble pairs, and

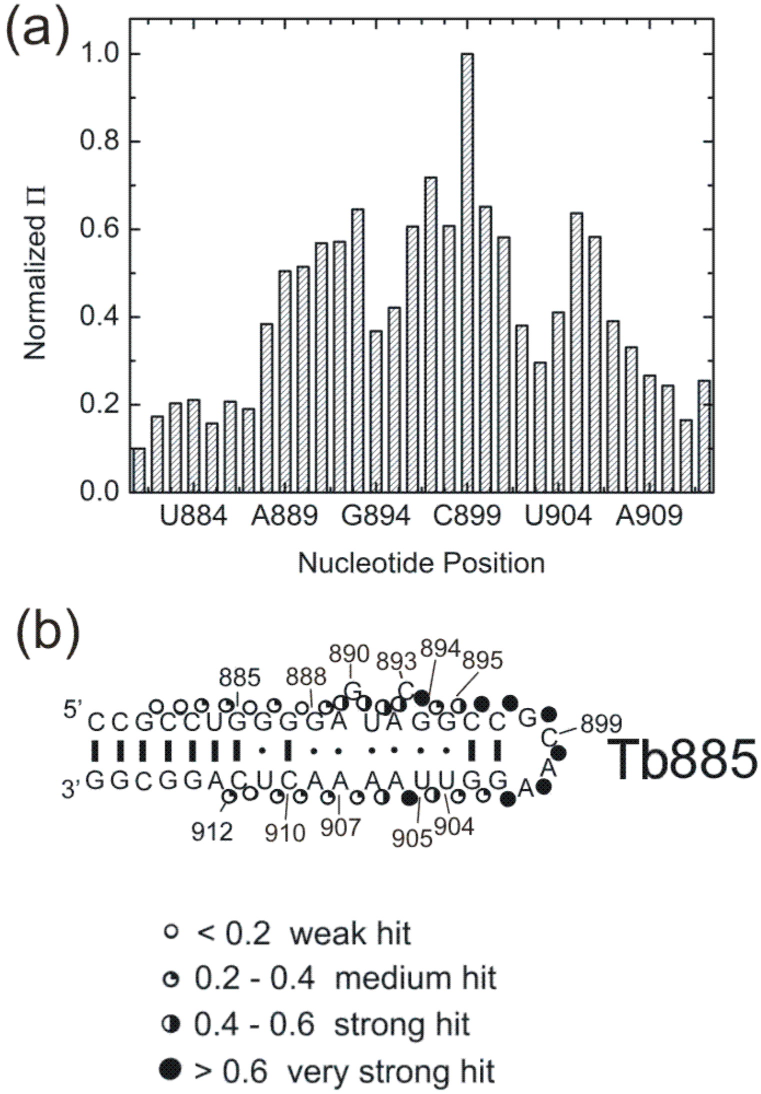


Figure 3-4. Terbium footprinting of the helix 27 Tb885 construct. a) Bar graph representing relative extents of scission 3' of each nucleotide, Π, as defined in Materials and Methods. Data was normalized relative to Π_{C899}. b) Secondary structure for the Tb885 construct used for terbium(III) footprinting studies. The relative strength of the Tb^{3+} hit 3' of each nucleotide is shown mapped onto the secondary structure.

GNRA tetraloop (Fig. 3-4). The loop E motif is characterized by an unusual, S-shaped backbone conformation (*98*). We find that this backbone geometry and perhaps its dynamics make the phosphodiester bonds in this region comparatively vulnerable to scission (Fig. 3-4), similar to observations for the related loop E motif of the hairpin ribozyme (*171*). Other strong Tb^{3+} hits within Tb885 are clustered towards the side of the loop E motif closed by the (weak) G894-U905 pair, consistent with metal ion binding in the vicinity of the G894-U905/G895-U904 tandem wobble pairs, in accord with our NMR data. Particularly intense Tb^{3+} hits are observed in the tetraloop region, a documented lanthanide metal binding site (*160*), in good agreement with our NMR data that suggest metal ion binding in the loop. Table 3-1 compares our terbium(III)-mediated footprinting data for nucleotides where NMR provides independent evidence of nearby metal ion binding.

In summary, our Tb^{3+} data support several conclusions from our NMR studies. In particular, scission patterns of the Tb885 construct are consistent with the proposed secondary structure, with weak cuts in the predicted helical regions and stronger cuts in the non-canonical loop E motif, tandem G894-U905 wobble pair, and tetraloop. These observations support the notion that Mg^{2+} binding sites reside within the non-Watson-Crick base paired regions of H27.

FRET reveals differences in distance upon Mg^{2+} binding

We find NMR to be a reliable reporter of local structure and metal binding in the H27 conformers, but the observed short-range (< 5 Å) NOE interactions are poor indicators of global conformation. To aid in assessment of more generalized Mg^{2+}-induced changes in H27 architecture, time-resolved (tr-)FRET was employed. tr-FRET is an established tool to determine, at angstrom resolution, long-range distances between donor and acceptor fluorophores attached to specific sites in an RNA (*165, 177*). A conformationally locked 885 construct was synthesized and labeled with donor fluorescein attached to nucleotide 899 in the tetraloop and acceptor tetramethylrhodamine conjugated to a linker on nucleotide 911 in the terminal helix, generating tr-FRET construct F885T. In the absence of Mg^{2+} (in 50 mM Tris-HCl, pH 7.5, 50 mM NH_4Cl; we

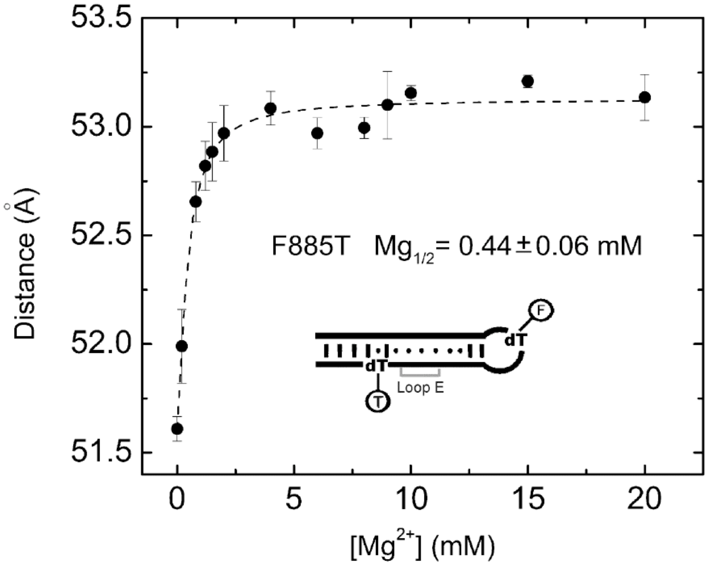


Figure 3-5. Time-resolved FRET monitored magnesium titrations of the helix 27 F885T construct demonstrates global structural changes upon Mg^{2+} binding. The mean distances between donor and acceptor fluorophore are shown as a function of Mg^{2+} concentration, and the curve has been fit to a modified Hill equation, yielding the apparent Mg^{2+} half-titration value, $Mg_{1/2}$, as indicated (Hill coefficient (n) = 1.3 for 885, which is close to unity, indicating a lack of metal binding cooperativity). Fluorophores are attached to the F885T construct at the dT nucleotides, positions 899 and 911, as indicated.

also repeated these experiments, with similar results, in 10 mM NaP_i, pH 7.0, 50 mM NaCl, 0.1 mM EDTA, close to our NMR buffer except for a higher pH to enhance the fluorescein fluorescence, see Materials and Methods), the mean of the tr-FRET derived donor-acceptor fluorophore distance distribution is 51.6 ± 0.2 Å. Upon Mg^{2+} addition, this mean distance increases to a maximum of 53.1 ± 0.2 Å with a Mg^{2+} half-titration point of $Mg_{1/2} = 0.44 \pm 0.06$ mM, as extracted from a modified Hill fit of the distance profile (see Materials and Methods) (Fig. 3-5). The full width at half maximum (FWHM) of the distance distribution also slightly increases (by less than 10%). The overall lengthening observed for F885T at high Mg^{2+} concentrations is consistent with our NMR data, which suggest that the 885 conformation is stabilized by the uptake of Mg^{2+}.

3.4 Discussion

We here have characterized, by several solution probing techniques, Mg^{2+}-RNA interactions within a conformationally locked construct representing the 885 conformation of H27 from *E. coli* 16S rRNA. Our model construct appears to diffusely bind magnesium hexahydrate throughout its full length, as evidenced by the slight shifting of nearly every imino proton resonance upon Mg^{2+} addition, and undergoes small-scale global conformational rearrangements and/or stabilization upon Mg^{2+} titration, as suggested by an increase in tr-FRET detected end-to-end distance. NMR NOESY spectra provide evidence for site-specific association of $Co(NH_3)_6^{3+}$ with the widened major groove of the G886-U911 pair, as is seen for other RNAs containing G-U wobbles (*158, 168, 169*). NMR chemical shift analyses support Mg^{2+} binding adjacent to tandem G894-U905/G895-U904 wobble pairs, while Tb^{3+} footprinting indicates enhanced susceptibility to metal induced cleavage in this region. NMR chemical shift data and Tb^{3+} footprinting also support the notion of a Mg^{2+} binding site within the GCAA tetraloop of the molecule. Neither of the latter two sites detectably accommodates $Co(NH_3)_6^{3+}$.

Helix 27 in the context of the ribosome. In phylogenetic comparisons of 16S rRNAs, H27 shows very little variation among species (*178*). In all ribosomal crystal structures to date the loop E motif of the 885 conformation is observed (*44-46, 156, 157*); the H27 GCAA tetraloop forms tertiary hydrogen bonds with a tetraloop receptor in helix 24 of 16S rRNA (*179*) and is involved in an inter-subunit bridge; the bulged C893 is involved in a cis-Watson-Crick pair with U244; and A909 forms a cis sugar edge/sugar edge interaction with A1413 (*46*). The alternate 888 conformation, first proposed in 1995 based on comparative analysis of 16S rRNA sequences (*53*), was later demonstrated not to be a requisite for protein synthesis in *E. coli* (*54*). However, an isolated H27 adopts a dynamic equilibrium between the 885 and 888 conformations (*55*). It is therefore possible that the structural and electrostatic environment surrounding H27 within the ribosome may slightly alter its structural and metal binding properties from those observed in our solution studies.

Our NMR and terbium footprinting data indicate that the loop E motif of the 885 conformation of H27 is an example of an RNA segment that exhibits small-scale motions. Terbium footprinting reveals that a large portion of the RNA has a relatively susceptible backbone that is prone to Tb^{3+}-induced scission, indicative of sufficient flexibility to accommodate attack of the deprotonated 2’ hydroxyl on the adjacent phosphodiester bond. In addition, low concentrations of Mg^{2+} result in the sharpening of several imino proton resonances, consistent with structural stabilization upon addition of metal. Finally, tr-FRET analyses demonstrate a lengthening of our H27 construct with increasing Mg^{2+}, consistent with metal ion-induced stabilization of the helix.

Some, but not all Mg^{2+} interactions in the 885 conformation are depicted in the relevant 30S ribosomal subunit crystal structures. Crystal structures of the 30S ribosomal subunit that describe coordinates for site-bound Mg^{2+} ions do not depict any direct contacts between helix 27 and the small ribosomal subunit proteins (*44-46, 156, 157*) , making it one of the few small subunit helices that do not directly interact with protein. Its only near-contact is with the relatively unstructured cationic N-terminal tail of S12 that extends into a pocket adjacent to H27, where lysine 21 lies within 3.7 Å of each one non-bridging phosphate oxygen of A908 and A909 (*156*). The associated long-range electrostatic interaction may stabilize the 885 over the 888 conformation in the context of the ribosome, while they are close to isoenergetic in isolation (*55*). Recent studies on the 50S subunit have revealed that metal ions are particularly concentrated in conserved

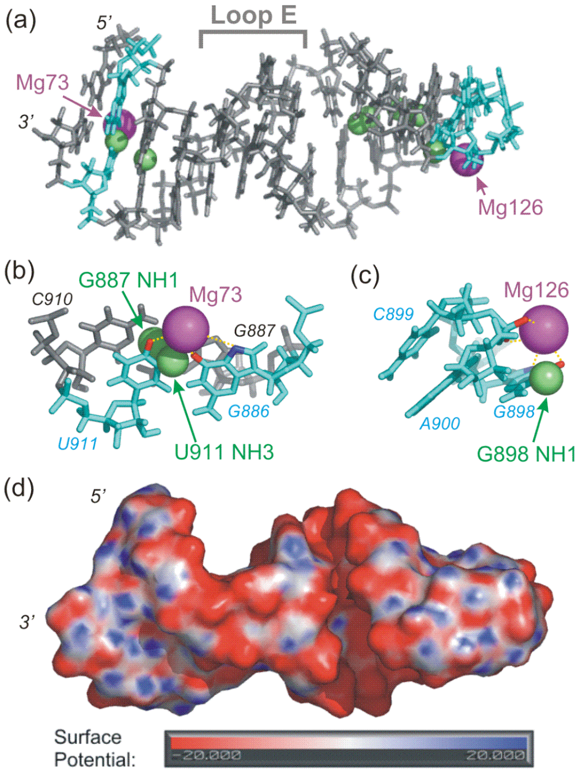


Figure 3-6. Map of protons in the 885 conformation of helix 27 that demonstrate either NOEs with cobalt(III) hexammine and/or significant spectral shifts upon addition of Mg^{2+}. a) Crystal structure of helix 27 (nucleotides G885-C912) from 16S rRNA in the 30S ribosomal subunit of *Thermus thermophilus* (PDB ID 1FJG, (*156*)). There are two site-bound Mg^{2+} ions assigned in the crystal structure, Mg73 and Mg126, shown in magenta. Nucleotides that are shown to bind the metal ions in crystal structures are colored aqua, and protons that NMR data suggest may be interacting with metal ions (U911 NH3, G887 NH1, U904 NH3, G895 NH1, G903 NH1, G902 NH1, and G898 NH1) are depicted as green spheres. b) Close-up view of the Mg73 binding pocket. c) Enlarged view of the tetraloop binding pocket occupied by Mg126. d) Electrostatic surface potential map of H27 from the 30S subunit crystal structure, shown in the same orientation as panel a. The S-turn in the backbone creates an unusually deep and narrow major groove that is dense with negative charge, and is likely to be a strong cation binding sink. The U904 NH3 and G895 NH1 protons within this groove (indicated in (a) by green spheres), shift considerably in resonance upon addition of Mg^{2+}. This figure was generated using pyMOL software (*180*).

regions of an RNA where proteins are absent (*181*), suggesting that H27 may represent a strong metal ion binding region. In addition, thermal denaturation studies of loop E motifs revealed a particularly strong Mg^{2+} ion dependence for the melting profiles of these RNA constructs, implying that loop E motifs bind Mg^{2+} ions specifically (*182*). Relevant crystal structures of the 30S ribosomal subunit indeed depict two Mg^{2+} ions bound to H27 (*44-46, 156, 157*) ; one (Mg73) in the major groove adjacent to the G886-U911 wobble pair, and one (Mg126) nestled in the GCAA tetraloop (Figs. 3-6a-c).

How do solution probing data compare with these crystallographically resolved Mg^{2+} ions? Crystallographic conditions have typically employed 10 to 15 mM Mg^{2+}, introduced as magnesium acetate and/or magnesium chloride salts (*183*). These concentrations are comparable to the concentrations of free Mg^{2+} used in our solutions studies (0-20 mM $MgCl_2$). Table 3-1 and Figure 3-6 summarize nucleotides where we observe strong NOEs upon the addition of cobalt(III) hexammine as a fully hydrated (outer-sphere coordinated) Mg^{2+} mimic and/or significant chemical shifts in NMR spectra upon addition of Mg^{2+}. We find three different scenarios when comparing solution with crystallographic data. First, the imino protons G887 NH1 and U911 NH3 show cobalt(III) hexammine NOEs as well as Mg^{2+} induced downfield chemical shifts (of 0.12 ppm and 0.05 ppm, respectively) in solution (Table 3-1). Both protons are proximal to Mg73 of the ribosomal crystal structures (Figs. 3-6 a and b), which has its closest contacts with G886 O6, G886 N7, and U911 O4 at 2.82 Å, 4.09 Å, and 3.14 Å, respectively (Fig. 3-6b). A subset of 30S subunit crystal structures, such as PDB ID 1IBM (*16*), assign an additional Mg^{2+} ion, Mg406, coordinated to G887 N7 at a distance of 3.26 Å. Both the NMR and crystallographic data are therefore consistent with outer-sphere coordination of a Mg^{2+} ion in the major grove of this region, next to the G886-U911 wobble pair.

Second, G898 NH1 in the GCAA tetraloop does not detectably coordinate with cobalt(III) hexammine in solution, but shows one of the most pronounced downfield chemical shifts (0.13 ppm) upon titration with Mg^{2+}. In the crystal structures, Mg126 is coordinated to G898 O6 and G898 N7 at distances of 3.28 Å and 3.77 Å, respectively (Fig. 3-6c). Again, the NMR chemical shift and crystallography data are both consistent with outer-sphere coordination of a Mg^{2+} ion in the major grove of the GNRA tetraloop, in agreement with previous observations in other RNAs (*154, 158-160*). The fact that no

NOE to the cobalt(III) hexammine ligands is observed suggests that the high-charge-density, exchange-inert cobalt(III) complex may bind at a (slightly) different location or with lower affinity than Mg^{2+}.

In the third case, strong evidence for metal ion binding within the major groove of the helix between the loop E motif and the GCAA tetraloop was observed in our solution studies, while such a metal ion binding site was not detected in ribosomal 30S subunit crystal structures that describe magnesium ion binding (*16, 44-46, 91, 156, 157*). More specifically, Tb^{3+} footprinting studies revealed that nucleotides G888 through A907 of Tb885 are cut with significantly (up to 2-fold) greater probability than regions G881 through G887 and A908 through C912, consistent with metal ion association with the GCAA tetraloop, as well as the loop E motif and tandem G-U pairs (Figs. 3-4 a and b). NMR data are also suggestive of metal ion binding in the vicinity of the tandem G-U pairs, as the G895 NH1, G902 NH1, G903 NH1, and U904 NH3 imino proton resonances all display significant titratable shifts upon addition of Mg^{2+} (Fig. 3-3, Table 3-1). In particular, the U904 NH3 resonance shifts downfield more than any other resonance in the molecule (0.16 ppm), while the other three resonances shift upfield. While there are no metal ions resolved crystallographically around G895 NH1, G902 NH1, G903 NH1, and U904 NH3, these protons are positioned deeply within the highly negatively charged major (deep) groove adjacent to the loop E motif, which is expected to be a particularly strong metal ion binding site (Figs. 3-6 a and d) (*138, 141, 184*). We therefore propose that Mg^{2+} either binds near the tandem G894-U905/G895-U904 wobble pairs in solution and not in the crystal or, alternatively, that binding is transient under both sets of conditions (possibly explaining why no NOE with $Co(NH_3)_6^{3+}$ hexammine protons was observed) so that the metal ion occupancy at this site is too low to produce an assignable signal in crystallographic electron density maps. Additionally, if we consider the slightly higher monovalent cation concentration used in crystallization (50 mM – 200 mM) than utilized in our solution studies (50 mM – 100 mM), we cannot rule out the possibility that this particular site is occupied by (undetected) monovalents in the 30S crystal structures.

In summary, we have gathered experimental evidence for metal-ion interactions within the distorted major groove near tandem G-U wobble pairs of an isolated helix 27 from *E. coli* 16S rRNA, which are not depicted in any 30S ribosomal subunit crystal structures. In addition, we find that even in a conformationally stabilized 885 structure small-scale dynamics occur, and that millimolar Mg^{2+} concentrations attenuate these motions. We propose that the interplay of local H27 base and backbone dynamics, as detected here by NMR and terbium footprinting studies, respectively, may contribute to differences in detectable metal ion interactions between dynamic solution structures and relatively static crystal structures. Our studies thus highlight the necessity to continue to complement crystallographic with solution phase detection of metal ion binding in RNA.

CHAPTER 4:

siRNA-LIKE 21 AND 24-NUCLEOTIDE DOUBLE-STRANDED RNAs ARE SPECIFICALLY PROTECTED IN CELL EXTRACT[3]

4.1 Introduction

RNA interference (RNAi), first described in the nematode *C. elegans* (*57*), is a conserved post-transcriptional gene-silencing pathway present in higher eukaryotes. Following the first demonstration that RNAi is functional in human cells and receptive to synthetic small interfering (si)RNA effector molecules (*59*), significant progress has been made in harnessing the RNAi pathway for functional genomics studies and therapies targeting previously "undruggable" genetic targets. Inside the cell, siRNAs specifically interact with a number of proteins including Dicer, R2D2, and Argonaute, and are ultimately incorporated into a multiprotein complex referred to as the RNA-induced silencing complex (RISC). A single strand of the siRNA is retained in RISC and acts as a template for the sequence-specific identification and site-specific cleavage of messenger (m)RNAs and concomitant reduction in gene expression. The inherent lability of RNA molecules due to the presence of ribonuclease enzymes both in and outside the cell forced an immediate focus on development of chemically modified siRNAs that can withstand nucleolytic extracellular environments and reach the cellular milieu intact. It has been

---

[3] John Hoerter conducted all assays with the exception of several serum and HeLa cell extract FRET assays carried out by Troy Lionberger.

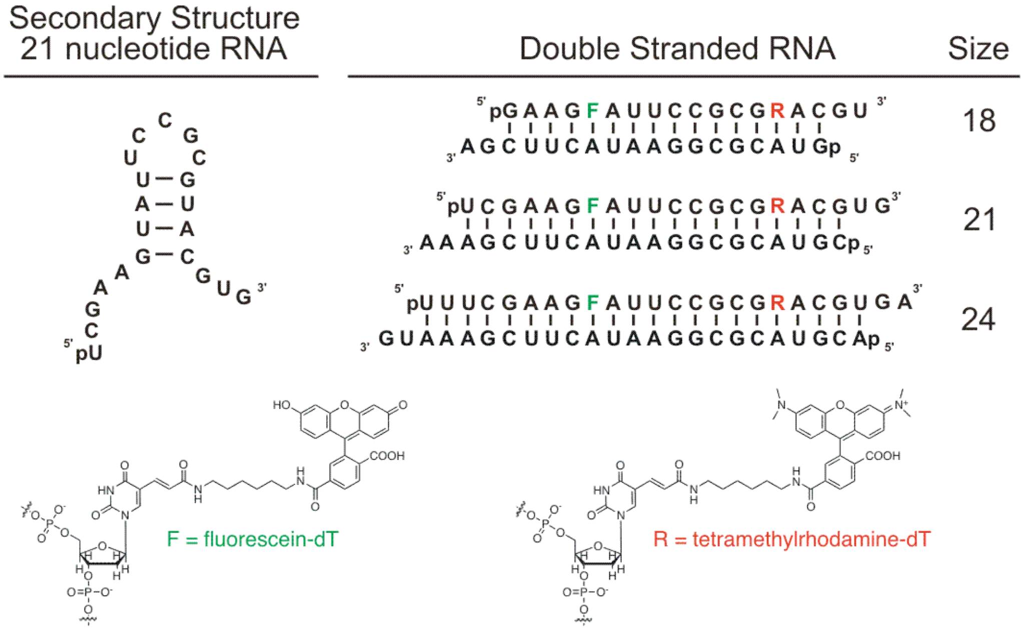


Figure 4-1. Design, secondary structure, sequence, and modifications of the RNAs used in this study. The sequence for the RNAs are derived from the firefly luciferase gene and the siRNA previously reported(*185*). The double stranded RNAs consist of an identical core sequence and internal fluorescein and tetramethylrhodamine modifications, and the flanking sequence is derived from the luciferase gene. All double stranded RNAs contain the two nucleotide 3’ overhangs characteristic of siRNAs.

established that chemical modifications are absolutely necessary for systemic delivery of siRNA, both to mediate binding to serum proteins to increase circulating half-life as well as to block sites vulnerable to nuclease cleavage (*82, 83*).

In order to understand the determinants of siRNA stability, research has focused in on the following areas: extracellular stability; intracellular stability; and potency of the silencing effect. To mimic the extracellular environment, several groups have used blood serum and purified nucleases to determine the reactivity of siRNAs and probed the stabilizing effect of chemical modifications. Research has now shown that chemical modifications can impose a considerable degree of nuclease resistance on siRNAs while retaining the ability to induce silencing through the RNAi pathway (*69, 71, 72, 74, 77, 82, 83, 186, 187*).

By monitoring the expression of a target gene, several groups have been able to assess the potency and duration of siRNA effects, including a number of studies that have focused on determining if chemically modified siRNAs are more potent than unmodified siRNAs (*69, 71, 72, 74, 77, 186*). After administration of an siRNA drug, there are at least two environments to which the siRNA is exposed, extracellular and intracellular. Although extracellular fluids vary in nuclease content, it is clear that siRNAs introduced into the blood and most other regions of the body will require chemical modifications. One exception to this trend of RNA-hostile extracellular environments is found in the lungs where the environment requires minimal or no chemical modification to the siRNA (*81*). Another exception is siRNAs encapsulated in liposomes which have a protective effect on the siRNA (*81*). The intracellular milieu, the functional environment of the siRNA, is the second environment to which the siRNA is exposed. Potential benefits to siRNA stability and potency by chemical modifications in this environment are less clear. Position-specific 2’-O-Me modification has been shown to reduce off target effects, ameliorating a negative consequence of siRNA adminstration (*66*), but this is not a nuclease sensitivity issue. A recent study has shown that enhanced intracellular nucleolytic stability is not necessarily correlated with increased duration of the silencing effect (*69*). In fact, silencing in non-dividing cells can persist for up to one month from a single dose of unmodified siRNA. This is an exciting result and suggests that siRNAs are quite stable inside the cell. However, broad generalizations based on this result are unfounded as the modified siRNA used in this study contained proprietary modifications (*69*).

By introducing fluorophore and/or FRET modified siRNAs into the cell, several groups have directly addressed the questions of siRNA stability and localization using various microscopy techniques. It has been reported that siRNAs are actively exported from the nucleus (*188*), except in cases where the RNA target is actually located in the nucleus (*80*). Fluorescence fluctuation spectroscopy has been utilized in a separate study to assess the integrity of labeled intracellular RNA, revealing that FRET labeled single stranded RNA is unstable compared to the same labeled RNA strand incorporated into an siRNA duplex (*189*). In addition, FRET labeled RNAs have been used to address RNA reactivity as a function of RNA secondary structure, degradation by cell extract, or purified nuclease (*190*).

The precise origin of the intracellular protection of siRNAs is unclear. Long duration of silencing in non-dividing cells may be indicative of stable single-stranded guide strand incorporated into the RISC complex. Prior to RISC maturation, it is known that some stability is simply derived from the decreased nuclease sensitivity of double-stranded RNA compared to single stranded RNA (*189, 190*). However, it is unknown if the siRNA is protected by some mechanism upon entry into the cell before advancement into holo-RISC. The observation that the RNAi pathway utilizes double-stranded RNAs larger than ~21 nucleotides (*191*) suggests a basis for intracellular discrimination between double-stranded RNAs based on size.

Characterization of the real-time kinetics of degradation under these conditions is necessary, and should include rapid and precise assessment of RNA stability as a function of bodily fluid, base paring partner, and RNA size. In order to address these issues, we developed a series of RNAs modified for fluorescence resonance energy transfer (FRET) measurements, where two uridines in positions 7 and 16 of the 21-nucleotide guide strand were replaced with amino-modified deoxythymidines to introduce two fluorophores, see Figure 4-1. We varied the size of the labeled strand to generate double-stranded RNAs ranging in size from 18-24 nucleotides. We found that our double-labeled RNAs interact with relevant cellular enzymes, including RNase H1 (21-nucleotide DNA/RNA hybrid) and the Dicer enzyme (24-nucleotide dsRNA). The DNA/RNA hybrid degradation by

RNase H1 wsa quite efficient, with observed rate constants exceeding even those for the single stranded RNA. Furthermore, our experimental design allowed us to correlate the previously studied potency/size relationships for siRNAs (*185, 191*) with a real-time readout of RNA stability. Our results demonstrated that 21 and 24-nucleotide dsRNAs are specifically protected in cell extract when compared to the 18-nucleotide dsRNA and when compared to degradation profiles in blood serum.

4.2 Materials and Methods

Nucleic Acid Synthesis and Labeling. All RNAs were synthesized by the HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory at the Yale University School of Medicine, deprotected, tetramethylrhodamine labeled (as necessary), and purified as previously reported (*55*). Sequences are derived from the firefly luciferase gene (*185*) and are shown in Figure 4-1. All RNAs were purchased with a 5' PO_4 group, except that a second fluorophore-labeled 21-nucleotide RNA was purchased with a 5'OH. The DNA oligos used in these studies (Sequence = V-2, VI-2, and S1-DNA) were synthesized by Invitrogen, were used as received, and had the following sequence: S1-DNA, CGTACGCGGAATACTTTGAAA; VI-2, CGGTCGCTCCGTGTGGCTTGGGTTGGGTGTGGCAGTGAC; V-2, GCTGGTCTCTGCGGGTTGTTGCGCCGCGGCACCCTTGGCA.

Fluorometer Assays of RNA degradation. FRET labeled RNAs were annealed either alone (single stranded RNA) or with an appropriate DNA or RNA complement by heating to 70º C for 2 minutes and cooling to room temperature over 10 minutes at a final concentration of 50 nM labeled strand and 100 nM complement in a close-to physiologic standard buffer of 50 mM Tris-HOAc, pH 7.4, 80 mM KCl, 20 mM NaCl, and 1 mM $MgCl_2$. After the 10 minute cooling period at room temperature, RNA solutions were incubated in a circulating water bath at 37º C for 5 minutes before the samples were added to the fluorometer cuvette, also controlled at 37º C. After the initiation of data collection which was conducted as described previously (*55*), a baseline measurement was recorded for approximately 50-100 seconds before a fraction of either HeLa S100 cytoplasmic cell extract (a gift from Danny Reinberg, Department of Biochemistry, Rutgers University, prepared following published protocols (*192*)), fetal bovine serum (Gibco), or purified Dicer (Stratagene, 2 Units) was added and the resulting FRET profiles recorded. The FRET ratio Q ($F_{585}/(F_{585}+F_{520})$) was plotted and fit where appropriate to either the single or double exponential form of the expression below to extract the observed rate constants.

$$y = y_0 + A_1 e^{-k_{obs,1}X} + A_2 e^{-k_{obs,2}X} \quad \textbf{Equation 4-1}$$

These assays were carried out over a range of HeLa and serum concentrations, ranging from 0-2.5 mg/ml HeLa total protein (concentration measured by Bradford assay) and 0-11.1 mg/ml total serum protein (concentration provided by Gibco). The observed rate constants were plotted as a function of total protein content and fit to a modified form of the hyperbolic binding equation below where no cooperativity was assumed.

$$y = \frac{V_{\max}[protein]}{protein_{1/2} + [protein]} \quad \textbf{Equation 4-2}$$

RNase H1 Inhibition Assays. Fluorometer assays for RNase H1 inhibition by DNA aptamers VI-2 and V-2 (*193*) were conducted identically to the fluorometer assay described above with the following exceptions. The HeLa cell extract used in these assays was pre-incubated with a defined concentration of the pre-annealed DNA aptamer or a negative control DNA for 30 minutes at 37º C before addition to the DNA/RNA hybrid in the fluorometer. In addition, the DNA/RNA hybrid was prepared with the aptamer at a concentration equal to that in the HeLa cell extract. Again, the FRET Ratio Q for each trace was extracted and fit to a single exponential expression (above) and the observed rate constants were plotted as a function of aptamer concentration and fit to the hyperbolic binding equation below where no cooperativity was assumed.

$$y = V_0 - \frac{(V_0 - V_{\min})[aptamer]}{aptamer_{1/2} + [aptamer]} \qquad \textbf{Equation 4-3}$$

FRET Gel Electrophoresis. Denaturing gel electrophoresis was paired with fluorescence detection of the RNA to gain qualitative insight into the interaction of the labeled RNAs with Dicer, cell extract, and serum. The labeled RNA strand was prepared at 500 nM final concentration, and as necessary, the complementary RNA or DNA was added to 1 μM concentration in the standard buffer listed above and annealed as described above. After cooling, Dicer, cell extract, or serum was added and the samples were incubated at 37° C until the reactions were stopped by mixing with a final concentration of 10% (v/v) Contrad 70 at pH 9.3 (Decon Labs) (*190*). For the FRET gel in Figure 4-2A, serum was added to a final concentration of 11.1 mg/ml for the DNA/RNA hybrid and the double-stranded RNA, and 0.93 mg/ml for the single stranded RNA. Time points for these reactions were as follows: single stranded RNA, 1, 5, and 10 minutes; DNA/RNA hybrid and double-stranded RNA, 0.5, 1 and 2.5 hours. For the FRET gel in Figure 4-2C, HeLa cell extract was added to a final concentration of 2.49 mg/ml and time points were taken at 5, 10, and 20 minutes for the single stranded RNA and the DNA/RNA hybrid, while time points of 5, 20, and 60 minutes were taken for the double-stranded RNA. The FRET gel in Figure 4-4B was prepared by exposing each of the 18, 21, and 24-nucleotide double-stranded RNAs to Dicer enzyme (1 unit), HeLa cell extract (2.49 mg/ml), or buffer alone for 3.3 hours. The buffer for the HeLa and buffer only reactions were composed of the standard buffer listed above plus 10 mM DTT, and the Dicer reactions contained the buffer supplied and recommended by the manufacturer whose final composition was 20 mM Tris-HCl pH 8.0, 150 mM NaCl, and 2.5 mM $MgCl_2$. T1 and OH- reference ladders were prepared according to (*187*) except that no tRNA was added to the T1 digest. For all reactions, 10 pmol/lane of labeled RNA was mixed with loading buffer, added to a final concentration of 1×TBE, 0.025% bromophenol blue, and 40% formamide, and the products separated on a denaturing, 8 M urea, 20% (w/v) polyacrylamide gel. Fluorescence detection of the RNA was accomplished as previously reported (*23, 194*).

Radiolabeling and Cleavage Assays. The 21-nucleotide fluorophore-labeled RNA was 5'-^{32}P labeled with T4 polynucleotide kinase (PNK) and γ-^{32}P-ATP, mixed with loading buffer to a final concentration of 1×TBE, 0.025% bromophenol blue, and 40% formamide, and gel purified on a denaturing, 8 M urea, 20% (w/v) polyacrylamide gel. The labeled RNA was excised from the gel, diffusion eluted into 1 mM EDTA overnight, and ethanol precipitated. The dried RNA was dissolved in water.

5'-^{32}P labeled RNA was mixed with 50 nM non-5'^{32}P labeled fluorophore containing strand and 100 nM of the complementary 21-nucleotide RNA, annealed as described above in a close-to physiologic standard buffer of 50 mM Tris-HOAc, pH 7.4, 80 mM KCl, 20 mM NaCl, and 1 mM $MgCl_2$. Degradation was initiated by the addition of 11.1 mg/ml serum or 2.49 mg/ml cell extract at 37°C. Time points were taken after 5, 10, 20, 40 , 80, 120, and 180 min for the serum assay and 2, 5, 10, 20, 40, and 80 minutes for cell extract. All time points were quenched by mixing with a final concentration of 10% (v/v) Contrad 70 at pH 9.3 (Decon Labs) (*190*). Loading buffer was added to a final concentration of 1×TBE, 0.025% bromophenol blue, 0.025% xylene cyanol, and 40% formamide, and the degradation products (50,000 cpm per lane) separated on a denaturing, 8 M urea, 20% (w/v) polyacrylamide gel. Individual bands were identified by comparison with sequencing ladders from partial digestion with G-specific RNase T1 and alkali as described (*164*), quantified and normalized to the sum of all bands in a lane using a PhosphorImager Storm 840 with Image Quant software (Molecular Dynamics).

