AD_____

AWARD NUMBER:
W81XWH-16-1-0176

TITLE:
Chemical Library Screening for Potential Therapeutics Using Novel Cell-Based Models of ALS

PRINCIPAL INVESTIGATOR:
Keith T. Gagnon, Ph.D.

RECIPIENT:
Southern Illinois University, Carbondale, Illinois, 62901

REPORT DATE:
September 14, 2019

TYPE OF REPORT:
Final Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
09/14/19	Annual	05/15/16 - 05/14/19

4. TITLE AND SUBTITLE

Chemical Library Screening for Potential Therapeutics Using Novel Cell-Based Models of ALS

5a. CONTRACT NUMBER
W81XWH-16-1-0176

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Keith T. Gagnon, Ph.D.

E-Mail: ktgagnon@siu.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Southern Illinois University
1263 Lincoln Dr.
Carbondale, IL 62901

Stanford University
1291 Welch Road
CCSR-0133
Stanford, CA 94305

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The leading cause of inherited amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) is a microsatellite repeat expansion in the C9ORF72 gene. This disorder is referred to as C9FTD/ALS. The disease mechanisms are still poorly understood. However, it is clear that the repeat expansion sequence is made into RNA that can aggregate in the nucleus of patient cells. The expansion RNA can also be translated into repetitive polypeptides in the cytoplasm of patient cells. These two processes are expected to play key roles in the initiation and progression of disease at the molecular and cellular level. Drugs that can block or reverse these processes would hold promise as therapeutics to treat C9FTD/ALS. The overall goal of this project is to develop new cell-based models of C9FTD/ALS that recapitulate these two disease processes. RNA foci and repetitive polypeptides will be visible through fluorescence microscopy. These cells will then be used for high throughput chemical library screening to identify and characterize molecules with therapeutic potential. For this Final Report, we report on progress for years 1 and 2, as well as our no-cost extension. We will discuss our attempts to complete the original proposed tasks in the Statement of Work and our alternative strategies that we employed to continue moving toward therapeutic development for ALS. Although we made some changes to our research plan, we still expect to identify chemical compounds that will represent lead molecules for therapeutic development of C9FTD/ALS, the leading genetic cause of ALS, in the coming months and years as we try to find funding to continue this important project.

15. SUBJECT TERMS
ALS, repeat expansion, RNA, fluorescence, cell-based models, chemical library, high throughput screening

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	69	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. INTRODUCTION:

We are developing cell-based models of C9FTD/ALS, the most common inherited form of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), for chemical library screening to identify compounds that have promise for therapeutic development. The premise of this project was that cell models should recapitulate the two most basic disease processes, formation of repetitive RNA foci and production of repetitive poly-dipeptides, and these processes must be rapidly detectable through microscopy of live cells to enable high throughput cell-based screening. In this final report we discuss progress toward the original tasks, unexpected technical challenges that arose, some minor adjustments that were made in year 2, and the results of those experiments. During our no-cost extension phase, which was requested due to technical challenges, we made further experimental adjustments to continue moving forward and those will also be discussed. The final outcome can be summarized as 1) development of new and optimized methods for building cell models for C9FTD/ALS, 2) ongoing cell-based model development for future studies, 3) development of a new biochemical screen for small molecule discovery that is currently undergoing final screening, and 4) novel insight into cell-based model tools that led to three publications.

2. KEYWORDS:

Amyotrophic lateral sclerosis, frontotemporal dementia, C9FTD/ALS, repeat expansion, RNA, RAN translation, fluorescence, cell-based models, chemical library, high throughput screening

3. ACCOMPLISHMENTS:

We accomplished four important tasks which were either directly proposed as part of the original Statement of Work and research plan or emerged as solutions to unexpected problems. These include:

1) Development of new and optimized methods for building cell models for C9FTD/ALS
2) Ongoing cell-based model development for future studies
3) Development of a new biochemical screen for small molecule discovery
4) Novel insight into tools for cell-based model development

Major Goals of the Project Based on the Statement of Work (SOW).

Major Task 1: Establish neuronal cells that inducibly express fluorescent expRNA and poly-dipeptides.

Milestone Expected: Generate C9FTD/ALS cell-based model that recapitulates molecular pathology and is compatible with high throughput screening. (**80% complete**)

Subtask 1: Build custom expRNA expression vectors. (**100% complete**)
Subtask 2: Stably transform nH9 cells with TetR. (**100% complete**)
Subtask 3: Stably transform nH9-TR cells with custom vectors for expRNA expression. (**80% complete**)
Subtask 4: Validate c9FTD/ALS molecular pathology of new cell-based models. (**20% complete**)

Major Task 2: High throughput screening of chemical libraries at Stanford HTBC.

Milestone Expected: Identification of lead molecules for further validation and therapeutic development. (**50% complete**)

Subtask 1: Cell culture establishment at HTBC. (**0% complete**)
Subtask 2: Assay development and high throughput screen workflow protocol. (**100% complete**)
Subtask 3: Primary high throughput screen and titrations of top 1000 compounds. (**25% complete**)
Subtask 4: Data analysis, lead compound ranking, and chemical and structural analysis of lead compounds. (**0% complete**)

Major Task 3: Validate promising lead molecules using biochemical and cell-based assays.

Milestone Expected: Detailed characterization of several lead candidates and identification of top compounds to move into preclinical testing. (**0% complete**)

Subtask 1: Independently test lead compound effects on expRNA and poly-dipeptide expression/aggregation. (**0% complete**)
Subtask 2: Determine effect of lead compounds on other disease-associated molecular defects. (**0% complete**)

Accomplishments under these goals.

Major Task 1.

Our first major activity was the design and engineering of a custom inducible plasmid for expressing fluorescently-labeled repeat expansion RNA and poly-dipeptides. The specific objective was to create a plasmid with features to enhance the likelihood of developing successful cell-based models. These included i) ability to easily introduce or remove fluorescent tags using traditional restriction enzymes, ii) *in vitro* transcription if needed, iii) a designated site for cloning repeat expansions, and iv) inducible by doxycycline.

Significant results and key outcomes were the accomplishment of all the initial objectives of Subtask 1. However, as described below, significant changes had to be made once these vectors were constructed. We initially designed these vectors to inducibly express the Broccoli RNA aptamer fused to the repeat expansion followed by LUMIO tags. These were designed to facilitate detection of nuclear foci and poly-dipeptide translation inside of live cells, respectively, both of which are key markers of disease. We named this vector pINC-3G (**Figure 1**).

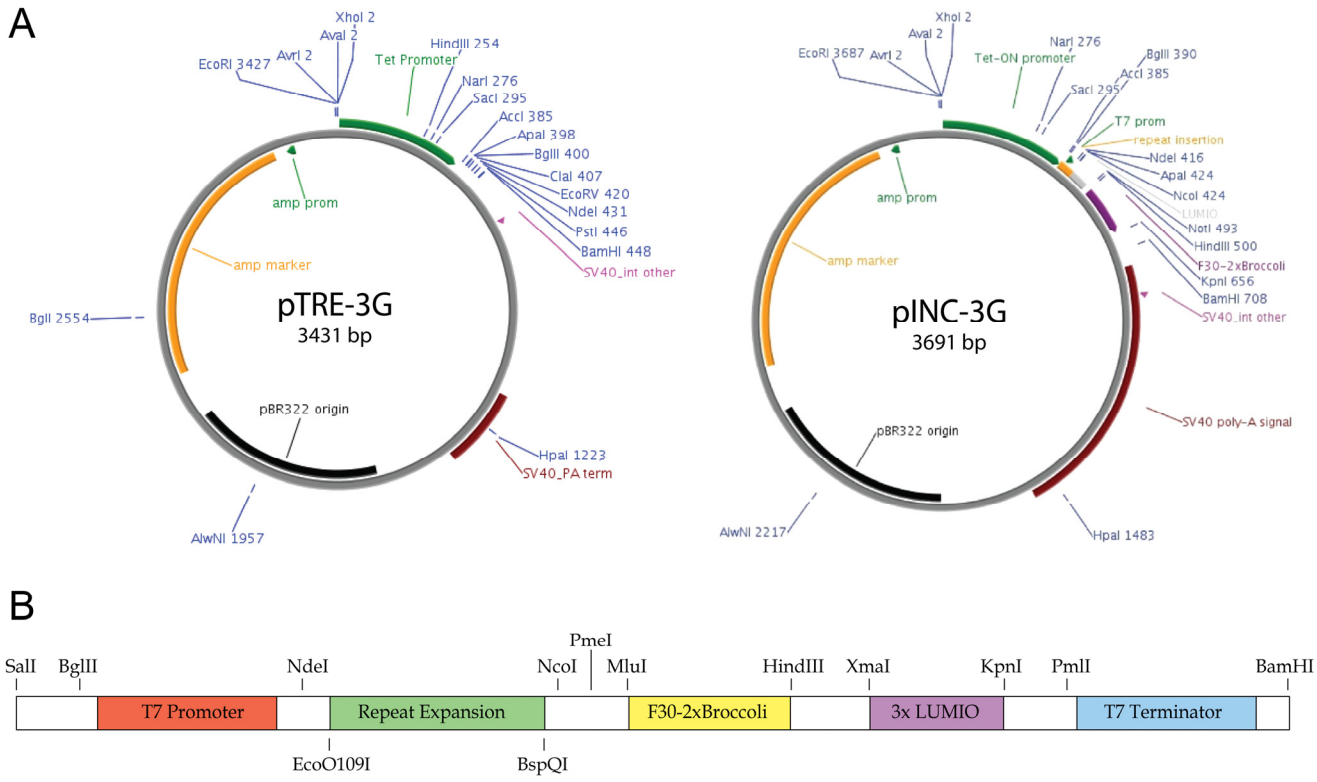


Figure 1. (**A**) Parent plasmid, pTRE-3G, is shown on left and the custom inducible plasmid we have engineered, pINC-3G, is shown on the right. (**B**) Schematic layout of custom MCS used to make pINC-3G and key functional elements.