4.3 Results

We developed a series of FRET modified fluorescein (FL) and tetramethylrhodamine (TMR) modified RNAs (Figure 4-1) for use in studies of intracellular and extracellular RNA stability. We first sought to establish the relative reactivities of single stranded RNA compared to constructs where the relatively unstructured single stranded RNA (Figure 4-1) is protected by base pairing to fully complementary strands composed of either RNA or DNA. In addition, we constructed a series of double-stranded RNAs, ranging in size from 18 to 24 nucleotides (Figure 4-1),

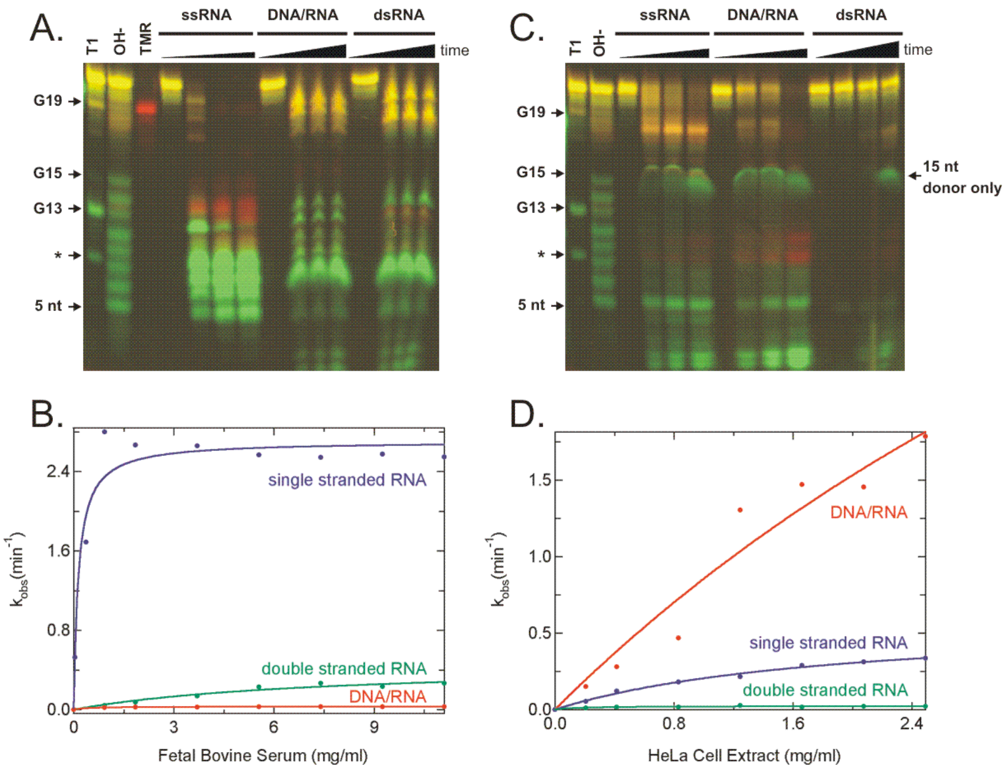


Figure 4-2. Degradation profiles of the labeled 21 nucleotide RNA when studied as a ssRNA, dsRNA, and a DNA/RNA hybrid in both HeLa cell extract and blood serum. A) FRET gel revealing the serum degradation pattern of the RNA species listed above as a function of time (details listed in the Materials and Methods section). This gel describes the rapid degradation of single stranded RNA in serum, the protective effect of secondary structure, and support the relative trends described in B). B) Graph of observed degradation rate constant as a function of serum protein concentration. These data clearly show the rapid degradation of the single stranded RNA and the relative inhibition of degradation for the dsRNA and the DNA/RNA hybrid. C) FRET gel revealing the HeLa cell extract degradation pattern of the RNA species listed above as a function of time (details listed in the Materials and Methods section). This gel qualitatively supports the data in D) where the DNA/RNA hybrid is found to be more efficiently degraded than the single stranded RNA of the dsRNA. Solid lines denote fits to a modified form of the Hill equation where no cooperativity is assumed. In both A) and C), T1 and OH- denote marker ladders, and the arrows to the left of the gel denote the nucleotide and position of the band in the T1 ladder. The * denotes a double hit at both G13 and G3.

derived from the 21-nucleotide luciferase siRNA validated in earlier studies of RNAi (*185, 187*), to determine the relative stabilities of these double-stranded RNAs. All constructs were examined in both S100 HeLa cytoplasmic cell extract (HeLa) and fetal bovine serum.

Formation of secondary structure stabilizes RNA

To establish the relative reactivities of a 21-nucleotide dsRNA compared to a DNA/RNA hybrid and single stranded RNA, we applied our FRET assay to monitor the degradation of these complexes upon the addition of either HeLa S100 cell extract or fetal bovine serum. Cell extract contains the protein components of the RNAi pathway, whereas the bovine blood serum is primarily nucleolytic, containing a significant fraction of RNase A.

Applying the fluorometer based real-time FRET degradation assay to these constructs over a range of concentrations of serum and cell extract, we accumulate the curves of observed rate versus protein concentration of either serum or HeLa (Figure 4-2 B or D, respectively). The fastest observed rates are found with the decay of single stranded RNA in serum (V_{max}= 2.7 min^{-1}, $FBS_{1/2}$= 0.14 mg/ml). The traces of the dsRNA degradation in serum exhibit a profound increase in the FRET ratio followed by a decrease. The rising phase of these data were fit to a double exponential function to extract the rate constants. The rate constant describing the fast phase was found to be reflective of RNA degradation, as determined in the subsequent section (Figure 4-5a). Observed serum rates decrease significantly with the formation of secondary structure, where the dsRNA (V_{max}= 0.47 min^{-1}, $FBS_{1/2}$= 7.7 mg/ml) and the DNA/RNA hybrid (V_{max}= 0.03 min^{-1}, $FBS_{1/2}$= 0.46 mg/ml) are relatively stable.

The observed kinetics in S100 cytosolic HeLa extract in Figure 4-2D are significantly different from those in serum. The DNA/RNA hybrid is most efficiently degraded (V_{max}= 7.4 min^{-1}, $HeLa_{1/2}$= 7.6 mg/ml) and the single stranded RNA is still quite efficiently degraded (V_{max}= 0.60 min^{-1}, $HeLa_{1/2}$= 1.9 mg/ml). The traces of the dsRNA degradation exhibit a slight increase in the FRET ratio followed by a decrease. These data were fit to a double exponential function to extract the rate constants. The rate constant describing the decreasing phase was found to be reflective of RNA degradation, as determined in the subsequent section (Figure 4-5b). After this analysis, we found that the dsRNA was protected by secondary structure making it the most stable of the constructs in HeLa (V_{max}= 0.02 min^{-1}, $HeLa_{1/2}$= 0.18 mg/ml).

We showed by FRET gel analysis that serum (Figure 4-2A) manifested more efficient cleavage of the ssRNA compared to either the dsRNA or the DNA/RNA hybrid. While the cleavage patterns of the DNA/RNA hybrid and the dsRNA are quite similar, the cleavage pattern of the ssRNA is notably different. These differences include accumulation of a slight TMR only band (note the reference TMR only control lane), which indicates selective removal of the donor FL fluorophore. There is significant accumulation of intermediate size (5-13 nucleotides) products that are either singly TMR or FL labeled. The absence of donor only products smaller than 5 nucleotides in the ssRNA lanes indicate minimal 5' exonuclease activity. Some accumulation of these short donor-only products is also seen in degradation of the dsRNA and DNA/RNA hybrid, suggesting limited 5' exonuclease activity.

By contrast, the degradation patterns in HeLa cell extract are notably different (Figure 4-2C). For example, the DNA/RNA hybrid is more completely degraded in HeLa cell extract, whereas the single stranded RNA was more completely degraded in serum. Another difference between serum and HeLa is the significant accumulation of short (<5 nucleotide) products in the HeLa gels, indicating that there is terminal 5' and 3' exonucleolytic degradation at work in these samples. The gels also show that single stranded RNA in HeLa cell extract generates long (15-20 nucleotide) exonucleolytic products that are not present in either of the base-paired constructs. There is also a donor-only labeled band present in all HeLa samples that co-migrates with an ~13–nucleotide RNA fragment, which may be evidence of an artificial interaction of the fluorophores with a component of the cell extract.

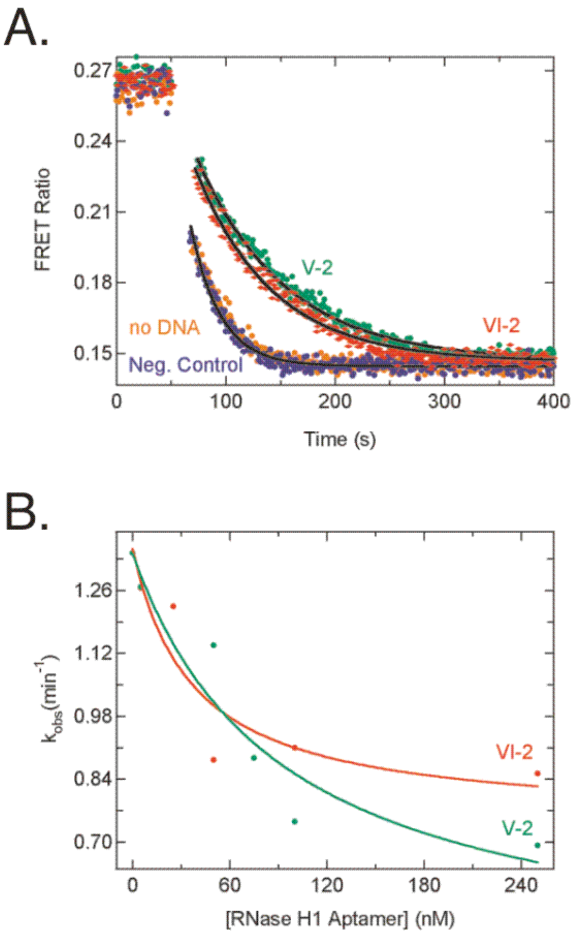


Figure 4-3. Inhibition of RNase H1 degradation of a fluorophore labeled DNA/RNA hybrid by DNA aptamers. A) Specific inhibition of RNase H1, as detected in FRET degradation assays, by the DNA aptamers VI-2 (red) and V-2 (green) at 500 nM concentration compared to a degradation assay in the absence of inhibitor (orange) or an assay with a negative control DNA primer (blue). Solid black lines represent a single exponential fit to the data. B) Dose dependent suppression of the observed rate constant of DNA/RNA hybrid degradation by RNase H1 as a function of VI-2 or V-2 aptamer concentration. The solid lines are fits to a modified form of the Hill equation where no cooperativity is assumed, and reveal nanomolar binding constants for the DNA aptamers.

DNA/RNA hybrid is cleaved by RNase H1 in HeLa cell extract

We hypothesized that the efficient degradation of the DNA/RNA hybrid in HeLa cell extract resulted from the action of the human RNase H enzymes. RNase H enzymes degrade the RNA in a DNA/RNA hybrid and are important in nucleic acid processing in the cell, including in DNA replication. Two DNA aptamers have been *in vitro* selected against human RNase H1 (*193*), and were used in the current study to probe possible RNase H1 degradation of the fluorophore-labeled DNA/RNA hybrid. Figure 4-3A shows the time evolution of the FRET ratio of several DNA/RNA hybrid degradation assays. In the absence of any added DNA or a negative control DNA, the degradation is the same. When the RNase H1 aptamers VI-2 and V-2 were added, we observed a reduction in the observed rate of decay of the FRET ratio, indicating specific inhibition of RNase H1 in the degradation of the fluorophore-labeled DNA/RNA hybrid. These experiments were repeated over a range of aptamer concentrations, and a dose dependent decrease in observed rate was observed for both aptamers (Figure 4-3B: VI-2, V_0= 1.35 min^{-1}, V_{min}= 0.75 min^{-1}, VI-$2_{1/2}$= 40 nM; V-2, V_0= 1.34 min^{-1}, V_{min}= 0.4 min^{-1}, VI-$2_{1/2}$= 90 nM).

Dicer cleaves FRET labeled 24-nucleotide dsRNA

To address specific questions of how double-stranded RNAs may be discriminated between in the cell, we developed three dsRNAs (Figure 4-1) spanning the useful region of siRNA lengths. These dsRNAs are derived from the luciferase siRNA developed, validated, and used in the literature (*185*). Interaction with the Dicer enzyme is one of the earliest steps in the RNAi pathway, where larger dsRNAs are cleaved into 21-nucleotide fragments that then go on to be incorporated into the holo-RISC complex (*62, 195*). Clearly Dicer is intimately involved and interacts with siRNAs from the initial cleavage, to guide strand selection, on to RISC assembly.

The 24-nucleotide dsRNA is expected to be a substrate for Dicer, and we tested this hypothesis by both FRET assays and gel based analysis. Interaction of FRET labeled dsRNA with Dicer is important to establish in order to validate our use of these substrates as probes of siRNA stability in cell extract, as the Dicer enzyme is one of the initial

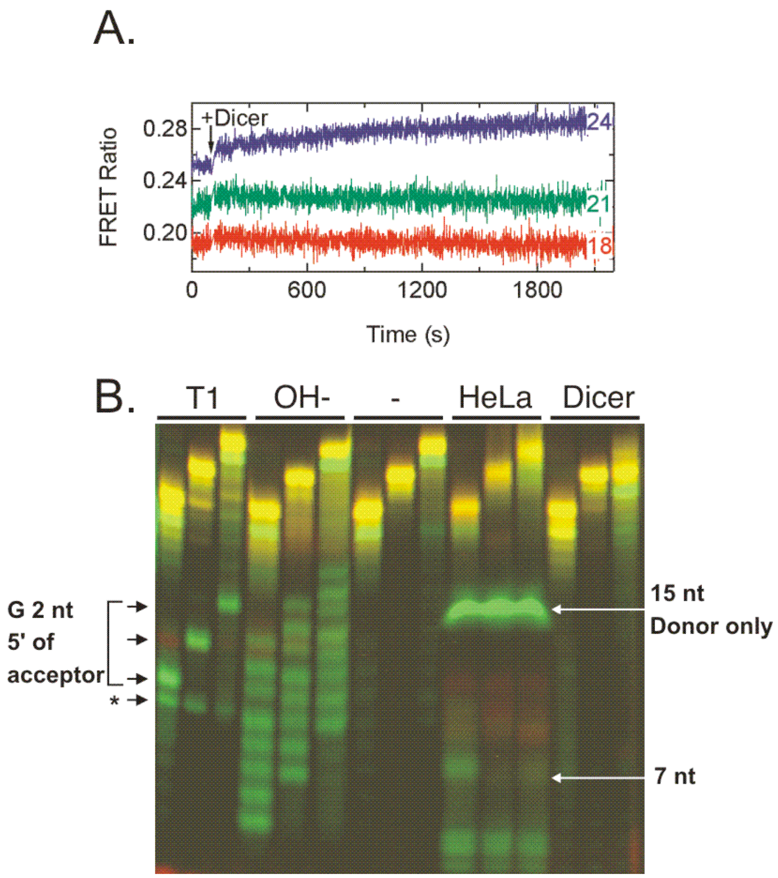


Figure 4-4. Dicer cleavage of the fluorophore labeled 24 nucleotide dsRNA. A) FRET traces of 18, 21, and 24 nucleotide dsRNAs (red, green, and blue respectively) upon Dicer addition, where the there is a small increase in the FRET ratio of the 24 nucleotide dsRNA indicative of RNA cleavage by Dicer. B) A FRET gel comparing the interaction of 18, 21, and 24 nucleotide dsRNAs with Dicer and HeLa cell extract. T1 and OH- denote marker ladders, "-" indicates incubation in buffer only, while the last two panels are incubation in HeLa cell extract and purified Dicer enzyme, respectively. The black arrows to the left of the gel mark the bands in the T1 ladder. The white arrows on the right of the gel mark 15 and 7 nucleotide bands in the HeLa reactions. The 24 nucleotide dsRNA is clearly cleaved to a 21 nucleotide RNA upon incubation with Dicer. In HeLa cell extract, the dsRNAs are all relatively stable, with the exception of the 18 nucleotide dsRNA that shows enhanced accumulation of short (<9 nucleotide) donor only products.

components of the RNAi pathway. Using purified Dicer enzyme, we conducted fluorometer assays with all the dsRNAs and the results are shown in Figure 4-4A. Only with the 24-nucleotide dsRNA did we see evidence for an interaction. This result is not unexpected as the 24-nucleotide dsRNA is the only dsRNA in this study which is a substrate for Dicer cleavage. These results do not rule out interaction of Dicer with the shorter dsRNAs, as the fluorometer assay may not be sensitive to binding of Dicer alone.

To further test the nature of this interaction, we incubated purified Dicer enzyme and HeLa cell extract with the three dsRNAs and analyzed the results using a FRET gel (Figure 4-4A). As expected, the 24-nucleotide dsRNA is cleaved by Dicer to 21 nucleotides. In HeLa cell extract, the 18-nucleotide dsRNA exhibits a slightly stronger cleavage band at approximately 7 nucleotides than either the 21 or 24-nucleotide dsRNAs, but with that exception, the lanes are quite similar. We did not observe a discrete 21-nucleotide Dicer-cleaved band in the 24-nucleotide dsRNA lane, but this does not mean that cleavage is absent, as downstream processing maybe rapid and prevent observation of this intermediate. This downstream processing seems to be common to all the dsRNAs in HeLa cell extract as we see a common strong donor only band that runs at approximately 15 nucleotides.

21 and 24-nucleotide dsRNAs are protected in cell extract

When we applied the real-time fluorometer assay to explore the degradation kinetics of the 18, 21, and 24-nucleotie FRET labeled RNAs, we again utilized both serum and HeLa cell extract. In general, a decrease in FRET ratio should be indicative of degradation of the labeled strand, as cleavage between the fluorophore labels leads to a decrease in FRET. Unexpectedly, when the dsRNAs are incubated in serum, we observed a large phase representing an increase in FRET ratio. Assays using purified RNase A (the major nuclease enzyme present in serum) also qualitatively reproduced an increase in FRET ratio similar to the serum trace. In cell extract, the dominant phase decreased, although there was a small fast rising phase in the degradation traces for the 18 and 21-nucleotide dsRNA. To determine precisely which phase and corresponding rate constants from the

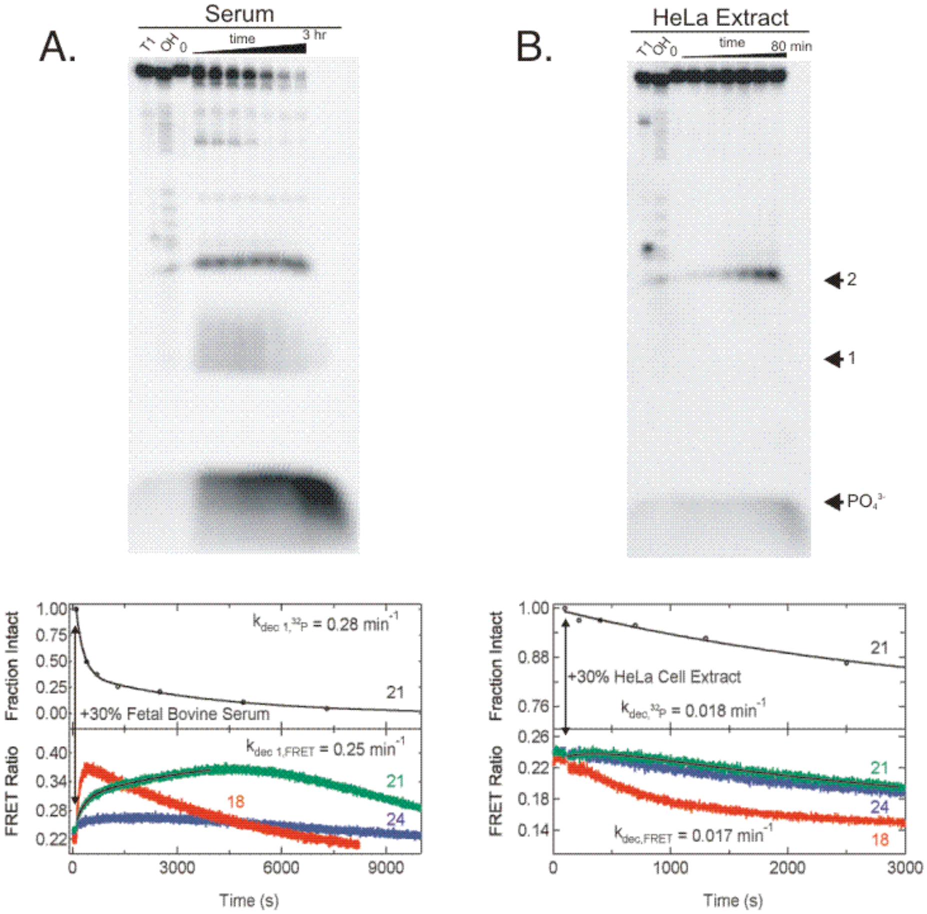


Figure 4-5. Determination of the rate constants representative of RNA degradation in the serum and HeLa FRET traces. The multiplicity of rate constants in the FRET traces required a control degradation assay of the 21 nucleotide dsRNA using ^{32}P to determine which are the relevant rate constants, where T1 and OH ladders were run in denaturing gels alongside time points of RNA cleavage in serum (A) and HeLa cell extract (B). Arrows to the right of the gels denote bands corresponding to position 1, 2, and PO_4^{3-}. A) The serum traces reveal that the rate constants associated with the rising phase of the FRET trace characterize the degradation of the RNA. The solid lines represents double exponential fits to the degradation data. B) Determination of the degradation rate constants in HeLa cell extract using a ^{32}P degradation assay of the 21 nucleotide dsRNA. The decreasing phase in the FRET traces corresponds to the degradation rate in HeLa cell extract. For the ^{32}P data, the solid black line represents a single exponential fit to the data, where as the solid black line in the FRET data represents a double exponential fit.

FRET traces correspond to RNA degradation, we used radiolabeling and denaturing gel analysis to determine the degradation rate-constants from the real-time output of our FRET traces (Figure 4-5). We 5'-^{32}P labeled the 21-nucleotide fluorophore-modified antisense strand and followed the degradation as a function of time using denaturing gel analysis (*187*). Gels were quantified for the fraction of the antisense strand that was intact, therefore quantifying only primary hits. These data were plotted and fit to exponential decrease functions (double exponential for serum, Figure 4-5A, and a single exponential for HeLa cell extract, Figure 4-5B.). These data unambiguously identified the relevant degradation rate constants from the FRET traces as the rising phase in serum and the decreasing phase in HeLa cell extract. In serum degradation of the 5-^{32}P-labeled 21-nucleotide dsRNA followed a double exponential function, where the increasing portion of the 21-nucleotide dsRNA FRET trace was also a double exponential (Figure 4-5B). The rising phases of the 18 and 24-nucleotide dsRNAs were characterized by single exponentials. In the HeLa cell extract traces, there is a rising phase seen in the 18 and 21-nucleotide dsRNAs, but not the 24-nucleotide dsRNA.

The rate constants from the FRET traces that were identified above and in Figure 4-5 as being reflective of the RNA degradation kinetics were plotted to give the graphs in Figure 4-6 A and B, describing the degradation rate behavior of each of the 3 double-stranded RNAs as a function of concentration of either serum or HeLa cell extract. The results from serum show that the rate of degradation of the RNA decreases with length (Serum: 18 nt dsRNA, V_{max}= 1.2 min^{-1}, Serum$_{1/2}$= 4.5 mg/ml; 21 nt dsRNA, V_{max}= 0.47 min^{-1}, Serum$_{1/2}$= 7.7 mg/ml, 24 nt dsRNA, V_{max}= 0.09 min^{-1}, Serum$_{1/2}$= 1.7 mg/ml).

The rates in cell extract were significantly different from the trend in serum. In general, the rates observed in cell extract were slower. The 18-nucleotide dsRNA was the most rapidly degraded of the three constructs (HeLa: 18 nt dsRNA, V_{max}= 0.11 min^{-1}, HeLa$_{1/2}$= 1.1 mg/ml). In contrast to the serum result, the 21 and 24-nucleotide dsRNAs exhibited the same V_{max} values, signifying that they were degraded at the same rate in cell extract (Hill coefficients HeLa: 21 nt dsRNA, V_{max}= 0.02 min^{-1}, HeLa$_{1/2}$= 0.14 mg/ml; 24 nt dsRNA, V_{max}= 0.02 min^{-1}, HeLa$_{1/2}$= 0.04 mg/ml).

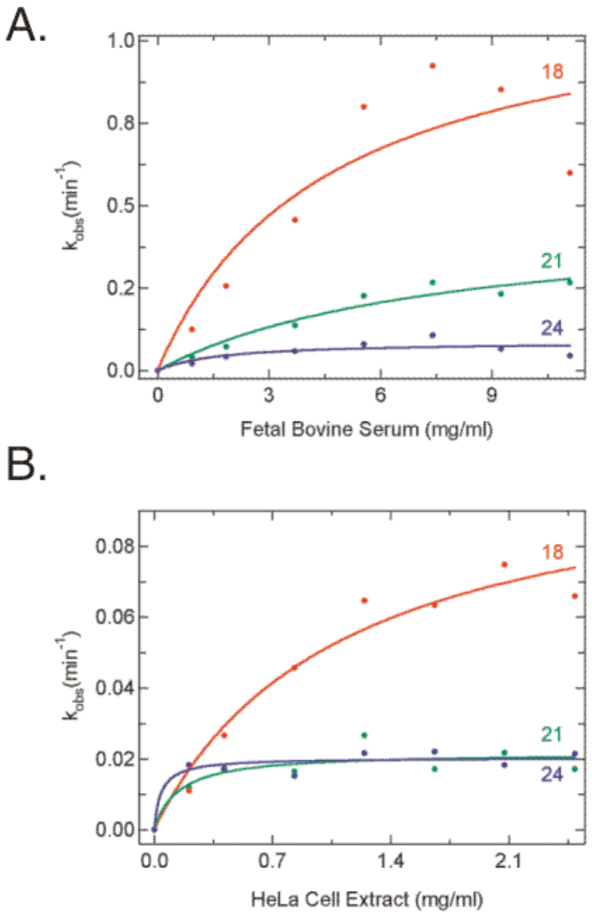


Figure 4-6. Degradation rates of 18, 21, and 24 nucleotide dsRNAs as a function of protein concentration in both serum and HeLa cell extract. A) The rate constants describing the degradation of the 18, 21, and 24 nucleotide dsRNAs (red, green, and blue respectively) in serum plotted as a function of serum protein concentration. B) The rate constants describing the degradation of the 18, 21, and 24 nucleotide dsRNAs (red, green, and blue respectively) in HeLa cell extract plotted as a function of cell extract protein concentration. Solid lines indicate fits to hyperbolic binding equations (Materials and Methods).

4.4 Discussion

Potent and specific induction of mRNA degradation is a key determinant for harnessing RNAi for future reverse genetic studies and therapeutic applications. A number of chemical modifications of RNA have been shown to increase siRNA stability both *in vitro* and *in vivo* (*69, 71, 72, 74, 77, 186*), while chemical modifications at certain positions have proven valuable to bringing specificity to the intracellular recognition of mRNA in RNAi (*66*). It remains unclear precisely how stable siRNAs are inside the cell, or if there is a benefit to intracellular stability from chemically modified siRNA. Our group and others have established the decreased nuclease sensitivity of double-stranded RNAs compared to single stranded RNA, including recent work that has employed FRET fluctuation spectroscopy to further demonstrate this trend in cells (*189, 190*). It has been reported that potency of antisense oligonucleotides correlates with chemical modification and a resulting increase in intracellular stability (*196*), however, recent work in the RNAi field suggests that no such trend exits for siRNAs. Utilizing mathematical modeling studies to understand the impact of siRNA modifications on RNAi performance, it has been suggested that the duration of silencing is the same for modified and unmodified siRNA (*69*). Further study is required, however, to generalize this effect as the chemical modifications used in this study are proprietary (*69*). This apparent contrast concerning the benefit of chemical modifications on antisense versus siRNA therapeutics may arise from the fact that antisense oligonucleotides do not harness an endogenous pathway in a manner similar to siRNA therapeutics. We hypothesized that there is a degree of protection afforded to RNAi effector molecules that does not extend to molecules too short to be assembled into a RISC particle. To address this hypothesis, establish stability profiles, and measure kinetics of degradation of small RNAs in relevant biological fluids, we developed a modified guide strand based on the *pp-luc* siRNA validated in earlier studies (*185, 187*). By creating such a reporter system, we have been able to show that our fluorophore-labeled dsRNAs interact with relevant cellular enzymes including RNase H1 and Dicer, and that 21 and 24-nucleotide double-stranded RNAs are specifically protected in HeLa cell extract.

In both HeLa cell extract and blood serum, we found that the siRNA was more stable that the single stranded RNA. These results are consistent with earlier reports, and highlight the differences in reactivity of single stranded and double-stranded RNA. We further extended earlier studies and assess the impact of secondary structure formation on RNA stability by using a DNA strand as the base-pairing partner. In blood serum, this pairing further stabilized the RNA, perhaps because of the perturbation to the A-form helical structure, or perhaps because the lack of 2'-OH functionality on an entire strand of the duplex rendered RNase A activity to a minimum. The DNA/RNA hybrid in cell extract revealed the reactivity of RNase H1. In humans, there are two known enzymes, RNase H1 and RNase HII, that specifically process DNA/RNA hybrids. This reactivity is conserved across all three kingdoms of life, and is a particularly interesting drug target because of the importance of the RNase H domain of HIV reverse transcriptase. The endogenous human enzymes are distinguished based on differential reactivity derived from substrate length and sequence, as well as sub-cellular localization where RNase HII is primarily nuclear, and RNase H1 is more widely distributed in the cell (*193*). Given the observed efficient degradation of our fluorophore-labeled DNA/RNA hybrid in cell extract, the obvious candidate in this case is RNase H1. Two DNA aptamers have been raised as inhibitors of RNase H1 (*193*), and we applied both to establish inhibition of the degradation of the DNA/RNA hybrid and implicate RNase H1. We observed dose dependent inhibition of degradation for both the V-2 and VI-2 aptamers with $\text{Aptamer}_{1/2}$ values of tens of nanomolar, in reasonable agreement with the published development of these aptamers(*193*).

We developed 18, 21, and 24-nucleotide fluorophore-labeled dsRNAs and recorded the degradation of these complexes in cell extract and blood serum. The size range of the RNAs was chosen to bridge the well characterized ~21-nucleotide cutoff that is observed for siRNAs in the induction of the RNAi response. The 21-nucleotide siRNA is bound directly to the RLC, a heterodimeric protein complex composed of Dicer and R2D2, and incorporated into RISC (*195*), whereas double-stranded RNA substrates larger than ~21 nucleotides are cleaved by Dicer into siRNAs with characteristic two-nucleotide 3' overhangs during the initial processing of the substrate. In concert with Dicer, R2D2 binds the more strongly base paired end of the siRNA helix termini, strongly associating with the 5’ phosphate of the passenger strand and priming the RLC for interaction with the catalytic protein component of the RISC complex, Argonaute (*62*). As Dicer is one of the first proteins of the RNAi pathway to interact with double-stranded RNA substrates, we tested to see if our fluorophore-labeled RNAs interact with purified Dicer enzyme in both a fluorometer assay and the related FRET gel. In both cases, we find evidence for the interaction of Dicer with the 24-nucleotide double-stranded RNA. The fluorometer assay reveals a slight increase in the FRET ratio upon Dicer addition (Figure 4-4), which is revealed to be cleavage of the 24-nucleotide substrate to a 21-nucleotide siRNA in the FRET gel. These results also demonstrate that the fluorophore-labeled antisense strand does not prevent these RNAs from interacting with Dicer, the gatekeeper to the RNAi response.

By examining the degradation rates as a function of protein concentration in serum, we found that the efficiency of degradation of the RNAs in serum decreases with size and overall thermodynamic stability of the duplex. The 18-nucleotide dsRNA was the most rapidly degraded whereas the 24-nucleotide dsRNA exhibits the slowest degradation rates. These results are easily explained based on knowledge of RNase A function, were the limiting steps of enzyme action on double-stranded substrates is oligomerization, destabilization, and invasion of duplex RNA to access the in-line attack conformation and cleave the RNA backbone (*197*). The analogous experiment was performed in HeLa cell extract, where V_{max} values from the fits of the data describe the efficiency of degradation of each duplex. In this case, the 18-nucleotide dsRNA was the most rapidly degraded, and the 21 and 24-nucleotide dsRNAs exhibited a common upper limit to the degradation rate, characterized in our system by a V_{max} of 0.02 min^{-1}. These results suggest that the 21 and 24-nucleotide dsRNAs are processed in a similar way, which, based on our finding that Dicer does interact with the fluorophore-labeled RNAs, suggests that these constructs are specifically protected in cell extract.

Our results are consistent with a recent report demonstrating that siRNA substrates remained intact and were not processed into active RISC in the absence of target RNA (*80*). This suggests that siRNAs may be bound by the RLC awaiting processing into RISC on the appearance of target RNA. The molecular basis for this finding is not known, but the phenomenon is interesting, in that it suggests a basis for interpreting our findings of a specific mechanism for protection of appropriately sized double-stranded RNAs.

CHAPTER 5:

CHEMICAL MODIFICATIONS RESOLVE THE ASYMMETRY OF siRNA STRAND DEGRADATION IN HUMAN BLOOD SERUM[4]

5.1 Introduction

Originally described in *C. elegans* (*57*), RNA interference (RNAi) is a conserved set of pathways in eukaryotes that mediates potent gene silencing triggered by the interaction of small interfering (si) and related RNA effector molecules with the RNA induced silencing complex (RISC) (for a recent review see (*67*)). RNAi holds promise for treatment of human disease because it is broadly and specifically applicable to any undesired gene whose sequence is known. While recent studies have successfully utilized systemically administered siRNAs in nonhuman primates via intravenous injection (*83*), at least two major obstacles remain before the therapeutic promise of RNAi can be fully realized. First, the guide strand that, together with the passenger strand, makes up an siRNA duplex must be stabilized against the ubiquitous exo- and endonucleases found in blood serum and elsewhere in the body (*68*). Second, adverse off-target activity due to cross-reactivity between RNAi pathways must be suppressed (*198, 199*). We demonstrate here that the same modest chemical double modification found to decrease off-target

---

[4] Adapted with permission from Hoerter, J.A., and Walter, N.G., *RNA* 13, 1887-93. Copyright 2007 Cold Spring Harbor Press.

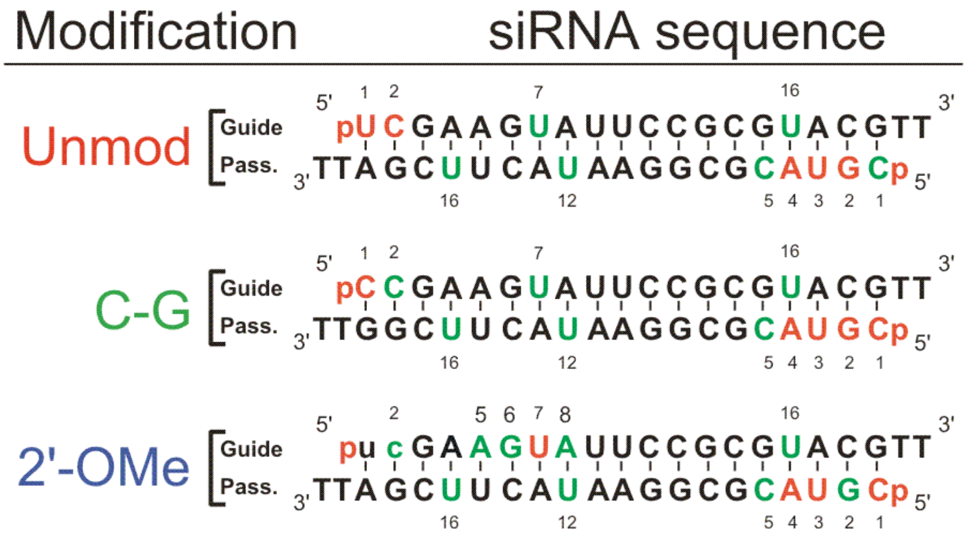


Figure 5-1. siRNAs used in this study as derived from reference (Martinez et al., 2002), with guide and passenger strands indicated. All three constructs carry 5’-phosphates as well as two deoxythymidines on the 3’-end of each strand. In the C-G siRNA a C-G base pair replaces the closing U-A base pair at the 5’-guide strand terminus. The 2’-OMe siRNA harbors two 2’-OMe modifications at positions 1 and 2 of the guide strand (depicted in lower case letters). A red nucleotide indicates a nucleolytic hit 3’ to that position and represents ≥5% of the total lane intensity, while a green nucleotide indicates a hit 3’ to that position that is >10-fold of background, but <5% of the total lane intensity.

toxicity (*66*) also substantially stabilizes the siRNA guide strand against exonucleolytic attack.

A number of studies have previously examined the stability of siRNAs in blood serum (*70-75, 77-79*). Most of this work relied on assays that could not identify the strand and/or specific site of cleavage, which is a prerequisite for designing chemical modifications to specifically protect vulnerable sites. Such information in principle is accessible through cleavage product analysis by either mass spectroscopy (*78*) or radioactive labeling combined with denaturing polyacrylamide gel electrophoresis (PAGE) (*79*). A robust kinetic analysis of nucleolytic cleavage products of both the guide and passenger strands at nucleotide resolution has, however, not yet been reported. We here have developed such an assay, which allows us to show that specific chemical 2'-O-methylation, already known to reduce off-target activity of siRNAs, selectively protects the vulnerable 5'-end of the guide strand against exonucleolytic degradation in human blood serum. A simple chemical modification of the guide strand thus serves dual functions essential for the efficient application of siRNA therapeutics.

5.2 Materials and Methods

RNA synthesis. All RNAs were synthesized by the HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory at the Yale University School of Medicine, were deprotected and purified as previously reported (*55*). Sequences are derived from the firefly luciferase gene (*185*) and are shown in Figure 5-1.

RNA labeling. RNAs were first treated with calf intestine alkaline phosphatase to remove the synthetic 5'-$PO_4^{2-}$and prepare the 5'-OH necessary for labeling. Strands were 5'-^{32}P labeled with T4 polynucleotide kinase (PNK) and γ-^{32}P-ATP at an RNA concentration of ~800 nM. PNK was inactivated by heating to 90 °C for 10 min. The appropriate complementary RNA strand was then added to the reaction at a final concentration of 2 μM and annealed by heating to 70 °C for 2 min, followed by cooling to room temperature over 10 min. Cold non-denaturing loading buffer was added to a final concentration of 1×TBE , 0.025% bromophenol blue, and 10% glycerol and the annealed, labeled siRNA purified at 4°C by electrophoresis on a non-denaturing, 1×TBE, 12% (w/v) polyacrylamide gel run at ~20 V/cm, excision of the radiolabeled duplex band, diffusion elution into 1 mM EDTA overnight at 4°C and ethanol precipitation. The purified siRNA was dissolved in 25 mM Tris-HCl, pH 7.5, 25 mM NaCl. Single-stranded 5'-^{32}P labeled RNA for the T1 and alkali ladders was prepared similarly and dissolved in water.

siRNA degradation assays. 5'-^{32}P labeled, purified siRNA duplex was prepared in a close-to physiologic standard buffer of 50 mM Tris-HOAc, pH 7.4, 80 mM KCl, 20 mM NaCl, and 1 mM $MgCl_2$. Degradation was initiated by the addition of 3% (v/v) human blood serum (Innovative Research, single healthy donor) at 37°C. RNase induced siRNA degradation assays were identical to the serum assays except that purified RNase A was added to a final concentration of 290 nM. Time points were taken after 5, 10, 20, 40 , 80, 120, and 180 min and quenched by mixing with a final concentration of 10% (v/v) Contrad 70 at pH 9.3 (Decon Labs) (*190*). Loading buffer was added to a final concentration of 1×TBE, 0.025% bromophenol blue, 0.025% xylene cyanol, and 40% formamide, and the degradation products (50,000 cpm per lane) separated on a denaturing, 8 M urea, 20% (w/v) polyacrylamide gel. Individual bands were identified by comparison with sequencing ladders from partial digestion with G-specific RNase T1 and alkali as described (*164*), quantified and normalized to the sum of all bands in a lane using a PhosphorImager Storm 840 with Image Quant software (Molecular Dynamics). Error bars are derived from at least two independent determinations. Time traces of the loss of the full-length RNA were fit with the double-exponential first-order rate equation $y = y_0 + A_1e^{-tk1} + A_2e^{-tk2}$, employing Marquardt-Levenberg nonlinear least-squares regression (Igor Pro 5.03), where A_1 and A_2 are the fractions cleaved and $\ln2/k_{1,2}$ are the reported half-lives $t_{1/2}$ of the fast and slow phases, respectively. The $t_{1/2}$ value for the fast phase of the reaction is estimated to be faster than 3 minutes based on an interpolated lower estimate of cleavage at a hypothetical 2.5 minute time point. The 2'-O-Me guide strand data were fit with the above equation with only a single exponential.

RNase induced siRNA degradation assays were identical to the serum assays except that purified RNase A was added to a final concentration of 290 nM, yielding the data shown in Table 5-1.

Thermal denaturation. siRNA duplexes at a concentration of 1 μM per strand were annealed in standard buffer (50 mM Tris-OAc pH 7.4, 80 mM KCl, 20 mM NaCl, and 1 mM $MgCl_2$) by heating to 70 °C for 2 min and cooled at room temperature over 10 min. Solutions were filtered through a 0.45 μM centrifugal filter device (Millipore Ultrafree-MC) and degassed by spinning under vacuum in an Eppendorf Vacufuge for 5 min. Samples were placed in the Micro Auto 6 T_m cell holder of a Beckman DU640B UV-Vis spectrophotometer. Samples were equilibrated at 30 °C for 15 min, and then heated to 95 °C at a rate of 0.5^oC/min while taking 1-s absorbance readings at 260 nm every min, which were corrected against a 320 nm background measured in standard buffer at the corresponding temperature. Each melting curve was measured in duplicate and independently analyzed to extract the melting temperature (T_m) and thermodynamic parameters(*200*). Reported values are the average ± error from the two melting curves except T_m values where only the average value is reported, as the error in all cases was less than 0.1°C.

Target Cleavage. RISC-mediated, siRNA-induced target cleavage of cap labeled *pp-luc* RNA was examined essentially as described (*185*). Purified *Pp*-luciferase RNA was 5'-cap ^{32}P labeled using α-^{32}P-GTP and guanylyltransferase enzyme (Ambion) as recommended by the supplier. Cap labeled RNA was purified on a denaturing, 8 M urea, 6% (w/v) polyacrylamide gel run at 27V/cm. The RNA was excised, eluted into 1 mM EDTA at 4°C overnight, ethanol precipitated, and dissolved in water.

We used each of our three siRNAs (Unmodified, C-G, 2'-O-Me) as well as an siRNA that was 2'-O-Me modified at positions 1 and 2 of both the guide and passenger strands (termed OMe(2) for distinction from the guide strand only modified OMe(1)). siRNA strands were annealed at 1:1 ratio such that the final concentration of RNA was 100 nM in the target cleavage reaction and the final $MgCl_2$ concentration was 1.1 mM. After annealing, GTP, ATP, RNasin, creatine phosphate, creatine kinase, and HeLa S100 cytoplasmic extract (Jena Bioscience) were added as specified previously(*185*), and incubated at 37°C for 15 min, at which point 100,000 cpm of cap-^{32}P labeled *pp-luc* target RNA was added. Incubation was continued at 37 °C for 2.5 h, when proteinase K and the corresponding buffer were added(*185*). The solution was incubated at 65 °C for 15 min, phenol/chloroform extracted, chloroform extracted, ethanol precipitate

Degradation experiment	Guide		Passenger	
	$t_{1/2, fast}$ *(min) [%]*	$t_{1/2, slow}$ *(min) [%]*	$t_{1/2, fast}$ *(min) [%]*	$t_{1/2, slow}$ *(min) [%]*
Unmod	< 3 [56 ± 6]	15 ± 9 [43 ± 7]	< 3 [36 ± 1]	57 ± 23 [12 ± 1]
Unmod + RNase A	< 3 [77 ± 3]	17 ± 4 [21 ± 3]	< 3 [58 ± 5]	31 ± 14 [24 ± 4]
C-G	< 3 [41 ± 4]	60 ± 4 [20 ± 2]	< 3 [42 ± 1]	35 ± 4 [6 ± 2]
2’-OMe	< 3 [53 ± 1]	-	< 3 [39 ± 2]	101 ± 100 [10 ± 1]

Table 5-1. Half-lives ($t_{1/2}$) describing the observed biphasic siRNA degradation kinetics in 3% (v/v) human blood serum and purified RNase A under standard conditions.

overnight, and the RNA dissolved in formamide loading buffer, and loaded onto a 37 cm denaturing, 8 M urea, 10% (w/v) polyacrylamide gel run at 25 W/cm. The gel was wrapped in plastic foil and exposed to a phosphor screen, which was recorded on a Storm PhosphorImager Storm 840, quantified using Image Quant software (Molecular Dynamics), and analyzed in IgorPro. Data reported results from 2-3 independent experiments.

5.3 Results and Discussion

Guide strand degraded more efficiently than passenger strand

We selectively labeled each strand of a canonical siRNA against firefly luciferase (Figure 5-1) (*185*) and followed cleavage of the full-length RNA into specific shorter products over time by gel electrophoretic separation (Figure 5-2). Dephosphorylation and cleavage of both the guide and passenger strands of our siRNA is substantial after only 5 min incubation with 3% (v/v) human blood serum in a close-to physiologic buffer (50 mM Tris-HOAc, pH 7.4, 80 mM KCl, 20 mM NaCl, and 1 mM $MgCl_2$, at 37^oC) and increases over 3 h (Figure 5-2). (As expected, the degradation kinetics accelerate with the volume fraction of serum; 3% (v/v) serum was chosen as optimal for resolving the cleavage products by gel electrophoresis.) Strong cleavage at the 5'-terminus of the guide strand, 3' to positions 1 and 2, causes markedly lower stability of the guide strand compared to the passenger strand (Figures 5-2 B and C). The kinetics are strongly biphasic so that 56% of the guide strand degrades with a half-life < 3 min, while the remaining 43% degrade with a half-life of 15 min (Table 5-1). Thus, the guide strand is completely degraded over the course of 2-3 h while 50% of the passenger strand persists after an initial, rapid, biphasic decay (Figure 5-2 and Table 5-1). Given the strict length requirement in RNA interference (*201*), the observed loss of even a single nucleotide on the 5'-end of the guide strand is expected to be functionally detrimental. The distribution of the remaining cleavage sites on the individual strands is generally not predictable (Figures 5-1 and 5-2), except that at least one cleavage site is observed in all clusters of

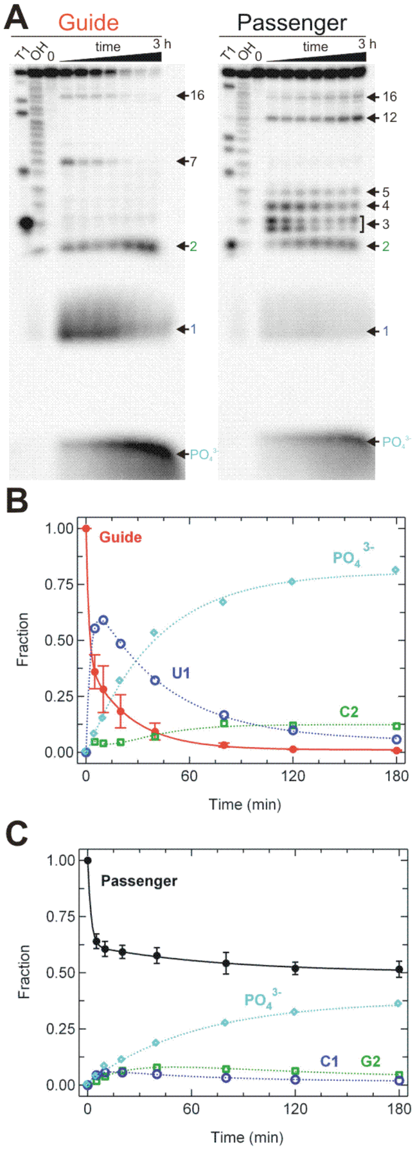


Figure 5-2. Comparison of the degradation of guide and passenger strands of the unmodified siRNA in the presence of human blood serum. A) Gel electrophoretic analysis of the time courses of serum induced cleavage of both the guide and passenger strands alongside RNase T1 and alkali (OH) sequencing ladders. Significant cleavage at positions 1 and 2 of the guide strand leads to a marked instability of the guide strand compared to the passenger strand. B) and C) Quantification of the gels in panel A where the fraction of intact siRNA, position 1, position 2, and PO_4^{3-} band intensity are plotted as a function of time. The guide strand is completely degraded over the course of three hours, whereas the passenger strand persists at ~50% over the full course of the assay.

A-U base pairs. A significant fraction of cleavage events may be secondary hits on partially degraded strands, although the time dependent decrease in full-length RNA as measured here quantifies only primary hits. The observed preference for cleavage 3' to pyrimidines is consistent with earlier studies suggesting that siRNA degradation in serum is at least in part due to RNase A-type enzymes (*75*). Human pancreatic ribonuclease (hpRNase) is the likely candidate as it is the most competent tetrapod RNase A-type enzyme to target dsRNA substrates and accounts for 70-80% of total ribonuclease activity in blood serum (*202-204*). Despite the biphasic nature of the observed degradation kinetics, the results are consistent with the action of a single enzyme, as we do not observe any clearly distinct products at early and late stages of the time course (Figure 5-2) and controls utilizing purified RNase A also manifest biphasic kinetics (Table 5-1); still, we cannot rule out the combined action of (groups of) faster and slower nucleases. Similar biphasic degradation kinetics were observed in earlier work monitoring the amount of intact siRNA duplex upon exposure to mouse plasma, suggesting that both strands survive together (*70*).

Asymmetry defined by weaker terminal U-A base-pairing

The activity of RNase A-type nucleases on double-stranded RNAs depends on the ability of the enzyme to invade the duplex and thus access the in-line attack conformation required for RNA cleavage (*197, 203*) . Our observation of a cleavage preference 3' to U over C is consistent with this requirement for strand invasion, as the weaker base pairing of U-A compared to C-G would overcome the reported preference of RNase A type enzymes for poly-C over poly-U (*204*). We therefore hypothesized that the same thermodynamic asymmetry that leads to the siRNA end fraying used by RISC to select a 5'-phosphate and thus define the guide strand (*62*) predisposes this strand to ribonuclease exposure.

To test this hypothesis, we mutated the 5'-terminal U-A base pair of our guide strand to the more stable C-G. This U-to-C transversion may be expected to enhance reactivity towards hpRNase due to the enzyme's C-over-U preference (*204*). In contrast, yet consistent with our hypothesis, we observe protection of the guide strand 5'-terminus

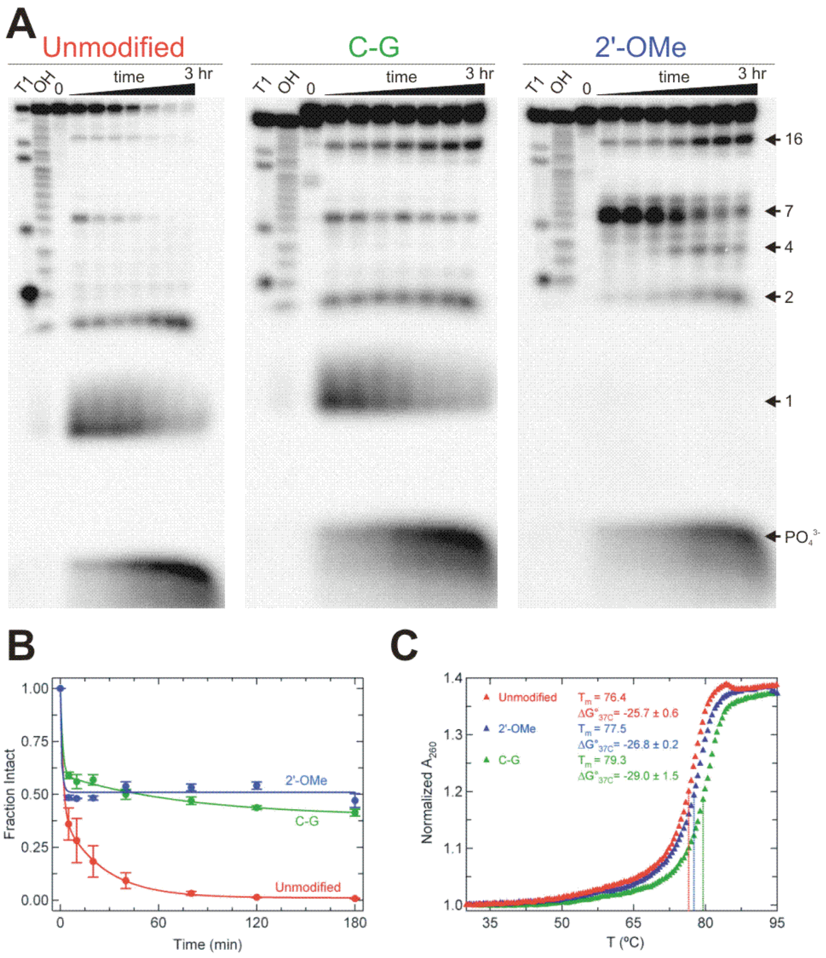


Figure 5-3. Stabilization of the guide strand by mutation and chemical modification in the presence of human blood serum under standard conditions. A) Gel electrophoretic analysis of the time course of serum induced cleavage of the guide strand of our three different siRNA constructs, as indicated, alongside RNase T1 and alkali (OH) sequencing ladders. Both terminal C-G mutation and 2'-OMe modification substantially stabilize the guide strand. B) Quantification of the gels in panel A where the fraction of intact guide strand is plotted as a function of time. C) Thermal denaturation of our three siRNA constructs as monitored by UV melting. Derived melting temperatures (T_m) and ΔG^o_{37C} are indicated on the plot.

of the C-G modified compared to the unmodified siRNA (Figures 5-3 A and B, Table 5-1). A control experiment using a mismatch duplex with the U-to-C modified guide strand and the unmodified passenger strand exhibits degradation profiles essentially the same as the unmodified duplex. Passenger strand cleavage in the C-G modified siRNA remains essentially unchanged (Table 5-1). Given the increased stability of the C-G modified siRNA (ΔG^{o}_{37C}=-29.0 ± 1.5 kcal/mol, Figure 5-3C) relative to the unmodified siRNA (ΔG^{o}_{37C}=-25.7 ± 0.6 kcal/mol, Figure 5-3C), we therefore suggest that thermodynamic stabilization of the duplex terminus protects it from RNase degradation.

2'-O-Me modification of positions 1 and 2 stabilizes the guide strand

Selection rules for effective siRNA design dictate full complementarity to the target sequence and asymmetric melting of the termini for proper selection of the guide strand by RISC (*62, 205-207*), typically ruling out a terminal U-A to C-G mutation as employed above. We therefore sought to introduce protective chemical modifications into the ribonucleolytically most vulnerable positions 1 and 2 at the 5'-end of the guide strand. A number of modifications, especially of the RNA backbone, have been previously explored for broadly stabilizing siRNAs in gene therapy applications (*67, 208-212*). In addition, recent studies have found that 2'-O-methyl (2'-O-Me) modification of positions 1 and 2 at the 5'-ends reduces the widespread problem of off-target silencing effects induced by siRNAs (*66, 213*). We thus set out to test whether this same double modification may serve the added purpose of extending the lifetime of an siRNA in human blood serum by applying our quantitative degradation assay. We find that this modest chemical double modification completely suppresses ribonucleolytic cleavage 3' to position 1. It also significantly reduces, but does not completely block cleavage 3' to position 2 (Figure 5-3A). The residual cleavage upon 2'-O-Me modification of position 2 suggests either that hpRNase uses an alternate cleavage mechanism not requiring the 2'-OH moiety or that less abundant nucleases partake in degradation (also not requiring the 2'-OH). The latter possibility is supported by the unusual cleavage sites 3' to purine residues that are induced upon 2'-O-Me modification of the guide strand (Figure 5-3A). It is possible that interference with preferred cleavage sites by chemical modification

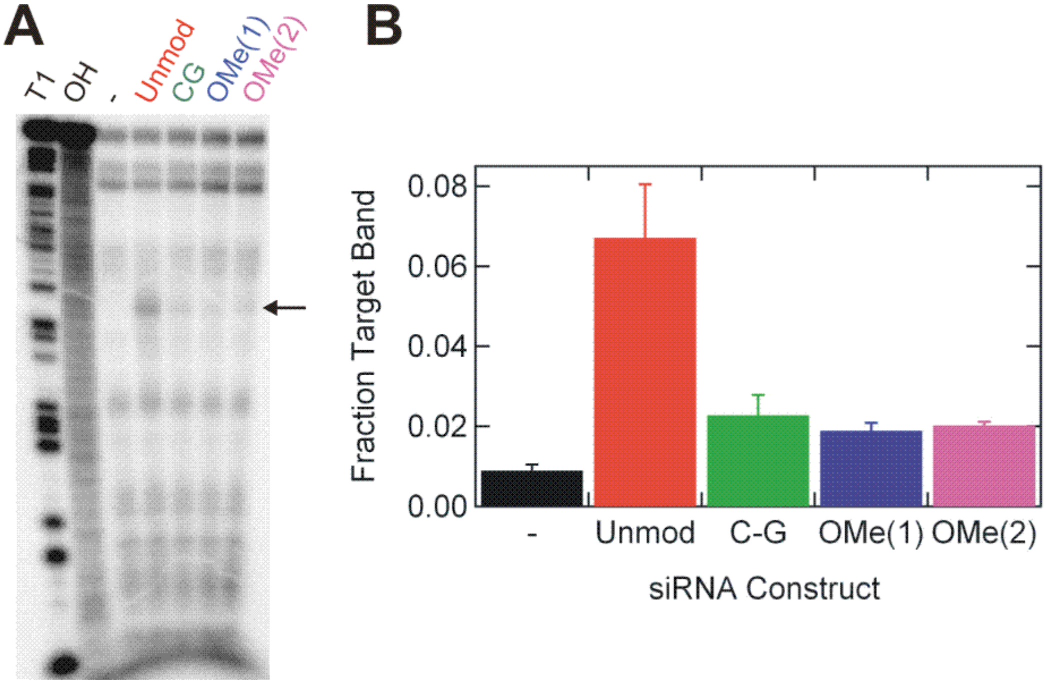


Figure 5-4. RISC-mediated, siRNA-induced cleavage of pp-luc target RNA by the siRNAs examined in this study. A) Gel electrophoretic analysis of a representative target cleavage assay, where the "-" lane is target RNA incubated with cell extract in the absence of siRNA, "Unmod" is incubated in the presence of our unmodified siRNA, "C-G" is incubated in the presence of the C-G modified siRNA, "OMe(1)" is incubated in the presence of the siRNA with guide strand only double 2'-OMe modification, and "OMe(2)" is incubated in the presence of the siRNA with symmetric 2'-OMe double modification (in positions 1 and 2 from the 5'-end) of both the guide and passenger strands. "T1" and "OH" indicate RNase T1 and alkali sequencing ladders, respectively. The arrow indicates the siRNA induced cleavage site on the target RNA. B) Quantification of the target band intensity from panel A.

enhances slower, less optimal ribonucleolytic activity at purine residues. The overall degradation kinetics of the doubly 2’-O-Me protected guide strand in human blood serum is monophasic (with the time resolution of the data) such that 52% of the RNA is still rapidly degraded with a half-life of 0.8 min, but the remaining RNA persists for long periods of time (Figure 5-3B and Table 5-1). Serving as a control, passenger strand degradation is again largely unaffected by modification of the guide strand (Table 5-1). Thermal denaturation experiments indicate only a slight increase in stability of the 2’-O-Me modified siRNA duplex relative to the unmodified duplex (ΔG^{o}_{37C}= -26.8 ± 0.2 kcal/mol versus ΔG^{o}_{37C}= -25.7 ± 0.6 kcal/mol, Figure 5-3C), which is a lesser increase in stability than afforded by the terminal U-A to C-G mutation (ΔG^{o}_{37C}=-29.0 ± 1.5 kcal/mol). These observations suggest that both thermodynamic and chemical stabilization likely contribute to the observed prolonged half-life of our 2’-O-Me modified siRNA duplex.

Interestingly, we observed reduced target RNA cleavage by RISC (*185*) upon introduction of our 2’-O-Me modifications into the siRNA guide strand, as we do for the C-G mutant (Figure 5-4). We investigated the possibility that the guide strand only 2’-O-Me modifications perturbed RISC function by conducting target cleavage assays with an siRNA carrying 1 and 2 position 2’-O-Me modifications on both the guide and passenger strands (denoted OMe(2) in the Figure 5-4), better reflecting the experiments described in the study identifying the off-target amelioration of 1 and 2 position 2’-O-methylation (*66*). Interestingly, this symmetrically modified siRNA did not rescue activity, at least in our target cleavage assay. Since the mechanism by which 2’-O-Me modification reduces off-target activity is not well understood (and target RNA cleavage was not described in the earlier study (*66*)), our finding raises the intriguing possibility that RNAi mechanisms not involving target RNA cleavage, including target sequestration, may be involved in enhancing siRNA specificity.

5.4 Conclusions

In this report we have found that a relatively modest double 2'-O-Me modification at the 5'-end significantly prolongs the half-life of a large fraction of the siRNA guide strand when exposed to human blood serum. The promise of siRNA drugs that can be specifically tailored to downregulate previously “undruggable” gene products via the RNAi pathway may usher in a new era of pharmaceuticals. There are still several problems such as targeted delivery and cellular uptake of the siRNA drug in addition to the issues of off-target effects and nucleolytic degradation to be overcome before these therapies are widely applicable. Showing here that a simple double 2'-O-Me modification, which previously has been shown to enhance siRNA specificity, selectively protects the vulnerable 5'-end of the guide strand against exonucleolytic degradation in human blood serum may help simultaneously address two of the barriers to successful implementation of siRNA therapeutics for human disease.

CHAPTER 6:

SUMMARY AND FUTURE DIRECTIONS

ncRNAs play a number of important roles in biological systems, metaphorically and literally bridging the divide between DNA information storage and the functional repertoire of proteins. RNA carries the genetic message and, at the same time, provides for recognition elements, structural scaffolds, and catalysis. Biophysical studies of RNA highlight a number of general attributes of RNA, including its ability to fold into complex and dynamic three-dimensional structures that enable function (*24, 33*). Even in cases of functional RNAs like siRNAs, a duplex composed of two perfectly base-paired strands, the dynamics of the helix termini plays a role in which strand progresses as the guide strand in RNAi (*62*). The two core systems investigated in the preceding chapters, H27 from 16S rRNA and siRNAs, are examples of the diverse array of structural and functional cellular RNAs.

6.1 H27 Dynamics and Translation

The three prokaryotic rRNAs, 16S, 23S, and 5S, are classified as primarily structural RNAs (*4*). They serve as scaffolds for binding and positioning the ribosomal proteins, tRNAs, and translation factors required for the translational cycle. These ancient RNA structures also enable a number of other functions. Peptidyl-transferase activity is harbored in 23S rRNA, where the peptidyl and aminoacyl tRNAs are ideally juxtaposed, which catalyzes the formation of the new peptide bond (*18, 43*). The decoding process is carried out in the 30S subunit primarily by 16S rRNA and involves the interaction of two highly conserved adenosines and a guanosine with the three codon:anticodon base pairs (*16*). The absence of these interactions inhibits progress on the decoding pathway. The wide variety of conformations of rRNAs, from ribosomal assembly to translation, illustrates the dynamic nature of RNA structures (*30-32*). An intriguing region of 16S rRNA, H27, was thought to be particularly important in this respect. Several lines of evidence suggest that H27 is an important player in ribosome function. H27 is highly conserved across all kingdoms of life, forms an important intersubunit bridge, and harbors sites that when mutated influence susceptibility to antibiotics (*46, 51*). In addition, it was noted that H27 harbored a sequence at the helix termini that could allow for at least two different base-pairing patterns, either the 885 or 888 conformations (*53, 93*). The original evidence of a functionally dynamic H27 provided an exciting hypothesis, suggesting that the structural variation available to a small RNA helix could profoundly influence the accuracy of the decoding process (*52*). The H27 "switch helix hypothesis" was based on a series of mutant ribosomes that exhibited biases in decoding fidelity. Ribosomes stabilized in the 885 conformation were error-prone, while 888 ribosomes were hyperaccurate. These differences led to the assertion that there was an active conformational switch in H27 of 16S rRNA, suggesting that H27 cycled between the 885 and 888 conformations during each round of amino acid incorporation, impacting the decoding process (*52*). The work with models of H27 described in Chapter 2 of this thesis establishes the co-existence of the 885 and 888 conformations and provides a kinetic and thermodynamic framework that describes their interconversion (*55*). We further showed that the antibiotic tetracycline impacts that switching process. Our H27 models also provided data for an interesting comparison of metal ion binding in crystal structures and solution phase, described in Chapter 3 (*214*).

As our results were being finalized, a new study was published that provided strong evidence against the "switch helix hypothesis", citing that a "synergistic effect" between marker mutations and the H27 mutations that led to effects not seen in the absence of the marker mutations (*54*). A "pQuad" mutant, where it is impossible for H27 to assume the 888 conformation, afforded an *E. coli* strain the same doubling time as did the WT ribosome, suggesting that the 888 conformation is not an obligatory functional structure in translation (*54*).

A number of possible explanations exist that may enable the reconciliation of the data from our H27 models showing a very dynamic isolated H27 and conclusions that the 888 conformation is not functionally relevant during translation. It is possible that H27 dynamics are an evolutionary relic from a proto-ribosome that required a dynamic H27 for translation. This possibility implies that the H27 dynamics inherent to the sequence are suppressed in modern, actively translating ribosomes. Another possibility is that H27 dynamics are important during assembly of the 30S particle, facilitating structural changes to 16S rRNA that allow for the elaborate and sequential progression of steps that enable the ribosomal proteins to bind correctly.

Another possibility for H27 function is in the process of translocation. Cryo-EM reconstructions of the ribosome have at least "circumstantially" implicated H27 in translocation, demonstrating that H27 lies on the axis that defines the ratchet-like reorganization of the 30S subunit with respect to the 50S subunit (*48, 49*). There is also additional evidence in the literature that suggests that H27 may be involved in large-scale conformational changes that could resemble the observed subunit dynamics from cryo-EM reconstructions.

The antibiotic streptomycin (Sm) binds in the 30S subunit in the region of protein S12 (*143, 156, 215*). In addition to its involvement in decoding, S12 is known to play an important structural role by binding together disparate regions of the 30S subunit, including the 530 loop, H44, and H27 (*88*). Crystallographic studies show that Sm disrupts a number of S12-rRNA contacts including H27 contacts (*156*). DMS footprinting data essentially agree with the crystallographic findings regarding Sm, in that there are a number of perturbations in the footprinting pattern of the 30S subunit centered around nucleotides 910-915 (*143, 215*). Another region of 16S rRNA that shows protections to DMS after Sm addition are nucleotides 1468-1469 in H44, which are ~60 Å away from the Sm binding site in the crystal structure (*156, 215*). A number of rRNA mutations have been reported that attenuate the effect of Sm. The first to note is that the mutation C912U in H27 confers Sm resistance (*51*). Furthermore, Allen and Noller reported that the mutation C1469U in *E.coli* 16S rRNA H44 suppresses Sm dependence

(*216*). Given that crystallography identifies a single Sm binding site, the 1469 mutation and the DMS protections suggest that there is some communication between the Sm binding region of the 30S subunit and a region along H44 that is ~60 Å away. Additional evidence links H27 to the distant 1468-1469 region of H44. Wollenzien and co-workers found a number of UV-induced crosslinks between 16S rRNA bases in the 30S subunit in the absence of Sm. G894, a base in H27, is observed to make two such crosslinks (*120, 121*). The first crosslink is to U244, which is consistent with their proximity in the crystal structures. However, G894 also crosslinks to A1468. This evidence is consistent with the notion that H27 is involved in a large (> 35 Å) conformational change that brings it in contact with the 1468-1469 region of H44. (This crosslink has been noted to depend on the ribosome preparation method used (*217*).)

Analysis of the conservation of the nucleotide identity at selected positions in these regions provides further evidence for a potential interaction. The base pairs 894-905 and 1432-1468 are highly conserved across all kingdoms of life, with the exception of mitochondrial small subunit RNA (*218*). In mitochondria, H44 is dramatically shortened. Only in this case, when position 1468 is deleted, do we find any significant variability in the nucleotide identity at positions 894-905.

All these data taken together suggest that there is a large conformational change in the 30S subunit involving H27. If H27 were the "axis" of this conformational change, then the movement of H44 would have to be in the direction of the P/E site, which is in fact what has been reported (*86*). In addition, independent data show that proteins S12 and S13 regulate translocation (*219*). These findings are consistent with the role of S12 proposed here, as well as the idea that translocation is inherent to the RNA. It is still difficult to determine any role that 885-888 H27 conformational rearrangement would have in this hypothetical large-scale motion associated with translocation. Given that translocation is catalyzed by the translation factor EF-G, it is conceivable that the inhibitory effects of the "pQuad" mutant are masked by the action of EF-G. Furthermore, perhaps only a disruption of the base pairing pattern in H27 is required, and that a formal establishment of an alternate arrangement is not necessary.

Future experiments to test this hypothesis should focus on defining compensatory mutations between the H27 894 site and the H44 1468-1469 region, examining Sm sensitivity and EF-G independent translocation in these mutant ribosomes. An ideal system with which to test these mutant ribosomes has been developed by our collaborator at Wayne State, Prof. Phil Cunningham. The system employs ribosomes with a modified Shine-Dalgarno sequence that enables ribosomes to translate specific reporter messages, separating out the mutant ribosomes from the background of WT ribosomes present in the cell (*220*). In fact, this system has already been utilized to study mutations in H27 (*179, 221, 222*). Strategies could include creating deleterious mutations at either the 894 position of H27 or the 1468-9 position in H44 followed by PCR mutagenesis of 16S rRNA to find compensatory mutants that rescue ribosome function. If any mutants were found, they could be assayed for translocation using a toeprinting assay (*223*). A number of other potential experiments exist for this hypothesis, but the *in vivo* reporter system from the Cunningham lab is a facile and high throughput method for testing a functional interaction between H27 and H44.

6.2 Stability of siRNAs and Their Utility as Therapeutics

As effector molecules of the RNAi pathway, enormous effort has focused on modifying the pharmacokinetic properties of siRNAs to enhance their utility as therapeutics (*67, 68*). The great promise of siRNA therapeutics has yet to be fulfilled, and challenges include stability, tissue specific delivery, and specificity of the siRNA drug (*81*). A primary problem with siRNA therapeutics is the inherent nuclease sensitivity of RNA. Regardless of the techniques of delivery, siRNA drugs are exposed to the extracellular environment, a region that is rich in nuclease enzymes that can quickly degrade RNA. Chemical modifications have provided excellent stability to siRNAs in the extracellular environment, as demonstrated in Chapter 5 of this thesis and elsewhere (*67, 76*). Important questions persist regarding the utility of chemical modifications in the intracellular environment. A recent study has shown the importance of 2'-O-Me chemical modifications in the 1 and 2 positions of the siRNA for reducing "off-target" effects of siRNAs, and we have further demonstrated in Chapter 5 that this same pattern of modification greatly enhances siRNA stability in the extracellular environment (*66, 76*). Work in Chapter 4 shows that RNAi active dsRNA substrates are specifically protected in cell extract, consistent with independent fluorescence and modeling studies of intracellular siRNAs (*69, 80*). The precise mechanism of this protection is unknown. Future studies should focus on which components of the cell or of the RNAi machinery are responsible for this specific protection of siRNA-like duplexes. An obvious hypothesis is that the RISC loading complex (RLC) is responsible for binding siRNAs immediately on entry into the cell and thereby affording protection from degradation.

The studies in Chapter 5 have found a discrepancy between the action of siRNAs that are 2'-O-Me modified at the 1 and 2 positions (*66, 76*). Our data show reduced target cleavage efficiency, whereas a previous study did not note any reduction in gene silencing activity of the modified siRNAs as evaluated by microarray assay (*66*). The differences in the target cleavage assay and the microarray analysis are consistent with a mechanism whereby 2'-O-Me modification of the 1 and 2 positions influences whether the target is directly cleaved or sequestered in a manner similar to the miRNA pathway.

The first step in resolving this discrepancy would be to utilize the same siRNA between the two studies, that is, conduct microarray analysis and target cleavage assays utilizing the same siRNA sequence. In the event that a discrepancy still exists, the 1 and 2 position 2'-O-Me modified siRNA and a control without the modifications would have to be assayed for translational repression, an effect resembling miRNA repression. It is also necessary to determine if the chemically modified guide strand sill targets the same region of the mRNA, or perhaps the siRNA becomes less specific and also targets other regions. Another possibility includes inhibition of RISC catalytic efficiency by the chemical modifications, which could be determined by binding assays paired with target cleavage assays. It is clear that the RNAi pathway is still being characterized, and future efforts to understand the crossover between siRNA cleavage mechanisms and miRNA repression will significantly advance the field.

APPENDIX 1

DEVELOPMENT OF A PARTICLE TRACKING ASSAY FOR PROKARYOTIC TRANSLATION[5]

Introduction

There has been enormous progress recently in understanding the structure and dynamics of the ribosome. Recent crystal structures of the ribosome have illuminated a number of questions concerning ribosome function, including how the ribosome selects for the correct codon:anticodon pairing (*96*) and how peptidyl transfer works (*44*). Cryo-EM studies of ribosomes have provided additional insight into structure and dynamics, particularly into the dramatic changes associated with translocation (*30*). However, these studies have centered on local ribosome-mRNA interactions. A translating ribosome interacts with many more codons of mRNA than those pictured in the snapshots of x-ray or cryo-EM structures, leaving a gap in our understanding of how the translating ribosome moves over and interacts with mRNA. It is well known that structural features of mRNAs can induce frameshifts, changes in the mRNA-ribosome register, and thereby profoundly impact the nature of the translated protein. Additionally, recent studies have demonstrated that the ribosome harbors helicase activity, allowing the ribosome to unfold structures in mRNA encountered during the course of translation (*224*). To address

---

[5] This work is a collaboration between the following individuals: John Hoerter, Dr. Anthony Manzo, Ryan Murphy (rotation student), Prof. Nils Walter, Prof. Michael Morris, Prof. Mark Burns, Dr. Peizhi Zhu, Prof. Phil Cunningham, and Tek Lamichhane. Rotation students Annette Casiano, Blake Erickson, and Dan Wahl have also contributed during the early stages of this work.

processes of this nature, we are developing materials and methods that will allow for particle tracking of single translating ribosomes and mRNAs.

Particle tracking methods (*225*) have been successfully developed for use in tracking biological complexes both *in vitro* and *in vivo* (*226, 227*). To date, particle tracking efforts have focused primarily on the use of organic dyes as fluorescent probes, with notable success in application to the molecular motors kinesin and myosin. Quantum dots (QDots) are semiconductor nanocrystals whose unique luminescent properties including high photostability and broad excitation but narrow emission spectra make them ideal for particle tracking applications where long observation windows and multiple different probes are required. QDots have already been applied as particle tracking probes (*228*), and we seek to extend the utility of QDots to assays of prokaryotic translation. We have developed a protocol (based on an existing Invitrogen protocol) for derivatizing carboxylate coated QDots with DNA primers. With these materials, we have established specific binding of QDots to surface tethered nucleic acids and finally used QDot probes targeted to different regions of an mRNA to study diffusive motion. We also discuss initial *in vitro* translation assays and control experiments exploring the binding of 30S subunits to QDot-DNA conjugates.