We designed our expression plasmid based upon a commercially available Tet-On 3G plasmid, called pTRE-3G (Clontech). We modified this plasmid by removing its existing multiple cloning site (MCS) and inserting a custom MCS that was chemically synthesized (**Figure 1A**). This new MCS contained a broader range of compatible restriction endonuclease cleavage sites and a T7 RNA polymerase promoter and terminator sequence for *in vitro* transcription, if needed later (**Figure 1B**). This MCS swapping experiment required site-directed mutagenesis of an existing restriction site elsewhere on the plasmid. We have named the resulting custom plasmid that we have generated pINC3G. To test pINC3G for its ability to support inducible gene

expression, we cloned an EGFP protein into the MCS. Upon transient transfection of pINC3G-EGFP into HeLa Tet-On 3G cells (which express the appropriate tetracycline receptor, or TR, protein), we observed strong fluorescence in the presence of doxycycline (**Figure 2A**).

To prepare the plasmid for fluorescent labeling of RNA foci and poly-dipeptides, which are the hallmarks of disease that will be used as readouts for chemical screening, we inserted a dimeric Broccoli RNA aptamer sequence (Filonov et al., 2014) downstream of the future repeat cloning site (**Figure 1B**). The aptamer tag should enable visualization of repeat expansion RNA foci in live cells. We then inserted a 3x LUMIO tag downstream of the Broccoli tag (**Figure 1B**). The 3x LUMIO tag is in three different reading frames and should produce a tagged poly-dipeptide no matter what reading frame is used during translation of the repeat expansion RNA (Adams et al., 2002; Irtegun et al., 2011). In order to test the Broccoli aptamer tag, we cloned into pINC3G-dBroc-LUMIO a sequence encoding the non-coding RNAs U24 (snoRNA) or 7sK (snRNA). Upon transfection into cells, however, we saw very poor green fluorescence of the RNA. Expecting that this result is due to improper folding of dimeric Broccoli, we removed dimeric Broccoli from the plasmid and instead inserted two Broccoli aptamers embedded in a stably-folding three-way junction RNA called F30 (Filonov et al., 2015). We have expressed F30-2xBroccoli in HeLa TetOn 3G cells and still observed low fluorescence (**Figure 2B**). Upon closer inspection, we found that brightness issues were due to very rapid photobleaching of the broccoli RNA aptamer. This property has since been revealed in the literature by others (Autour et al., 2018), but was not known to us at the time. So, despite proper cloning and validation, the aptamer itself was intrinsically unstable for live cell imaging.

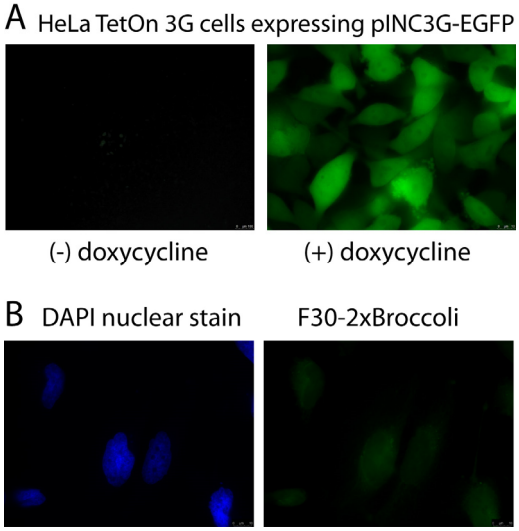


Figure 2. (**A**) Successful testing of pINC-3G plasmid using EGFP expression. (**B**) Successful expression of F30-2xBroccoli pINC-3G plasmid.

In an attempt to understand why the Broccoli RNA aptamer was unstable and to improve its activity, we investigated its properties. Since no other reasonable aptamers were available at the time of project initiation, we felt compelled to try and understand and improve the Broccoli RNA aptamer. We performed a straightforward biochemical investigation of its structure and stability. We performed a systematic mutagenesis screen as well. We uncovered key factors that controlled stability and fluorescence and discovered mutations that made Broccoli aptamer work better, which we named enhanced Broccoli, or eBroccoli. This application-driven project was published during the project period (Ageely et al., 2016) and represents one of the products (**Figure 3A**). We have incorporated these changes into our F30 scaffold and show that the new F30-2x-eBroccoli aptamer provided fluorescence (**Figure 3B**), but it was still not substantially better than previous designs. Now that an additional aptamer is available, the Mango RNA aptamer (Autour et al., 2018; Dolgosheina et al., 2014), we will consider testing this aptamer in the future for its fluorescence when fused to repeat expansion RNA.

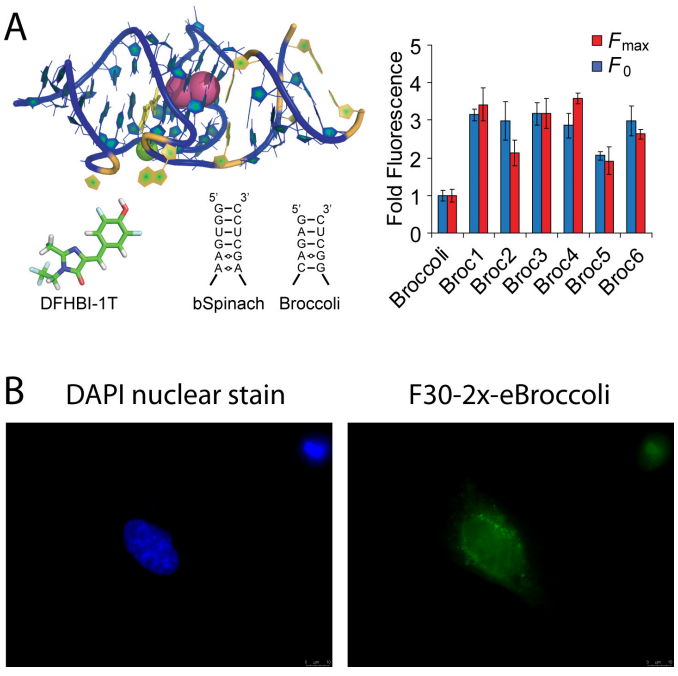


Figure 3. (**A**) Table of content graphic from our published investigation of fluorescent Broccoli RNA aptamer structure and stability. (**B**) Successful expression of F30-2x-eBroccoli pINC-3G plasmid.

Testing of protein tags requires non-canonical translation of repeat expansions in all three frames. Therefore, we began cloning C9FTD/ALS repeat expansions into pINC3G. To do so, we turned to a method known as recursive directional ligation, or RDL (Grabczyk and Usdin, 1999; Meyer and Chilkoti, 2002). In this method, a small 5 repeat sequence flanked by type II restriction enzyme sites is cloned into the plasmid. These repeats were either the sense expansion $(GGGGCC)_5$ or the antisense expansion $(CCCCGG)_5$. The type II

restriction sites were chosen based on a previous publication (Mizielinska et al., 2014) and required extensive site-directed mutagenesis of our pINC3G plasmid to remove redundant restriction sites. After cloning, the type II restriction enzymes are used to cut the plasmid and double the number of repeats with each round of RDL. Using this technique, we have now generated pINC3G plasmids with 80 repeats for sense expansions (**Figure 4A**). The minimum number needed to form foci and generate poly-dipeptide translation products is expected to be between 30-50 repeats (Green et al., 2016; Jain and Vale, 2017).

After initial cloning and testing of the vectors, we observed some unexpected challenges. We realized that our plasmids contained a canonical (standard) AUG start codon upstream of the repeat expansion cloning locus (**Figure 4B**). This will not allow us to recapitulate repeat-associated non-AUG (RAN) translation, which produces the disease poly-dipeptides, but instead will use standard translation mechanisms (Green et al., 2016). We also discovered that LUMIO tags were not easily detectable in a manner that was suitable for live cell imaging. The fluorescent dyes that must be added were relatively toxic to the cells, despite testing various conditions (data not shown), and were also not bright enough. As noted above, the Broccoli RNA aptamer also was not bright enough or reliable enough for tagging. Therefore, we have elected to omit the presence of a fluorescent RNA aptamer and the LUMIO tag and replace these with a single mCherry fluorescent protein tag (**Figure 4C**). If RAN translation occurs, this will result in poly-glycine-proline (poly-GP) dipeptide repeats being fused to the N-terminus of mCherry. Our rationale is that a simpler and proven protein tagging system can initially be used for screening. We plan to perform the screens using protein fluorescence of translated repetitive poly-dipeptides then follow with FISH as a secondary assay (**Figure 4D**). We also converted the AUG codon to a CUG start codon using site-directed mutagenesis (**Figure 4B**).

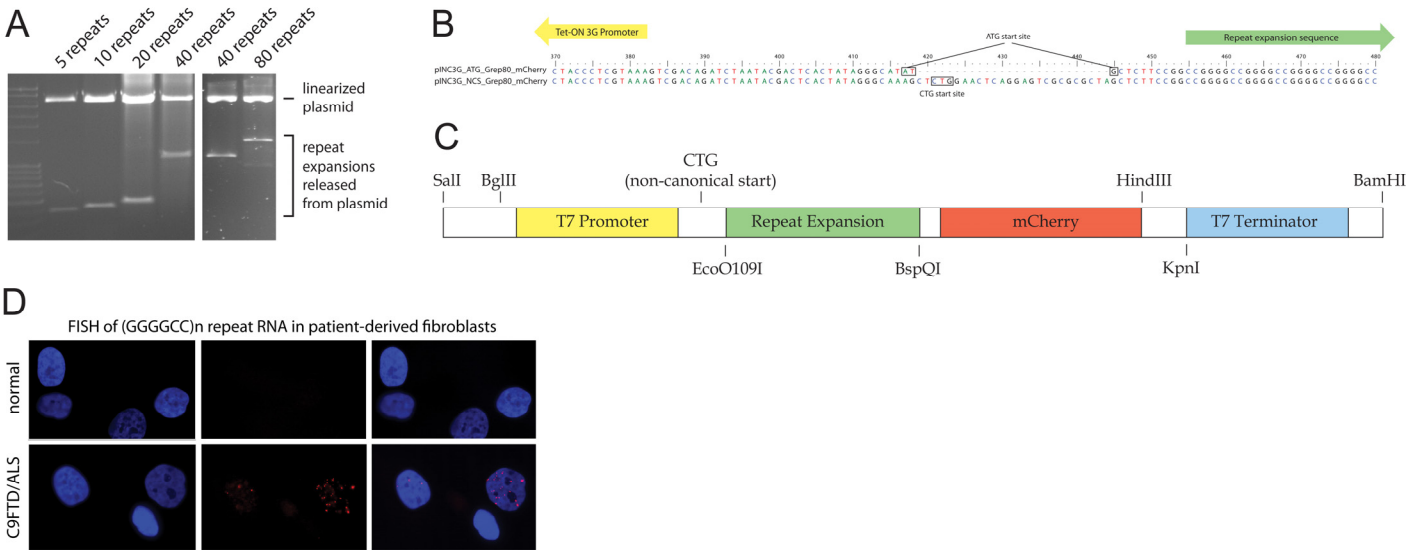


Figure 4. Cloning of additional expansions and second-generation modifications to vector. (**A**) pINC3G repeat expansion plasmids were treated with restriction enzymes to release a fragment containing the repeats. Separation of repeat fragments on an agarose gel reveals a step wise increase in size, indicating successful cloning of repeat expansions up to 80 repeats. (**B**) The pINC3G_NCS_Grep80_mCherry vector contains sequence derived directly from that C9ORF72 first intron that replaces the canonical ATG start site with a non-canonical CTG start site to support RAN translation. (**C**) The repeat construct fuses mCherry in-frame with a poly-GP dipeptide repeat sequence. (**D**) Example FISH data detecting focal aggregates of expansion RNA in patient-derived fibroblasts.