Materials and Methods

Conjugation of Amine Coated QDots and SH-modified DNA

Our initial strategy for generating QDot-DNA conjugates relied on a protocol from Evident Technology that utilized the company's amino modified QDots and 5' SH-modified DNA oligonucleotides from Keck. We performed the coupling reaction several times, noting that during the ultracentrifugation purification process the QDots tended to form aggregates. We did use the material from these coupling reactions and found evidence for mRNA binding, including good correspondence between the 565 and 705 QDots hybridized to the same mRNA. Upon storage of several months, we noted that the QDots lost their fluorescence properties. This finding, combined with the fact that the QDots are cationic and the mRNA target to which they bind is anionic, led us to investigate other QDot conjugation strategies and suppliers.

Conjugation of Carboxylate QDots and NH_2-modified DNA

5' amino modified DNA primers were purchased from Invitrogen Corp. and used as received from the supplier. The sequences are as follows: 32-5'DNA, CAGTTGGTGCTATGACACTTTACC; 32-3'DNA, TCAAAAGGTCATCCAGGTCC; 60-5'DNA, TTTCATCAGGAATCCAACCG; 60-3'DNA, GCTCATCCATTCTTTACGAAGG; hp5 DNA, GGGAGATCAGGATA. The conjugation protocol is based on an Invitrogen Corp. protocol. Briefly, 5 µl of 565 nm or 705 nm carboxylate modified QDots (Invitrogen Corp.) were mixed with a 0.8 µl solution of EDC (1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride, Pierce) (1 mg/ml EDC concentration for 565 nm QDots and 10mg/ml EDC concentration for 705 nm QDots) and incubated at room temperature for 5 min protected from light. The QDot EDC solution was then mixed with 35 µl of the appropriate DNA primer that had previously been dissolved in 10 mM sodium borate pH 7.4 at final concentration of 200 µM. The QDot/EDC/DNA solution was mixed and incubated overnight at room temperature while protected from light. The solutions were then loaded into a 50,000 molecular weight cut-off (MWCO) VivaSpin (VWR) centrifugal filter and spun at 10,000 RPM for 5 minutes. The retained QDot-DNA conjugate was then washed four times with 400 µl of 50 mM sodium borate pH 8.3. The concentration of the QDot-DNA conjugates was then measured by UV absorbance using the QDot extinction coefficients supplied by Invitrogen (for 705 QDots, ε_{350}= 12.9x10^6 $M^{-1}cm^{-1}$, ε_{405}= 8.3x10^6 $M^{-1}cm^{-1,}$ ε_{488}= 3.0x10^6 $M^{-1}cm^{-1}$, chosen such that the absorbance at the wavelength specified by the extinction coefficient was less than 1 AU) and diluted out to a standard working concentration of 1 µM.

Purification of T4 bacteriophage genomic DNA

To approximately 2 mg of whole T4 bacteriophage (ACTC, lyophilized powder), 150 µl of 100 mM Tris-HOAc, pH 7.4, 160 mM KCl, 40 mM NaCl, and 2 mM $MgCl_2$ (other buffers can be and were also used here), 2 ml water, and 5 ml of 1:1 phenol:chloroform were added and the resulting suspension was vortexed and then centrifuged at 9,000 RPM. The resulting top aqueous phase was separated and mixed with 0.1 volumes of 3 M NaOAc pH 5.2 and 3 volumes of ethanol and allowed to precipitate overnight at -20ºC. The precipitated DNA was recovered by centrifugation at 9,000 RPM for 30 minutes, decanted, washed with 80% ethanol, dried, and dissolved in water. This stock solution served as the template for PCR reactions generating the gene 60 and gene 32 mRNAs and plasmids.

Initial cloning of gene 60 and gene 32 mRNAs

The initial cloning strategy for gene 32 (*229*) and gene 60 (*230*) consisted of a single round of Taq PCR were a T7 promoter was added upstream of the gene of interest, followed by digestion by Sph I and Kpn I, and ligation into the pUC19 plasmid. The primers are as follows: 32-left, CGATACTAGATAGCATGCTAATACGACTCACTATAGGGCTATGAGGTAAAGTGTCATAGCACCAACTGTT; 32-right, CCTTAGAACATGGGTACCGGGGGACCTCTAGGGTCCCCAATTAA; 60-left, CGATACTAGATAGCATGCTAATACGACTCACTATAGGGCGGTTGGATTCCTGATGAAAAGTTCTAT; 60-right, CCTTAGAACATGGGTACCGAAATTCTTCGGCTTCTCCCATATCGAAAAGC. Transformation and sequencing of these ligation reactions yielded no gene 32 clones and a number of multiply mutated gene 60 clones. These results suggested that the phage mRNAs may have a detrimental effect on *E.coli* growth, and therefore our strategy may not be fruitful. Consequently, we utilized the above primers to generate large amounts of the PCR product to use as a template for T7 transcription reactions to generate the mRNA for particle tracking experiments. This is the sequence depicted in Appendix Figure 1.

Design and Development of mRNAs compatible with Cunningham Lab Ribosomes

The mutant 30S ribosomal subunits utilized in the Cunningham lab contain a mutated Shine-Dalgarno sequence at the 3'-end of 16S rRNA (*220*). In order to utilize this system in our *in vitro* translation assay, we developed new gene 32 and gene 60 mRNAs with two modifications including an altered ribosome binding site in the 5' UTR of each mRNA. The Cunningham lab uses genes that have been further modified to have the same nucleotide sequence in the 5'-UTR to normalize the frequency of translation

Gene 60 Primer Binding Sites

5'-bitoin
GGGCGGTTGGATTCCTGATGAAAAGTTCTATGAGGTGTATAATGAAATTTGT
AAAAATTGATTCTTCTAGCGTTGATATGAAAAAATATAAATTGCAGAACAAT
GTTCGTCGTTCTATTAAATCCTCTTCAATGAACTATGCGAATGTCGCTATTAT
GACAGACGCAGATCACGATGGATAGCCTTCGGGCTATCTATAGAAATACCTC
ATAATTAAGAGATTATTGGATTAGGTTCTATTTATCCTTCTCTGCTCGGATTTT
TTAGTAATTGGCCAGAATTGTTTGAGCAAGGACGAATTCGCTTTGTCAAAACT
CCTGTAATCATCGCTCAGGTCGGTAAAAAACAAGAATGGTTTTATACAGTCG
CTGAATATGAGAGTGCCAAAGATGCTCTACCTAAACATAGCATCCGTTATATT
AAGGGACTTGGCTCTTTGGAAAAATCTGAATATCGTGAGATGATTCAAAACC
CAGTATATGATGTTGTTAAACTTCCTGAGAACTGGAAAGAGCTTTTTGAAATG
CTCATGGGAGATAATGCTGACCTTCGTAAAGAATGGATGAGCCAGTAGTTTA
CTTTACCACAAGGATGTGGTATAATTAATTGGGCAAATGAGGATATTGAAAT
GAATCATATAAAGTAAATTTAGAACTTTTTGATAAAGCAGTTCATCGAGAAT
ATAGAATCATTCAACGCTTTTTCGATATGGGAGAAGCCGAAGAATTT
3'

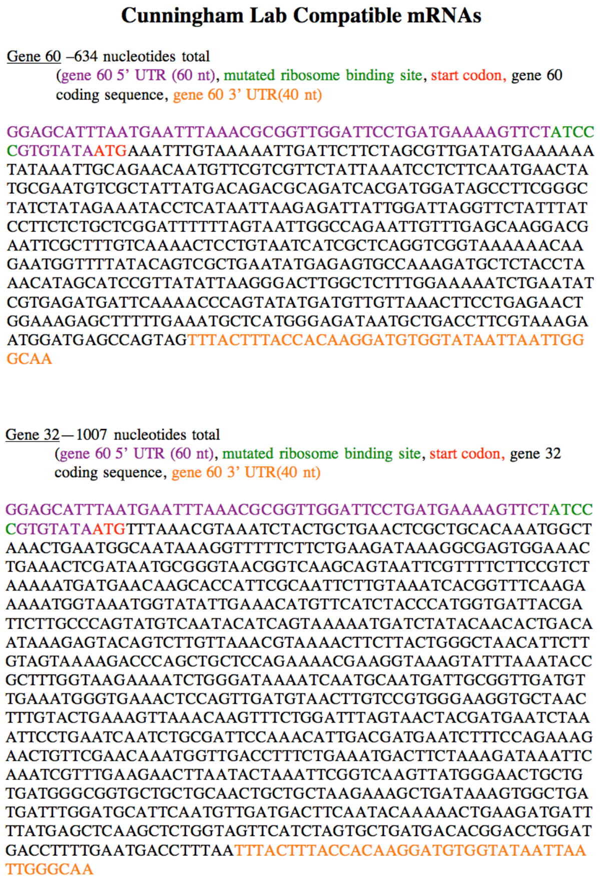
5' labeling site
3' labeling site

729 nucleotides total
523 nucleotides between primer binding sites

Appendix Figure 1. Sequence and primer binding sites for PCR transcribed gene 60 mRNA. The start codon and stop codons are underlined, and the 5'-UTR primer binding site is highlighted in red, and the 3' binding site is highlighted in blue.

initiation. For this reason, the gene 32 mRNA was redesigned to carry the gene 60 5’ and 3’ UTRs. The PCR strategy required 3 rounds, and the primers are as follows: gene 60 left-1, TGATGAAAAGTTCTATCCCGTGTATAATGAAATTTGTAAAAATTGATTCTTCTAG; gene 60 right-1, TAATTATACCACATCCTTGTGGTAAAGTAAACTACTGGCTCATCCATTCTTTACG; gene 32 left-1, TGATGAAAAGTTCTATCCCGTGTATAATGTTTAAACGTAAATCTACTGCTGAAC; gene 32 right-1, TAATTATACCACATCCTTGTGGTAAAGTAAATTAAAGGTCATTCAAAAGGTCATCC; common left-2, ATTTAATGAATTTAAACGCGGTTGGATTCCTGATGAAAAGTTCTATCCCGTG; common right-2, GCTACTAAGCTTTTGCCCAATTAATTATACCACATCCTTGTGGTAAAG; common left-3, CGTAGTTCTAGATAATACGACTCACTATAGGAGCATTTAATGAATTTAAACGCGGTTG. The restriction enzymes used for successful cloning were HindIII and XbaI.

Clones of both gene 32 and gene 60 were isolated using this strategy, but each isolate had a number of mutations. The Quick Change Multi Mutagenesis kit (Stratagene) was used to correct these mutations. The clones utilized were 864419 and 864420 (UofM Sequencing Core ID numbers) for gene 32 and gene 60, respectively. The primers utilized for the Quick Change mutagenesis are listed as follows: Q32-1, GAAGGTAAAGTATTTAAATACCGCTTTGGTAAGAAAATCTGGG; Q32-2, GGATTTAGTAACTACGATGAATCTAAATTCCTGAATCAATCTGC; Q32-3, TTCTGAAATGACTTCTAAAGATAAATTCAAATCGTTTGAAGAAC; Q32-4, GCTAAGAAAGCTGATAAAGTGGCTGATGATTTGGATGCATTC; Q32-5, ATTTACTTTACCACAAGGATGTGGTATAATTAATTGGGCAAAAG; Q60-1, GAAAAAATATAAATTGCAGAACAATGTTCGTCGTTCTATTAAATC; Q60-2, CCTTCGGGCTATCTATAGAAATACCTCATAATTAAGAGATTATTG; Q60-3/4, TCTATTTATCCTTCTCTGCTCGGATTTTTTAGTAATTGGCCAGAATTGTTTGAGC.



Cunningham Lab Compatible mRNAs

Gene 60 –634 nucleotides total
(gene 60 5’ UTR (60 nt), mutated ribosome binding site, start codon, gene 60 coding sequence, gene 60 3’ UTR(40 nt)

GGAGCATTTAATGAATTTAAACGCGGTTGGATTCCTGATGAAAAGTTCTATCC
CGTGTATAATGAAATTTGTAAAAATTGATTCTTCTAGCGTTGATATGAAAAAA
TATAAATTGCAGAACAATGTTCGTCGTTCTATTAAATCCTCTTCAATGAACTA
TGCGAATGTCGCTATTATGACAGACGCAGATCACGATGGATAGCCTTCGGGC
TATCTATAGAAATACCTCATAATTAAGAGATTATTGGATTAGGTTCTATTTAT
CCTTCTCTGCTCGGATTTTTTAGTAATTGGCCAGAATTGTTTGAGCAAGGACG
AATTCGCTTTGTCAAAACTCCTGTAATCATCGCTCAGGTCGGTAAAAAACAA
GAATGGTTTTATACAGTCGCTGAATATGAGAGTGCCAAAGATGCTCTACCTA
AACATAGCATCCGTTATATTAAGGGACTTGGCTCTTTGGAAAAATCTGAATAT
CGTGAGATGATTCAAAACCCAGTATATGATGTTGTTAAACTTCCTGAGAACT
GGAAAGAGCTTTTTGAAATGCTCATGGGAGATAATGCTGACCTTCGTAAAGA
ATGGATGAGCCAGTAGTTTACTTTACCACAAGGATGTGGTATAATTAATTGG
GCAA

Gene 32—1007 nucleotides total
(gene 60 5’ UTR (60 nt), mutated ribosome binding site, start codon, gene 32 coding sequence, gene 60 3’ UTR(40 nt)

GGAGCATTTAATGAATTTAAACGCGGTTGGATTCCTGATGAAAAGTTCTATCC
CGTGTATAATGTTTAAACGTAAATCTACTGCTGAACTCGCTGCACAAATGGCT
AAACTGAATGGCAATAAAGGTTTTTCTTCTGAAGATAAAGGCGAGTGGAAAC
TGAAACTCGATAATGCGGGTAACGGTCAAGCAGTAATTCGTTTTCTTCCGTCT
AAAAATGATGAACAAGCACCATTCGCAATTCTTGTAAATCACGGTTTCAAGA
AAAATGGTAAATGGTATATTGAAACATGTTCATCTACCCATGGTGATTACGA
TTCTTGCCCAGTATGTCAATACATCAGTAAAAATGATCTATACAACACTGACA
ATAAAGAGTACAGTCTTGTTAAACGTAAAACTTCTTACTGGGCTAACATTCTT
GTAGTAAAAGACCCAGCTGCTCCAGAAAACGAAGGTAAAGTATTTAAATACC
GCTTTGGTAAGAAAATCTGGGATAAAATCAATGCAATGATTGCGGTTGATGT
TGAAATGGGTGAAACTCCAGTTGATGTAACTTGTCCGTGGGAAGGTGCTAAC
TTTGTACTGAAAGTTAAACAAGTTTCTGGATTTAGTAACTACGATGAATCTAA
ATTCCTGAATCAATCTGCGATTCCAAACATTGACGATGAATCTTTCCAGAAAG
AACTGTTCGAACAAATGGTTGACCTTTCTGAAATGACTTCTAAAGATAAATTC
AAATCGTTTGAAGAACTTAATACTAAATTCGGTCAAGTTATGGGAACTGCTG
TGATGGGCGGTGCTGCTGCAACTGCTGCTAAGAAAGCTGATAAAGTGGCTGA
TGATTTGGATGCATTCAATGTTGATGACTTCAATACAAAAACTGAAGATGATT
TTATGAGCTCAAGCTCTGGTAGTTCATCTAGTGCTGATGACACGGACCTGGAT
GACCTTTTGAATGACCTTTAATTTACTTTACCACAAGGATGTGGTATAATTAA
TTGGGCAA

Appendix Figure 2. Cunningham lab-compatible mRNA design. These new gene 60 and gene 32 mRNAs incorporate the modified ribosome binding site required my the Cunningham lab mutant ribosomes. In addition, the gene 60 5’ and 3’ UTRs were grafted onto the gene 32 coding sequence so that both mRNA should share similar sequence in the region of the initiating ribosome as well as sharing common UTR binding sites for DNA primers.

Finally, these clones were HindIII digested for run-off transcription for use in both *in vitro* transcription reactions as well as *in vitro* translation reactions. The final sequences after the Quick Change mutagenesis were the exact constructs outlined in Appendix Figure 2.

PCR for Generation of Complementary RNAs for gene 60 and gene 32

The purified round 1 PCR product taken from the cloning process that generated the of the Cunningham compatible mRNAs was used as template to generate PCR products for transcription of RNAs that are complementary to gene 32 and gene 60. The primers used are as follows: c32 left, TACTGCTGAACTCGCTGCAC; c32 right, GCATGCTAATACGACTCACTATAGGCTTGAGCTCATAAAATCATCTTC; c60 left, AATTGATTCTTCTAGCGTTGATATG; c60 right, GCATGCTAATACGACTCACTATAGAAAAAGCTCTTTCCAGTTCTCAGG. The underlined regions of the primers denote the T7 promoter region including the G^{+1}.

Transcription of RNAs

Transcription reactions of RNAs were carried out using the PCR DNA as template in all cases, including gene 60 and gene 32 mRNAs and the gene 60 and gene 32 complementary (c)RNAs. Template concentration ranged from 80-100 nM with 120 mM HEPES-KOH pH 7.5, 30 mM $MgCl_2$, 2 mM spermidine, 0.01% Triton X-100, 4 mM each NTP, 40 mM DTT and, in select cases where generation of a 5’ biotin was required, 0.4 mM of biotin-GG (Dharmacon). The reactions were run at sufficiently small volumes (20-100 µl) so they were loaded directly on the gel. Reactions were mixed with loading buffer, added to a final concentration of 1×TBE, 0.025% bromophenol blue, and 40% formamide, and the products separated on a denaturing, 8 M urea, 4% (w/v) polyacrylamide gel (6% gel for cRNAs), visualized using UV shadowing, excised, eluted in 1 mM EDTA, and ethanol precipitated.

Ribosome Design, Development, and Purification

Puglisi and co-workers have developed a strategy for site-specific fluorophore labeling of the *E. coli* ribosome (*231*). The basic idea is a genetic swap of a surface exposed, isolated rRNA loop with an elongated meta-stable helix that enables binding of a fluorophore-labeled DNA primer. The authors initially designed two hairpin sequences and introduced those sequences into the 30S subunit in a number of places. They observed that several of these mutants encoded functional ribosomes. However, these original hairpin designs did not have high affinity for the complementary DNA primer. The authors then designed three new hairpin sequences (hairpins 3, 4, and 5), but only introduced these sequences into a single site (helix 33a). The new hairpin designs displayed a much higher affinity for the complementary DNA. The authors conducted single molecule studies on the helix 33a hairpin 5 mutant.

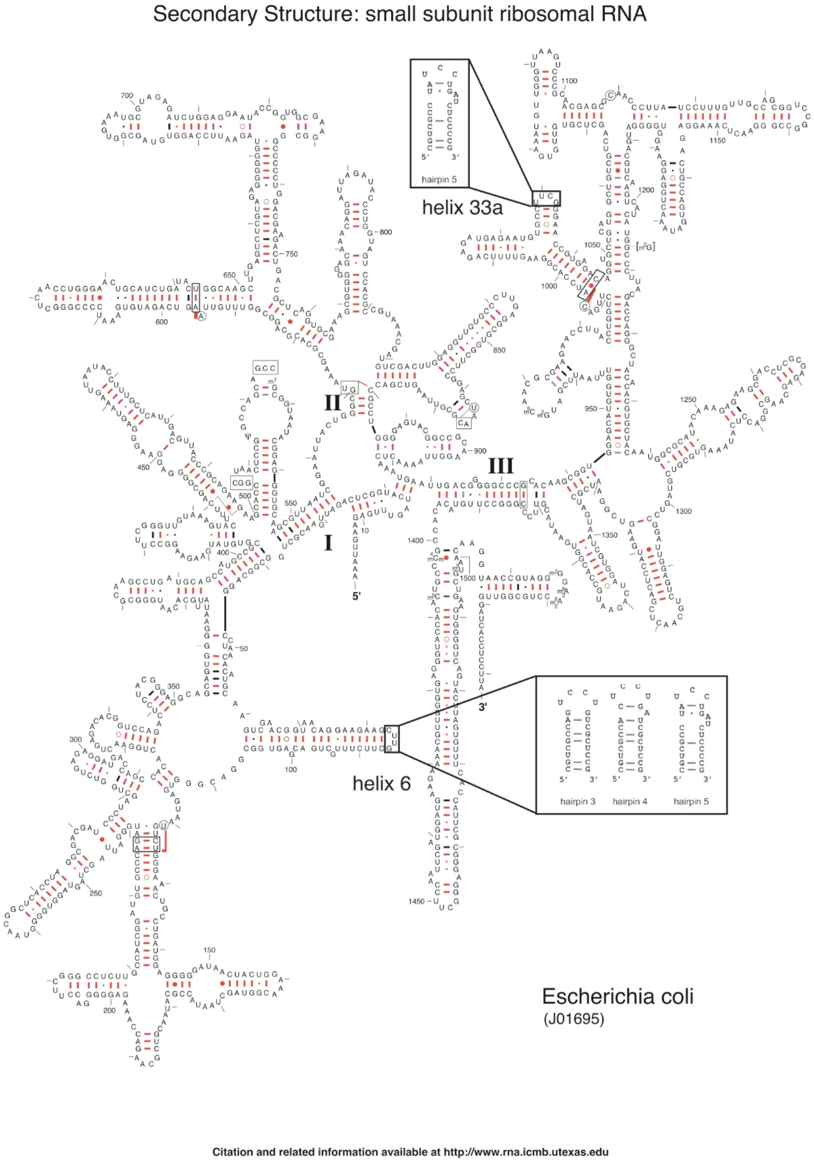
In collaboration with the Cunningham lab, we worked towards using this scheme to introduce specific fluorophore or QDot labels (via complementary DNA probes) to the *E. coli* ribosome. We have selected two candidate-labeling positions, helix 33a and helix 6 of the 30S subunit (both were viable mutants in the Puglisi study).

A technician in the Cunningham lab made a total of 4 mutants: hairpin 5 mutant of helix 33a and hairpin 3, 4, and 5 mutants of helix 6. This strategy exploits two different positions on the ribosome, both of which were shown by Puglisi and co-workers to tolerate these hairpin-type modifications. Helix 33a-hairpin 5 is well characterized, but the helix 6 mutation was not explored with the high affinity hairpins (not hairpins 3, 4, and 5). Because helix 6 is not as well studied as helix 33a, we introduced all 3 of the high affinity mutant hairpins. This strategy is spelled out below on a secondary structure depiction of *E. coli* 16S rRNA (Appendix Figure 3).

A technician in the Cunningham lab assayed each of these mutants for translational efficiency in an *in vivo* assay. The results are in the following table

30S subunit mutation	% GFP production relative to wild-type
Helix 6, hairpin 3	34
Helix 6, hairpin 4	46
Helix 6, hairpin 5	10
Helix 33a, hairpin 5	65

Appendix Table 1. Activities of mutant ribosomes**.**



Appendix Figure 3. Location and sequence of mutated 16S rRNA constructs developed in the Cunningham lab. Of each of these constructs, helix 33a hairpin 5 mutant was the most active. The secondary structure diagram is adapted from reference (*90*).

Based on these results we chose the most active ribosome helix 33a hairpin 5 mutant for our studies. Briefly, *E. coli* containing the mutant 30S subunits were grown up, pelleted, resuspended in Buffer A (50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 100 mM NH_4Cl, and 0.5 mM EDTA) lysed using a French Press, subjected to a DNase I (final concentration of 5 μg/ml) digestion step on ice for 5 minutes, and the cell debris was pelleted in a centrifuge for 15 minutes at 15,000 rpm. The supernatant was adjusted to 0.5 M NH_4Cl, and the ribosomes were pelleted by ultracentrifugation through a 10% sucrose cushion. The pellet was collected and resuspended in Buffer B (50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, and 100 mM NH_4Cl) overnight at 4 ºC. The ribosomes were then passed through a sucrose gradient to separate the large and small subunits and separated using a fraction collector and UV detector. Fractions were pooled, yielding 50S subunits, 30S subunits with mutant Shine-Dalgarno sequences, and 30S subunits with both mutant Shine-Dalgarno sequences and the helix 33a hairpin 5 mutation. The 30S subunit pools were further treated to remove wild-type 30S subunits by incubation with a biotinylated DNA oligo (FISH primer, 5`biotin- ATC$(ATC)_{14}$ATCTAAGGAGGTGAT) designed to bind the wild-type Shine-Dalgarno sequence and be irreversibly trapped on Tetravidin beads (Promega). This yielded the final stocks of 30S ribosomal subunits. Detailed versions of these protocols are available from the Cunningham lab at Wayne State University.

PEGylation of microscope slides

The fused silica microscope slides used for imaging experiments were covalently modified with poly(ethyleneglycol) (PEG) to minimize non-specific binding of both QDots and proteins. This was accomplished according to protocols modified from previous publications (*232-234*). Briefly, the slides were thoroughly cleaned and flamed off, reacted with aminopropyltriethoxysilane (APTES) in acetone, rinsed, dried, allowed to react overnight with OMe-PEG-OSu and biotin-PEG-OSu (the PEG reagent suppliers are currently in flux, but the MW used ranged from 3.000 to 5,000 MW), rinsed, dried, and allowed to react with sulfo-disuccinimidyltartarate (sulfo-DST, Soltec Ventures) for 30 minutes. The slides were finally rinsed, dried, and assembled into flow channels.

Annealing of mRNA and QDot-Conjugates

QDot-DNA conjugates were annealed with target (mRNAs or cDNA) at a concentration of 20 nM each in T50 buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl) and 100 mM DTT. Additionally, complementary (c)RNA for gene 60 or gene 32 (prepared above) can be added at the desired concentration to hybridize to the mRNA and create a region of long dsRNA. The resulting solution is heated at 70 ºC for 2 minutes followed by cooling to room temperature over 10 minutes. Samples are then diluted out to a working concentration of ~50 pM in T50 buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl) plus 100 mM DTT before loading on the microscope slide.

Loading of microscope slide for imaging

Microscope slides were prepared by first adding a stock solution of 0.2 mg/ml streptavidin in T50 buffer to the slide followed by an incubation of 10 min at room temperature. The slide is then rinsed with 80 µl of T50 + DTT buffer followed by application of a blocking buffer (1 mg/ml BSA, 1 mg/ml sheared salmon sperm DNA in T50 + DTT buffer) with an incubation time of 10 min. The slide is again rinsed with 80 µl of T50 + DTT buffer followed by application of the QDot containing sample and incubation for 10 min. The slide is finally rinsed with 160 µl of T50 + DTT buffer. The external regions of the slide are wiped clean using water and allowed to dry.

Fluorescence Microscopy and Data Analysis

Data were collected on a prism-based total internal reflection fluorescence (TIRF) microscope constructed as described in (*24*). For our experiments, the filter wheel on the 532 nm laser was set at 0.04, the MCP on the I-Pentamax CCD camera was set at 60%, and the integration time was ~0.1 s. For the data sets examining the specificity of QDot binding, only the first frame of each movie was analyzed, and the number of spots from 2-3 different fields of view under each condition was reported. For data sets where samples were tracked as a function of time, 2-3 movies (different fields of view) each totaling of 600 frames (60 s) were analyzed. The data analysis was carried out either in Diatrack (Semasopht) or Matlab.

In Vitro Translation

The Ambion Active-Pro *In Vitro* translation kit was used for *in vitro* translation assays. Lab made recombinant T7 RNA polymerase was exchanged with the T7 polymerase supplied in the kit without problems. The kit was used as described by the manufacturer, except that the final reaction volume was 25 µl instead of 50 µl. Initially, SDS PAGE gel analysis (4-20% Tris-glycine and 4-12% bis-tris gels (Invitrogen) and corresponding buffers were used interchangeably) was conducted using Coomassie Brilliant Blue G250 staining to visualize proteins. However, only the positive control was visible, so ^{35}S-methionine was used and the gels were visualized autoradiographically. The ^{35}S assays were identical to the assay above except for the addition of ^{35}S-methionine and the fact that the gel was not stained at all after electrophresis.

The templates for the reaction were the chloramphenicol acetyltransferase plasmid (pCAT) supplied by Ambion, the gene 60 and gene 32 PCR DNA templates (PCR templates are valid templates according to Ambion) and the HindIII digested Cunningham lab compatible gene 32 and gene 60 plasmids. In experiments where the Cunningham lab plasmids were used, purified 30S ribosomal subunits with the modified Shine-Dalgarno sequence were added to the reaction.

QDot binding to the 30S particle and construction of a composite gel

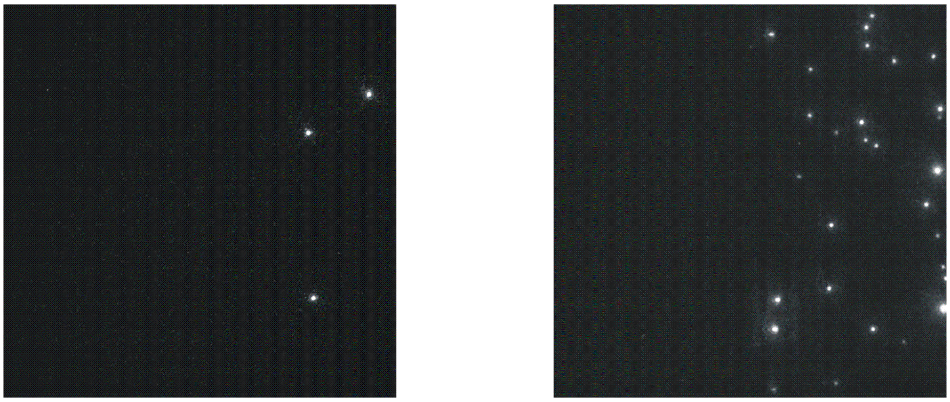
Gel based assays were used to asses the nature of the binding interaction between the QDot-DNA conjugate of the hp5 DNA to the helix 33a hairpin 5 (hp5) mutant or to a wild-type (WT) ribosome that does not carry the complementary binding site for the DNA oligonucleotide. Initial assays were run on a 0.5 X TBE, 0.8% (w/v) agarose gels, but a composite gel (see below) turned out to be more appropriate. 30S particles and QDot-DNA conjugates were incubated at 32 nM, while the 30S particles were at a concentration of 271 nM in RPE buffer (10 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 60 mM NH_4Cl) at 37 ºC for variable periods of time. The results of these assays revealed binding of the QDot-DNA conjugate to both the mutant and, unexpectedly, WT 30S ribosomal particles.

A composite gel (*231, 235*) contains the following components: 0.5% (w/v) agarose, 2.75% (w/v) acrylamide, 1% sucrose, 25 mM Tris-OAc pH 7.5, 6 mM KOAc, 2 mM $Mg(OAc)_2$, 1 mM DTT. Briefly, the agarose was mixed with 100 ml water, heated in the microwave to dissolve the agarose, and set to cool in a 37 ºC water bath. The other components were mixed together in a separate container and heated in the 37 ºC water bath. The agarose and acrylamide solutions were combined (take care not to make the agarose solution gel prematurely), 200 µl each of 50% APS and TEMED were added under stirring, and the solution was poured into a horizontal agarose gel cast and allowed to polymerize. The running buffer consisted of 25 mM Tris-OAc pH 7.5, 6 mM KOAc, and 2 mM $Mg(OAc)_2$. A non-denaturing loading buffer that does not contain loading dyes should be used, especially for assays which will be imaged on the Trio Imager.

Results

Particle Tracking

We sought to validate our QDot-DNA conjugates by establishing a pattern of specific binding of these materials using fluorescence microscopy. We first minimized surface non-specific binding by covalently modifying our microscope slides with polyethylenglycol (PEG, see materials and methods). Further reduction in surface binding was accomplished by incubation of the slide with a blocking buffer prior to introduction of the QDots to the slide. The combined action of the surface PEG and the blocking buffer serves to virtually eliminate non-specific binding to the surface (7 ± 5 spots/field of view), as seen in the QDot only panel of Appendix Figure 4. We also found that the QDot-DNA conjugates become more prone to non-specific surface binding as they age. We therefore only used freshly derivatized (< 1 month old) QDot-DNA conjugates in our assays. Appendix Figure 4 also shows the result of another negative control where we incubated 5'-biotinylated gene 60 mRNA with a mismatched QDot-DNA conjugate (32-3' 705) and again found minimized surface binding (8 ± 4 spots/field of view). To the contrary, when we combined QDot-DNA conjugates with either a small biotinylated cDNA or a matched 5'-biotinylated mRNA, we found an enrichment in surface binding (30 ± 4 spots/field of view for the cDNA and 85 ± 23 spots/field of view for the matched mRNA). These results demonstrate that QDot-DNA conjugates targeted our mRNA or cDNA constructs quite specifically.

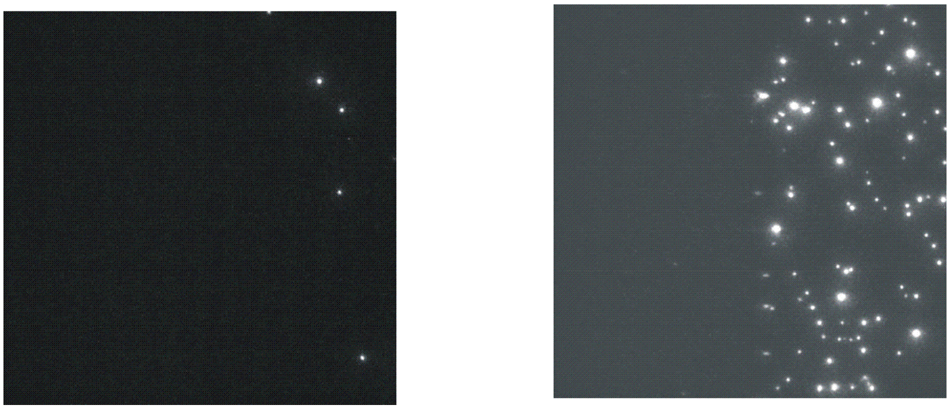


QDot only

(7 ± 5 spots)

cDNA

(30 ± 4 spots)



QDot mRNA mismatch

(8 ± 4 spots)

QDot mRNA match

(85 ± 23 spots)

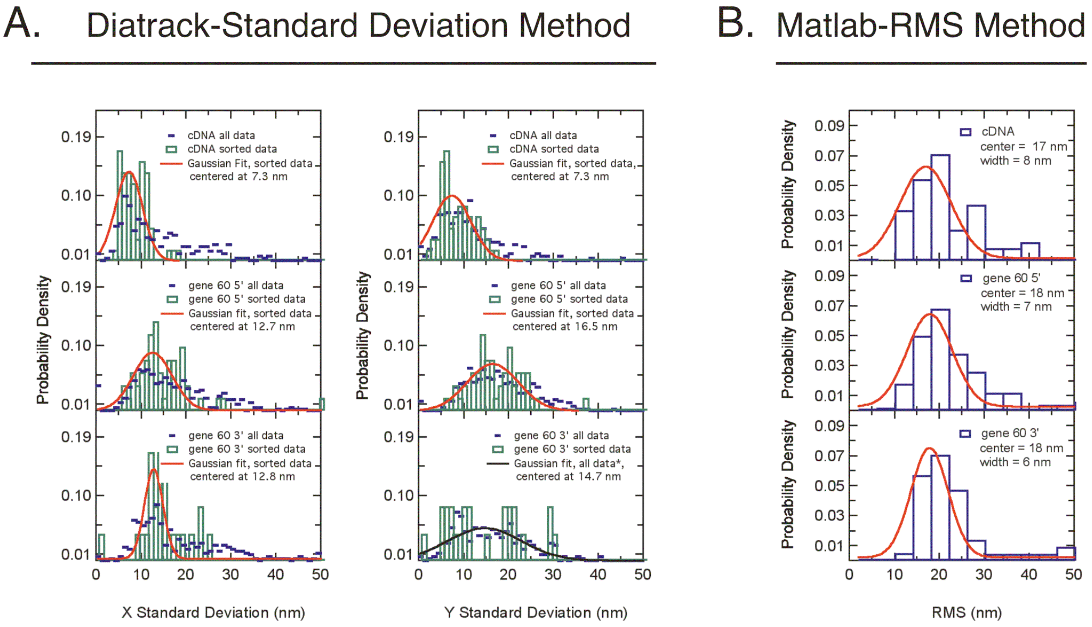
Appendix Figure 4. Imaging controls demonstrating specific QDot binding. smTIR images of QDot-DNA conjugates demonstrating that the presence of sequence matched, biotinylated binding partner is required for surface QDot binding. Negative controls are on the left column, and positive controls are on the right column.

We then conducted experiments designed to compare the observed mobilities of QDot probes tethered to the short cDNA, the 5' end of gene 60 mRNA, or the 3' end of Gene 60 mRNA. We collected two or three 60-s movies of different fields of view for each sample set. Initial data analysis was made with the commercial Diatrack program (Semasopht) followed by the calculation of the standard deviation (SD) value for each trajectory; all SD values were plotted as a histogram, and a representative value for each histogram was determined by fitting the distribution to a Gaussian function. These data were further analyzed by sorting the traces, where the only trajectories utilized originated from particles that were present in the first frame of the data set. These analyses yielded the histograms in Appendix Figure 4A.

Because of the fact that Diatrack's algorithm does not deal well with pixel boundaries (makes discrete jumps at pixel boundaries), we applied a Matlab routine to the Gaussian fitting and trajectory generation steps of the data analysis. The data were also smoothed using a rolling average and output was reported as root mean square (RMS) deviation rather than a raw SD value for each X and Y coordinate. The trajectories in the Matlab analysis were also joined together based on positional identity time linking to join the many short trajectories generated because of QDot blinking. The RMS values were plotted as histograms and fit to a Gaussian as described for the Diatrack data sets (Appendix Figure 4B).

These two different analysis methods (Diatrack and Matlab) yield very different results when comparing the different experimental constructs. The Diatrack results reveal a general trend where the cDNA has restricted mobility (~7 nm) when compared to either the gene 60 5' or 3' end (13-17 nm). This result is consistent with our initial hypothesis, in that we expect that the length of the designed tether would correlate with the observed motion. However, in the case of the Matlab analysis, this trend is no longer observed and the cDNA, gene 60 5' and 3' appear virtually identical with RMS values centered around ~18 nm. Further analysis is necessary to determine the source of these discrepancies, but possibilities include weaknesses in the Diatrack output due to discrete jumps at pixel boundaries or an influence from the rolling-average in the Matlab data set.

One possible explanation for the similarity of the gene 60 5' and 3' end mobility data is the influence of secondary structure on the effective tether length of these regions



A. Diatrack-Standard Deviation Method

B. Matlab-RMS Method

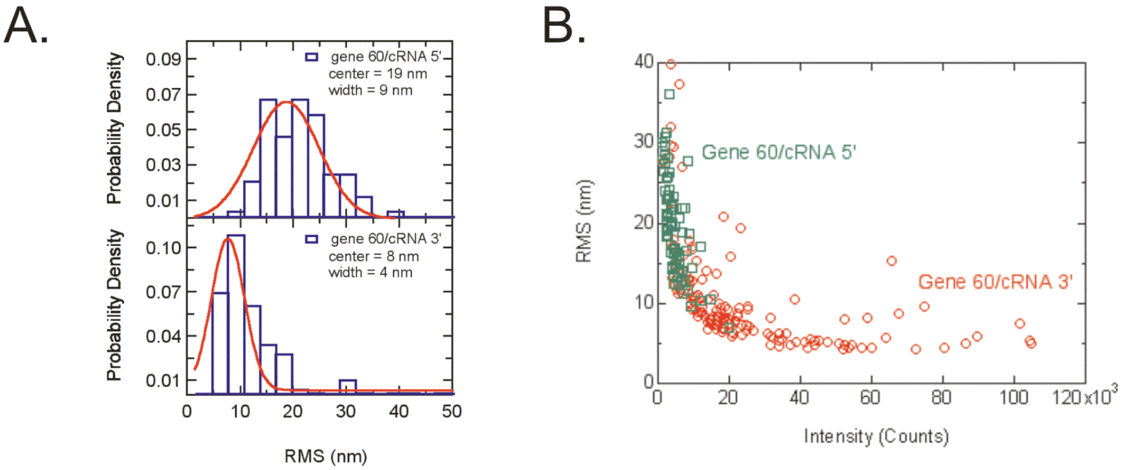
Appendix Figure 5. A comparison of observed QDot mobilities as a function of construct and analysis method. A) The cDNA, gene 60 5', and gene 60 3' data sets were analyzed using Diatrack and a basic SD based assessment of mobility which yielded representative mobilities for each sample that suggested the mRNA samples exhibited larger motions. B) The same data sets were analyzed using a lab-written Matlab routine that also performed sorting and smoothing of the trajectories. These data manifest no difference between the cDNA, gene 60 5', and gene 60 3' data sets.

of the mRNA. To test this hypothesis, we synthesized a complementary (c)RNA to the region of gene 60 between the regions of the 5' and 3' labels, with the expectation that this cRNA would base pair to the gene 60 mRNA, disrupt secondary structure present in the mRNA, and greatly increase the effective tether length of the 3' region of the mRNA. We conducted this experiment and found precisely the opposite result. For the 5' end of the gene 60/cRNA complex, we again found a similar mobility as in earlier experiments with an RMS value centered around 18 nm. For the 3' end however, we find much more restricted motion, manifested in RMS values centered around 7 nm (Appendix Figure 5A). We have not yet determined the source of this unexpected result. It should be noted that the gene 60/cRNA 3' data are characterized by a considerably higher average intensity than the corresponding 5' data set (Appendix Figure 5B). It is possible that these differences in intensity arise from differential interactions with the surface of the microscope slide, as the gene 60 3' cRNA construct should have a much larger separation from the surface. Another possibility is a simple difference between the quality of the 60 5' DNA-QDot conjugate and the 60 3' DNA-QDot conjugate. Controls using the same preparation of the 60 3' DNA-QDot conjugate should be conducted, comparing two gene 60 samples with and without the gene 60 cRNA.

In Vitro Translation

Initial *in vitro* translation assays utilized only Coomassie stains to detect the translation products. It proved impossible to detect gene 32 or gene 60 translation products with the Coomassie stain, as the background of proteins from the extract was too great. The control experiment on the left in Appendix Figure 6 shows a Coomassie stained gel where the pCAT (25 kDa) positive control and the gene 60 PCR product DNA template are shown independently, and then as a mixture. The lack of inhibition of pCAT protein expression by the gene 60 PCR product indicates that the gene 60 PCR product does not contain inhibitors. It suggests, rather, that the problem with gene 60 lies inherently in the sequence of the DNA template.

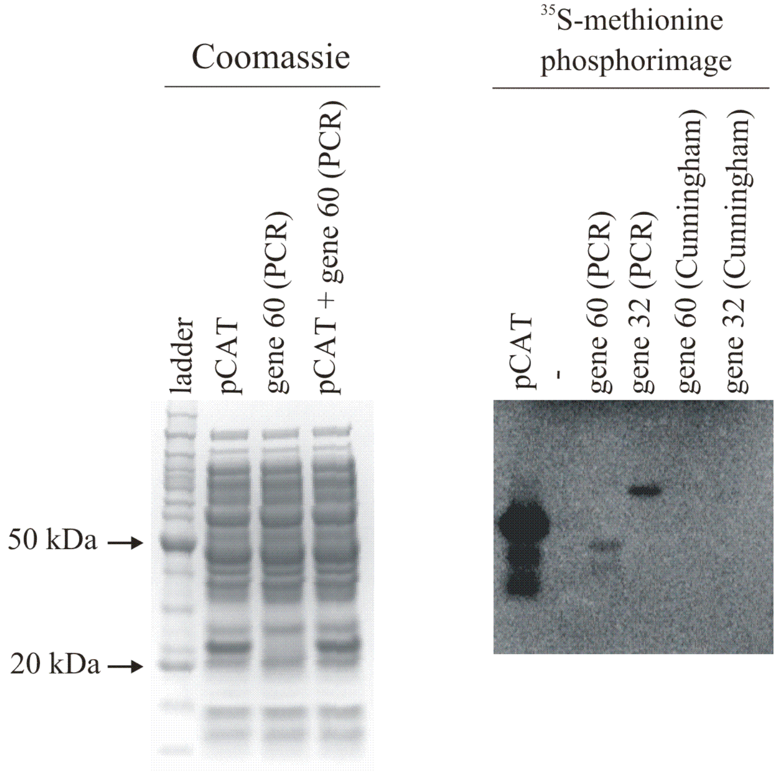
The gel on the right in Appendix Figure 6 shows *in vitro* translation assays run using ^{35}S-methionine labeling for the purposes of detection. This assay reveals that gene 32 (34 kDa) and gene 60 (18 kDa) are both expressed at low levels from the PCR



Appendix Figure 6. Mobilities and intensities of gene 60-cRNA 5' and 3' ends. A) Hybridization of cRNA to gene 60 yields a gene 60 5’ mobility that is indistinguishable from gene 60 with out the cRNA. The gene 60/cRNA 3’ mobility is drastically restricted, contrary to our hypothesis. B) Another discrepancy between the samples in (A) is the very different intensity values found for the 5’ and 3’ end samples.

templates in this system. No evidence is seen for expression from the Cunningham lab compatible gene 32 and gene 60 plasmids (N.B. assays were conducted in the presence of the Cunningham lab compatible 30S subunits).

Future directions in these studies should focus on using a positive control for the expression of the Cunningham lab compatible translation system to determine if the problem rests with the gene 60 and gene 32 plasmids or with the translation system itself.



Appendix Figure 7. *In vitro* translation with gene 32 and gene 60 samples. The left panel is a Coomassie stained gel that addresses the possibility of translational inhibition by some contaminant in the gene 60 PCR DNA template. No inhibition of pCAT translation is seen. Note the high background in the Coomassie stain, originating from the *in vitro* translation components. The panel on the right is a similar *in vitro* assay, except using ^{35}S-methionine to detect the translation products. The assay reveals that both the gene 32 and gene 60 PCR templates are translated, but at a level much lower than the pCAT positive control. The gene 32 and gene 60 plasmid DNA coding for the Cunningham lab compatible mRNAs are not translated under these conditions in the presence of the Cunningham lab mutant 30S subunits.

APPENDIX 2

INVESTIGATING THE MECHANISM OF TRANSCRIPTIONAL REGULATION IN THE MINOR *E. COLI* tRNA ARGU[6]

The recent discovery that non-coding genome regions in bacteria can regulate gene expression by presenting RNA motifs that are sensitive to metabolites (*20, 236*) represents a true paradigm shift in how we think about the roles RNA can play in the cell. These "riboswitches" are found at the untranslated 5' ends of mRNAs and act by modulating either transcription or translation (*237*). The 5' leader of ArgU, the minor *E.coli* tRNA_{UCU} is intriguing in this respect, as it has been implicated in reducing the level of the mature transcript by approximately 100-fold (*238*). ArgU is in a region of the genome derived from a defective lambdoid phage, where the phage integrase gene overlaps ArgU and the attachment site is located within the mature tRNA (*238*). Other distinguishing features of ArgU include its transcription as a discrete unit, and the fact that it is often the limiting factor in expression of proteins using the AGA codon (*239*). There are several mechanisms by which the ArgU 5' leader may potentially reduce levels of the mature transcript. These include ρ-independent transcription termination, a riboswitch like mechanism, and an HIV-Tar like mechanism.

Analysis of the *E. coli* K12 genome for the frequency of the AGA codon reveals that this codon is most widely found in prophages, transposons, and integrases (Appendix Figure 8A). We have also classified these data by % AGA codon usage and a similar

---

[6] This work has been a collaboration with Prof. Orkun Soyer (University of Trento, Italy) and Prof. Nils G. Walter. Prof. Soyer conducted bioinformatic analyses of UCU codon usage in *E. coli* and T4 bacteriophage.

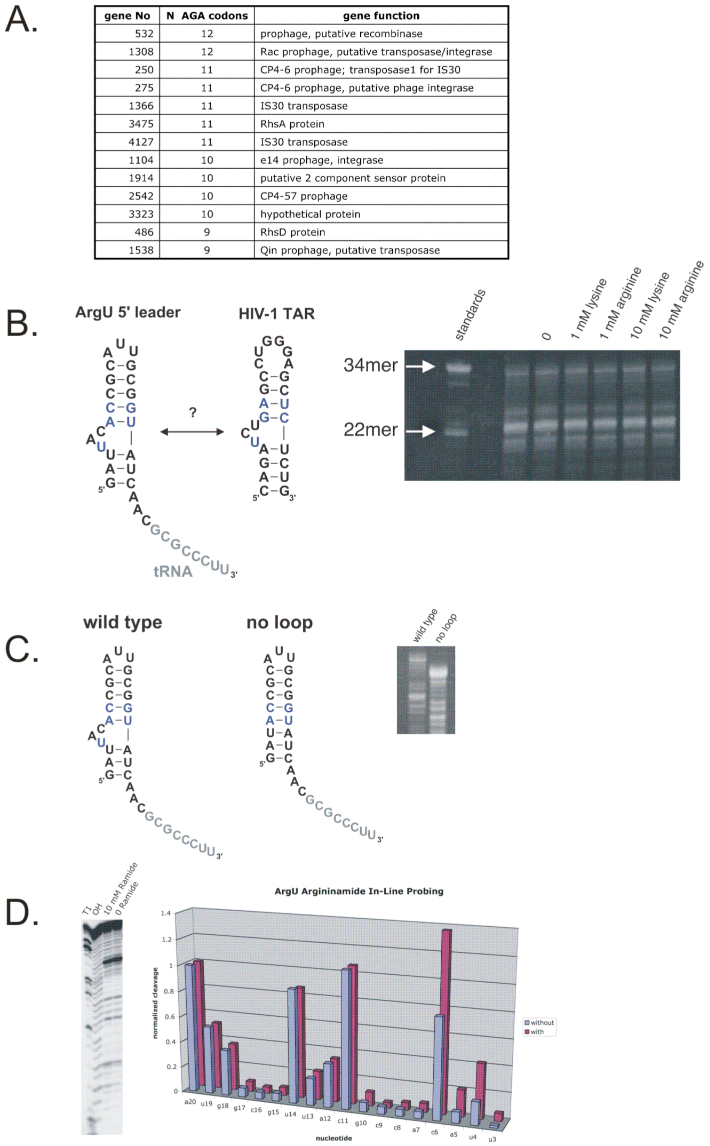
trend emerges. The viral origin of the ArgU gene and the high frequency of the AGA codon in proteins associated with lysogeny and DNA mobility are intriguing. In addition, the potential for an HIV-Tar like mechanism in transcriptional control raises interesting evolutionary issues (*240*) concerning common structures and mechanisms involved in transcriptional control of a prophage and a human retrovirus.

Our initial investigations of ArgU have been with double-stranded synthetic DNA templates and T7 RNA polymerase. Transcription of the WT 5' leader ArgU sequence yields full length (33-nucleotide) product as well as a significant amount of an approximately 22-nucleotide abortive product. In an initial experiment depicted in Appendix Figure 8B, we investigated the possibility that the 5' leader binds the amino acid arginine leading to increased production of full length transcript. This appears not to be the case.

One possibility for formation of the 22-nucleotide transcription product is degradation of the full length transcript. To address this possibility, we conducted transcriptions using [α-32P] GTP to "body label" the RNA and took time points to study the role of degradation in formation of the shorter transcript. There is a small amount of background degradation, but this is not sufficient to account for all abortive transcript formed.

Among the explanations for our observation of an abortive transcript is ρ-independent transcription termination, characterized by formation of a hairpin in the nascent transcript and destabilization of the DNA-RNA hybrid, leading to transcription termination. If this were the case, deletion of the putative internal loop in the Arg U 5' leader would stabilize the hairpin and lead to an increase in the relative amount of the abortive transcript. The results in Appendix Figure 8C show that deletion of the three-nucleotide internal loop actually leads to an increase in the amount of the full length transcript, which is the opposite result predicted if ρ-independent transcription termination were operative.

To investigate the possibility of specific binding of an arginine residue by ArgU, we used the arginine analogue argininamide and an in-line probing assay (*236*). Incubating 5'-^{32}P-labeled ArgU RNA with 10 mM argininamide in 30 mM $MgCl_2$ at 37° C for 48 h yields a cleavage pattern that, on comparison, yields information on changes in



gene No	N AGA codons	gene function
532	12	prophage, putative recombinase
1308	12	Rac prophage, putative transposase/integrase
250	11	CP4-6 prophage; transposase1 for IS30
275	11	CP4-6 prophage, putative phage integrase
1366	11	IS30 transposase
3475	11	RhsA protein
4127	11	IS30 transposase
1104	10	e14 prophage, integrase
1914	10	putative 2 component sensor protein
2542	10	CP4-57 prophage
3323	10	hypothetical protein
486	9	RhsD protein
1538	9	Qin prophage, putative transposase

Appendix Figure 8. Studies on the minor *E. coli* tRNA argU. A) Genes with the highest utilization of the UCU codon in *E. coli*. B) T7 transcriptions of the argU leader as a function of added amino acid. No effect is seen in the ratio of ~22 to ~33 nucleotide product. C) Deletion of the putative internal loop in the argU leader leads to greater accumulation of full-length transcript. D) A single inline probing assay of the argU leader with argininamide reveals differences in the region of the putative internal loop.

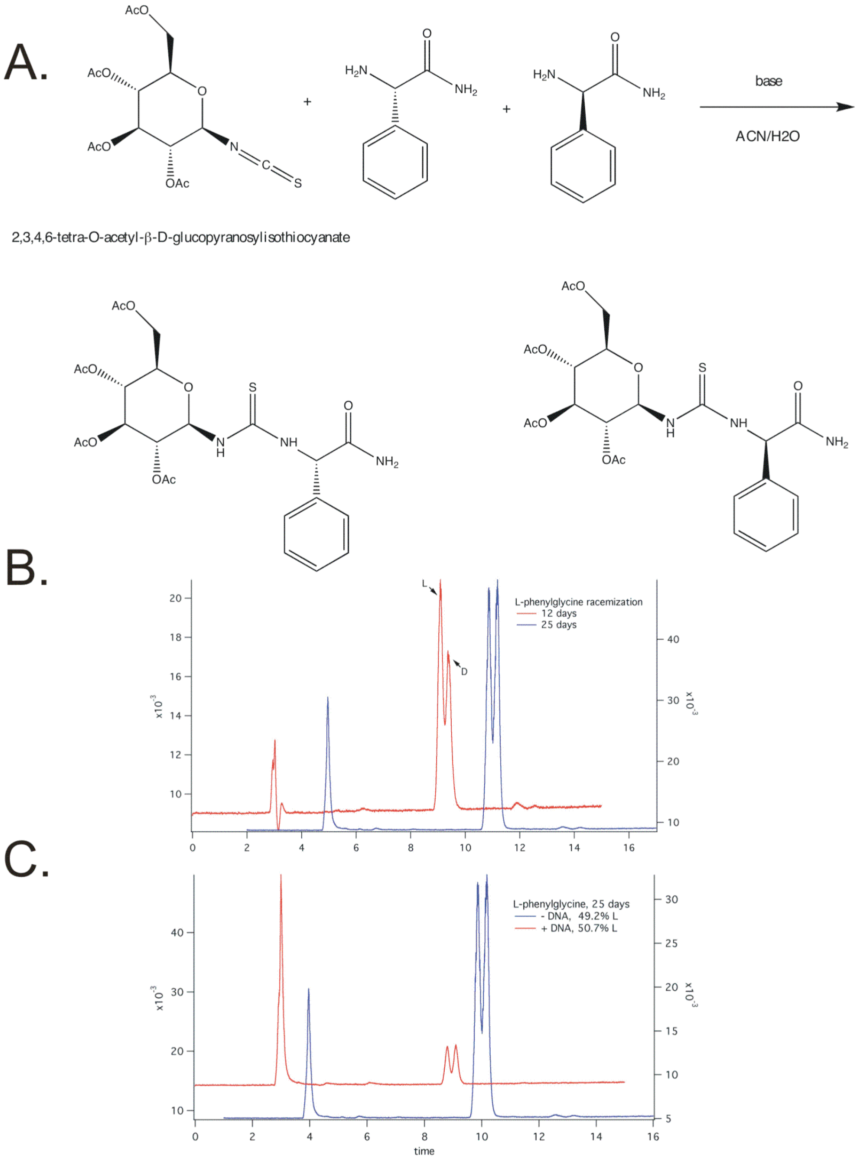
RNA structure. In very preliminary results, comparison of reactions containing no amino acid, 10 mM lysine, and 10 mM argininamide suggests there may be a change in the cleavage pattern of ArgU RNA on argininamide addition in the region of the internal loop that in HIV1-TAR is essential to specific arginine binding and in ArgU is responsible for decreasing production of full length transcript (Appendix Figure 8D). This result has not been reproduced.

These results are not conclusive regarding the nature of transcriptional control in ArgU. Future experiments should include *in vitro* transcription assays using *E. coli* RNA polymerase rather than T7 polymerase, in-line probing assays that utilize 3’ radiolabeling as well as an ArgU mutant as a negative control, and *in vivo* assays of ArgU abundance as a function of mutation in the 5’ leader as probed by Northern blot analysis.

APPENDIX 3

RACEMIZATION OF AN AMINO ACID: HYPOTHETICAL CHIRAL INDUCTION BY DNA

The origin of homochirality in biological systems continues to challenge scientists. The essentially exclusive use of levorotary (L) amino acids and dextrorotary (D) sugars in the construction of proteins, DNA, and RNA poses interesting questions regarding the origin, propagation, and purpose for the phenomenon. The RNA world hypothesis implies that existing homochiral nucleic acids may have influenced the establishment of an L-amino acid protein world. A potential mechanism for the selection of L-amino acids may be chiral induction of amino acid racemization by nucleic acids. I have not found an example of this phenomenon in the literature. Reports of chiral separations of phenylalanine by DNA (*241*) may provide evidence to support a hypothesis of non-specific chiral induction in the racemization of an amino acid by nucleic acid. To test this hypothesis, I selected L-phenylglycine amide (H-Phg-NH_2) as the amino acid derivative to racemize. The amide functionality and the phenyl group directly attached to the α-carbon will serve to stabilize the carbanion generated on deprotonation and therefore accelerate the racemization reaction. GITC (2,3,4,6-tetra-O-acetyl-β -D-glucopyranosyl isothiocyanate) was selected to derivatize the H-Phg-NH_2 to diasteriomers that can be separated and quantified by reversed phase (RP)-HPLC (*242*). An ~ 60-nucleotide synthetic DNA duplex was used as the DNA in the racemization reactions. The buffer was 250 mM K_iPO_4 pH 8, the H-Phg-NH_2 concentration was 50 mM, the duplex DNA concentration was ~ 12 μM. Samples were incubated for 42 °C



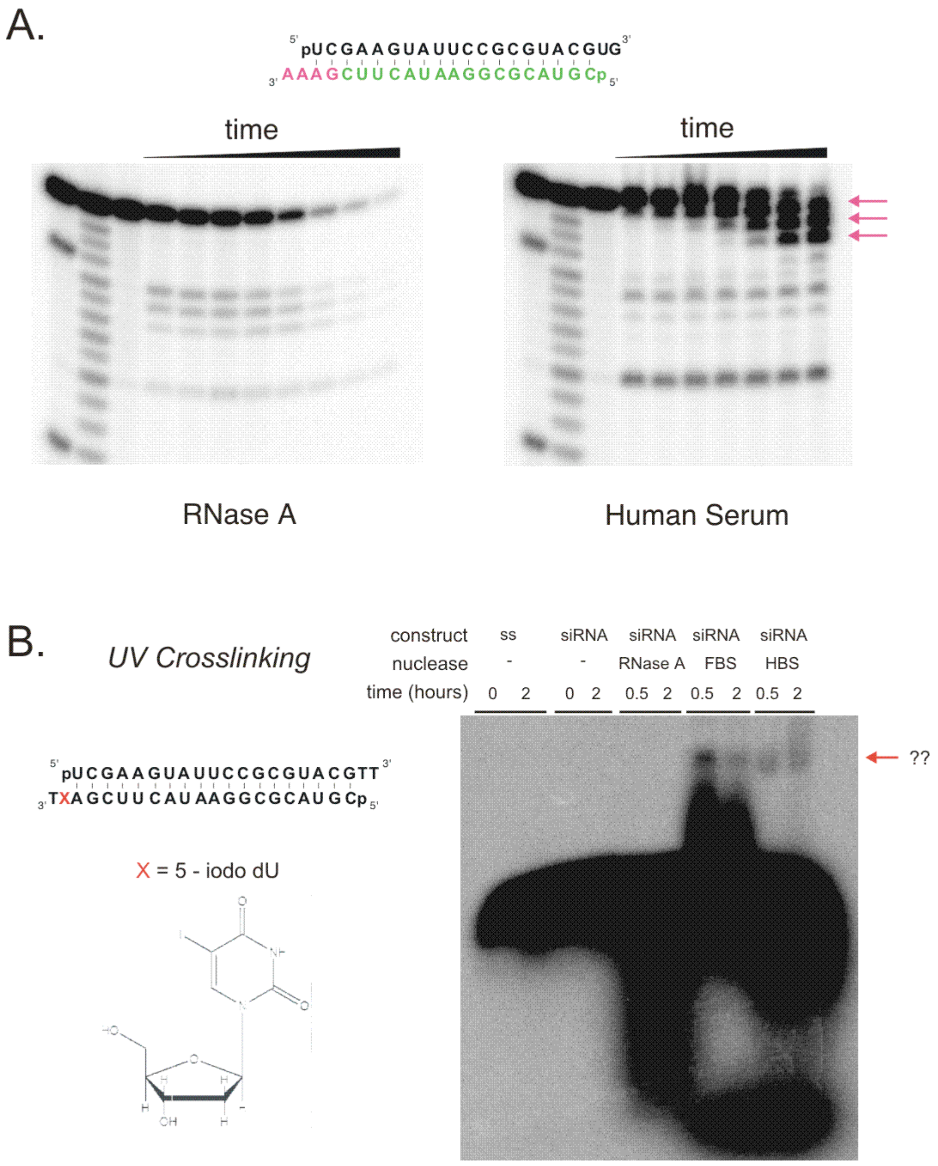
Appendix Figure 9. Racemization of H-Phg-NH_2. A) The reaction between GITC and D and L H-Phg-NH_2. B) Chromatograms of the (L)-H-Phg-NH_2 racemization reaction as a function of time. C) Chromatograms of the (L)-H-Phg-NH_2 racemization reaction plus or minus DNA.

and time points were sampled, derivatized with GITC (Appendix Figure 9A), and frozen until analysis on the HPLC. Appendix Figure 9B shows that the GITC method works for the H-Phg-NH_2 system and that the progress of the racemization reaction is monitored and approaches completion after approximately 25 days. Appendix Figure 9C compares the samples with and without DNA. The DNA containing sample is suspect because there is relatively little signal for the sample compared to the sample without DNA. The data were analyzed and revealed that for this particular data set there is a slightly higher percentage of the L amino acid in the DNA containing sample. While this is the result that the hypothesis would predict, no conclusions can be made because there is certainly large error due to the incomplete resolution of the two diasteriomer peaks and potential uncertainties regarding the completion of the racemization reaction. Future experiments would need to address why the DNA containing sample does not exhibit a strong signal when compared to the other sample. An internal standard should also be used to decrease error in the calculation of the relative peak areas, and a different HPLC system that does not use dedicated independent pumps for each solvent should be employed so the gradient and flow could be smoother and perhaps allow for better resolution of the two diasteriomer peaks.

APPENDIX 4

PHOTOCROSSLINKING AN siRNA TO A COMPONENT OF BLOOD SERUM

Data from siRNA degradation experiments indicated that there is nuclease activity in serum that is not consistent with the pyrimidine preference of the known human pancreatic type nucleases (Appendix Figure 10A). Based on these data, an siRNA was designed that incorporated a photoactived nucleotide analog, 5-iodo uracil. Early assays showed the presence of a UV crosslink between the siRNA and a component of human serum (Appendix Figure 10B). Affinity purification of the crosslinked species was undertaken, but no component above the background of the negative control was detected. More stringent wash steps are required to reduce the non-specific protein background for detection of potential crosslinked species.



Appendix Figure 10. siRNA degradation by and crosslinking to a component of blood serum. A) Degradation assays showing the 3' end of the passenger strand degraded by purified RNase A and human blood serum. The serum gel reveals cleavage after purine residues, reactivity that is not seen in the purified RNase A. B) The crosslinking strategy and an early gel showing evidence for siRNA crosslinking to a component of blood serum

APPENDIX 5

MATLAB COMPUTER CODE FOR THE ANALYSIS OF THE PARTICLE TRACKING DATA IN APENDIX A[7]

Program “Tony_simpleGfit_tif_070503_findpeaks_v1p3.m”

```
% 070517 -- This version (v1p3) thresholds the peaks, selects peaks that are within
% the fitting box, slects peaks that are more than N pixels away from each
% other and displays the the peaks for visual inspection.

%pbfitframe centroid-determination algorithm.  No distribution without
%permission from Prof. Sunney Xie, Harvard University.  Contact:
%xie@chemistry.harvard.edu.  Technical inquiries to Paul Blainey,
%blainey@fas.harvard.edu.  February 28, 2005

%sample command line: pbfitframe('sampleim.txt',200,3,1400)

   %g = filename (including extension),
   %c = pixel size (in nm),
   %number_iterations = number of fitting cycles,
   %fb_dim = fit box dimension (nm)

%This function fits a signal in an image with a 2-dimensional gaussian,
%reporting the fit parameters to the file xy.txt (description of output
%data below)

%function asdf = pbopen(g,c,number_iterations,fb_dim)

%Modified by AJM, 060223

clear

%Input data file
%g = 'C:\Documents and Settings\jhoerter\Desktop\work\oligodt705_2_16bit.TIF'
```