One of the ongoing challenges that we have faced is the lack of mechanistic insight regarding how RAN translation of repeat expansions actually works. Several publications have come out describing different mechanisms (Green et al., 2017; Rodriguez and Todd, 2019). RAN translation can initiate from many non-canonical start codons. Upon inspection of our pINC-3G expression vector, we found several NCS codons before the repeat expansion and in front of the mCherry gene. In fear that our high throughput screening might simply find inhibitors against normal translation and not RAN translation, we made a complete redesign, building from the lessons learned previously and using our existing repeat expansions. These new vectors, which we call pINC3G+, simply have a new MCS that contains a simple leader sequence, the start codon, the repeat cloning site, and an mCherry gene (**Figure 5**). All NCS codons except the initiation codon upstream of the repeat expansion have been removed. These plasmids produce mCherry fluorescence from an AUG and CUG start

codon but not a UAG codon (stop codon) (**Figure 5**). Now that we have an optimized system that we believe can recapitulate RAN translation, we will use these to finish the cell models that we initially set out to generate.

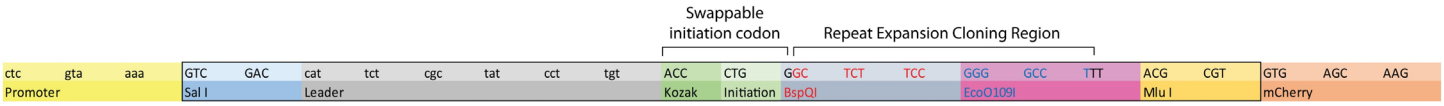


Figure 5. Third generation MCS and insert for pINC3G+ repeat expansion and RAN translation reporter vector.

Based on the lessons learned from generating cell-based models for C9FTD/ALS, we are unwilling to assume that these custom models will be sufficient. To that extent, we are exploring the use of CRISPR-Cas9 technology to embed fluorescent tags at endogenous genomic loci in patient-derived cells. To optimize this technology, we have established CRISPR in the laboratory and performed optimizations with chemically modified guides RNAs. This has led to two publications (Kartje et al., 2018; O'Reilly et al., 2019).

Our second major activity was to generate human neural stem cells that stably express a tetracycline receptor (TetR, or TR) protein. The specific objective was to create cells that would support doxycycline-inducible expression of the pINC3G plasmids we engineered. These cells are not commercially available and so must be custom made.

Significant results and key outcomes were the establishment of neural stem cell culture and demonstration of stable expression of the TR gene in these cells without significantly impact key gene expression markers of the neural cell lineage, proliferation and multipotency. We obtained pCMV-Tet3G (Clontech) plasmid containing the appropriate TR gene for Tet-ON 3G regulation. We obtained neural stem cells derived from human embryonic H9 stem cells. We have named these cells nH9 cells. We transfected the pCMV-Tet3G plasmid into nH9 cells and then performed antibiotic selection with neomycin. Surviving colonies were selected, expanded and tested for their ability to i) express sufficient levels of the TR gene and ii) maintain gene expression markers indicative of the parental cell lines. Upon isolation and qPCR we identified a few clonal populations that grow similarly to the parental nH9 cells and expressed the TR gene at sufficient levels (**Figure 6**). These cells are now prepared and ready for the next stage of cell model development.

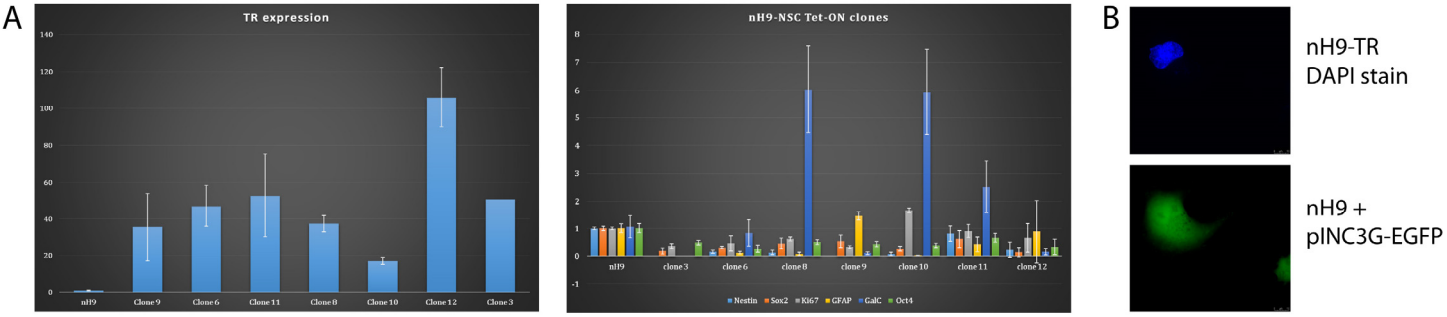


Figure 6. (**A**) Expression of the TR gene in nH9-TR clonal cell lines (left) and expression levels for genes that are markers for neuronal cell state, multipotent stem cell state, and proliferation state. Clone 11 was the most promising (**B**) Fluorescence microscopy demonstrating successful induction of expression from pINC-3G-EGFP in nH9-TR(11) cells.

Our third major activity was to generate nH9-TR cells (stably expressing the TR gene) that also contained stably-transfected pINC3G plasmids. The specific objective is to avoid the need to repeatedly transfect cells with the repeat expansion expressing plasmids, which is not feasible for high throughput chemical library screening. These cells, therefore, would contain all the necessary disease-associated genetic information, but would not present any of the disease-associated biomarkers, including foci and poly-dipeptides, until these are induced with doxycycline.

Significant results and key outcomes are the successful testing of pINC3G-G40-mCherry and pINC3G-G80-mCherry plasmid expression in nH9 cells by transient transfection (**Figure 7**). These experiments demonstrated that our nH9-TR cells are compatible with our pINC3G plasmids and that the mCherry expression was dependent on the addition of doxycycline. However, nH9 cells stably expressing our custom vectors have not been prepared yet due to the delay of generating our newest generation vector, pINC3G+. We now plan to extract the entire expression construct and insert into a lentivirus backbone for future stable cell line selections.

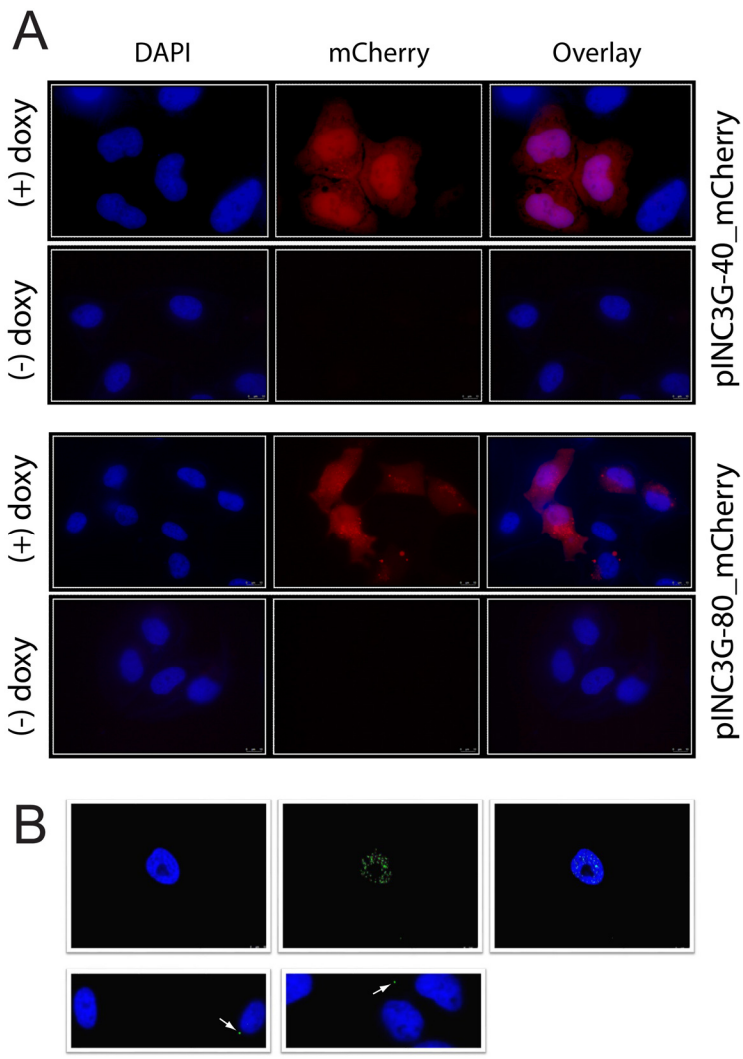


Figure 7. (**A**) Fluorescence microscopy demonstrating expression of 40 repeat and 80 repeat expression constructs when induced with doxycycline. (**B**) Foci detected by FISH in patient-derived fibroblasts (upper panels) and in cells expressing 80-repeat expansions fused to mCherry. A green (fluorescein) probe was used in these experiments.

Our fourth major activity of this task was to validate the cell-based models prepared in Subtasks 1-3 of Aim 1. The specific objective for this activity is to ensure that cell-based models are faithful representations of cellular disease. Positive results for other disease markers, including rRNA defects or TDP-43 mislocalization to the cytoplasm, would be desirable. However, these markers are not required and will not prevent successful screening with chemical libraries as long as fluorescent disease biomarkers established in Subtasks 1-3 are successful. Due to the delay in optimizing expression vectors, this Subtask 4 remains largely incomplete.

Major Task 2.

The first major activity was to establish cell culture at the Stanford High Throughput Bioscience Center (HTBC) for screening. However, because cell models were not generated in time, this subtask was not completed. Because we knew that our cell model generation was falling behind during our no-cost extension phase, we decided to develop an alternative screening method that did not rely on cell-based models. That system is described below.

Our second major activity of Major Task 2 was to perform a high-throughput screen at Stanford's HTBC. We have successfully initiated a high throughput screen at the HTBC using an alternative biochemical assay (**Figure 8**). The rationale behind this screen is the discovery that dimerization between two RNA polymerase II (RNAPII)-associated proteins, Supt4h1 and Supt5h, is essential for efficient transcription of the repeat expansion RNA from the C9ORF72 genomic locus (Kramer et al., 2016). Indeed, we stumbled upon this literature and many more publications supporting this mechanism when we performed a deep literature survey of repeat expansion RNA biology. We later published this survey as an extensive review (Rohilla and Gagnon, 2017).

To develop this screen, we prepared two synthetic genes. One was a fusion of Supt4h1 to the first half of a β-lactamase protein and the other was a fusion of Supt5h to the second half of β-lactamase protein. If the Supt4h1 and Supt5h proteins dimerize, then they will reconstitute full activity of the β-lactamase protein. The β-lactamase activity can then be measured by colorimetric change of nitrocefin, a molecule that contains a β-lactam bond (**Figure 9**). The colorimetric change is easily detected as yellow to red color shift (Galarneau et al., 2002). We have validated this assay in our laboratory and sent it to the Stanford HTBC.

The third major activity was to perform high throughput screening on the cell-based models we developed. Instead, we have performed an initial round of high throughput screening using our alternative biochemical assay. The Stanford HTBC as done this step and we are awaiting the results. We expect that this will identify a few compounds of interest and establish the feasibility of expanding this screen for more compounds.

Major Task 3.

Our remaining major activities are to validate lead molecules and their effects on disease markers. The stated goals for these activities have not been met yet and there are no other achievements to discuss for these tasks. We expect to identify some lead molecules from our initial biochemical screen that will serve as initial steps in future studies.

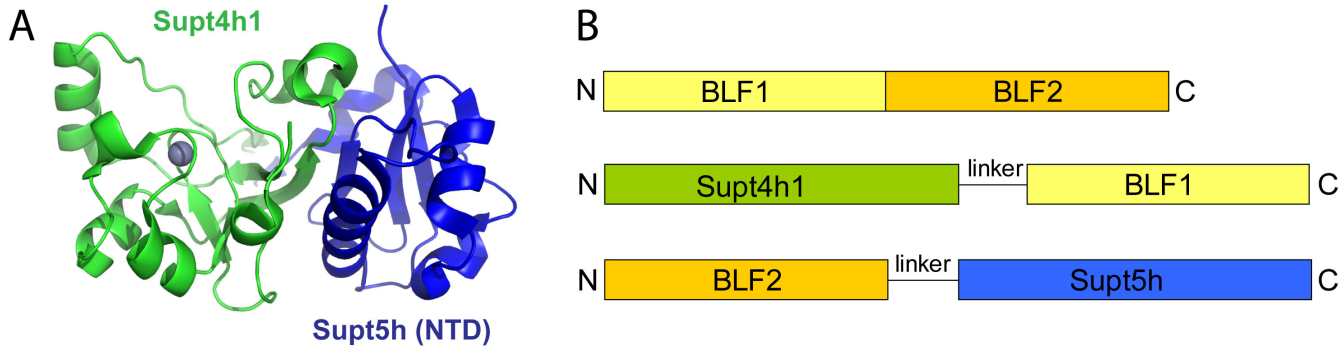


Figure 8. (**A**) Crystal structure of the Supt4h1-Supt5h protein dimer. (**B**) Gene fusions to β-lactamase enzyme halves.

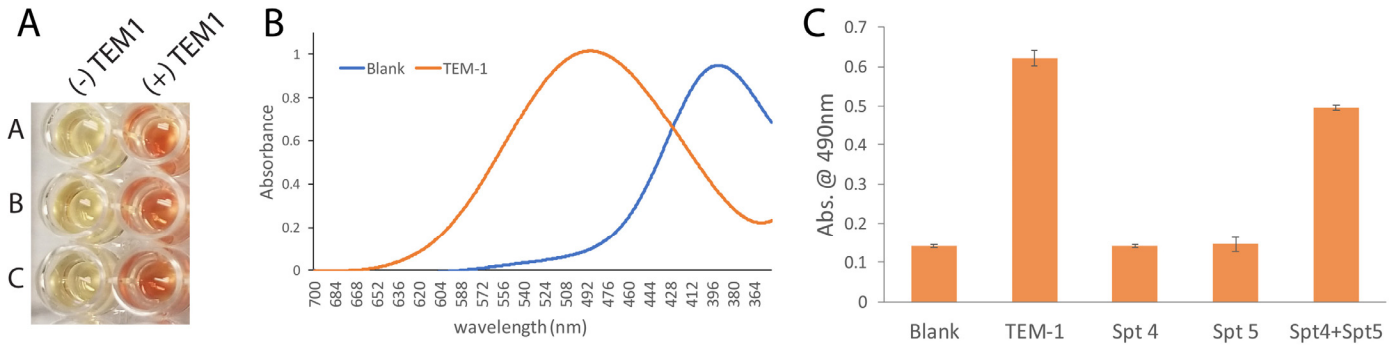


Figure 9. (**A**) Example of the colorimetric readout from nitrocefin in our biochemical assay screen. (**B**) The yellow to red shift has about a 20% overlap in spectrum at 490 nm, the wavelength for absorbance reading. (**C**) Quantified absorbance after dimerization of Supt4h1 with Supt5h fusions. Sufficient signal-to-noise is achieved.

Opportunities for training and professional development.

A variety of students and young professionals have participated in this project. They include two undergraduate students (Mr. James Donohue and Mr. Anthony Henke), seven graduate students (Mr. Zachary Kartje, Mr. Christopher Barkau, Mrs. Kushal Rohilla, Mrs. Ayomi Hewavidana, Ms. Katy Ovington, and Ms. Ramadevi Chilamkurthy, Ms. Courtney Slavich), and a postdoctoral fellow (Dr. Maria Barton). Mr. Kartje and Mr. Barkau developed new approaches with CRISPR to be applied toward developing future models of C9FTD/ALS. Dr. Barton, Mr. Barkau, Mrs. Rohilla, Mr. Slavich, and Ms. Ovington developed cell models. Dr. Barton, Mrs. Rohilla, Mrs Hewavidana, Ms. Chilamkurthy, and Ms. Ovington have constructed repeat expression vectors. The undergraduates Mr. Donohue and Mr. Henke have participated in cloning repeat expansions.

Training of undergraduates by graduate students provided a fantastic opportunity for graduate students to learn how to teach molecular biology techniques. The undergraduates learned important research skills that will serve them well in their future careers. Mr. Donohue is now pursuing his M.S. at St. Louis University and Mr. Henke is now pursuing a Ph.D. at the University of Texas at Dallas.

Ms. Kushal Rohilla attended an international conference, RNA Metabolism in Neurological Disorders, in San Diego, California in 2016 and 2018 and was able to present our progress on this project. She also presented her research at a local SIU School of Medicine symposium and won first place poster prize. Another one of the graduate students, Mr. Zachary Kartje, attended a conference, The Oligonucleotide Therapeutics Society Meeting, in Montreal, Canada in 2017 and was able to present our progress on this project. Zachary also won a poster prize at this conference. Mr. Barkau and Dr. Barton both attended an international conference, the RNA Society Meeting, in Berkeley, California, in 2018. They were able to present their progress on this project and

related projects. Mr. Kartje presented his project at a local SIU School of Medicine Graduate Symposium and won first place for his oral presentation.

Dissemination of results to communities of interest.

Results were disseminated by conference presentations, abstracts, and publications (see *Appendix* material).

4. IMPACT:

Impact on the development of the principal disciplines of the project.

Our progress on cell-based models was presented as a poster at the recent RNA Society Meeting in Berkeley, California. Colleagues were interested in our custom expression vectors and cell-based models and their eventual availability. When we publish our cell-based models and the subsequent chemical library screening results, we believe we will make an important impact by shifting research practices. Currently there are no easy-to-access or easy-to-use cell-based models that are representative of the disease. We believe our work will fill this gap. There are also no informative studies using chemical libraries to identify lead compounds for C9FTD/ALS, so we believe we will add a tremendous contribution to the principal discipline when this project is complete by providing detailed insight into potential therapeutics.

We believe our innovative approach of inducible RAN translation coupled to fluorescent live cell imaging in neurons will help transform the standard for basic research in the field. Our characterization of the many moving parts involved in engineering successful cell-based models will provide a foundation for researchers in the principal discipline to build upon. Our cell-based models will also make it significantly easier for researchers outside of the field to enter in and make important contributions.

Impact on other disciplines.

We have published an aspect of this project regarding the function and optimization of the Broccoli RNA aptamer and on optimization of CRISPR-Cas9 for efficient catalysis by the Cas9 enzyme. We intend to use our experience in this area to build even better models of C9FTD/ALS in the future and to facilitate the use of patient-derived cells to make models that are more amenable to drug screening and biochemistry.

Impact on technology transfer.

There is nothing to report for this section.

What was the impact on society beyond science and technology?

There is nothing to report for this section.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change.

We did not propose any significant changes to our approach or SOW. To meet our objectives and timeline, we made minor changes in the use of aptamers for fluorescence imaging and in downstream validation of potential lead hits from screening. Changes in our approach have been necessitated by technical challenges, primarily centered around generation of custom expression vectors and the custom cells to use for screening. These have been described above in the Accomplishments section. We met these challenges with appropriate modifications and put a plan in place to finish this project during our no-cost extension. However, ongoing and unexpected challenges with establishment of faithful RAN translation expression vectors hampered our efforts and forced a complete redesign of our expression system. Nevertheless, determined to perform the screening that we promised, we developed a new biochemical assay for small molecule discovery, which we described above in the Major Task 2 section.