---

[7] This code was written in Matlab by Dr. Anthony Manzo, a postdoctoral researcher in the Walter lab.

```
%g = 'C:\Documents and Settings\amanzo\Desktop\Work\060308_beads_10s_16bit.TIF'
%g = 'W:\Tony\060317\060317_spider_withZinc_mcp75_1.6x_5sec_15sec_1hour.TIF'
%g =
'X:\amanzo\060416\060416_spiders+705qd_nozinc_2si_0w_2.5x_box_gain1_mcp75_600frames.TIF'
%g = 'X:\amanzo\Kurt\060626\qdot1s235g-int20s-60f.tif'
%g = 'D:\Ribosome Project\pz\Matlab_test_file\streptavidin705dot100ms1000f.tif';
g = 'E:\Ryan\primary\cDNA705_MCP60_1pt6x_PEG+BB_2.TIF';
%g = 'E:\Tony\070215\video1.TIF';
%g =
'X:\amanzo\060317Walter\060317\060317_spider_withZinc_mcp75_1.6x_5sec_15sec_1hour_16bit.TIF';
%g =
'E:\Tony\070519\070519_60x_mcp70_spiderDNA_705spider2w4mMZn_1mg1_1p6x_1si_4sd_360frames
_position2.TIF';

thresh = 125;
c=132; % pixel length (nm)
number_iterations=25; %number of iterations for search routine
fb_dim=2112; %fit box dimension (nm)
fb_dim_pix = round(fb_dim/c);
hold off; format short;

%Set inital parameter values here
guessintensity = 2000;
guesswidth = 160/c;
guessoffset = 500;

nametoopen = g;

% Read first three frames and average them
[A1]=imread(nametoopen, 1); %read data file, 1st image (Reads 16 bit .tif)
[A2]=imread(nametoopen, 2); %read data file, 2nd image (Reads 16 bit .tif)
[A3]=imread(nametoopen, 3); %read data file, 3rd image (Reads 16 bit .tif)

A = (A1+A2+A3)/3;

% mirror image across x-axis
Anew = zeros(512,512);
for ii = 1:512;
   for jj = 1:512;
      Anew(ii,jj) = A(513-ii,jj);
   end;
end;
A = Anew;

%[A]=imread(nametoopen, 1); %read data file, 1st image (Reads 16 bit .tif)

imnum=1; %reference for current image number
info = imfinfo(nametoopen); %Get file information
d = size(info);

%xsize = info.Width;
%ysize = info.Height;
Axy=size(A);
xsize = Axy(2);
ysize = Axy(1);
%xsize = info(1,1).Width;
```

```
%ysize = info(1,1).Height;
xdim = xsize;
ydim = ysize;

amax=max(A);
amin=min(A);
disp(amax);
disp(amin);

imindexval= d(1,2); %Number of images (frames)
%imindexval = 10;

disp('Number of frmaes is: ');
disp(imindexval);
disp('The x and y length in pixels is:');
disp(xsize);
disp(ysize);

NTOT=zeros(1,imindexval);

%load -ascii nametoopen;

    hold off;
    contour (A(:,:,1),20); axis equal; % Show image of 1st frame
    %Adjust the last number to show more or less detail in contour map.

hold on;
%pause;   %070507

idata = A;

%BW1 = edge(A,'sobel');
%BW2 = edge(A,'canny');
%imshow(BW1)
%figure, imshow(BW2)

%idata = BW1;


% Blob analysis routine

% variables
%dfm = 1;  % deviations from mean used in thereshold determination
%mf = 2; % multiplicative factor of mean used in threshold determination
min_bs = 1; % minimum bead size
max_bs = 20; % maximum bead size
%thresh = 400;
peaks = [];

%determine threshold
s = size(idata);
border = [idata(1:s(1),1)' idata(1:s(1),s(2))' idata(1,2:(s(2)-1)) idata(s(1),2:(s(2)-1))];
%bavg = 100; %mean(border)
%bavg = mean(border);
bavg = mean(mean(idata(257:512,1:512)));
bstd = std(std(idata(257:512,1:512)));
```

```
%bstd = std(border);
%thresh = mf*bavg + dfm*bstd;
%thresh = bavg;
thresh = bavg+3*bstd;
%threshold image
tdata = (idata > thresh); % This places a 1 in all locations that have values > thresh.

%disp('check tdata'); pause;

b = 0;
while any (tdata(:)) % Do while any tdata value is > 0.
   x = find (tdata > 0);  % This makes a single row of index values for tdata = 1.
   %disp('check x'); pause;
   [bead,IDX] = BWSELECT(tdata,ceil(x(1)/s(1)),mod(x(1)-1,s(1))+1,4);
   %disp('check bead'); pause;
      %The awkward expression for the row is necessary to because
      %indexing is for 1 to s(1), not 0 to s(1)-1
   tdata(IDX) = 0; % Set tdata corresponding to IDX locations = 0.
   %disp(idata(IDX));
   %disp('check idata(IDX) output'); pause;
   %if (length (IDX) >= min_bs)
   if ((length (IDX) >= min_bs) && (length(IDX) <= max_bs))
      b = b + 1;
      avgx = sum([1:s(2)].*sum(bead))/sum(sum(bead));
      avgy = sum([1:s(1)]'.*sum(bead,2))/sum(sum(bead));
      %sdx = sqrt(sum(vx.*(abs(x-cx).^2))/sum(sum(bead)));
      peaks(b,1) = round (avgx);
      peaks(b,2) = round (avgy);
      peaks(b,3) = length(IDX);

      %[sizey sizex] = size(m);
      %vx = sum(m); %Note that m is a matrix and vx will be vector
      %vx = vx.*(vx>0); % .* is element by element multiplication
      %x = [1:sizex];
      %cx = sum(vx.*x)/sum(vx);
      %sx = sqrt(sum(vx.*(abs(x-cx).^2))/sum(vx));

   end
end

npeaks = b;

%scatter(X,Y,S,C)
for ii = 1:npeaks
   scatter(peaks(ii,1),peaks(ii,2),10,'sr');
end

disp('done with blob analysis');

%pause;     %070507

% Sort peaks and delete those with a fitting box extending betond the CCD image.
halfbox = round(fb_dim_pix/2);
newpeaks = [0 0 0];
for ii = 1:npeaks
   if (peaks(ii,1) > (halfbox) && peaks(ii,1) < 512-halfbox)
```

```
        if (peaks(ii,2) > (halfbox) && peaks(ii,2) < 512-halfbox)
            %if (peaks(ii,3) >= 10)   % This if statement checks how big the blob is and makes a selection.
                newpeaks = [newpeaks;peaks(ii,:)];
            %end
        end
    end
end

newpeaks(1,:) = []; % Delete first row of zeros
peaks = newpeaks;
[px,py] = size(peaks);

npeaks = px;

%Here I plot another scatter plot with the peaks selected for the proper fit box.
%scatter(X,Y,S,C)
for ii = 1:npeaks
    scatter(peaks(ii,1),peaks(ii,2),10,'sg');
end


%pause;  %070509


% List all the peak positions and sort them
% Discard any peaks that are within N pixels of each other.

peaktest = peaks;

peakcheck(1:npeaks)=1;
kk = npeaks;
for ii = 1:npeaks
    testpeak1 = peaks(ii,:);
    for jj = 1 :npeaks
        if (jj == ii)
            % Do nothing
        else

                %testpeak2 = peaktest(jj,:);
                testpeak2 = peaks(jj,:);
                %if (testpeak2 == [0,0])
                    % Do nothing
                %else
                peakdist = sqrt((testpeak2(1,1)-testpeak1(1,1))^2+(testpeak2(1,2)-testpeak1(1,2))^2);
                disp(peakdist);
                if (peakdist < 14)
                    peakcheck(ii)=0;

                    %peaktest(jj,:)=[0,0];
                     peaks(ii,:) = [0,0,0];
                     peaks(jj,:) = [0,0,0];   % Both peaks of the comparison should be marked and deleted.

                end


                %end     % End for the testpeak2 = [0,0] if statement.
```

```
        end

    end
end

peaks3 = peaks;

peaks2 = [0,0,0];



% Get rid of all peaks that fail test.

jj = 0;
for ii = 1:npeaks

if (peaks3(ii,:) ~= [0,0,0])
    peaks2 = [peaks2;peaks3(ii,:)];
    jj = jj + 1;
end

end


peaks2(1,:)=[]; % delete first row of zeros;


peaks = peaks2;
npeaks = jj;


% Check the std of peaks
L=19;
Lm=L-1;
BW = zeros(npeaks,L,L);
for ii = 1:npeaks

BW(ii,:,:) = idata( peaks(ii,2)-Lm/2 : peaks(ii,2)+Lm/2, peaks(ii,1)-Lm/2:peaks(ii,1)+Lm/2 );

%BW2 = bwselect(BW,6,6,4);
%[sigm,sIDX] = BWSELECT( idata(peaks(ii,1)-5,peaks(ii,1)+5) );

%disp('check BW2'); pause;

bwc(:,:)=BW(ii,:,:);

m = bwc;
%[sizey sizex] = size(bwc);
sizex = L;
sizey = L;
vx = sum(m);
vy = sum(m');

vx = vx.*(vx>0);
vy = vy.*(vy>0);
```

```
%rx=round((sizex/2)-1);%%
%ry=round((sizey/2)-1);%%

x = [1:sizex];
y = [1:sizey];

%x = [-rx:rx];
%y = [-ry:ry];

cx = sum(vx.*x)/sum(vx);
cy = sum(vy.*y)/sum(vy);

%cx = 7;
%cy = 7;

sx = sqrt(sum(vx.*(abs(x-cx).^2))/sum(vx));
sy = sqrt(sum(vy.*(abs(y-cy).^2))/sum(vy));

%disp(cx);
%disp(cy);
%disp(sx);
%disp(sy);


%sparam = (sy/sx);

sparam = sx/sy;

disp('sparam 1 and 2 are');
disp(sparam);


vvx=0;
vvy=0;
ccx=0;
ccy=0;

%for ll = 1:13
  for mm = 1:L
    vvx = vvx + m(mm,mm);
    vvy = vvy + m(mm,L-mm+1);
  end
%end

%for ll = 1:13
  for mm = 1:L
    ccx = ccx + mm*m(mm,mm);
    ccy = ccy + (L-mm+1)*m(mm,L-mm+1);
  end
%end

ccx = ccx / vvx;
ccy = ccy/ vvy;

%disp(ccx);
```

```
%disp(ccy);

%sp2 = ccx-ccy;
sp3=ccx/ccy;

%disp(sp2);
disp(sp3);

surf(m);
view(-15,40)
disp('check');

if(abs(sp3-1) > 0.04)
    peaks(ii,:)= [0,0,0];
end

if(abs(sparam-1) > 0.04)
    peaks(ii,:) = [0,0,0];
end

end


%Get rid of selected peaks
peaks3 = peaks;
peaks2 = [0,0,0];

% Get rid of all peaks that fail test.

jj = 0;
for ii = 1:npeaks

if (peaks3(ii,:) ~= [0,0,0])
    peaks2 = [peaks2;peaks3(ii,:)];
    jj = jj + 1;
end

end

peaks2(1,:)=[]; % delete first row of zeros;

peaks = peaks2;
npeaks = jj;


figure(1);
contour (A(:,:,1),20); axis equal; % Show image of 1st frame
hold on;
%scatter(X,Y,S,C)
for ii = 1:npeaks
    scatter(peaks(ii,1),peaks(ii,2),20,'sm');
end

hold off;
```

```
% Now show 3D image of locations (except if it is too close to boundary of
% CCD (the fitting box will go over boundary and there will be an error).


%fb_dim = 2000;
%c = 132;
for peaknum = 1:npeaks

% get X and Y from peaks array
      X = peaks(peaknum,1);
      Y = peaks(peaknum,2);
      %XY = peaks(peaknum,3); % = length(IDX);



      xx = round(X);
                     yy = round(Y);
      %xx=350;
      %yy=300;

                     fitboxdimension = fb_dim/c;
      fb_dim_pix = round(fb_dim/c);
                     R = fitboxdimension/2;
                     r = round(R);


%
      % Sort peaks and delete those with a fitting box extending betond the CCD image.
halfbox = round(fb_dim_pix/2);
%newpeaks = [0 0 0];
%for ii = 1:npeaks
   if (peaks(peaknum,1) > (halfbox) && peaks(peaknum,1) < 512-halfbox)
      if (peaks(peaknum,2) > (halfbox) && peaks(peaknum,2) < 512-halfbox)
         %if (peaks(ii,3) >= 10)   % This if statement checks how big the blob is and makes a selection.
            %newpeaks = [newpeaks;peaks(ii,:)];
         %end


                     %bestcoeffs = zeros(1,5);

      %B = A((yy-r):(yy+r),(xx-r):(xx+r));
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r))); %Forced double to satisfy fitting functions
      %contour (B); axis equal;

      [x,y] = meshgrid(-r:r,-r:r);
      figure(2);
      hold off;
      surf(x,y,B(:,:));
      %contour (B); axis equal;


[cx,cy,sx,sy] = centerofmass(B);
disp(peaknum);
disp(xx); disp(yy);
disp('sx and sy are '); disp(sx); disp(sy);
```

```
pause;
      end

   end

end


disp('Done displaying 3D images.');
pause;

                     %disp('--->  select x0, y0')

      %[X,Y] = ginput (1); %Get coordinates selected by mouse


      % Here I will select the peak.  The ROI is determined by peak and
      % fit_dim (fit box dimension set above).

for peaknum = 1:npeaks
%for peaknum = 4:4

   disp('Initiating peak '); disp(peaknum);
   %pause;     %070507

  hold off;
      %contour (A(:,:,1),10); axis equal; % Show image of 1st frame.

                     %disp('--->  select x0, y0')

      %[X,Y] = ginput (1); %Get coordinates selected by mouse

% get X and Y from peaks array
      X = peaks(peaknum,1);
      Y = peaks(peaknum,2);
      XY = peaks(peaknum,3); % = length(IDX);


      xx = round(X);
                     yy = round(Y);
      %xx=350;
      %yy=300;

                     fitboxdimension = fb_dim/c;
                     R = fitboxdimension/2;
                     r = round(R);

                     %bestcoeffs = zeros(1,5);

      %B = A((yy-r):(yy+r),(xx-r):(xx+r));
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r))); %Forced double to satisfy fitting functions
      %contour (B); axis equal;

      [x,y] = meshgrid(-r:r,-r:r);
      surf(x,y,B(:,:));
      %contour (B); axis equal;
```

```
% Comment out all input queries
%{

        [X,Y,button] = ginput(1);
        disp([button]);
 while ([button] ~= 115);
        disp('In loop');
        [X,Y,button] = ginput(1);
        method = [button];

switch lower(method);
     case {30}; % up arrow key
      disp('up');
      yy=yy-1;
      if ((yy-r) < 1) %Check to see if outside image dimensions.
          yy=yy+1;
      end
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
      surf(x,y,B(:,:));
   case {114}; % r key
      disp('up');
      yy=yy+1;
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
      surf(x,y,B(:,:));
   case {29}; % right arrow key
      disp('right');
      xx=xx-1;
      if ((xx-r) < 1) %Check to see if outside image dimensions.
          xx=xx+1;
      end
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
      surf(x,y,B(:,:));
   case {100}; % d key
      disp('right');
      xx=xx-1;
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
      surf(x,y,B(:,:));
   case {28}; % left arrow key
      disp('left');
      xx=xx+1;
      if ((xx+r) > xdim) %Check to see if outside image dimensions.
          xx=xx-1;
      end
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
      surf(x,y,B(:,:));
   case {102}; % f key
      disp('left');
      xx=xx+1;
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
      surf(x,y,B(:,:));
   case  {31}; % down arrow key
      disp('down');
      yy=yy+1;
      if ((yy+r) > ydim) %Check to see if outside image dimensions.
          yy=yy-1;
```

```
   end
   B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
   surf(x,y,B(:,:));
 case  {99}; % c key
   disp('down');
   yy=yy-1;
   B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
   surf(x,y,B(:,:));

   case {105}; % i key
   disp('up 10');
   yy=yy+10;
   if ((yy-r) < 1) %Check to see if outside image dimensions.
      yy=yy-10;
   end
   B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
   surf(x,y,B(:,:));
 case {106}; % j key
   disp('left 10');
   xx=xx-10;
   if ((xx+r) > xdim) %Check to see if outside image dimensions.
      xx=xx+10;
   end
   B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
   surf(x,y,B(:,:));
 case {107}; % k key
   disp('right 10');
   xx=xx+10;
   B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
   surf(x,y,B(:,:));
 case  {109}; % m key
   disp('down 10');
   yy=yy-10;
   if ((yy+r) > ydim) %Check to see if outside image dimensions.
      yy=yy+10;
   end
   B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
   surf(x,y,B(:,:));

case  {43}; % - key
  disp([button]);
  fitboxdimension = fitboxdimension-2;
  if(fitboxdimension <= 2)
     fitboxdimension=2;
  end
  R = fitboxdimension/2;
  r = round(R);
  [x,y] = meshgrid(-r:r,-r:r);
  B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
  surf(x,y,B(:,:));
case  {45}; % + key
  disp([button]);
  fitboxdimension = fitboxdimension+2;
  R = fitboxdimension/2;
  r = round(R);
  [x,y] = meshgrid(-r:r,-r:r);
```

```
  B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
  surf(x,y,B(:,:));

case (46); % right mouse button

    disp([button]);
    disp([X,Y]);

    oldfitboxdimension=fitboxdimension;

    fitboxdimension = fitboxdimension+4;

   if (fitboxdimension >= 44)
       fitboxdimension = 8;
   end

    R = fitboxdimension/2;
    r = round(R);

      if(yy-r < 0)
         fitboxdimension=oldfitboxdimension;
      elseif(yy+r > 512)
         fitboxdimension=oldfitboxdimension;
      elseif(xx-r < 0)
         fitboxdimension=oldfitboxdimension;
      elseif(xx+r > 512)
         fitboxdimension=oldfitboxdimension;
      end
    R = fitboxdimension/2;
    r = round(R);

     disp(r);
  [x,y] = meshgrid(-r:r,-r:r);
  disp([xx,yy]);
  disp([x,y]);
  B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
  surf(x,y,B(:,:));

  case{1}; % left mouse button
      disp([button]);
      xold = xx;
      yold = yy;

      if (round(X) < 0)
      xx = xx - 4;
      end
      if( round(X) > 0)
      xx = xx + 4;
      end
      if(round(Y) < 0)
      yy = yy - 4;
      end
      if(round(Y) > 0)
                          yy = yy + 4;
      end
```

```
        if(yy-r < 0)
            yy=yold;
        elseif(yy+r > 512)
            yy=yold;
        elseif(xx-r < 0)
            xx=xold;
        elseif(xx+r > 512)
            xx=xold;
        end
      B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
    surf(x,y,B(:,:));


      case  {110}; % n key

          imnum=imnum+1;
        if(imnum > imindexval)
            imnum=imnum-1;
        end
        [A]=imread(nametoopen, imnum)
        R = fitboxdimension/2;
        r = round(R);
        [x,y] = meshgrid(-r:r,-r:r);
        B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
        surf(x,y,B(:,:));

    case  {112}; % p key

          imnum=imnum-1;
        if(imnum == 0)
            imnum=imnum+1;
        end
        [A]=imread(nametoopen, imnum)
        R = fitboxdimension/2;
        r = round(R);
        [x,y] = meshgrid(-r:r,-r:r);
        B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
        surf(x,y,B(:,:));

   otherwise;
      disp('Unknown key.');
      disp([button]);
   end;    %End of switch


    %   B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));
    %   surf(x,y,B(:,:));
      %contour (B); axis equal;
       disp([xx,yy]);


        end;  %End of while

%}
% End of input selection code
```

```
%R = fitboxdimension/2;
%r = round(R);
%[x,y] = meshgrid(-r:r,-r:r);
%B = double(A((yy-r):(yy+r),(xx-r):(xx+r)));

%C=cool(2000);
surfc(x,y,B(:,:));
%surfc(x,y,B);
%colormap cool;
%axis([-x x -y y -1000 4000])

%surf(x,y,B(:,:),'FaceColor','yellow','EdgeColor','none');
          %camlight left; lighting gouraud
          %view(-15,40)



%contour(B);
%contour (A(:,:,1),4); axis equal; % Show image of 1st frame
%pause;    %070507

for imindex = 1:imindexval;
[A]=imread(nametoopen, imindex);

% mirror image across x-axis
Anew = zeros(512,512);
for ii = 1:512;
  for jj = 1:512;
    Anew(ii,jj) = A(513-ii,jj);
  end;
end;
A = Anew;

%close;

hold off;   %put back on later
          %xx = round(X);
          %yy = round(Y);
          %xx = 61;
          %yy = 276;


          %fitboxdimension = fb_dim/c;
          %R = fitboxdimension/2;
          %r = round(R);

%xx=xx+r;
%yy=yy+r;

          bestcoeffs = zeros(1,5);

  %B = A((yy-r):(yy+r),(xx-r):(xx+r));
  B = double(A((yy-r):(yy+r),(xx-r):(xx+r))); %Forced double to satisfy fitting functions
  contour (B); axis equal;

%This sets initial guess to the previous frames fitted values
```

```
     if (imindex == 1);
        guessx = 0;
        guessy = 0;
     else
     guessx = xyval(imindex-1,1);
     guessy = xyval(imindex-1,2);
     end;

                         [x,y] = meshgrid(-r:r,-r:r);
                         coeffs = zeros(number_iterations,5);

                         for iteration_number = 1:number_iterations;

                                   best_offstht = fminsearch('pbgaussoffstht',[guessintensity
guessoffset],[],x,y,B(:,:),guessx,guessy,guesswidth);
                                   coeffA = zeros(1,2); coeffA(1,1) = best_offstht(1,1); coeffA(1,2) =
best_offstht(1,2);
                                   guessintensity = coeffA(1,1); guessoffset = coeffA(1,2);

                                   best_xy = fminsearch('pbgaussxy',[guessx
guessy],[],x,y,B(:,:),guessintensity,guesswidth,guessoffset);
                                   coeffB = zeros(1,2); coeffB(1,1) = best_xy(1,1); coeffB(1,2) =
best_xy(1,2);
                                   guessx = coeffB(1,1); guessy = coeffB(1,2);

                                   best_width =
fminsearch('pbgausswidth',[guesswidth],[],x,y,B(:,:),guessintensity,guessx,guessy,guessoffset);
                                   coeffC = zeros(1,1); coeffC(1,1) = best_width(1,1);    guesswidth =
coeffC(1,1);

                                   coeffs(iteration_number,1) = guessintensity;
                                   coeffs(iteration_number,2) = guessx;
                                   coeffs(iteration_number,3) = guessy;
                                   coeffs(iteration_number,4) = guesswidth;
                                   coeffs(iteration_number,5) = guessoffset;

                                   coeffs(iteration_number,1) = 0.001*coeffs(iteration_number,1);
                                   coeffs(iteration_number,2) = c*coeffs(iteration_number,2);
                                   coeffs(iteration_number,3) = c*coeffs(iteration_number,3);
                                   coeffs(iteration_number,4) = c*coeffs(iteration_number,4);
                                   coeffs(iteration_number,5) = 0.001*coeffs(iteration_number,5);
                         end

                         bestcoeffs = coeffs(number_iterations,:);
        origbestcoeffs=bestcoeffs;
                         bestcoeffs(1,1) = 1000 * bestcoeffs(1,1);
                         bestcoeffs(1,2) = (1/c) * bestcoeffs(1,2);
                         bestcoeffs(1,3) = (1/c) * bestcoeffs(1,3);
                         bestcoeffs(1,4) = (1/c) * bestcoeffs(1,4);
                         bestcoeffs(1,5) = 1000 * bestcoeffs(1,5);


        xyval(imindex,1) = guessx;
        xyval(imindex,2) = guessy;
        Ival(imindex,3)  = guessintensity;
        width(imindex,4) = guesswidth;
```

```
%Blocked out all intermediate graphing 060218
            surf(x,y,B(:,:));
axis equal;
            [z1,z2] = meshgrid(-r:r,-r:r);
            B_fit = bestcoeffs(1)*exp(-0.5*(((z1-bestcoeffs(2)).^2  + (z2-bestcoeffs(3)).^2)/(bestcoeffs(4).^2)  )   ) + bestcoeffs(5) ;
            hold off;
            surf(x,y,B(:,:));
%axis equal;
            camlight left; lighting gouraud
            hidden off;
            hold on;
            surf(z1,z2,B_fit,'FaceColor','yellow','EdgeColor','none');
%axis equal;
            camlight left; lighting gouraud
            view(-15,40)
            pause(0.005);


%keyboard;
%[X,Y,button] = ginput(1);

partsum=sum(sum(B_fit-bestcoeffs(5)));
NTOT(imindex)=partsum;

chi=0;
for cx=1:(2*r+1);
   for cy=1:(2*r+1);
chi=chi+((B(cx,cy)-B_fit(cx,cy))^2)/(B_fit(cx,cy));
   end;
end;
chi=chi/((2*r+1)*(2*r+1)-5);
%chi=chi/5.0;

disp(chi);
%pause;

            absolute_bestcoeffs = zeros(1,7);
            absolute_bestcoeffs(1,1:5) = bestcoeffs;
            absolute_bestcoeffs(1,2) = (coeffs(1,2) + (c * xx));
            absolute_bestcoeffs(1,3) = (coeffs(1,3) + (c * yy));
            absolute_bestcoeffs(1,4) = coeffs(1,4);
            absolute_bestcoeffs(1,6) = (c * xx);
            absolute_bestcoeffs(1,7) = (c * yy);

%Blocking all intermediate disp commands 060218
            %display_coeffs = zeros(1,7);
            %display_coeffs(1,1:5) = coeffs(number_iterations,1:5);
            %display_coeffs(1,6:7) = (absolute_bestcoeffs(1,6:7)/1000);


if (imindex == 1);
%xpixel = round(absolute_bestcoeffs(1,2)/c);
%ypixel = round(absolute_bestcoeffs(1,3)/c);
```

```
        xpixel = round(xx+origbestcoeffs(1,2)/c);
                        %ypixel = 512-round(yy+origbestcoeffs(1,3)/c);
        ypixel = round(yy+origbestcoeffs(1,3)/c);
        end;



        xy = absolute_bestcoeffs(1,1:5);
        %xywindex = absolut_bestcoeffs(1,number_iterations,1:5);

        %disp(' intensity    x      y      FWHM    offset  ')
                        %disp('   (cts)     (nm)     (nm)     (nm)     (cts)')
                        %disp('                                          ')
                        %disp(xy)

        %sdata(imindex)=xy;
        totxy(imindex,1:5)=xy;

        %totxyout=[xval;totxy];
%save('xy.txt','xy','-ascii'); disp('output saved to current directory as xy.txt'), save;
%save('xy.txt', disp(xy(1,1), xy(1,2), xy(1,2), xy(1,3), xy(1,4), xy(1,5)),'-ascii');


end
%load -ascii xyfile.txt
%save('C:\Tony\060212\spider1_mcp70_.5s_4.5s_matlab_qd2.dat','totxy','-ascii');
%save('C:\Tony\060212\spider1_mcp70_.5s_4.5s_matlab_qd1.dat','totxy','-ascii');
%Previously used the save command to write data.  Switched to fprintf
%command.

xval = 0:1:imindexval;

%outfilename = strcat('D:\Ribosome
Project\pz\Matlab_test_file\out_x',num2str(xpixel),'_y',num2str(ypixel),'.dat');
outfilename =
strcat('E:\Ryan\cDNA705_MCP60_1pt6x_PEG+BB_2.dat',num2str(peaknum),'_',num2str(xpixel),'_',num2
str(ypixel),'.dat');
%outfilename = strcat('E:\Tony\070210\video4_out1.dat');
fid = fopen(outfilename, 'wt');
%fid = fopen('C:\Documents and Settings\amanzo\Desktop\Work\060308_beads_10s_16bit_p1.dat', 'wt');

for iterationnum = 1:imindexval;
fprintf(fid, '%6.2f %12.8f %12.8f %12.8f %12.8f %12.8f %12.8f\n', xval(iterationnum),
NTOT(iterationnum), totxy(iterationnum,1:5));
end
fclose(fid)
% The data is saved with collumns as follows:
% frame#   Total counts   intensity max    x      y      FWHM    offset
%    #        (cts)          (cts)        (nm)   (nm)     (nm)    (cts)

end % End for the peaknum for statement (number of peaks).

%save;
%close;  %Use this to close figure windows after fitting is done.
```