Actual or anticipated problems or delays, and actions or plans to resolve them.

We have encountered many unexpected problems during the course of this project that have resulted in delays that were not anticipated. These include technical challenges that were enumerated and described in the Accomplishments section above.

Changes that had a significant impact on expenditures.

The technical and personnel issues we encountered during year 1 (long delay in hiring lead postdoctoral associated) had a significant impact on expenditures. They resulted in less funds being spent. This is due to the delayed hiring of a postdoctoral fellow and a delay in sending cells to the HTBC for preliminary assay development. The technical challenges in year 2 and the no-cost extension phase resulted in an inability to produce the necessary cell-based models for screening. Nonetheless, we made wise use of the funds to develop a biochemical assay in parallel in case cell models were not prepared in time. We are happy to report that the biochemical assay was successful and underwent a round of initial screening at the HTBC. We are awaiting results.

Significant changes in use or care of human subjects

There is nothing to report for this section. Human subjects are not applicable to this proposal.

Significant changes in use or care of vertebrate animals.

There is nothing to report for this section. Vertebrate animals are not applicable to this proposal.

Significant changes in use of biohazards and/or select agents

There is nothing to report for this section.

6. PRODUCTS:

Journal publications (see full manuscripts attached).

Ageely, E.A., Kartje, Z.J., Rohilla, K., Barkau, C.L., and Gagnon, K.T. (2016) Quadruplex-flanking stem structures modulate the stability and metal ion preferences of RNA mimics of GFP. *ACS Chem. Biol.*, 11:2398-2406. Support acknowledged.

Rohilla, K.J., and Gagnon, K.T. (2017) RNA Biology of Disease-Associated Microsatellite Repeat Expansions. *Acta Neuropathologica Communications*. 5:63. Support Acknowledged.

Kartje, Z.J., Barkau, C.L., Rohilla, K.J., Ageely, E.A., and Gagnon, K.T. (2018) Chimeric Guides Probe and Enhance Cas9 Biochemical Activity. *Biochemistry*. 57:3027-3031. Support Acknowledged.

O'Reilly, D., Kartje, Z.J., Ageely, E.A., Malek-Adamian, E., Habibian, M., Schofield, A., Barkau, C.L., Rohilla, K.J., DeRossett, L.B., Weigle, A.T., Damha, M.J., and Gagnon, K.T. (2018) Extensive CRISPR RNA Modification Reveals Chemical Compatibility and Structure-Activity Relationships for Cas9 Biochemical Activity. *Nucl. Acids Res.*, 47:546-558. Support Acknowledged.

Books or other non-periodical, one-time publications.

There is nothing to report for this section.

Other publications, conference papers, and presentations.

4th RNA Metabolism in Neurological Diseases Conference, November 10-11, 2016, San Diego, CA. Poster presentation. "Cell-based models of repeat expansion disease for cellular and molecular biochemistry."

Invited Seminar, Department of Chemistry, St. Louis University, St. Louis, Missouri, February, 2017. Oral presentation. "Broccoli and CRISPR and What's Cooking in the RNA Kitchen."

27th Annual Trainee Research Symposium, Southern Illinois University School of Medicine, Carbondale, IL, April 21, 2017. Poster presentation. "Cell-Based Models of Repeat Expansion Disease for Cellular and Molecular Biochemistry."

22nd Annual Meeting of the RNA Society, May 30 - June 4th, Prague, Czech Republic. Poster presentation. "Structure-Function Studies of Broccoli RNA Aptamer for Better Fluorescence."13th Annual Meeting of the Oligonucleotide Therapeutics Society, Bordeaux, France, IL, September 24 – September 27, 2017. Invited Lecture. "A 'Guided' Tour of an Early Career in RNA and Nucleic Acid Therapeutics."

28th Annual Trainee Research Symposium, Southern Illinois University School of Medicine, Carbondale, IL, April 21, 2017. Poster presentation. "Nuclear turnover and export mechanisms for repeat expansion RNA in C9FTD/ALS."

28th Annual Trainee Research Symposium, Southern Illinois University School of Medicine, Carbondale, IL, April 21, 2017. Oral presentation. "Chimeric Guides in CRISPR-Cas9."

23rd Annual Meeting of the RNA Society, May 29 - June 3, Berkeley, California. Poster presentation. "Cell-based models of C9FTD/ALS for discovery of small molecule inhibitors of repeat-associated non-AUG (RAN) translation."

23rd Annual Meeting of the RNA Society, May 29 - June 3, Berkeley, California. Poster presentation. "Nuclear Export Mechanisms of Tandem Repeat Expansion RNA in C9FTD/ALS Neurological Disease."

Website(s) or other Internet site(s).

There is nothing to report for this section.

Technologies or techniques

There is nothing to report for this section.

Inventions, patent applications, and/or licenses

There is nothing to report for this section.

Other Products

There is nothing to report for this section.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

Individuals that have worked on the project.

Name:	Keith T. Gagnon
Project Role:	PI
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	1
Contribution to Project:	Dr. Gagnon supervised and oversaw research progress, experimental troubleshooting, and dissemination of findings.
Funding Support:	Southern Illinois University School of Medicine, DoD ALSRP (this award)

Name:	Kushal J. Rohilla
Project Role:	Graduate Student
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	14
Contribution to Project:	Ms. Rohilla has performed design and cloning of inducible repeat expansion expression plasmids.
Funding Support:	Judith and Jean Pape Adams ALS Research Grant, DoD ALSRP (this award)

Name: Christopher L. Barkau
Project Role: Graduate Student
Researcher Identifier (e.g. ORCID ID): N/A
Nearest person month worked: 8

Contribution to Project: Mr. Barkau has performed cloning of inducible repeat expansion expression plasmids.

Funding Support: SIU Graduate Research Fellowship, DoD ALSRP (this award)

Name: Zachary J. Kartje
Project Role: Graduate Student
Researcher Identifier (e.g. ORCID ID): N/A
Nearest person month worked: 19

Contribution to Project: Mr. Kartje has performed cloning of inducible repeat expansion expression plasmids, stem cell culture, and selection of nH9-TR stable cells. Also developed CRISPR systems for cell-based model development.

Funding Support: SIU Teaching Assistantship, DoD ALSRP (this award)

Name: Ayomia Hewavidana
Project Role: Graduate Student
Researcher Identifier (e.g. ORCID ID): N/A
Nearest person month worked: 9

Contribution to Project: Mrs. Hewavidana performed cloning of expression plasmids, cell culture, and FISH.

Funding Support: DoD ALSRP (this award)

Name: Maria Barton
Project Role: Postdoctoral Fellow
Researcher Identifier (e.g. ORCID ID): N/A
Nearest person month worked: 27

Contribution to Project: Dr. Barton has performed stem cell culture, cloning of repeat expansion expression plasmids, and FISH.

Funding Support: DoD ALSRP (this award) Name:

Name: Ramadevi Chilamkurthy
Project Role: Graduate Student
Researcher Identifier (e.g. ORCID ID): N/A
Nearest person month worked: 8

Contribution to Project: Ms. Chilamkurthy has performed cloning for vector preparation for this project.

Funding Support: SIU Teaching Assistantship, DoD ALSRP (this award)

Name: Courtney Slavich
Project Role: Graduate Student
Researcher Identifier (e.g. ORCID ID): N/A
Nearest person month worked: 3

Contribution to Project: Ms. Slavich assisted with FISH analysis of repeat RNA.

Funding Support: SIU Graduate Research Fellowship, DoD ALSRP (this award)

Name: Katy Ovington

Project Role:	Graduate Student
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	8
Contribution to Project:	Ms. Ovington has performed cloning for vector preparation for this project.
Funding Support:	SIU Teaching Assistantship, DoD ALSRP (this award)

Changes in the active other support of the PI since project start.

Previously funded grants that have completed:

Judith and Jean Pape Adams Foundation ALS Research Grant
Gagnon (PI) 0 CY person-months $60,000 02/01/17 - 01/31/18
Title: "C9ORF72 transcription and splicing as therapeutic targets for a genetic form of ALS"

SIU School of Medicine Discovery Science Grant
Gagnon (PI) 0 CY person-months $15,000 01/01/18 - 12/31/18
Title: "Discovering chromatin-associated long noncoding RNAs acting as mitotic bookmarks in human stem cells"

Other partnering organizations.

We have partnered with Stanford University's High Throughput Bioscience Center (HTBC) to provide high throughput chemical library screening as a service. They will provide us this service as part of the proposed research. The contact person at HTBC is the facility director Dr. David Solow-Cordero. The HTBC did not provide any services in this year 1 reporting period. However, they will be providing their service in the next reporting phase.

Organization Name:	Stanford University High Throughput Bioscience Center
Organization Location:	Stanford, California
Partners Contribution:	
Financial Support:	None.
In-kind Support:	None.
Facilities:	Provides high throughput robotics facility as a service for the chemical library screening phase of this project.
Collaboration:	The facility staff will help the project staff rank results and interpret results to identify promising lead compounds.
Personnel Exchanges:	One project staff from the PI institution may travel to the HTBC facility in reporting year 2 to help establish cell culture and assay conditions.

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS:

Nothing to report.

QUAD CHARTS:

Nothing to report.

9. APPENDICES:

See Appendix I below for publication and meeting abstracts. See Appendix II below for literature cited in the final report. Finally, published manuscripts funded by this research are attached at the end.

APPENDIX I.

Ageely, E.A., Kartje, Z.J., Rohilla, K., Barkau, C.L., and Gagnon, K.T. (2016) Quadruplex-flanking stem structures modulate the stability and metal ion preferences of RNA mimics of GFP. *ACS Chem. Biol.*, 11:2398-2406. Support acknowledged.

Abstract:
The spinach family of RNA aptamers are RNA mimics of green fluorescent protein (GFP) that have previously been designed to address the challenges of imaging RNA inside living cells. However, relatively low levels of free intracellular magnesium limited the practical use of these aptamers. Recent cell-based selections identified the broccoli RNA aptamer, which requires less magnesium for fluorescence, but the basis for magnesium preference remained unclear. Here, we find that the broccoli RNA structure is very similar to that of baby spinach, a truncated version of the spinach aptamer. Differences in stability and metal ion preferences between these two aptamers, and among broccoli mutants, are primarily associated with the sequence and structure of predicted quadruplex-flanking stem structures. Mutation of purine−purine pairs in broccoli at the terminal stem−quadruplex transition caused reversion of broccoli to a higher magnesium dependence. Unique duplex-to-quadruplex transitions in GFP-mimic RNAs likely explain their sensitivity to magnesium for stability and fluorescence. Thus, optimizations designed to improve aptamers should take into consideration the role of metal ions in stabilizing the transitions and interactions between independently folding RNA structural motifs.

4th RNA Metabolism in Neurological Diseases Conference, November 10-11, 2016, San Diego, CA. Poster presentation. "Cell-based models of repeat expansion disease for cellular and molecular biochemistry."

Abstract:
Access to simple cell-based models of neurological repeat expansion disease is critical for investigating biochemical mechanisms and for early therapeutic discovery. Most cell-based models for neurological disease, in particular repeat expansion diseases, are hard to access, can be challenging to use, or else do not sufficiently recapitulate disease at the cellular level. To help bridge this gap, we are engineering straightforward cell-based models of c9FTD/ALS, the leading genetic cause of frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS), designed to allow inducible expression of repeat expansions that can be tracked at the RNA and protein level by fluorescent and affinity tags. This model system should be readily amenable to other repeat expansion disorders. Our cell-based models incorporate the tetracycline receptor gene into commercially available human neural stem cells. These neural stem cells can then be transiently or stably transfected with vectors expressing repeat expansion sequences behind a tetracycline inducible promoter. We have built custom inducible plasmids and are establishing reliable protocols for repeat expansion cloning and expression. Repeat RNA and poly-dipeptides can be fused to modular tags, including small Broccoli RNA aptamers and tetra-cysteine peptide tags for fluorescence or affinity tags for purification. Expression levels are then controlled by doxycycline, which should allow studies like temporal expression and localization as well as step-by-step characterization of disease mechanism at the cellular and biochemical level.

27th Annual Trainee Research Symposium, Southern Illinois University School of Medicine, Carbondale, IL, April 21, 2017. Poster presentation. "Cell-Based Models of Repeat Expansion Disease for Cellular and Molecular Biochemistry."

Abstract:
Access to simple cell-based models of neurological repeat expansion disease is critical for investigating biochemical mechanisms and for early therapeutic discovery. Most cell-based models for these diseases are hard to access, can be challenging to use, or else do not sufficiently recapitulate disease at the cellular level. To help bridge this gap, we are engineering straightforward cell-based models of c9FTD/ALS, designed to allow inducible expression of repeat expansions that can be tracked at the RNA and protein level by fluorescent and affinity tags. This model system should be readily amenable to other repeat expansion disorders. Our cell-based model incorporates the tetracycline receptor gene into commercially available and relatively easy-to-use human neural stem cells. These neural stem cells can then be transiently or stably transfected with vectors expressing repeat expansion sequences behind a tetracycline inducible promoter. We have built custom inducible plasmids and are establishing reliable protocols for repeat expansion cloning and expression. Repeat RNA and poly-dipeptides can be fused to modular tags, including small Broccoli RNA aptamers and tetra-cysteine peptide tags for fluorescence or affinity tags for purification. Expression levels are then controlled by doxycycline, which should allow temporal studies of expression and localization as well as step-by-step characterization of disease mechanism at the cellular and biochemical level.

22nd Annual Meeting of the RNA Society, May 30 - June 4th, Prague, Czech Republic. Poster presentation. "Structure-Function Studies of Broccoli RNA Aptamer for Better Fluorescence."

Abstract:

The Spinach family of aptamers are RNA mimics of green fluorescent protein (GFP) that have previously been designed to address the challenges of imaging RNA inside living cells. However, relatively low levels of free intracellular magnesium limit the practical use of these aptamers. New cell-based selections identified the Broccoli RNA aptamer, which required less magnesium for fluorescence. However, the structure of Broccoli and the basis for lower magnesium dependence were unknown. Here we find that Broccoli RNA shares the same core quadruplex structure as Spinach and is nearly identical to the structure of a truncated version, Baby Spinach. Differences in stability and metal ion preferences between these two aptamers, and among Broccoli mutants tested, are primarily associated with the sequence, structure and stability of predicted quadruplex-flanking stem and stem-loop structures. Mutation of purine-purine pairs in Broccoli at the terminal stem-to-quadruplex transition caused reversion of Broccoli to a higher magnesium dependence. Unique duplex-to-quadruplex transitions in GFP-mimic RNAs likely explain their sensitivity to magnesium and certain other metal ions. Thus, optimizations designed to improve aptamers should pay careful attention to the role of transitions between distinct or independently folding RNA structural motifs. Systematic mutagenesis and comparative structure-function analyses have allowed us to rationally design an enhanced Broccoli aptamer, called eBroccoli, that exhibits better folding and higher stability than the original Broccoli aptamer.

Rohilla, K.J., and Gagnon, K.T. (2017) RNA Biology of Disease-Associated Microsatellite Repeat Expansions. *Acta Neuropathologica Communications*. 5:63. Support Acknowledged.

Abstract:

Microsatellites, or simple tandem repeat sequences, occur naturally in the human genome and have important roles in genome evolution and function. However, the expansion of microsatellites is associated with over two dozen neurological diseases. A common denominator among the majority of these disorders is the expression of expanded tandem repeat-containing RNA, referred to as xtrRNA in this review, which can mediate molecular disease pathology in multiple ways. This review focuses on the potential impact that simple tandem repeat expansions can have on the biology and metabolism of RNA that contain them and underscores important gaps in understanding. Merging the molecular biology of repeat expansion disorders with the current understanding of RNA biology, including splicing, transcription, transport, turnover and translation, will help clarify mechanisms of disease and improve therapeutic development.

Kartje, Z.J., Barkau, C.L., Rohilla, K.J., Ageely, E.A., and Gagnon, K.T. (2018) Chimeric Guides Probe and Enhance Cas9 Biochemical Activity. 57:3027-3031. Support Acknowledged.

Abstract:

DNA substitutions in RNA can probe the importance of A-form structure, 2′-hydroxyl contacts, and conformational constraints within RNA-guided enzymes. Using this approach, we found that Cas9 biochemical activity tolerated significant substitution with DNA nucleotides in the clustered regularly interspaced short palindromic repeat RNA (crRNA). Only minimal RNA content was needed in or near the seed region. Simultaneous substitution at all positions with predicted crRNA−Cas9 2′-hydroxyl contacts had no effect on enzyme activity. The trans-activating crRNA (tracrRNA) also tolerated >50% substitution with DNA. DNA substitutions in the tracrRNA-pairing region of crRNA consistently enhanced cleavage activity while maintaining or improving target specificity. Together, results point to a prominent role for guide:target A-form-like helical structure and a possible regulatory role for the crRNA−tracrRNA pairing motif. A model chimeric crRNA with high activity did not significantly alter RNP assembly or target binding but did reduce Cas9 ribonucleoprotein stability, suggesting effects through conformation or dynamics. Cas9 directed by chimeric RNA−DNA guides may represent a cost-effective synthetic or molecular biology tool for robust and specific DNA cleavage.

13th Annual Meeting of the Oligonucleotide Therapeutics Society, Bordeaux, France, IL, September 24 – September 27, 2017. Invited Lecture. "A ‘Guided’ Tour of an Early Career in RNA and Nucleic Acid Therapeutics."

Abstract:

This talk will discuss some of the history and future of RNA-guided enzymes in research and therapeutics. It is from the vantage point of a young investigator who has worked on RNA-guided enzymes throughout his career. These opinions may offer different perspectives and spark new ideas for future nucleic acid therapeutics. From RNAi to CRISPR-Cas9, and even RNase H-mediated ASOs, nucleic acid-guided enzymes have been an engine

of inspiration and innovation. Nucleic acid-guided enzymes have translated into ground-breaking laboratory tools and a new generation of therapeutics. The more we understand nucleic acid-guided enzymes, the more innovations we can envision. Before RNAi was discovered, complexes of small nucleolar noncoding RNAs and their associated proteins, known as snoRNPs, were the best-known examples of RNA-guided enzymes. Investigations revealed important insight into the structure-function relationships between the RNA guide and the enzyme and established biochemical principles for activity. Their manipulation for therapeutics, however, has never been fully realized. Nonetheless, the knowledge that was gained and the tools developed facilitated the subsequent characterization of RNAi. The utility of RNAi was quickly recognized because it was reminiscent of ASOs, where only a small nucleic acid guide was needed. Chemical modification of the guide RNA has since been the focus of tremendous therapeutic development. More recently, CRISPR-Cas systems have been discovered and their potential to revolutionize gene therapy has become obvious. Knowledge and tools from previous RNA-guided research contributed to rapid progress. However, these systems require introduction of a large RNP. Thus, a logical question was whether such systems would fit into the world of nucleic acid therapeutics? But a little creativity and some nucleic acid chemistry are suggesting ways in which nucleic acid therapeutics and CRISPR can join together for new therapeutic approaches. With siRNAs destined for the clinic, the excitement of CRISPR, and the ability to engineer new RNA-guided systems, nucleic acid-guided enzymes are expected to continue shaping therapeutics for the foreseeable future.

28th Annual Trainee Research Symposium, Southern Illinois University School of Medicine, Carbondale, IL, April 21, 2017. Poster presentation. "Nuclear turnover and export mechanisms for repeat expansion RNA in C9FTD/ALS."

Abstract:
The expansion of microsatellites is associated with over two dozen neurological diseases. A common denominator among the majority of these disorders is the expression of expanded tandem repeat-containing RNA, referred to as xtrRNA, which can mediate molecular disease pathology in multiple ways. To understand disease mechanism, we are focusing on xtrRNA export in C9FTD/ALS, a model repeat expansion disorder that is the leading genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). In C9FTD/ALS, the repeat expansion occurs in the first intron of the C9ORF72 gene yet xtrRNA is somehow exported and translated into toxic repetitive polypeptides. We hypothesize that xtrRNA escapes into the cytoplasm by hitch-hiking with a fraction of intron-retained mRNA. Two main mRNA export pathways may be utilized: nuclear RNA export factor 1 (NXF1)-mediated and chromosome region maintenance 1 (CRM1)-mediated export. We are knocking-down nucleocytoplasmic transport factors in these pathways, including ALY/REF, NXF1, TREX, CRM1 and UPF1, and monitoring xtrRNA nuclear focal aggregation (by fluorescence in situ hybridization), translation (by Western blot and immunofluorescence) and cellular distribution (by qPCR and North blot). Results and progress thus far will be presented. Identification of the pathway or specific proteins responsible for C9FTD/ALS xtrRNA export should shed light on natural RNA metabolism pathways, disease mechanisms and potential strategies for therapeutic intervention.