Program “Read_traj_and_link_070517_v6_indexf.m”

```
% Reads data files from Gauss fitting program.

% This version (070717 v6) is a beta version of the trajectory analysis
% program that analyzes intensities also.
%
% This v3 also incorporates a rolling average calculation and substracts
% this from the raw data.

% This v4 will ask for the trajectories to be checked for pixel jumoing
% before any rolling average is done.  In v3, the rolling average was done
% before screening for pixel jumping, but the averaging made it harder to
% distinguish which were bad trajectories undergoing severe pixel jumping.

clear  %clear memory
clear all;
close all;

dist1 = 1.1;     % Parameters that determine the maximum distance between points for the purpose of
ordering
dist2 = dist1;
dist3 = 1.1;

trajlen = 50;

% The following code blocked out 070717--used by Diatrack version

%{

% Open coordinate file

%xyname = 'traj_gene32_705-32-5pDNA_MCP60_BB+PEG_10';
%iname = 'intensity_gene32_705-32-5pDNA_MCP60_BB+PEG_10';
%path1 = ('E:\John\070406\TIFFS\');

%path1 = ('E:\John\070418\Trajectory Intensity Sorting\-Mg\');
%xyname = 'traj_gene60_5-705_MCP60_1pt6x_PEG+BB_4';
%iname = 'inten_gene60_5-705_MCP60_1pt6x_PEG+BB_4';

xyname = 'traj_gene60_3-705_MCP60_1pt6x_PEG+BB_2';
iname = 'inten_gene60_3-705_MCP60_1pt6x_PEG+BB_2';
path1 = ('E:\John\070418\Trajectory Intensity Sorting\-Mg\');

extension = '.txt';
xycoordfilename = strcat(path1,xyname,extension);
icoordfilename = strcat(path1,iname,extension);

%AA = importdata(coordfilename,'\t'); %For tab deliminated files
%AA = importdata(xycoordfilename);
AA = importdata(xycoordfilename,' ');
BB = importdata(icoordfilename);

xyDATA = AA.data; %put xy data into xyDATA
iDATA = BB.data;  %put intensity data into iDATA
```

```
%[dummy,totalcolumns] = size(xyDATA); % Want to only get the total number of columns here.

%Clear these arrays to free memory
clear AA;
clear BB;


%}
% End of code block


% This program reads the data files from the Gauss fitting program and
% displays them.

clear  all; %clear memory

close all; % Close all non-hidden figures
close all hidden;

%Define Path to OnOff data File
%path1 = ('E:\Tony\test\spider\');
%path1 = ('X:\amanzo\070519\spider2w4mMZnposition2\');
%path1 = ('X:\amanzo\070519\705spidernoZn\');
%path1 = ('X:\amanzo\070622\070622\');
path1 = ('D:\Ribosome Project\files\gene60_3-705_MCP60_1pt6x_PEG+BB_2\');
%path1 = ('X:\amanzo\070519\spider2wZn\');
%path1 = ('X:\amanzo\070519\spider2wZnposition2\');
%filename = 'QDS2_271_onoffB_p2cutoff.dat';
%file = strcat(path1,filename);
%extension = '.dat';
cd(path1);
%files = dir('docktime1*.dat');
%files = dir('060317_spider_withZinc_mcp75_1.6x_5sec_15sec_1hour_16bit_*.dat');
%files =
dir('070519_60x_mcp70_spiderDNA_705spider2w4mMZn_1mg1_1p6x_1si_4sd_360frames_position2_*.
dat');
%files =
dir('070519_5mg3_60x_mcp70_spiderDNA_705spider_1mg1_1p6x_1si_4sd_360frames_run070531_*.dat
');
%files =
dir('070622_slide2_6leg_wOSS_BSA_wZn_pos2_MCP70_1mg1_1si_4sd_360frames_f0p4_peak*.dat');

files = dir('gene60_3-705_MCP60_1pt6x_PEG+BB_2_*.dat');

%files = dir('070519_60x_mcp70_spiderDNA_705spider2wZn_1mg1_1p6x_1si_4sd_360frames_*.dat');
%files =
dir('070519_60x_mcp70_spiderDNA_705spider2wZn_1mg1_1p6x_1si_4sd_360frames_position2_*.dat');

g = 'D:\Ribosome Project\files\gene60_3-705_MCP60_1pt6x_PEG+BB_2\gene60_3-
705_MCP60_1pt6x_PEG+BB_2.TIF';


nf = length(files); % lf stores the number of data files (traces).

BigCell = cell(size(files));  % Creates a cell structure to store the data from traces
%Bigoodm = cell(size(files));  % Creates a cell structure to store the on/off resutls for the segments
```

```
%Bigavg = cell(size(lf)); %Stores the overall average on/off times for the
%traces (which are recorded for a particular buffer).
%DockCell = cell(size(files),numints); %A cell for the dock times of each trace.


for i = 1:nf;

        %filename = [rootname,num2str(data), extension];
        filename = [files(i).name];
        variable = load(filename); % Reads the data file and places data into variable array
        [AAy,AAx] = size(variable); %get size (AAy is number of rows, AAx is number of columns)
        %var(data) = variable;
        %n = int2str(i);
        disp(i); %Displays the current filenumber
        %CellArray = cell(variable);
        %BigCell(i,1) = variable;
        BigCell(i,1) = mat2cell(variable,AAy,AAx);  %Place data and size into one cell of BigCell
end ;

%This code averages the datapoints

frames2avg = 20;

for ii = 1:nf
   datatemp  = cell2mat(BigCell(ii));
   [sx,sy]=size(datatemp);
   datatemp2 = zeros(sx-frames2avg,2);
   for jj = 1:sx-frames2avg
      adatatempx = 0;
      adatatempy = 0;
      for kk = jj:jj+frames2avg
         adatatempx=adatatempx+datatemp(kk,4);
         adatatempy=adatatempy+datatemp(kk,5);
      end
      adatatempx = adatatempx/frames2avg;
      adatatempy = adatatempy/frames2avg;
      datatemp2(jj,1) = adatatempx;
      datatemp2(jj,2) = adatatempy;
   end
   datatempnew = [datatemp2(:,1),datatemp2(:,2)];

   BigCell2(ii,1)= mat2cell(datatempnew);

end

%{

% Now plot and select the trajectories you want

kk = 0;

for ii = 1:nf;
datatemp = cell2mat(BigCell2(ii));

datatemp2 = datatemp;
datatemp2(:,1) = datatemp2(:,1)-datatemp(1,1);
```

```
datatemp2(:,2) = datatemp2(:,2)-datatemp(1,2);

figure(1);
hold on;
plot(datatemp2(:,1),datatemp2(:,2));
axis equal;

axis([-300 300 -300 300]);

pause;

end


pause;

%}


% Now plot and select the trajectories you want

selectpeaks = [0 0];
kk = 0;

for ii = 1:nf;
datatemp = cell2mat(BigCell(ii));

datatemp2 = datatemp;
datatemp2(:,4) = datatemp2(:,4)-datatemp(1,4);
datatemp2(:,5) = datatemp2(:,5)-datatemp(1,5);

figure(1);
hold off;
plot(datatemp2(:,4),datatemp2(:,5),'r');
axis equal;


answer1 = input('Do you want to keep this trace?');

  if (answer1 == 0)

  else
    kk = kk + 1;
    currentxyCell(kk)=mat2cell(datatemp);
    %idatatemp = cell2mat((newiCell2(ii)));
    %currentiCell(kk) = mat2cell(idatatemp);

    %selectpeaks = [selectpeaks;[peaks(ii,1),peaks(ii,2)]];
    selectpeaks = [selectpeaks;[round(datatemp(1,4)),round(datatemp(1,5))]];

  end

  %pause;
end

selectpeaks(1,:) = [];
```

```
peaks = selectpeaks;
traj = kk;
xyCell = currentxyCell;
%iCell = currentiCell;


%pause;


%{
% Write peaks file

path1 = 'E:\Tony\test\';
%xyname = 'peakoutput';

path2 = strcat(path1,xyname,'\');
mkdir(path2);

     %filenum = num2str(1);
     outfile = strcat(path2,xyname,'.txt');
     disp(outfile);
     fid = fopen(outfile,'wt');
    for jj = 1:traj;
        fprintf(fid, '%12.8f %12.8f %12.8f\n', j, peaks(jj,1), peaks(jj,2) );
    end
    fclose(fid);


end

%}


%{

%This code averages the datapoints
% Now we do it here with revised Cell and subtract the drift

frames2avg = 20;

for ii = 1:traj
    datatemp  = cell2mat(xyCell(ii));
    [sx,sy]=size(datatemp);
    datatemp2 = zeros(sx-frames2avg,2);
    for jj = 1:sx-frames2avg
        adatatempx = 0;
        adatatempy = 0;
        for kk = jj:jj+frames2avg
            adatatempx=adatatempx+datatemp(kk,4);
            adatatempy=adatatempy+datatemp(kk,5);
        end
        adatatempx = adatatempx/frames2avg;
        adatatempy = adatatempy/frames2avg;
        datatemp2(jj,1) = adatatempx;
        datatemp2(jj,2) = adatatempy;
    end
```

```
    datatempnew = [datatemp2(:,1),datatemp2(:,2)];

    xyCellavg(ii,1)= mat2cell(datatempnew);

    beforeaveragex = datatemp(:,4);
    beforeaveragey = datatemp(:,5);
    beforeaverage = [beforeaveragex beforeaveragey];

    %for jj = frames2avg/2:sx-frames2avg/2

        averagecorrected = zeros(sx-frames2avg,2);
        initv = frames2avg/2;
        finalv = sx-frames2avg/2;
        for jj = 1:sx-frames2avg
        averagecorrected(jj,:) = beforeaveragex(initv-1+jj,:) - datatempnew(jj,:);
        end

end

%}

disp('We are done with displaying all the data');

%pause;

%xyDATA = [4 5 6]; %This is a dummy xyDATA

% Get all data and just work with all data.



%pause;

%For Diatrack
%[sx,sy] = size(xyDATA);
%traj = sy/3;

%For Gauss fit

% traj is defined earlier
% sx and sy are also defined already earlier

% Not needed for Gauss fit
%{
% intensity data
[ix,iy] = size(iDATA);
% iy should be equal to traj.
%}

%The following was used to determine the number of frames in each
%trajectory for the Diatrack data
%{
framecount=zeros(9,traj);
% row 1 = cell #
% row 2 = frame start
% row 3 = frame end
```

```
% row 4 = number of frames
% row 5 = x start value
% row 6 = x end value
% row 7 = y start
% row 8 = y end
% row 9  = d1 = sqrt(x1^2 + y1^2)
% row 10 = d2 = sqrt(x2^2 + y2^2)
for ii = 1:traj
   framecount(1,ii)=ii;
   framecount(2,ii)=xyDATA(1,ii*3);
end

pause;

%}

% Separate columns and get rid of all the zeros in columns (used for
% Diatrack)
%{
% Delete first row of xyDATA
xyDATA(1,:) = [];
[sx,sy]=size(xyDATA);

for ii = 1:traj %(columns)
   datatemp = [0,0];
   for jj=1:sx; %(rows)
      if xyDATA(jj,ii*3-2) ~= 0  %Check x value
         datatemp = [datatemp;[xyDATA(jj,ii*3-2),xyDATA(jj,ii*3-1)]];
         %disp('in while');
      end
   end
   datatemp(1,:) = []; %delete first row
   xyCell(ii) = {datatemp};
end

% For Intensity data:
% Separate columns and get rid of all the zeros in columns

% Delete first row of iDATA
iDATA(1,:) = [];
[ix,iy]=size(iDATA);

for ii = 1:iy %(columns)
   datatemp = [0];
   for jj=1:ix; %(rows)
      if iDATA(jj,ii) ~= 0  %Check x value
         datatemp = [datatemp;iDATA(jj,ii)];
         %disp('in while');
      end
   end
   datatemp(1,:) = []; %delete first row
   iCell(ii) = {datatemp};
end

disp('Done with intensity sorting');
```

```
pause;


for ii = 1:traj %(columns)
   datatemp = cell2mat(xyCell(ii));
   [xycx,xycy] = size(datatemp);
   framecount(3,ii) = framecount(2,ii)+xycx-1;
   framecount(4,ii) = xycx;
   framecount(5,ii) = datatemp(1,1);       %x1
   framecount(6,ii) = datatemp(xycx,1);    %x2
   framecount(7,ii) = datatemp(1,2);       %y1
   framecount(8,ii) = datatemp(xycx,2);    %y2
   framecount(9,ii)  = sqrt(framecount(5,ii)^2+framecount(7,ii)^2); %d1
   framecount(10,ii) = sqrt(framecount(6,ii)^2+framecount(8,ii)^2); %d1
end

disp('Finished framecount');

pause;

%}

% Do I still need this turned off for Gauss fit?
%warning off MATLAB:divideByZero
warning off all

% The following code is used to place xy and intensity data together for the Diatrack data.
%{
for ii = 1:traj
   test1 = framecount(10,ii);
   for jj = ii+1:traj
      test2 = framecount(9,jj);

      cc1 = cell2mat(xyCell(ii));
      cc2 = cell2mat(xyCell(jj));
      if ( cc1(1,1) == 0 || cc2(1,1)==0)  %If the cell (array) is a zero cell, then do nothing.

      else

         if (abs(test2 - test1) <=1 )    % Check the last row with the first row.
            if (ii ~= jj)
               datatemp1 = cell2mat(xyCell(ii));
               datatemp2 = cell2mat(xyCell(jj));
               % Check if x and y values agree
               if (framecount(6,ii)-framecount(5,jj)<=1)
                  if (framecount(8,ii)-framecount(7,jj)<=1)
                     datanew = [datatemp1;datatemp2];
                     xyCell(ii) = mat2cell(datanew); % Place the two matching data sets in the same cell.
                     xyCell(jj) = mat2cell([0,0]);   % Place a zero value in the cell that was moved.
                     %TEMP = cellfun(@timeszero, xyCell(jj));  % Place zero by using cellfun instead.


                     % Do the same for the intensity
                     idatatemp1 = cell2mat(iCell(ii));
                     idatatemp2 = cell2mat(iCell(jj));
                     idatanew   = [idatatemp1;idatatemp2];
```

```
                    iCell(ii)  = mat2cell(idatanew);
                    iCell(jj)  = mat2cell([0,0]);

                    disp(ii); disp('  '); disp(jj);
                end
            end
        end
    end
  end
 end
end

disp('Finished xyCell and iCell link');
pause;

% Get rid of all zero value cells in xyCell and iCell
%newxyCell(1) = {[0 0]};
jj = 0;
kk = 0;
for ii = 1:traj
   datatemp = cell2mat(xyCell(ii));
   if(datatemp(1,1) ~= 0 && datatemp(1,2) ~= 0)
      jj = jj + 1;
      newxyCell(jj) = {datatemp};
     %  jj = jj + 1;
   end

   idatatemp = cell2mat(iCell(ii));
   if(idatatemp(1) ~= 0)
      kk = kk + 1;
      newiCell(kk) = {idatatemp};
     %  jj = jj + 1;
   end

end

newtraj = jj;

disp('End of newxyCell and newiCellmake');
pause;
%}


% This calculation was done by Gauss fit program already
%{
% List all the peak positions and sort them
% Discard any peaks that are within N pixels of each other.
peaks = zeros(newtraj,2);

for ii = 1:newtraj
   datatemp = cell2mat(newxyCell(ii));
   peaks(ii,:) = datatemp(1,:);
end

peaktest = peaks;
```

```
peakcheck(1:newtraj)=1;
kk = newtraj;
for ii = 1:newtraj
   testpeak1 = peaks(ii,:);
   for jj = 1 :newtraj
      if (jj == ii)
         % Do nothing
      else

            %testpeak2 = peaktest(jj,:);
            testpeak2 = peaks(jj,:);
            %if (testpeak2 == [0,0])
               % Do nothing
            %else
            peakdist = sqrt((testpeak2(1,1)-testpeak1(1,1))^2+(testpeak2(1,2)-testpeak1(1,2))^2);
            disp(peakdist);
            if (peakdist < 14)
               peakcheck(ii)=0;

               %peaktest(jj,:)=[0,0];
                peaks(ii,:) = [0,0];
                peaks(jj,:) = [0,0];   % Both peaks of the comparison should be marked and deleted.

            end


            %end     % End for the testpeak2 = [0,0] if statement.

      end

   end
end

peaks3 = peaks;

peaks2 = [0,0];

% Get rid of all peaks that fail test (in xyCell and iCell)
%newxyCell(1) = {[0 0]};

jj = 0;
for ii = 1:newtraj

if (peaks3(ii,:) ~= [0,0])

   datatemp = cell2mat(newxyCell(ii));
   idatatemp = cell2mat(newiCell(ii));

   jj = jj + 1;
   newxyCell2(jj) = {datatemp};
   newiCell2(jj) = {idatatemp};
   peaks2 = [peaks2;peaks3(ii,:)];
end

end
```

```
%}

%{
% Get rid of all peaks that fail test (in xyCell and iCell)
%newxyCell(1) = {[0 0]};
jj = 0;
kk = 0;
for ii = 1:newtraj
   datatemp = cell2mat(newxyCell(ii));
   if(peakcheck(ii)==1)
      jj = jj + 1;
      newxyCell2(jj) = {datatemp};
      peaks2 = [peaks2;peaks(jj,:)];
     %  jj = jj + 1;
   end

   idatatemp = cell2mat(newiCell(ii));
   if(peakcheck(ii)==1)
      kk = kk + 1;
      newiCell2(kk) = {idatatemp};
     %  jj = jj + 1;
   end

end

peaks2(1,:)=[]; % delete first row of zeros;

newtraj2 = jj;

disp('End of newxyCell2 and newiCell2');
pause;

peaks = peaks2;
traj = newtraj2;

%}

%g = 'E:\John\070418\gene60_3-705_MCP60_1pt6x_PEG+BB_2.TIF';
%g = 'D:\Ribosome Project\files\gene60_3-705_MCP60_1pt6x_PEG+BB_2\gene60_3-
705_MCP60_1pt6x_PEG+BB_2.TIF';
nametoopen = g;
[A1]=imread(nametoopen, 1); %read data file, 1st image (Reads 16 bit .tif)
[A2]=imread(nametoopen, 2); %read data file, 2nd image (Reads 16 bit .tif)
[A3]=imread(nametoopen, 3); %read data file, 3rd image (Reads 16 bit .tif)

A = (A1+A2+A3)/3;

% mirror image across x-axis
Anew = zeros(512,512);
for ii = 1:512;
   for jj = 1:512;
      Anew(ii,jj) = A(513-ii,jj);
   end;
end;
A = Anew;
```

```
imnum=1; %reference for current image number
info = imfinfo(nametoopen); %Get file information
d = size(info);

%xsize = info.Width;
%ysize = info.Height;
Axy=size(A);
xsize = Axy(2);
ysize = Axy(1);
%xsize = info(1,1).Width;
%ysize = info(1,1).Height;
xdim = xsize;
ydim = ysize;

amax=max(A);
amin=min(A);
disp(amax);
disp(amin);

imindexval= d(1,2); %Number of images (frames)
%imindexval = 10;

disp('Number of frmaes is: ');
disp(imindexval);
disp('The x and y length in pixels is:');
disp(xsize);
disp(ysize);

%NTOT=zeros(1,imindexval);

%%load -ascii nametoopen;

 %      hold off;
 %      contour (A(:,:,1),10); axis equal; % Show image of 1st frame
 %      %Adjust the last number to show more or less detail in contour map.

%hold on;

figure(1);
%contour(A);
contour (A(:,:,1),20); axis equal; % Show image of 1st frame
hold on;
%scatter(X,Y,S,C)
for ii = 1:traj
   scatter(peaks(ii,1),peaks(ii,2),20,'sr');
end

hold off;
%pause;

% Now show 3D image of locations (except if it is too close to boundary of
% CCD (the fitting box will go over boundary and there will be an error).

fb_dim = 2000;
c = 132;
```

```
ppeaks = peaks/c;

for peaknum = 1:traj

% get X and Y from peaks array
    X = ppeaks(peaknum,1); %Divide by c to get pixel value
    Y = ppeaks(peaknum,2);
    %XY = peaks(peaknum,3); % = length(IDX);

    xx = round(X);
              yy = round(Y);
    %xx=350;
    %yy=300;

              fitboxdimension = fb_dim/c;
    fb_dim_pix = round(fb_dim/c);
              R = fitboxdimension/2;
              r = round(R);

% The following routine is performed in Gauss fit already and is probably redundant
% here.

    % Sort peaks and delete those with a fitting box extending betond the CCD image.
halfbox = round(fb_dim_pix/2);
%newpeaks = [0 0 0];
%for ii = 1:npeaks
  if (ppeaks(peaknum,1) > (halfbox) && ppeaks(peaknum,1) < 512-halfbox)
    if (ppeaks(peaknum,2) > (halfbox) && ppeaks(peaknum,2) < 512-halfbox)
      %if (peaks(ii,3) >= 10)   % This if statement checks how big the blob is and makes a selection.
        %newpeaks = [newpeaks;peaks(ii,:)];
      %end


              %bestcoeffs = zeros(1,5);

    %B = A((yy-r):(yy+r),(xx-r):(xx+r));
    B = double(A((yy-r):(yy+r),(xx-r):(xx+r))); %Forced double to satisfy fitting functions
    %contour (B); axis equal;

    [x,y] = meshgrid(-r:r,-r:r);
    figure(2);
    hold off;
    surf(x,y,B(:,:));
    %contour (B); axis equal;
%pause;

[cx,cy,ssx,ssy] = centerofmass(B);

disp('ssx and ssy are ');
disp(ssx);
disp(ssy);

    end

  end
```

```
end

disp('Done displaying 3D images.');
pause;

%The following code is for screening for pixel jumping
%and is only needed in Diatrack (as far as I know now).
%{
% Here the trajectories are plotted in x-y coordinates and the user may
% select which ones to keep and which ones to discard.
kk = 0;
selectpeaks = [0,0];
for ii = 1:traj;
datatemp = cell2mat(newxyCell2(ii));
figure(1);
hold off;
plot(datatemp(:,1),datatemp(:,2));
axis equal;

%answer1 = input('Do you want to keep this trace?');
answer1 = 1;    % Set to 1 or 0 for auto response.

   if (answer1 == 0)

   else
      kk = kk + 1;
      currentxyCell(kk)=mat2cell(datatemp);
      idatatemp = cell2mat((newiCell2(ii)));
      currentiCell(kk) = mat2cell(idatatemp);

      selectpeaks = [selectpeaks;[peaks(ii,1),peaks(ii,2)]];

   end

%pause;
end

selectpeaks(1,:) = [];
peaks = selectpeaks;
traj = kk;
xyCell = currentxyCell;
iCell = currentiCell;

%{
% Write peaks file

path1 = 'E:\Tony\test\';
%xyname = 'peakoutput';

path2 = strcat(path1,xyname,'\');
mkdir(path2);

    %filenum = num2str(1);
    outfile = strcat(path2,xyname,'.txt');
    disp(outfile);
```

```
     fid = fopen(outfile,'wt');
    for jj = 1:traj;
        fprintf(fid, '%12.8f %12.8f %12.8f\n', j, peaks(jj,1), peaks(jj,2) );
    end
    fclose(fid);


end

%}

disp('Done with pixel jumping screening.');

pause;

%}

% Now plot and select the intensity trajectories you want

%{
for ii = 1:jj;
datatemp = cell2mat(newiCell(ii));
[tx,ty] = size(datatemp);
figure(1);
hold off;
plot(1:tx,datatemp);
%axis equal;
pause;
end
%}

% Now plot and select the trajectories you want
%{
check = 0;

if (check == 1)

    % Read peaks files

    path1 = 'E:\Tony\test\';
    xyname = 'peakoutput';
    path2 = strcat(path1,xyname,'\');
    PA = importdata(path2,' ');

    [pax,pay]=size(PA);

    traj = pax;

else
%}

%Remarked out on 070718
%traj = kk;
%newxyCell2 = currentxyCell;
%newiCell2 = currentiCell;
```

```
% Get reducedxyCell

[numberofpeaks,dummyvar1] = size(peaks);
np=numberofpeaks;

datatemp4 = zeros(sx,2);

for ii = 1:np;
datatemp3 = cell2mat(xyCell(ii));

%datatemp4 = datatemp3;
datatemp4(:,1) = datatemp3(:,4);
datatemp4(:,2) = datatemp3(:,5);

redxyCell(ii)=mat2cell(datatemp4);

end

avgamt = 20;
halfavgamt = avgamt/2;

subtractarray = 1:traj;
cellcount = 0;

testrow = ones(1,20)/20;

for ii = 1:traj
    datatemp  = cell2mat(redxyCell(ii));
    [sx,sy]=size(datatemp);
    %datatemp2 = zeros(sx-avgamt,2);

cellcount=cellcount+1;

%data = [1:0.2:4]';
dataa = datatemp(:,1);
datab = datatemp(:,2);
windowSize = 20;
datatemp2a=filter(ones(1,windowSize)/windowSize,1,dataa);
datatemp2b=filter(ones(1,windowSize)/windowSize,1,datab);

datatempnew = [datatemp2a,datatemp2b];
redxyCellavg(cellcount,1)= mat2cell(datatempnew);

end

disp('done with filter');
pause;

% Here I use the averaging routine to get a rolling average of the
% trajectories and plot them:
% Right now it is set to ake a 20 frame average.

avgamt = 20;
halfavgamt = avgamt/2;

subtractarray = 1:traj;
```

```
cellcount = 0;
for ii = 1:traj
   %datatemp  = cell2mat(newxyCell2(ii));
   datatemp  = cell2mat(redxyCell(ii));
   [sx,sy]=size(datatemp);
   datatemp2 = zeros(sx-avgamt,2);
   if (sx > avgamt)    %Check if trajectory can be averaged for the specified number of points.
      cellcount = cellcount + 1;  % Used to keep track of number of valid trajectories (with valid lengths).
      subtractarray(ii) = 1;  %Use this as a marker to determine which cells past test.
   for jj = 1+avgamt:sx
      adatatempx = 0;
      adatatempy = 0;
      for kk = jj-avgamt:jj-1    % This is the points for averaging over.
         adatatempx=adatatempx+datatemp(kk,1);
         adatatempy=adatatempy+datatemp(kk,2);
      end
      adatatempx = adatatempx/avgamt;
      adatatempy = adatatempy/avgamt;
      datatemp2(jj,1) = adatatempx;
      datatemp2(jj,2) = adatatempy;
   end
   datatempnew = [datatemp2(:,1),datatemp2(:,2)];

   %for kk = 1:avgamt %Delete blank rows
      datatempnew(1:avgamt,:)=[];
   %end

   %newxyCell2avg(ii,1)= mat2cell(datatempnew);
   %newxyCell2avg(cellcount,1)= mat2cell(datatempnew);
   redxyCellavg(cellcount,1)= mat2cell(datatempnew);
   else
      % Do nothing if the trajectory isn't the required length for
      % meaningful averaging to be applied.
      subtractarray(ii) = 0;
   end
end


disp('Check datatemp and newxyCell2avg please...');
pause;

% Now plot the averaged trajectories

for ii = 1:cellcount;
datatemp = cell2mat(redxyCellavg(ii));

xdatatempavg = mean(datatemp(:,1));
ydatatempavg = mean(datatemp(:,2));

datatemp2 = datatemp;
%datatemp2(:,1) = datatemp2(:,1)-datatemp(1,1);
%datatemp2(:,2) = datatemp2(:,2)-datatemp(1,2);
datatemp2(:,1) = datatemp2(:,1)-xdatatempavg;
datatemp2(:,2) = datatemp2(:,2)-ydatatempavg;

disp('Check array');
```

```
%pause;

figure(3);
hold on;
plot(datatemp2(:,1),datatemp2(:,2));
axis equal;

%axis([-1.5 1.5 -1.5 1.5]);

pause;

end

%pause;

% Now subtract the averages from the true trajectories, neglecting the
% trajectories that didn't qualify for averaging.  Also do the intensities.

ccount = 0;
for ii = 1:traj

   if (subtractarray(ii) == 1)
      ccount = ccount + 1;
      %datatrue = cell2mat(newxyCell2(ii));
      %dataavg = cell2mat(newxyCell2avg(ccount));
      datatrue = cell2mat(redxyCell(ii));
      %datatrue(1:20,:)=[]; %Delete the corresponding avg count data
      dataavg = cell2mat(redxyCellavg(ccount));

      [sx,sy] = size(datatrue);
      [sxa,sya] = size(dataavg);

      datadiff = datatrue(avgamt+1:sx,:)-dataavg(:,:);
      newxyCelldiff(ccount) = mat2cell(datadiff);

      %idatatemp = cell2mat((newiCell2(ii)));
      %newiCell3(ccount) = mat2cell(idatatemp(11:sx-halfavgamt));



      disp('ok');
   else

   end
end

disp('Check newxyCelldiff');
pause;

% Now plot the trajectories

kk = 0;

for ii = 1:ccount;
datatemp = cell2mat(newxyCelldiff(ii));
```

```
%datatemp2 = datatemp;
%datatemp2(:,1) = datatemp2(:,1)-datatemp(1,1);
%datatemp2(:,2) = datatemp2(:,2)-datatemp(1,2);

disp('Check array');
%pause;

figure(4);
hold on;
%plot(datatemp2(:,1),datatemp2(:,2));
plot(datatemp(:,1),datatemp(:,2));
axis equal;

%axis([-1.5 1.5 -1.5 1.5]);

%pause;

end

disp('Done with plotting corrected data');
pause;

% Now use symmetry check to eliminate assymetric data sets:
% And plot data

symcount = 0;

for ii = 1:ccount;

%       datatrue = cell2mat(newxyCell2(ii));
%       [sx,sy] = size(datatrue);
%       datatemp2 = datatrue;

datatemp = cell2mat(newxyCelldiff(ii)); % Retrieve the data set.
origdatatemp = cell2mat(redxyCell(ii));

%datatemp2 = datatemp;
%datatemp2(:,1) = datatemp2(:,1)-datatemp(1,1);
%datatemp2(:,2) = datatemp2(:,2)-datatemp(1,2);

%%datatemp2(:,1) = datatemp2(:,1)+1000;
%%datatemp2(:,2) = datatemp2(:,2)+1000;

disp('In symmetry routine.');

[N,c] = size(datatemp(:,1));
m = 20;
disp('ii, N is:');
disp(ii);
disp(N);

sumx = 0;
sumy = 0;
sumxy = 0;
sumxx = 0;
sumyy = 0;
```

```
for kk = 1:N
    sumx = sumx + datatemp(kk,1);
    sumy = sumy + datatemp(kk,2);
    sumxy = sumxy + datatemp(kk,1)*datatemp(kk,2);
    sumxx = sumxx + datatemp(kk,1)*datatemp(kk,1);
    sumyy = sumyy + datatemp(kk,2)*datatemp(kk,2);
end

sigma11 = (1/N)*sumxx-(1/(N*N))*sumx*sumx;
sigma22 = (1/N)*sumyy-(1/(N*N))*sumy*sumy;
sigma12 = (1/N)*sumxy-(1/(N*N))*sumx*sumy;
sigma21 = (1/N)*sumxy-(1/(N*N))*sumx*sumy;

symA = zeros(2);
symA(1,1) = sigma11;
symA(1,2) = sigma12;
symA(2,1) = sigma21;
symA(2,2) = sigma22;

d = eig(symA);

disp('Eigenvalues are');
disp(d);

dmin = min(d);
dmax = max(d);
sparam = sqrt(dmax/dmin);

disp('sparam is');
disp(sparam);

paramtest = abs(sparam-1);

if (paramtest < 0.2)

    symcount = symcount + 1;
    symxyCelldiff(symcount) = mat2cell(datatemp);
    symorigdata(symcount) = mat2cell(origdatatemp);

    figure(5);
    hold on;
    plot(datatemp(:,1),datatemp(:,2));
    axis equal;
    %axis([-1.5 1.5 -1.5 1.5]);

end

pause;

end

disp('done with symmetry analysis');
pause;

% Adding autocorrelation analysis here:
```

```
for ii = 1:ccount;
for ii = 1:symcount;

%       datatrue = cell2mat(newxyCell2(ii));
%       [sx,sy] = size(datatrue);
%       datatemp2 = datatrue;


%datatemp = cell2mat(newxyCelldiff(ii)); % Retrieve the data set.
datatemp = cell2mat(symxyCelldiff(ii));
%datatemp = cell2mat(symorigdata(ii));

%datatemp2 = datatemp;
%datatemp2(:,1) = datatemp2(:,1)-datatemp(1,1);
%datatemp2(:,2) = datatemp2(:,2)-datatemp(1,2);

%%datatemp2(:,1) = datatemp2(:,1)+1000;
%%datatemp2(:,2) = datatemp2(:,2)+1000;

disp('In autocorrelation routine.');

[N,c] = size(datatemp(:,1));
m = 20;
disp('ii, N is:');
disp(ii);
disp(N);

sum1 = 0;
for n = 1:N
   sum1 = sum1 + datatemp(n,1);
end
sum1 = (sum1/N)^2;
disp('sum1 is:');
disp(sum1);

ac = zeros(1,m);
g = zeros(1,m);

for jj = 1:m;
   g(jj) = 0;
   sum2 = 0;
   for nn = 1:N-jj
      sum2 = sum2 + datatemp(nn,1)*datatemp(nn+jj,1);
   end
   g(jj) = sum2/(N-jj) - sum1;
end

gCell(ii) = mat2cell(g);

%g(:) = 1000000*g(:);
%gmax = max(g(:));
%g(:) = g(:)/gmax;

figure(6);
```

```
%hold on;
plot(1:m,g(:));
%axis equal;
%axis([-1.5 1.5 -1.5 1.5]);

pause;

end

disp('Done with autocorrelation routine.');
pause;

% Here we will calculate the MSD

totRMS = 0;

for ii = 1:symcount;

datatemp = cell2mat(symxyCelldiff(ii)); % Retrieve the data set.
%datatemp = cell2mat(symorigdata(ii));

%datatemp2 = datatemp;
%datatemp2(:,1) = datatemp2(:,1)-datatemp(1,1);
%datatemp2(:,2) = datatemp2(:,2)-datatemp(1,2);

%%datatemp2(:,1) = datatemp2(:,1)+1000;
%%datatemp2(:,2) = datatemp2(:,2)+1000;

disp('In MSD routine.');

[N,c] = size(datatemp);
m = 20;
disp('ii, N is:');
disp(ii);
disp(N);

tempxy = datatemp;
[xyrows,xycolumns] = size(tempxy); % Get dimensions of temp array (columns should = 2 all the time).

difftotal = 0;
diff = 1:N-1;
for kk = 2:N
diff(kk-1) = (datatemp(kk,1)-datatemp(kk-1,1))^2 + (datatemp(kk,2)-datatemp(kk-1,2))^2 ;
difftotal = difftotal + diff(kk-1);
end

difftotal = difftotal/(N-1);
RMS = sqrt(difftotal);

totRMS = [totRMS;RMS];

disp('RMS is');
disp(RMS);

end
```

```
totRMS(1)=[];

disp('done with RMS analysis');
pause;







%Prep for sending to ReadTrajWork
%selectpeaks(1,:) = [];
%peaks = selectpeaks;
traj = ccount;
xyCell = newxyCelldiff;
iCell = newiCell3;

%{
% Write peaks file

path1 = 'E:\Tony\test\';
%xyname = 'peakoutput';

path2 = strcat(path1,xyname,'\');
mkdir(path2);

    %filenum = num2str(1);
    outfile = strcat(path2,xyname,'.txt');
    disp(outfile);
    fid = fopen(outfile,'wt');
   for jj = 1:traj;
      fprintf(fid, '%12.8f %12.8f %12.8f\n', j, peaks(jj,1), peaks(jj,2) );
   end
   fclose(fid);


end

%}

disp('We are done with displaying all the data');

pause;

% The following was all code in the Read_traj_v7p2 program.  I placed the
% code above it it and will use the resutls to call ReadTrajWork

%{

% Use first row of data set to get the number of columns of data for the
% first frame.

index = 1;
test = xyDATA(1,3*index);
```

```
while (test < 1.1)
   disp(test);
   disp(index);
   index = index + 1;
   test = xyDATA(1,3*index);
end
index = index - 1;

% Get and store all fame results except the first, without deleting frame
% reference numbers (top row).
restofdata = xyDATA(:,(3*(index+1)-2):totalcolumns);
irestofdata = iDATA(:,index+1:(totalcolumns/3));
totiDATA = iDATA;

% Store first frame data in xyDATA (deleting all other contents of xyDATA)
origxyDATA = xyDATA(:,1:3*index);
xyDATA = origxyDATA;

origiDATA = iDATA(:,1:index);
iDATA = origiDATA;

%Comment (Uncomment) this line to leave (clear) this data when needed (not
%needed)
clear origxyDATA;
clear origiDATA;

xyDATA(1,:)=[]; %delete first row of data (set equal to empty array)
iDATA(1,:)=[]; %delete first row of data (set equal to empty array)
totiDATA(1,:) = [];

[AAx,AAy] = size(xyDATA); %get size
[BBx,BBy] = size(iDATA);  %get size

% Use the columns from the first fram data only, so set AAx accordingly
% Added (070509)
%AAx = index*3;

trajs = AAy/3;
tottrajs = totalcolumns/3;

row = AAx;
column = trajs;

textout = strcat('There are  ',' ',num2str(trajs),' trajectories and ',...
   ' ',num2str(AAx), ' frames in this file (before eliminating any zero values.');
disp(textout);

% No sorting of the data has been done yet.  This program does not link the
% trajectories.  This program calls the ReadTrajWork routine, which will
% sort the trajectories according to the following scheme: All trajs that
% occur int the first frame are separated into 1) those that last the full
% number of frames in the video, and 2) those that are shorter than the full number of frames.
% Then the rest of the traces that are initiated after the first frame are
% all bunched together.

% If we want to sort good and bad trajectories, then maybe we should place this
```

```
% here, then the trajs can be sorted by and selected for intensity.
% However, the number of trajectories to sort through will be huge if we
% include all the rest.  So maybe we should just limit ourselves to the
% first fram data.

% No, the ReadTrajWork program also selects for trajs that have aminimum
% number of frames.  So this can make thisngs easier.  THen it would be
% better to move the "quality check of traj" in to the ReadTrajWork
% routine.

% The following was added on recently and isn't part of the main program.
% 070509


% Now plot and select the trajectories you want

%{
for ii = 1:ii;
datatempx = cell2mat(xycCellx(ii));
datatempy = cell2mat(xycCelly(ii));
figure(1);
hold off;
plot(datatempx,datatempy);
axis equal;
pause;
end
%}

disp('Where are we?');

% For Intensity data:
% Separate columns and get rid of all the zeros in columns

% Delete first row of iDATA

[ix,iy]=size(iDATA);

for ii = 1:iy %(columns)
   datatemp = [0];
   for jj=1:ix; %(rows)
      if iDATA(jj,ii) ~= 0  %Check x value
         datatemp = [datatemp;iDATA(jj,ii)];
         %disp('in while');
      end
   end
   datatemp(1,:) = []; %delete first row
   iCell(ii) = {datatemp};
end

% PLace this here temporarily to check the intensities and compare with the
% linked intensities (070513)

% Get rid of all zero value cells in xyCell and iCell
%newxyCell(1) = {[0 0]};
jj = 0;
kk = 0;
```

```
for ii = 1:iy
%    datatemp = cell2mat(xyCell(ii));
%    if(datatemp(1,1) ~= 0 && datatemp(1,2) ~= 0)
%        jj = jj + 1;
%        newxyCell(jj) = {datatemp};
%    end

    idatatemp = cell2mat(iCell(ii));
    if(idatatemp(1) ~= 0)
        kk = kk + 1;
        newiCell(kk) = {idatatemp};
    end

end

disp('End of newxyCell and newiCellmake');
pause;

% Now plot and select the intensity trajectories you want

for ii = 1:kk;
datatemp = cell2mat(newiCell(ii));
[tx,ty] = size(datatemp);
figure(1);
hold off;
plot(1:tx,datatemp);
%axis equal;
pause;
end


disp('Did we plot?');

pause;


%}

AAx = 600; % Total number of frames in video
trajs = traj;

% Call ReadTrajWork but not using the intensity threshold yet.
inttrial = 0;  %Tells ReadTrajWork that we are not worrying about intensities yet.
ReadTrajWork_for_ReadTrajLink_v2(AAx, trajs, xyCell, trajlen, inttrial, path1, xyname);

% AAx is the maximum number of frames in the data (not needed in function
% call?).
% trajlen is the minumum trajectroy length specified
% trajs is the number of trajectories.

%{

% The following code sorts the intensities for all trajectories and
% histograms them.

% Separate columns and get rid of all the zeros in columns
```

```
kk = 1;
%while (ii < xynctraj);
for ii = 1:tottrajs %(columns)
    datatemp = [0];
    for jj=1:BBx; %(rows)
        %if iDATA(jj,ii) ~= 0  %Check x value
        if totiDATA(jj,ii) ~= 0  %Check x value
            %datatemp = [datatemp ; iDATA(jj,ii)];
            datatemp = [datatemp ; totiDATA(jj,ii)];
            %disp('in while');
        end
    end
    datatemp(1) = []; %delete first row
    [isx,isy]=size(datatemp);
    % The following if-else statement checks if the length of the
    % trajectory is long enough for my specifications ( determined by
    % trajlen ).  If long enough, then it puts it into xyncCell array.
    % This is giving me an error right now (April 29, 2007) because none of
    % the xync trajectories are my cpecified length.  This means that
    % xyncCell is not created and does not exist.
%    if (isx < trajlen)      %!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
    if (isx < 1)      %!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

        %xynctraj = xynctraj - 1;
    else
    IntensityCell(kk) = {datatemp};
    kk = kk + 1;
    end

end

%if ( xyncCellstat == 1 )    % Check if xyncCell exists.

% Get each trace out of the xyncCell and calculate the mean and std for
% each, and then store it in an array.
iave = 1:tottrajs; % used to be BBy
istd = 1:tottrajs;

for ii = 1:tottrajs
    tempi = cell2mat(IntensityCell(ii));
    [irows,icolumns] = size(tempi); % Get dimensions of temp array (columns should = 2 all the time).
        iave(ii)=mean(tempi(:,1));
        istd(ii)=std(tempi(:,1));

end

%end   % End of xyncCellstat check

bins = 40;
%bins = size(xystdtotx1)/20

[n,xout] = hist(iave,bins);
figure(2);
bar(xout,n);
xlabel('iave, DN');
ylabel('Distribution');
```

```
title('Histogram plot of iave');
%hmean = mean(xx);
%hstd = std(xx);
%text( 5, 100, ['mean is: ',num2str(hmean)]);
%text( 5, 50,['std is: ',num2str(hstd)]);

pause;

%}

% Now we are using an intensity threshold to select trajectorys.

% Separate columns and get rid of all the zeros in columns
kk = 1;
%while (ii < xynctraj);
for ii = 1:BBy %(columns)
   datatemp = [0];
   for jj=1:BBx; %(rows)
      if iDATA(jj,ii) ~= 0  %Check x value
      %if origiDATA(jj,ii) ~= 0  %Check x value
         datatemp = [datatemp ; iDATA(jj,ii)];
         %datatemp = [datatemp ; origiDATA(jj,ii)];
         %disp('in while');
      end
   end
   datatemp(1) = []; %delete first row
   [isx,isy]=size(datatemp);
   % The following if-else statement checks if the length of the
   % trajectory is long enough for my specifications ( determined by
   % trajlen ).  If long enough, then it puts it into xyncCell array.
   % This is giving me an error right now (April 29, 2007) because none of
   % the xync trajectories are my cpecified length.  This means that
   % xyncCell is not created and does not exist.
%    if (isx < trajlen)      %!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
   if (isx < 1)      %!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

      %xynctraj = xynctraj - 1;
   else
   IntensityCell(kk) = {datatemp};
   kk = kk + 1;
   end

end

%if ( xyncCellstat == 1 )    % Check if xyncCell exists.

% Get each trace out of the xyncCell and calculate the mean and std for
% each, and then store it in an array.
iave = 1:BBy; %
istd = 1:BBy;

for ii = 1:BBy
   tempi = cell2mat(IntensityCell(ii));
   [irows,icolumns] = size(tempi); % Get dimensions of temp array (columns should = 2 all the time).
      iave(ii)=mean(tempi(:,1));
      istd(ii)=std(tempi(:,1));
```

```
end

%end   % End of xyncCellstat check

bins = 40;
%bins = size(xystdtotx1)/20

[n,xout] = hist(iave,bins);
figure(2);
bar(xout,n);
xlabel('iave, DN');
ylabel('Distribution');
title('Histogram plot of iave');
%hmean = mean(xx);
%hstd = std(xx);
%text( 5, 100, ['mean is: ',num2str(hmean)]);
%text( 5, 50,['std is: ',num2str(hstd)]);

%lowint = input('Enter the low intensity value: ');
%highint = input('Enter the high intensity value: ');
intthresh = input('Enter the upper value for the intensity threshold: ');

% Create an array with the same length as the number of trajectories
% (and/or intensities) and then use a 1 or 0 value to denote if the trace
% satisfies the intensity threshold criteria.

goodtraj = 1:BBy; % BBy here is the number of intensity traces and should be
            % equal to the number of trajectories (all found in the first frame).
            % Here I just set up an array with the trajectory column
            % numbers in it, but these will be replaced by 1 and 0
            % values.

for ii = 1:BBy
   if (iave(ii) <= intthresh)
      goodtraj(ii) = 1;
   else
      goodtraj(ii) = 0;
      %xyDATA(:,ii) = [];
   end
end

pause;


% Here we need to determine which trajectorys to keep in the restofdata
% array.

[rdx,rdy] = size(irestofdata);

% Separate columns and get rid of all the zeros in columns
kk = 1;
for ii = 1:rdy %(columns)
   datatemp = [0];
   for jj=1:rdx; %(rows)
      if irestofdata(jj,ii) ~= 0  %Check x value
         datatemp = [datatemp ; irestofdata(jj,ii)];
```

```
      end
   end
   datatemp(1) = []; %delete first row
   [isx,isy]=size(datatemp);
   % The following if-else statement checks if the length of the
   % trajectory is long enough for my specifications ( determined by
   % trajlen ).  If long enough, then it puts it into xyncCell array.
   % This is giving me an error right now (April 29, 2007) because none of
   % the xync trajectories are my cpecified length.  This means that
   % xyncCell is not created and does not exist.
%    if (isx < trajlen)      %!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
   if (isx < 1)      %!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

   else
   IrodCell(kk) = {datatemp};
   kk = kk + 1;
   end

end

irodave = 1:rdy;
irodstd = 1:rdy;

for ii = 1:rdy
   tempi = cell2mat(IrodCell(ii));
   [irodrows,irodcolumns] = size(tempi); % Get dimensions of temp array (columns should = 2 all the
time).
      irodave(ii) = mean(tempi(:,1));
      irodstd(ii) = std(tempi(:,1));

end
goodrodtraj = 1:rdy; % BBy here is the number of intensity traces and should be
          % equal to the number of trajectories (all found in the first frame).
          % Here I just set up an array with the trajectory column
          % numbers in it, but these will be replaced by 1 and 0
          % values.

for ii = 1:rdy
   if (irodave(ii) <= intthresh)
      goodrodtraj(ii) = 1;
   else
      goodrodtraj(ii) = 0;
   end
end

% Now start eliminating the unwanted trajectories from the arrays (xtDATA
% and restofdata).




for ii = 1:BBy
   if (goodtraj(ii) == 0)
      % Delete all three columns
      xyDATA(:,ii)=[];
      xyDATA(:,ii)=[];
```

```
        xyDATA(:,ii)=[];
    else

    end
end

for ii = 1:rdy
    if (goodrodtraj(ii) == 0)
        restofdata(:,ii) = [];
        restofdata(:,ii) = [];
        restofdata(:,ii) = [];
    else

    end
end



[AAx,AAy] = size(xyDATA); %get size

trajs = AAy/3;


textout = strcat('There are  ',' ',num2str(trajs),' trajectories and ',...
    ' ',num2str(AAx), ' frames in this file (before eliminating any zero values.');
disp(textout);


% Call ReadTrajWork using the intensity selected data.
inttrial = 1;
ReadTrajWork_v2(AAx, trajs, xyDATA, trajlen, restofdata, inttrial, path1, xyname);
```

REFERENCES

(1) Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001) Initial sequencing and analysis of the human genome. *Nature 409*, 860-921.

(2) Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001) The sequence of the human genome. *Science 291*, 1304-51.

(3) Mattick, J. S. (2004) The hidden genetic program of complex organisms. *Sci Am 291*, 60-7.

(4) Mattick, J. S., and Makunin, I. V. (2006) Non-coding RNA. *Hum Mol Genet 15 Spec No 1*, R17-29.

(5) Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature 436*, 876-80.

(6) Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E.,

Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature 447*, 799-816.

(7) Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aidinis, V., Allen, J. E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L., Bansal, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crowe, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S. P., Kruger, A., Kummerfeld, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., et al. (2005) The transcriptional landscape of the mammalian genome. *Science 309*, 1559-63.

(8) Willingham, A. T., and Gingeras, T. R. (2006) TUF love for "junk" DNA. *Cell 125*, 1215-20.

(9) Wilson, F. H., Hariri, A., Farhi, A., Zhao, H., Petersen, K. F., Toka, H. R., Nelson-Williams, C., Raja, K. M., Kashgarian, M., Shulman, G. I., Scheinman, S. J., and Lifton, R. P. (2004) A cluster of metabolic defects caused by mutation in a mitochondrial tRNA. *Science 306*, 1190-4.

(10) Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H., Roix, J. J., Kathiresan, S., Hirschhorn, J. N., Daly, M. J., Hughes, T. E., Groop, L., Altshuler, D., Almgren, P., Florez, J. C., Meyer, J., Ardlie, K., Bengtsson Bostrom, K., Isomaa, B., Lettre, G., Lindblad, U., Lyon, H. N., Melander, O.,

Newton-Cheh, C., Nilsson, P., Orho-Melander, M., Rastam, L., Speliotes, E. K., Taskinen, M. R., Tuomi, T., Guiducci, C., Berglund, A., Carlson, J., Gianniny, L., Hackett, R., Hall, L., Holmkvist, J., Laurila, E., Sjogren, M., Sterner, M., Surti, A., Svensson, M., Svensson, M., Tewhey, R., Blumenstiel, B., Parkin, M., Defelice, M., Barry, R., Brodeur, W., Camarata, J., Chia, N., Fava, M., Gibbons, J., Handsaker, B., Healy, C., Nguyen, K., Gates, C., Sougnez, C., Gage, D., Nizzari, M., Gabriel, S. B., Chirn, G. W., Ma, Q., Parikh, H., Richardson, D., Ricke, D., and Purcell, S. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science 316*, 1331-6.

(11) Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Aza-Blanc, P., Hogenesch, J. B., and Schultz, P. G. (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science 309*, 1570-3.

(12) Newby, M. I., and Greenbaum, N. L. (2002) Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat Struct Biol 9*, 958-65.

(13) Valadkhan, S., and Manley, J. L. (2001) Splicing-related catalysis by protein-free snRNAs. *Nature 413*, 701-7.

(14) Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J. L. (2007) Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *Rna 13*, 2300-11.

(15) Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A 98*, 4899-903.

(16) Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science 292*, 897-902.

(17) Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell 108*, 557-72.

(18) Sievers, A., Beringer, M., Rodnina, M. V., and Wolfenden, R. (2004) The ribosome as an entropy trap. *Proc Natl Acad Sci U S A 101*, 7897-901.

(19) Tinsley, R. A., Furchak, J. R., and Walter, N. G. (2007) Trans-acting glmS catalytic riboswitch: locked and loaded. *Rna 13*, 468-77.

(20) Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature 428*, 281-6.

(21) Doudna, J. A., and Cech, T. R. (2002) The chemical repertoire of natural ribozymes. *Nature 418*, 222-8.

(22) Walter, N. G., and Engelke, D. R. (2002) Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist (London) 49*, 199-203.

(23) Pereira, M. J., Harris, D. A., Rueda, D., and Walter, N. G. (2002) Reaction pathway of the trans-acting hepatitis delta virus ribozyme: a conformational change accompanies catalysis. *Biochemistry 41*, 730-40.

(24) Zhuang, X., Kim, H., Pereira, M. J., Babcock, H. P., Walter, N. G., and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science 296*, 1473-6.

(25) Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science 294*, 858-62.

(26) Mayr, C., Hemann, M. T., and Bartel, D. P. (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science 315*, 1576-9.

(27) Bokinsky, G., and Zhuang, X. (2005) Single-molecule RNA folding. *Acc Chem Res 38*, 566-73.

(28) Walter, N. G. (2001) Structural dynamics of catalytic RNA highlighted by fluorescence resonance energy transfer. *Methods 25*, 19-30.

(29) Walter, N. G., Burke, J. M., and Millar, D. P. (1999) Stability of hairpin ribozyme tertiary structure is governed by the interdomain junction. *Nat Struct Biol 6*, 544-9.

(30) Frank, J., and Agrawal, R. K. (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature 406*, 318-22.

(31) Holmes, K. L., and Culver, G. M. (2004) Mapping structural differences between 30S ribosomal subunit assembly intermediates. *Nat Struct Mol Biol 11*, 179-86.

(32) Orr, J. W., Hagerman, P. J., and Williamson, J. R. (1998) Protein and Mg(2+)-induced conformational changes in the S15 binding site of 16 S ribosomal RNA. *J Mol Biol 275*, 453-64.

(33) Al-Hashimi, H. M. (2005) Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *Chembiochem 6*, 1506-19.

(34) Al-Hashimi, H. M., Gosser, Y., Gorin, A., Hu, W., Majumdar, A., and Patel, D. J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J Mol Biol 315*, 95-102.

(35) Bokinsky, G., Rueda, D., Misra, V. K., Rhodes, M. M., Gordus, A., Babcock, H. P., Walter, N. G., and Zhuang, X. (2003) Single-molecule transition-state analysis of RNA folding. *Proc Natl Acad Sci U S A 100*, 9302-7.

(36) Rueda, D., Bokinsky, G., Rhodes, M. M., Rust, M. J., Zhuang, X., and Walter, N. G. (2004) Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc Natl Acad Sci U S A 101*, 10066-71.

(37) Ermolenko, D. N., Majumdar, Z. K., Hickerson, R. P., Spiegel, P. C., Clegg, R. M., and Noller, H. F. (2007) Observation of intersubunit movement of the ribosome in solution using FRET. *J Mol Biol 370*, 530-40.

(38) Fredrick, K., and Noller, H. F. (2003) Catalysis of ribosomal translocation by sparsomycin. *Science 300*, 1159-62.

(39) Agalarov, S. C., Sridhar Prasad, G., Funke, P. M., Stout, C. D., and Williamson, J. R. (2000) Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain. *Science 288*, 107-13.