28th Annual Trainee Research Symposium, Southern Illinois University School of Medicine, Carbondale, IL, April 21, 2017. Oral presentation. "Chimeric Guides Probe and Regulate CRISPR-Cas9 Activity."

Abstract:
DNA substitutions in RNA can probe the importance of A-form structure, 2'-hydroxyl contacts, and conformational constraints within RNA-guided enzymes. Using this approach, we found that Cas9 endonuclease activity tolerated substantial substitutions in the CRISPR RNA (crRNA). Only minimal RNA content was needed in or near the seed region. The trans-activating crRNA (tracrRNA) also tolerated over 50% substitution with DNA. Substitutions in the tracrRNA-pairing region of the crRNA consistently enhanced biochemical cleavage activity while maintaining or improving target specificity. No specific 2'-hydroxyl contacts for crRNA were critical. Substitutions also suggested induced fit mechanisms during RNP assembly. Together, results point to a prominent role for guide:target A-form-like helical structure and a possible regulatory role for the crRNA-tracrRNA pairing motif. A highly active chimeric crRNA did not significantly alter RNP assembly or target binding but did reduce Cas9 ribonucleoprotein (RNP) stability, suggesting effects through conformation or dynamics. Unexpectedly, gene editing with many of these chimeric crRNAs was inefficient, a phenomenon that did not correlate with nuclease susceptibility or the stability of guide:target interaction. Partially restoring RNA into chimeric crRNAs confirmed the presence of a putative crRNA regulatory element in the tracrRNA-pairing region of crRNA that modulates enzyme activity. Chimeric crRNAs with significant DNA substitutions and robust gene editing were identified. Cas9 directed by chimeric RNA-DNA guides may represent a cost-effective synthetic or

molecular biology tool for robust and specific DNA cleavage. Further characterization of the putative regulatory element should help unlock more predictable tuning of CRISPR-Cas9 activity and gene editing outcomes.

23rd Annual Meeting of the RNA Society, May 29 - June 3, Berkeley, California. Poster presentation. "Cell-based models of C9FTD/ALS for discovery of small molecule inhibitors of repeat-associated non-AUG (RAN) translation."

Abstract:
Two dozen neurological repeat expansion disorders are known, many of which are caused by expression and translation of repeat expansion RNA. To better understand molecular mechanisms and screen for molecules with therapeutic potential, we are engineering relevant cell-based models focusing on a genetic form of frontotemporal dementia and amyotrophic lateral sclerosis called C9FTD/ALS. In C9FTD/ALS, large GGGGCC repeat expansions in the first intron of the C9ORF72 gene are transcribed into expanded tandem repeat-containing RNAs which are somehow translated through a non-canonical mechanism known as repeat-associated non-AUG (RAN) translation. Our cell-based models aim to express GGGGCC repeats of varying sizes fused to mCherry to monitor expression. We are collaborating with Stanford High Throughput Bioscience Center to screen chemical libraries and identify RAN translation inhibitors in these model cell lines, thus identifying potential lead compounds for therapeutic development or RAN translation research.

23rd Annual Meeting of the RNA Society, May 29 - June 3, Berkeley, California. Poster presentation. "Nuclear Export Mechanisms of Tandem Repeat Expansion RNA in C9FTD/ALS Neurological Disease."

Abstract:
The advent of high-throughput sequencing technologies has revealed that the human transcriptome comprises mostly non-protein-coding RNAs. A large class of these transcripts, the long noncoding RNAs (lncRNAs) have been implicated in a diverse array of cellular processes, but their many functions are only beginning to be appreciated. How these molecules operate mechanistically is even less well-understood, but many are believed to act in direct proximity to the chromatin. We are using a recently-developed technique called chromatin-associated RNA sequencing (ChAR-seq) to map RNA-chromatin contacts globally. To understand mechanisms of chromatin-associated lncRNA function, we are treating MCF-7 breast cancer cells with 17β-estradiol and will be presenting progress toward applying ChAR-seq to this well-characterized system. Our results will reveal how the RNA composition and localization of the chromatin correlates with a disease-relevant remodeling of gene expression.

O'Reilly, D., Kartje, Z.J., Ageely, E.A., Malek-Adamian, E., Habibian, M., Schofield, A., Barkau, C.L., Rohilla, K.J., DeRossett, L.B., Weigle, A.T., Damha, M.J., and Gagnon, K.T. (2018) Extensive CRISPR RNA Modification Reveals Chemical Compatibility and Structure-Activity Relationships for Cas9 Biochemical Activity. *Nucl. Acids Res.*, 47:546-558. Support Acknowledged.

Abstract:
CRISPR (clustered regularly interspaced short palindromic repeat) endonucleases are at the forefront of biotechnology, synthetic biology and gene editing. Methods for controlling enzyme properties promise to improve existing applications and enable new technologies. CRISPR enzymes rely on RNA cofactors to guide catalysis. Therefore, chemical modification of the guide RNA can be used to characterize structure-activity relationships within CRISPR ribonucleoprotein (RNP) enzymes and identify compatible chemistries for controlling activity. Here, we introduce chemical modifications to the sugar-phosphate backbone of Streptococcus pyogenes Cas9 CRISPR RNA (crRNA) to probe chemical and structural requirements. Ribose sugars that promoted or accommodated A-form helical architecture in and around the crRNA 'seed' region were tolerated best. A wider range of modifications were acceptable outside of the seed, especially D-2'-deoxyribose, and we exploited this property to facilitate exploration of greater chemical diversity within the seed. 2'-fluoro was the most compatible modification whereas bulkier O-methyl sugar modifications were less tolerated. Activity trends could be rationalized for selected crRNAs using RNP stability and DNA target binding experiments. Cas9 activity in vitro tolerated most chemical modifications at predicted 2'-hydroxyl contact positions, whereas editing activity in cells was much less tolerant. The biochemical principles of chemical modification identified here will guide CRISPR-Cas9 engineering and enable new or improved applications.

APPENDIX II.

References

Adams, S.R., Campbell, R.E., Gross, L.A., Martin, B.R., Walkup, G.K., Yao, Y., Llopis, J., and Tsien, R.Y. (2002). New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. J Am Chem Soc *124*, 6063-6076.

Ageely, E.A., Kartje, Z.J., Rohilla, K.J., Barkau, C.L., and Gagnon, K.T. (2016). Quadruplex-Flanking Stem Structures Modulate the Stability and Metal Ion Preferences of RNA Mimics of GFP. ACS chemical biology *11*, 2398-2406.

Autour, A., S, C.Y.J., A, D.C., Abdolahzadeh, A., Galli, A., Panchapakesan, S.S.S., Rueda, D., Ryckelynck, M., and Unrau, P.J. (2018). Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. Nature communications *9*, 656.

Dolgosheina, E.V., Jeng, S.C., Panchapakesan, S.S., Cojocaru, R., Chen, P.S., Wilson, P.D., Hawkins, N., Wiggins, P.A., and Unrau, P.J. (2014). RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. ACS chemical biology *9*, 2412-2420.

Filonov, G.S., Kam, C.W., Song, W., and Jaffrey, S.R. (2015). In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. Chem Biol *22*, 649-660.

Filonov, G.S., Moon, J.D., Svensen, N., and Jaffrey, S.R. (2014). Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. J Am Chem Soc *136*, 16299-16308.

Galarneau, A., Primeau, M., Trudeau, L.E., and Michnick, S.W. (2002). Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. Nature biotechnology *20*, 619-622.

Grabczyk, E., and Usdin, K. (1999). Generation of microgram quantities of trinucleotide repeat tracts of defined length, interspersion pattern, and orientation. Analytical Biochemistry *267*, 241-243.

Green, K.M., Glineburg, M.R., Kearse, M.G., Flores, B.N., Linsalata, A.E., Fedak, S.J., Goldstrohm, A.C., Barmada, S.J., and Todd, P.K. (2017). RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. Nature communications *8*, 2005.

Green, K.M., Linsalata, A.E., and Todd, P.K. (2016). RAN translation-What makes it run? Brain Res.

Irtegun, S., Ramdzan, Y.M., Mulhern, T.D., and Hatters, D.M. (2011). ReAsH/FlAsH labeling and image analysis of tetracysteine sensor proteins in cells. J Vis Exp.

Jain, A., and Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. Nature *546*, 243-247.

Kartje, Z.J., Barkau, C.L., Rohilla, K.J., Ageely, E.A., and Gagnon, K.T. (2018). Chimeric Guides Probe and Enhance Cas9 Biochemical Activity. Biochemistry *57*, 3027-3031.

Kramer, N.J., Carlomagno, Y., Zhang, Y.J., Almeida, S., Cook, C.N., Gendron, T.F., Prudencio, M., Van Blitterswijk, M., Belzil, V., Couthouis, J.*, et al.* (2016). Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts. Science *353*, 708-712.

Meyer, D.E., and Chilkoti, A. (2002). Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system. Biomacromolecules *3*, 357-367.

Mizielinska, S., Gronke, S., Niccoli, T., Ridler, C.E., Clayton, E.L., Devoy, A., Moens, T., Norona, F.E., Woollacott, I.O., Pietrzyk, J.*, et al.* (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science *345*, 1192-1194.

O'Reilly, D., Kartje, Z.J., Ageely, E.A., Malek-Adamian, E., Habibian, M., Schofield, A., Barkau, C.L., Rohilla, K.J., DeRossett, L.B., Weigle, A.T.*, et al.* (2019). Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity. Nucleic acids research *47*, 546-558.

Rodriguez, C.M., and Todd, P.K. (2019). New pathologic mechanisms in nucleotide repeat expansion disorders. Neurobiology of disease *130*, 104515.

Rohilla, K.J., and Gagnon, K.T. (2017). RNA biology of disease-associated microsatellite repeat expansions. Acta Neuropathol Commun *5*, 63.