(40) Recht, M. I., and Williamson, J. R. (2001) Central domain assembly: thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *J Mol Biol 313*, 35-48.

(41) Casiano-Negroni, A., Sun, X., and Al-Hashimi, H. M. (2007) Probing Na(+)-induced changes in the HIV-1 TAR conformational dynamics using NMR residual dipolar couplings: new insights into the role of counterions and electrostatic interactions in adaptive recognition. *Biochemistry 46*, 6525-35.

(42) Zhang, Q., Sun, X., Watt, E. D., and Al-Hashimi, H. M. (2006) Resolving the motional modes that code for RNA adaptation. *Science 311*, 653-6.

(43) Youngman, E. M., Brunelle, J. L., Kochaniak, A. B., and Green, R. (2004) The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell 117*, 589-99.

(44) Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science 289*, 905-20.

(45) Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature 407*, 327-39.

(46) Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science 292*, 883-96.

(47) Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science 310*, 827-34.

(48) Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S. M., Van Roey, P., Agrawal, R. K., Harvey, S. C., Sali, A., Chapman, M. S., and Frank, J. (2003) Study of the structural dynamics of the E coli 70S ribosome using real-space refinement. *Cell 113*, 789-801.

(49) Tama, F., Valle, M., Frank, J., and Brooks, C. L., 3rd. (2003) Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc Natl Acad Sci U S A 100*, 9319-23.

(50) Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell 103*, 1143-54.

(51) Montandon, P. E., Wagner, R., and Stutz, E. (1986) E. coli ribosomes with a C912 to U base change in the 16S rRNA are streptomycin resistant. *Embo J 5*, 3705-8.

(52) Lodmell, J. S., and Dahlberg, A. E. (1997) A conformational switch in Escherichia coli 16S ribosomal RNA during decoding of messenger RNA. *Science 277*, 1262-7.

(53) Lodmell, J. S., Gutell, R. R., and Dahlberg, A. E. (1995) Genetic and comparative analyses reveal an alternative secondary structure in the region of nt 912 of Escherichia coli 16S rRNA. *Proc Natl Acad Sci U S A 92*, 10555-9.

(54) Rodriguez-Correa, D., and Dahlberg, A. E. (2004) Genetic evidence against the 16S ribosomal RNA helix 27 conformational switch model. *Rna 10*, 28-33.

(55) Hoerter, J. A., Lambert, M. N., Pereira, M. J., and Walter, N. G. (2004) Dynamics Inherent in Helix 27 from Escherichia coli 16S Ribosomal RNA. *Biochemistry 43*, 14624-36.

(56) Napoli, C., Lemieux, C., and Jorgensen, R. (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell 2*, 279-289.

(57) Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature 391*, 806-11.

(58) Dykxhoorn, D. M., and Lieberman, J. (2006) Running interference: prospects and obstacles to using small interfering RNAs as small molecule drugs. *Annu Rev Biomed Eng 8*, 377-402.

(59) Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature 411*, 494-8.

(60) Zamore, P. D., Tuschl, T., Sharp, P. A., and Bartel, D. P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell 101*, 25-33.

(61) Macrae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Adams, P. D., and Doudna, J. A. (2006) Structural basis for double-stranded RNA processing by Dicer. *Science 311*, 195-8.

(62) Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P. D. (2004) A protein sensor for siRNA asymmetry. *Science 306*, 1377-80.

(63) Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science 305*, 1437-41.

(64) Jackson, A. L., Burchard, J., Schelter, J., Chau, B. N., Cleary, M., Lim, L., and Linsley, P. S. (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *Rna 12*, 1179-87.

(65) Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell 120*, 15-20.

(66) Jackson, A. L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J. M., Lim, L., Karpilow, J., Nichols, K., Marshall, W., Khvorova, A., and Linsley, P. S. (2006) Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *Rna 12*, 1197-205.

(67) Rana, T. M. (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol 8*, 23-36.

(68) de Fougerolles, A., Manoharan, M., Meyers, R., and Vornlocher, H. P. (2005) RNA interference in vivo: toward synthetic small inhibitory RNA-based therapeutics. *Methods Enzymol 392*, 278-96.

(69) Bartlett, D. W., and Davis, M. E. (2007) Effect of siRNA nuclease stability on the in vitro and in vivo kinetics of siRNA-mediated gene silencing. *Biotechnol Bioeng 97*, 909-21.

(70) Allerson, C. R., Sioufi, N., Jarres, R., Prakash, T. P., Naik, N., Berdeja, A., Wanders, L., Griffey, R. H., Swayze, E. E., and Bhat, B. (2005) Fully 2'-modified oligonucleotide duplexes with improved in vitro potency and stability compared to unmodified small interfering RNA. *J Med Chem 48*, 901-4.

(71) Braasch, D. A., Jensen, S., Liu, Y., Kaur, K., Arar, K., White, M. A., and Corey, D. R. (2003) RNA interference in mammalian cells by chemically-modified RNA. *Biochemistry 42*, 7967-75.

(72) Chiu, Y. L., and Rana, T. M. (2003) siRNA function in RNAi: a chemical modification analysis. *Rna 9*, 1034-48.

(73) Choung, S., Kim, Y. J., Kim, S., Park, H. O., and Choi, Y. C. (2006) Chemical modification of siRNAs to improve serum stability without loss of efficacy. *Biochem Biophys Res Commun 342*, 919-27.

(74) Czauderna, F., Fechtner, M., Dames, S., Aygun, H., Klippel, A., Pronk, G. J., Giese, K., and Kaufmann, J. (2003) Structural variations and stabilising modifications of synthetic siRNAs in mammalian cells. *Nucleic Acids Res 31*, 2705-16.

(75) Haupenthal, J., Baehr, C., Kiermayer, S., Zeuzem, S., and Piiper, A. (2006) Inhibition of RNAse A family enzymes prevents degradation and loss of silencing activity of siRNAs in serum. *Biochem Pharmacol 71*, 702-10.

(76) Hoerter, J. A., and Walter, N. G. (2007) Chemical modification resolves the asymmetry of siRNA strand degradation in human blood serum. *Rna 13*, 1887-93.

(77) Layzer, J. M., McCaffrey, A. P., Tanner, A. K., Huang, Z., Kay, M. A., and Sullenger, B. A. (2004) In vivo activity of nuclease-resistant siRNAs. *Rna 10*, 766-71.

(78) Turner, J. J., Jones, S. W., Moschos, S. A., Lindsay, M. A., and Gait, M. J. (2007) MALDI-TOF mass spectral analysis of siRNA degradation in serum confirms an RNAse A-like activity. *Mol Biosyst 3*, 43-50.

(79) Xia, J., Noronha, A., Toudjarska, I., Li, F., Akinc, A., Braich, R., Frank-Kamenetsky, M., Rajeev, K. G., Egli, M., and Manoharan, M. (2006) Gene silencing activity of siRNAs with a ribo-difluorotoluyl nucleotide. *ACS Chem Biol 1*, 176-83.

(80) Berezhna, S. Y., Supekova, L., Supek, F., Schultz, P. G., and Deniz, A. A. (2006) siRNA in human cells selectively localizes to target RNA sites. *Proc Natl Acad Sci U S A 103*, 7682-7.

(81) de Fougerolles, A., Vornlocher, H. P., Maraganore, J., and Lieberman, J. (2007) Interfering with disease: a progress report on siRNA-based therapeutics. *Nat Rev Drug Discov 6*, 443-53.

(82) Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., Elbashir, S., Geick, A., Hadwiger, P., Harborth, J., John, M., Kesavan, V., Lavine, G., Pandey, R. K., Racie, T., Rajeev, K. G., Rohl, I., Toudjarska, I., Wang, G., Wuschko, S., Bumcrot, D., Koteliansky, V., Limmer, S., Manoharan, M., and Vornlocher, H. P. (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature 432*, 173-8.

(83) Zimmermann, T. S., Lee, A. C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M. N., Harborth, J., Heyes, J. A., Jeffs, L. B., John, M., Judge, A. D., Lam, K., McClintock, K., Nechev, L. V., Palmer, L. R., Racie, T., Rohl, I., Seiffert, S., Shanmugam, S., Sood, V., Soutschek, J., Toudjarska, I., Wheat, A. J., Yaworski, E., Zedalis, W., Koteliansky, V., Manoharan, M., Vornlocher, H. P., and MacLachlan, I. (2006) RNAi-mediated gene silencing in non-human primates. *Nature 441*, 111-4.

(84) Frank, J. (2002) Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu Rev Biophys Biomol Struct 31*, 303-19.

(85) Rodnina, M. V., Daviter, T., Gromadski, K., and Wintermeyer, W. (2002) Structural dynamics of ribosomal RNA during decoding on the ribosome. *Biochimie 84*, 745-54.

(86) Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., and Frank, J. (2003) Locking and unlocking of ribosomal motions. *Cell 114*, 123-34.

(87) Savelsbergh, A., Katunin, V. I., Mohr, D., Peske, F., Rodnina, M. V., and Wintermeyer, W. (2003) An elongation factor G-induced ribosome rearrangement precedes tRNA-mRNA translocation. *Mol Cell 11*, 1517-23.

(88) Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature 407*, 327-39.

(89) Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell 102*, 615-23.

(90) Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics 3*, 2.

(91) Ogle, J. M., Murphy, F. V., Tarry, M. J., and Ramakrishnan, V. (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell 111*, 721-32.

(92) Ogle, J. M., Carter, A. P., and Ramakrishnan, V. (2003) Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci 28*, 259-66.

(93) Woese, C. R., Gutell, R., Gupta, R., and Noller, H. F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev 47*, 621-69.

(94) Gutell, R. R., Larsen, N., and Woese, C. R. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev 58*, 10-26.

(95) Gabashvili, I. S., Agrawal, R. K., Grassucci, R., Squires, C. L., Dahlberg, A. E., and Frank, J. (1999) Major rearrangements in the 70S ribosomal 3D structure caused by a conformational switch in 16S ribosomal RNA. *EMBO J. 18*, 6501-7.

(96) Ogle, J. M., Broderson, D. E., Clemons Jr., W. M., Tarray, M. J., Carter, A. P., and Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science 292*, 897-902.

(97) Vila-Sanjurjo, A., Ridgeway, W. K., Seymaner, V., Zhang, W., Santoso, S., Yu, K., and Cate, J. H. (2003) X-ray crystal structures of the WT and a hyper-accurate ribosome from Escherichia coli. *Proc Natl Acad Sci U S A 100*, 8682-7.

(98) Leontis, N. B., and Westhof, E. (1998) A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J Mol Biol 283*, 571-83.

(99) Snoussi, K., and Leroy, J. L. (2001) Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry 40*, 8898-904.

(100) Pan, Y., and MacKerell, A. D., Jr. (2003) Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucleic Acids Res 31*, 7131-40.

(101) Namy, O., Rousset, J. P., Napthine, S., and Brierley, I. (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol Cell 13*, 157-68.

(102) Zamel, R., and Collins, R. A. (2002) Rearrangement of substrate secondary structure facilitates binding to the Neurospora VS ribozyme. *J Mol Biol 324*, 903-15.

(103) Lynch, S. R., Gonzalez, R. L., and Puglisi, J. D. (2003) Comparison of X-ray crystal structure of the 30S subunit-antibiotic complex with NMR structure of decoding site oligonucleotide-paromomycin complex. *Structure (Camb) 11*, 43-53.

(104) Goody, T. A., Melcher, S. E., Norman, D. G., and Lilley, D. M. (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *Rna 10*, 254-64.

(105) He, B., Rong, M., Lyakhov, D., Gartenstein, H., Diaz, G., Castagna, R., McAllister, W. T., and Durbin, R. K. (1997) Rapid mutagenesis and purification of phage RNA polymerases. *Protein Expr Purif 9*, 142-51.

(106) Walter, N. G. (2002) Probing RNA structural dynamics and function by fluorescence resonance energy transfer (FRET). *Curr. Protocols Nucleic Acid Chem. 11.10*, 11.10.1-11.10.23.

(107) Czworkowski, J., Odom, O. W., and Hardesty, B. (1991) Fluorescence study of the topology of messenger RNA bound to the 30S ribosomal subunit of Escherichia coli. *Biochemistry 30*, 4821-30.

(108) Grzesiek, S., and Bax, A. (1993) Measurement of amide proton exchange rates and NOEs with water in 13C/15N-enriched calcineurin B. *J Biomol NMR 3*, 627-38.

(109) Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR 2*, 661-5.

(110) Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry 28*, 6150-6.

(111) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR 6*, 277-93.

(112) Jeong, S., Sefcikova, J., Tinsley, R. A., Rueda, D., and Walter, N. G. (2003) Trans-acting hepatitis delta virus ribozyme: catalytic core and global structure are dependent on the 5' substrate sequence. *Biochemistry 42*, 7727-40.

(113) Walter, N. G. (2002) Probing RNA structural dynamics and function by fluorescence resonance energy transfer, in *Current Protocols in Nucleic Acid Chemistry* pp 11.10.1-11.10.23.

(114) Pereira, M. B., Harris, D. A., Rueda, D., and Walter, N. G. (2002) The reaction pathway of the trans-acting hepatitis delta virus ribozyme: a conformational change accompanies catalysis. *Biochemistry 41*, 730-740.

(115) Heus, H. A., and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science 253*, 191-4.

(116) Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol 264*, 968-80.

(117) Szewczak, A. A., and Moore, P. B. (1995) The sarcin/ricin loop, a modular RNA. *J Mol Biol 247*, 81-98.

(118) Seggerson, K., and Moore, P. B. (1998) Structure and stability of variants of the sarcin-ricin loop of 28S rRNA: NMR studies of the prokaryotic SRL and a functional mutant. *Rna 4*, 1203-15.

(119) Hobartner, C., and Micura, R. (2003) Bistable secondary structures of small RNAs and their structural probing by comparative imino proton NMR spectroscopy. *J Mol Biol 325*, 421-31.

(120) Wilms, C., Noah, J. W., Zhong, D., and Wollenzien, P. (1997) Exact determination of UV-induced crosslinks in 16S ribosomal RNA in 30S ribosomal subunits. *Rna 3*, 602-12.

(121) Noah, J. W., Shapkina, T., and Wollenzien, P. (2000) UV-induced crosslinks in the 16S rRNAs of Escherichia coli, Bacillus subtilis and Thermus aquaticus and their implications for ribosome structure and photochemistry. *Nucleic Acids Res 28*, 3785-92.

(122) Shapkina, T., Lappi, S., Franzen, S., and Wollenzien, P. (2004) Efficiency and pattern of UV pulse laser-induced RNA-RNA cross-linking in the ribosome. *Nucleic Acids Res 32*, 1518-26.

(123) Pape, T., Wintermeyer, W., and Rodnina, M. V. (1998) Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome. *Embo J 17*, 7490-7.

(124) Yurke, B., Turberfield, A. J., Mills, A. P., Jr., Simmel, F. C., and Neumann, J. L. (2000) A DNA-fuelled molecular machine made of DNA. *Nature 406*, 605-8.

(125) Tung, C. S., Joseph, S., and Sanbonmatsu, K. Y. (2002) All-atom homology model of the Escherichia coli 30S ribosomal subunit. *Nat Struct Biol 9*, 750-5.

(126) Walter, N. G., and Burke, J. M. (1997) Real-time monitoring of hairpin ribozyme kinetics through base-specific quenching of fluorescein-labeled substrates. *RNA 3*, 392-404.

(127) Pioletti, M., Schlunzen, F., Harms, J., Zarivach, R., Gluhmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., Hartsch, T., Yonath, A., and Franceschi, F. (2001) Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *Embo J 20*, 1829-39.

(128) Gesteland, R. F., Cech, T. R., and Atkins, J. F. E. (1999) *The RNA World, Second Edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

(129) Katunin, V. I., Muth, G. W., Strobel, S. A., Wintermeyer, W., and Rodnina, M. V. (2002) Important contribution to catalysis of peptide bond formation by a single ionizing group within the ribosome. *Mol Cell 10*, 339-46.

(130) Fritzsche, H., Kan, L. S., and Ts'o, P. O. (1983) Nuclear magnetic resonance study on the exchange behavior of the NH-N protons of a ribonucleic acid miniduplex. *Biochemistry 22*, 277-80.

(131) Lovett, S. T. (2004) Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol Microbiol 52*, 1243-53.

(132) Hoffmann, B., Mitchell, G. T., Gendron, P., Major, F., Andersen, A. A., Collins, R. A., and Legault, P. (2003) NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site. *Proc Natl Acad Sci U S A 100*, 7003-8.

(133) Mulrooney, S. B., Fishel, R. A., Hejna, J. A., and Warner, R. C. (1996) Preparation of figure 8 and cruciform DNAs and their use in studies of the kinetics of branch migration. *J Biol Chem 271*, 9648-59.

(134) Yamada, K., Ariyoshi, M., and Morikawa, K. (2004) Three-dimensional structural views of branch migration and resolution in DNA homologous recombination. *Curr Opin Struct Biol 14*, 130-7.

(135) LeCuyer, K. A., and Crothers, D. M. (1994) Kinetics of an RNA conformational switch. *Proc Natl Acad Sci U S A 91*, 3373-7.

(136) Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol 288*, 911-40.

(137) Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res 31*, 3406-15.

(138) Correll, C. C., Freeborn, B., Moore, P. B., and Steitz, T. A. (1997) Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell 91*, 705-12.

(139) Kulinski, T., Bratek-Wiewiorowska, M. D., Zielenkiewicz, A., and Zielenkiewicz, W. (1997) Mg2+ dependence of the structure and thermodynamics of wheat germ and lupin seeds 5S rRNA. *J Biomol Struct Dyn 14*, 495-507.

(140) Reblova, K., Spackova, N., Stefl, R., Csaszar, K., Koca, J., Leontis, N. B., and Sponer, J. (2003) Non-Watson-Crick basepairing and hydration in RNA motifs: molecular dynamics of 5S rRNA loop E. *Biophys J 84*, 3564-82.

(141) Auffinger, P., Bielecki, L., and Westhof, E. (2004) Symmetric K+ and Mg2+ ion-binding sites in the 5S rRNA loop E inferred from molecular dynamics simulations. *J Mol Biol 335*, 555-71.

(142) VanLoock, M. S., Agrawal, R. K., Gabashvili, I. S., Qi, L., Frank, J., and Harvey, S. C. (2000) Movement of the decoding region of the 16 S ribosomal RNA accompanies tRNA translocation. *J Mol Biol 304*, 507-15.

(143) Moazed, D., and Noller, H. F. (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature 327*, 389-94.

(144) Spahn, C. M., Blaha, G., Agrawal, R. K., Penczek, P., Grassucci, R. A., Trieber, C. A., Connell, S. R., Taylor, D. E., Nierhaus, K. H., and Frank, J. (2001) Localization of the ribosomal protection protein Tet(O) on the ribosome and the mechanism of tetracycline resistance. *Mol Cell 7*, 1037-45.

(145) Moazed, D., and Noller, H. F. (1987) Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature 327*, 389-94.

(146) Oehler, R., Polacek, N., Steiner, G., and Barta, A. (1997) Interaction of tetracycline with RNA: photoincorporation into ribosomal RNA of Escherichia coli. *Nucleic Acids Res 25*, 1219-24.

(147) Murray, J. B., and Arnold, J. R. (1996) Antibiotic interactions with the hammerhead ribozyme:tetracyclines as a new class of hammerhead inhibitor. *Biochem J 317 (Pt 3)*, 855-60.

(148) Rogers, J., Chang, A. H., von Ahsen, U., Schroeder, R., and Davies, J. (1996) Inhibition of the self-cleavage reaction of the human hepatitis delta virus ribozyme by antibiotics. *J Mol Biol 259*, 916-25.

(149) Manning, G. S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys 11*, 179-246.

(150) Draper, D. E. (2004) A guide to ions and RNA structure. *Rna 10*, 335-43.

(151) Woodson, S. A. (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol 9*, 104-9.

(152) Draper, D. E., Grilley, D., and Soto, A. M. (2005) Ions and RNA folding. *Annu Rev Biophys Biomol Struct 34*, 221-43.

(153) Wedekind, J. E., and McKay, D. B. (2003) Crystal structure of the leadzyme at 1.8 angstrom resolution: Metal ion binding and the implications for catalytic mechanism and allo site ion regulation. *Biochemistry 42*, 9554-9563.

(154) Hermann, T., and Westhof, E. (1998) Exploration of metal ion binding sites in RNA folds by Brownian-dynamics simulations. *Structure 6*, 1303-14.

(155) Peske, F., Savelsbergh, A., Katunin, V. I., Rodnina, M. V., and Wintermeyer, W. (2004) Conformational changes of the small ribosomal subunit during elongation factor G-dependent tRNA-mRNA translocation. *J Mol Biol 343*, 1183-94.

(156) Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature 407*, 340-8.

(157) Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell 102*, 615-23.

(158) Rudisser, S., and Tinoco, I., Jr. (2000) Solution structure of Cobalt(III)hexammine complexed to the GAAA tetraloop, and metal-ion binding to G.A mismatches. *J Mol Biol 295*, 1211-23.

(159) Maderia, M., Horton, T. E., and DeRose, V. J. (2000) Metal interactions with a GAAA RNA tetraloop characterized by (31)P NMR and phosphorothioate substitutions. *Biochemistry 39*, 8193-200.

(160) Mundoma, C., and Greenbaum, N. L. (2002) Sequestering of Eu(III) by a GAAA RNA tetraloop. *J Am Chem Soc 124*, 3525-32.

(161) Lian, L.-Y., and Roberts, C. (1993) Effects of chemical exchange on NMR spectra, in *NMR of Macromolecules: A Practical Approach* (Roberts, G. C. K., Ed.) pp 153-181, Oxford University Press, Oxford.

(162) Gonzalez, R. L., Jr., and Tinoco, I., Jr. (1999) Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot. *J Mol Biol 289*, 1267-82.

(163) Tanaka, Y., and Taira, K. (2005) Detection of RNA nucleobase metalation by NMR spectroscopy. *Chem Commun (Camb)*, 2069-79.

(164) Harris, D. A., Tinsley, R. A., and Walter, N. G. (2004) Terbium-mediated footprinting probes a catalytic conformational switch in the antigenomic hepatitis delta virus ribozyme. *J Mol Biol 341*, 389-403.

(165) Rueda, D., Wick, K., McDowell, S. E., and Walter, N. G. (2003) Diffusely bound Mg2+ ions slightly reorient stems I and II of the hammerhead ribozyme to increase the probability of formation of the catalytic core. *Biochemistry 42*, 9924-36.

(166) Wijmenga, S., Mooren, M., and Hilbers, C. (1993) NMR of Nucleic Acids: From Spectrum to Structure, in *NMR of Macromolecules: A Practical Approach* (Roberts, G. C. K., Ed.) pp 217-288, Oxford University Press, Oxford.

(167) Varani, G., Aboul-ela, F., and Allain, F.-T. (1996) NMR investigation of RNA structure. *Progress in Nuclear Magnetic Resonance Spectroscopy 29*, 51-127.

(168) Kieft, J. S., and Tinoco, I., Jr. (1997) Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexammine. *Structure 5*, 713-21.

(169) Gdaniec, Z., Sierzputowska-Gracz, H., and Theil, E. C. (1998) Iron regulatory element and internal loop/bulge structure for ferritin mRNA studied by cobalt(III) hexammine binding, molecular modeling, and NMR spectroscopy. *Biochemistry 37*, 1505-12.

(170) Ciesiolka, J., Marciniec, T., and Krzyzosiak, W. (1989) Probing the environment of lanthanide binding sites in yeast tRNA(Phe) by specific metal-ion-promoted cleavages. *Eur J Biochem 182*, 445-50.

(171) Walter, N. G., Yang, N., and Burke, J. M. (2000) Probing non-selective cation binding in the hairpin ribozyme with Tb(III). *J Mol Biol 298*, 539-55.

(172) Hargittai, M. R., and Musier-Forsyth, K. (2000) Use of terbium as a probe of tRNA tertiary structure and folding. *Rna 6*, 1672-80.

(173) Matsumura, K., and Komiyama, M. (1997) Enormously fast RNA hydrolysis by lanthanide(III) ions under physiological conditions: eminent candidates for novel tools of biotechnology. *J Biochem (Tokyo) 122*, 387-94.

(174) Sigel, R. K., and Pyle, A. M. (2003) Lanthanide ions as probes for metal ions in the structure and catalytic mechanism of ribozymes. *Met Ions Biol Syst 40*, 477-512.

(175) Harris, D. A., and Walter, N. G. (2003) Probing RNA structure and metal-binding sites using terbium footprinting. *Curr. Protocols Nucleic Acid Chem. 6.8*, 6.8.1-6.8.8.

(176) Harris, D. A., and Walter, N. G. (2005) Terbium(III) footprinting as a probe of RNA structure and metal-binding sites, in *Handbook of RNA Biochemistry* (Hartmann, R. K., Bindereif, A., Schön, A., and Westhof, E., Eds.) pp 205-213, Wiley-VCH, Weinheim.

(177) Tinsley, R. A., Harris, D. A., and Walter, N. G. (2004) Magnesium dependence of the amplified conformational switch in the trans-acting hepatitis delta virus ribozyme. *Biochemistry 43*, 8935-45.

(178) Gutell, R. R., Weiser, B., Woese, C. R., and Noller, H. F. (1985) Comparative anatomy of 16-S-like ribosomal RNA. *Prog Nucleic Acid Res Mol Biol 32*, 155-216.

(179) Belanger, F., Gagnon, M. G., Steinberg, S. V., Cunningham, P. R., and Brakier-Gingras, L. (2004) Study of the functional interaction of the 900 Tetraloop of 16S ribosomal RNA with helix 24 within the bacterial ribosome. *J Mol Biol 338*, 683-93.

(180) DeLano, W. (2002), DeLano Scientific, San Carlos, CA.

(181) Klein, D. J., Moore, P. B., and Steitz, T. A. (2004) The contribution of metal ions to the structural stability of the large ribosomal subunit. *Rna 10*, 1366-79.

(182) Serra, M. J., Baird, J. D., Dale, T., Fey, B. L., Retatagos, K., and Westhof, E. (2002) Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *Rna 8*, 307-23.

(183) Clemons, W. M., Jr., Brodersen, D. E., McCutcheon, J. P., May, J. L., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2001) Crystal structure of the 30 S ribosomal subunit from Thermus thermophilus: purification, crystallization and structure determination. *J Mol Biol 310*, 827-43.

(184) Reblova, K., Spackova, N., Koca, J., Leontis, N. B., and Sponer, J. (2004) Long-residency hydration, cation binding, and dynamics of loop E/helix IV rRNA-L25 protein complex. *Biophys J 87*, 3397-412.

(185) Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell 110*, 563-74.

(186) Amarzguioui, M., Holen, T., Babaie, E., and Prydz, H. (2003) Tolerance for mutations and chemical modifications in a siRNA. *Nucleic Acids Res 31*, 589-95.

(187) Hoerter, J. A., and Walter, N. G. (2007) Chemical modification resolves the asymmetry of siRNA strand degradation in human blood serum. *Rna in press*.

(188) Ohrt, T., Merkle, D., Birkenfeld, K., Echeverri, C. J., and Schwille, P. (2006) In situ fluorescence analysis demonstrates active siRNA exclusion from the nucleus by Exportin 5. *Nucleic Acids Res 34*, 1369-80.

(189) Raemdonck, K., Remaut, K., Lucas, B., Sanders, N. N., Demeester, J., and De Smedt, S. C. (2006) In situ analysis of single-stranded and duplex siRNA integrity in living cells. *Biochemistry 45*, 10614-23.

(190) Uhler, S. A., Cai, D., Man, Y., Figge, C., and Walter, N. G. (2003) RNA degradation in cell extracts: real-time monitoring by fluorescence resonance energy transfer. *J Am Chem Soc 125*, 14230-1.

(191) Kim, D. H., Behlke, M. A., Rose, S. D., Chang, M. S., Choi, S., and Rossi, J. J. (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat Biotechnol 23*, 222-6.

(192) Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res 11*, 1475-89.

(193) Pileur, F., Andreola, M. L., Dausse, E., Michel, J., Moreau, S., Yamada, H., Gaidamakov, S. A., Crouch, R. J., Toulme, J. J., and Cazenave, C. (2003) Selective inhibitory DNA aptamers of the human RNase H1. *Nucleic Acids Res 31*, 5776-88.

(194) Ramirez-Carrozzi, V., and Kerppola, T. (2001) Gel-based fluorescence resonance energy transfer (gelFRET) analysis of nucleoprotein complex architecture. *Methods 25*, 31-43.

(195) Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W., and Sontheimer, E. J. (2004) A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. *Cell 117*, 83-94.

(196) Monia, B. P., Johnston, J. F., Sasmor, H., and Cummins, L. L. (1996) Nuclease resistance and antisense activity of modified oligonucleotides targeted to Ha-ras. *J Biol Chem 271*, 14533-40.
(197) Raines, R. T. (1998) Ribonuclease A. *Chem Rev 98*, 1045-1066.
(198) Scacheri, P. C., Rozenblatt-Rosen, O., Caplen, N. J., Wolfsberg, T. G., Umayam, L., Lee, J. C., Hughes, C. M., Shanmugam, K. S., Bhattacharjee, A., Meyerson, M., and Collins, F. S. (2004) Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc Natl Acad Sci U S A 101*, 1892-7.
(199) Birmingham, A., Anderson, E. M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., Marshall, W. S., and Khvorova, A. (2006) 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods 3*, 199-204.
(200) Plum, G. E., Breslauer, K. J., and Roberts, R. W. (1999) Thermodynamics and Kinetics of Nucleic Acid Association/Dissociation and Folding Processes, in *Comprehensive Natural Products Chemistry: DNA and Aspects of Molecular Biology* (Kool, E. T., Ed.) pp 15-53, Elsevier, New York.
(201) Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *Embo J 20*, 6877-88.
(202) Weickmann, J. L., and Glitz, D. G. (1982) Human ribonucleases. Quantitation of pancreatic-like enzymes in serum, urine, and organ preparations. *J Biol Chem 257*, 8705-10.
(203) Libonati, M., and Sorrentino, S. (2001) Degradation of double-stranded RNA by mammalian pancreatic-type ribonucleases. *Methods Enzymol 341*, 234-48.
(204) Sorrentino, S., Naddeo, M., Russo, A., and D'Alessio, G. (2003) Degradation of double-stranded RNA by human pancreatic ribonuclease: crucial role of noncatalytic basic amino acid residues. *Biochemistry 42*, 10182-90.
(205) Yuan, B., Latek, R., Hossbach, M., Tuschl, T., and Lewitter, F. (2004) siRNA Selection Server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Res 32*, W130-4.
(206) Pei, Y., and Tuschl, T. (2006) On the art of identifying effective and specific siRNAs. *Nat Methods 3*, 670-6.
(207) Patzel, V. (2007) In silico selection of active siRNA. *Drug Discov Today 12*, 139-48.
(208) Dorsett, Y., and Tuschl, T. (2004) siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov 3*, 318-29.
(209) Manoharan, M. (2004) RNA interference and chemically modified small interfering RNAs. *Curr Opin Chem Biol 8*, 570-9.
(210) Chen, X., Dudgeon, N., Shen, L., and Wang, J. H. (2005) Chemical modification of gene silencing oligonucleotides for drug discovery and development. *Drug Discov Today 10*, 587-93.
(211) Eckstein, F. (2005) Small non-coding RNAs as magic bullets. *Trends Biochem Sci 30*, 445-52.
(212) Zhang, H. Y., Du, Q., Wahlestedt, C., and Liang, Z. (2006) RNA Interference with chemically modified siRNA. *Curr Top Med Chem 6*, 893-900.

(213) Snove, O., Jr., and Rossi, J. J. (2006) Chemical modifications rescue off-target effects of RNAi. *ACS Chem Biol 1*, 274-6.

(214) Lambert, M. N., Hoerter, J. A., Pereira, M. J., and Walter, N. G. (2005) Solution probing of metal ion binding by helix 27 from Escherichia coli 16S rRNA. *Rna 11*, 1688-700.

(215) Spickler, C., Brunelle, M. N., and Brakier-Gingras, L. (1997) Streptomycin binds to the decoding center of 16 S ribosomal RNA. *J Mol Biol 273*, 586-99.

(216) Allen, P. N., and Noller, H. F. (1991) A single base substitution in 16S ribosomal RNA suppresses streptomycin dependence and increases the frequency of translational errors. *Cell 66*, 141-8.

(217) Wollenzien, P., personal communication.

(218) Gutell, R. R.

(219) Cukras, A. R., Southworth, D. R., Brunelle, J. L., Culver, G. M., and Green, R. (2003) Ribosomal proteins S12 and S13 function as control elements for translocation of the mRNA:tRNA complex. *Mol Cell 12*, 321-8.

(220) Lee, K., Holland-Staley, C. A., and Cunningham, P. R. (1996) Genetic analysis of the Shine-Dalgarno interaction: selection of alternative functional mRNA-rRNA combinations. *Rna 2*, 1270-85.

(221) Belanger, F., Leger, M., Saraiya, A. A., Cunningham, P. R., and Brakier-Gingras, L. (2002) Functional studies of the 900 tetraloop capping helix 27 of 16S ribosomal RNA. *J Mol Biol 320*, 979-89.

(222) Belanger, F., Theberge-Julien, G., Cunningham, P. R., and Brakier-Gingras, L. (2005) A functional relationship between helix 1 and the 900 tetraloop of 16S ribosomal RNA within the bacterial ribosome. *Rna 11*, 906-13.

(223) Spiegel, P. C., Ermolenko, D. N., and Noller, H. F. (2007) Elongation factor G stabilizes the hybrid-state conformation of the 70S ribosome. *Rna 13*, 1473-82.

(224) Takyar, S., Hickerson, R. P., and Noller, H. F. (2005) mRNA helicase activity of the ribosome. *Cell 120*, 49-58.

(225) Thompson, R. E., Larson, D. R., and Webb, W. W. (2002) Precise nanometer localization analysis for individual fluorescent probes. *Biophys J 82*, 2775-83.

(226) Kural, C., Kim, H., Syed, S., Goshima, G., Gelfand, V. I., and Selvin, P. R. (2005) Kinesin and dynein move a peroxisome in vivo: a tug-of-war or coordinated movement? *Science 308*, 1469-72.

(227) Yildiz, A., Forkey, J. N., McKinney, S. A., Ha, T., Goldman, Y. E., and Selvin, P. R. (2003) Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science 300*, 2061-5.

(228) Tada, H., Higuchi, H., Wanatabe, T. M., and Ohuchi, N. (2007) In vivo real-time tracking of single quantum dots conjugated with monoclonal anti-HER2 antibody in tumors of mice. *Cancer Res 67*, 1138-44.

(229) Shamoo, Y., Adari, H., Konigsberg, W. H., Williams, K. R., and Chase, J. W. (1986) Cloning of T4 gene 32 and expression of the wild-type protein under lambda promoter PL regulation in Escherichia coli. *Proc Natl Acad Sci U S A 83*, 8844-8.

(230) Huang, W. M., Ao, S. Z., Casjens, S., Orlandi, R., Zeikus, R., Weiss, R., Winge, D., and Fang, M. (1988) A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science 239*, 1005-12.

(231) Dorywalska, M., Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S., and Puglisi, J. D. (2005) Site-specific labeling of the ribosome for single-molecule spectroscopy. *Nucleic Acids Res 33*, 182-9.

(232) Blainey, P. C., and Xie, X. S. (Unpublished).

(233) Ha, T., Rasnik, I., Cheng, W., Babcock, H. P., Gauss, G. H., Lohman, T. M., and Chu, S. (2002) Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature 419*, 638-41.

(234) van Oijen, A. M., Blainey, P. C., Crampton, D. J., Richardson, C. C., Ellenberger, T., and Xie, X. S. (2003) Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder. *Science 301*, 1235-8.

(235) Peacock, A. C., and Dingman, C. W. (1968) Molecular weight estimation and separation of ribonucleic acid by electrophoresis in agarose-acrylamide composite gels. *Biochemistry 7*, 668-74.

(236) Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., and Breaker, R. R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science 306*, 275-9.

(237) Winkler, W. C., and Breaker, R. R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol 59*, 487-517.

(238) Saxena, P., and Walker, J. R. (1992) Expression of argU, the Escherichia coli gene coding for a rare arginine tRNA. *J Bacteriol 174*, 1956-64.

(239) Chen, K. S., Peters, T. C., and Walker, J. R. (1990) A minor arginine tRNA mutant limits translation preferentially of a protein dependent on the cognate codon. *J Bacteriol 172*, 2504-10.

(240) Baker, M. L., Jiang, W., Rixon, F. J., and Chiu, W. (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. *J Virol 79*, 14967-70.

(241) Higuchi, A., Higuchi, Y., Furuta, K., Yoon, B., Hara, M., Maniwa, S., Saitoh, M., and Sanui, K. (2003) Chiral separation of phenylalanine by ultrafiltration through immobilized DNA membranes. *Journal of Membrane Science 221*, 207-18.

(242) Pangilinan, D. R. M., Michael L.; Hughes, John; Stevenson, David. (1990) Derivatization of amino acids with 2,3,4,6-tetra-O-acetyl-β -D-glucopyranosyl isothiocyanate and resolution of their diastereomers by RPHPLC, in *Proc. Am. Pept. Symp.* (Rivier, J. E. M., Garland R., Ed.) pp 802-3, ESCOM, Leiden.