ACS chemical biology

Letters

pubs.acs.org/acschemicalbiology

Quadruplex-Flanking Stem Structures Modulate the Stability and Metal Ion Preferences of RNA Mimics of GFP

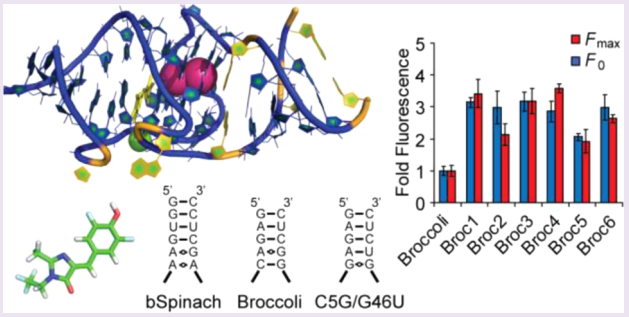
Eman A. Ageely,[†,§] Zachary J. Kartje,[†,§] Kushal J. Rohilla,[‡,§] Christopher L. Barkau,[‡] and Keith T. Gagnon*[,†,‡]

[†]Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, Illinois 62901, United States

[‡]Department of Biochemistry and Molecular Biology, Southern Illinois University, School of Medicine, Carbondale, Illinois 62901, United States

Supporting Information

ABSTRACT: The spinach family of RNA aptamers are RNA mimics of green fluorescent protein (GFP) that have previously been designed to address the challenges of imaging RNA inside living cells. However, relatively low levels of free intracellular magnesium limited the practical use of these aptamers. Recent cell-based selections identified the broccoli RNA aptamer, which requires less magnesium for fluorescence, but the basis for magnesium preference remained unclear. Here, we find that the broccoli RNA structure is very similar to that of baby spinach, a truncated version of the spinach aptamer. Differences in stability and metal ion preferences between these two aptamers, and among broccoli mutants, are primarily associated with the sequence and structure of predicted quadruplex-flanking stem structures. Mutation of purine−purine pairs in broccoli at the terminal stem−quadruplex transition caused reversion of broccoli to a higher magnesium dependence. Unique duplex-to-quadruplex transitions in GFP-mimic RNAs likely explain their sensitivity to magnesium for stability and fluorescence. Thus, optimizations designed to improve aptamers should take into consideration the role of metal ions in stabilizing the transitions and interactions between independently folding RNA structural motifs.



RNAs known as aptamers can be systematically evolved and selected to bind small molecules with high affinity and specificity.[1,2] Aptamers that bind fluorescent molecules can be used for visualizing RNA in synthetic biology applications, metabolite sensing, or monitoring dynamics inside living cells.[3−8] The unique three-dimensional structure of an aptamer confers its specificity and affinity for ligands, while positive counterions neutralize the negatively charged phosphate backbone to facilitate folding and stability.[9,10] Specific metal ions, such as magnesium and potassium, often bridge tertiary structures via hydrogen bonding networks and site-specific binding.[10−12] Thus, understanding the structure, stability, and metal ion interactions of aptamers promises to offer critical insight into their activity and design.

Fluorescent RNA aptamers have faced a number of challenges to their practical use as molecular tags for imaging RNA inside of cells. Early generation aptamers suffered from high background fluorescence and toxicity inside of cells, as well as low quantum yields.[13,14] More recently, RNA aptamers were selected to bind and activate fluorescence of chemically modified versions of HBI [(Z)-4-(4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one], the fluorophore that is generated within green fluorescent protein (GFP).[15] From these *in vitro* selections, the "spinach" RNA aptamer was identified, which fluoresces green similar to GFP. Additional versions of spinach, such as spinach2 (higher thermal stability) and baby spinach (minimal size), have been subsequently developed.[7,16] Diverse HBI fluorophore versions have also been reported, such as DFHBI-1T [(Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-1-(2,2,2-trifluoroethyl)-1H-imidazol-5(4H)-one)] (Figure 1A), which shifts the excitation and emission spectrum to better match that of GFP.[17] Despite low cellular toxicity, reduced background fluorescence, and relatively high quantum yields, a requirement for non-physiological levels of magnesium has complicated the use of spinach aptamers inside of cells.

To overcome magnesium requirements and low thermodynamic stability, new selections were developed using cell-based selection and directed evolution.[18] These selections resulted in "broccoli," an RNA aptamer that binds DFHBI-1T, possesses greater thermal stability in the presence of reduced magnesium, and is only 49 nucleotides, approximately the size of baby spinach. Our interest in using broccoli as a fluorescent tag for imaging RNA in living cells prompted us to investigate its structure and function. We reasoned that broccoli RNA structure was similar to that of the spinach RNA aptamers

Received: January 15, 2016
Accepted: July 28, 2016

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

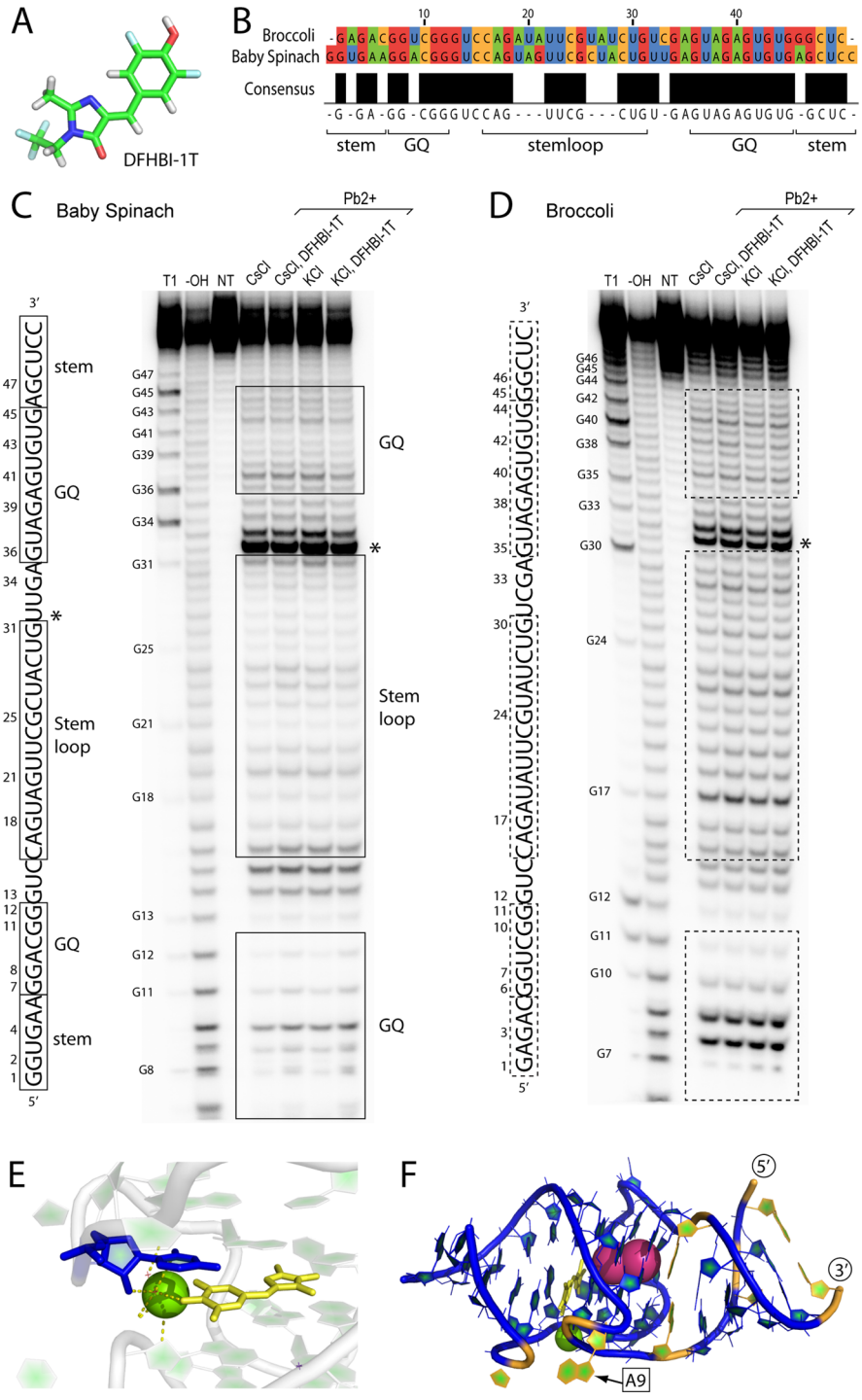


Figure 1. Sequence and structural comparisons support a strong similarity between baby spinach and broccoli RNA aptamers. (A) Structure of DFHBI-1T, the small molecule that is bound and fluorescently activated by baby spinach and broccoli aptamers. (B) Sequence alignment of broccoli and baby spinach RNAs. The consensus sequence is shown, and secondary structure alignment predicted from the spinach aptamer structure[16] (PDB ID: 4TS2) is indicated below. (C,D) Lead cleavage mapping of baby spinach and broccoli RNA structures. The sequence is shown to the left and labeled with structural alignment from the spinach aptamer structure[16] as well as nucleotide numbering. Lead cleavage mapping is shown, and the treatments for each lane are indicated above the gel. Bands on the gel that correspond to nucleotides in the sequence (shown to the left) are boxed. An asterisk indicates the base where magnesium ion binding is implicated. GQ, G-quadruplex; T1, T1 RNase cleavage ladder; −OH, alkaline hydrolysis ladder; NT, no treatment; Pb2+, addition of lead acetate. (E) Uridine U61 of spinach (blue), which corresponds to U32 of baby spinach,[16] is hydrogen-bonded to a water molecule coordinated by a magnesium ion (green), which also stabilizes DFHBI-1T (yellow) binding. (F) A fragment of the spinach RNA aptamer crystal structure[16] is shown that corresponds to the shared consensus sequence with baby spinach. Nucleotides that are identical between baby spinach and broccoli are shown in blue, whereas the few nucleotides that are different are shown in orange. A9, the only base found within the consensus quadruplex that differs between baby spinach and broccoli, is indicated. DFHBI-1T is shown in yellow. Potassium ions are shown as magenta spheres and coordinated magnesium ion as a green sphere.

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

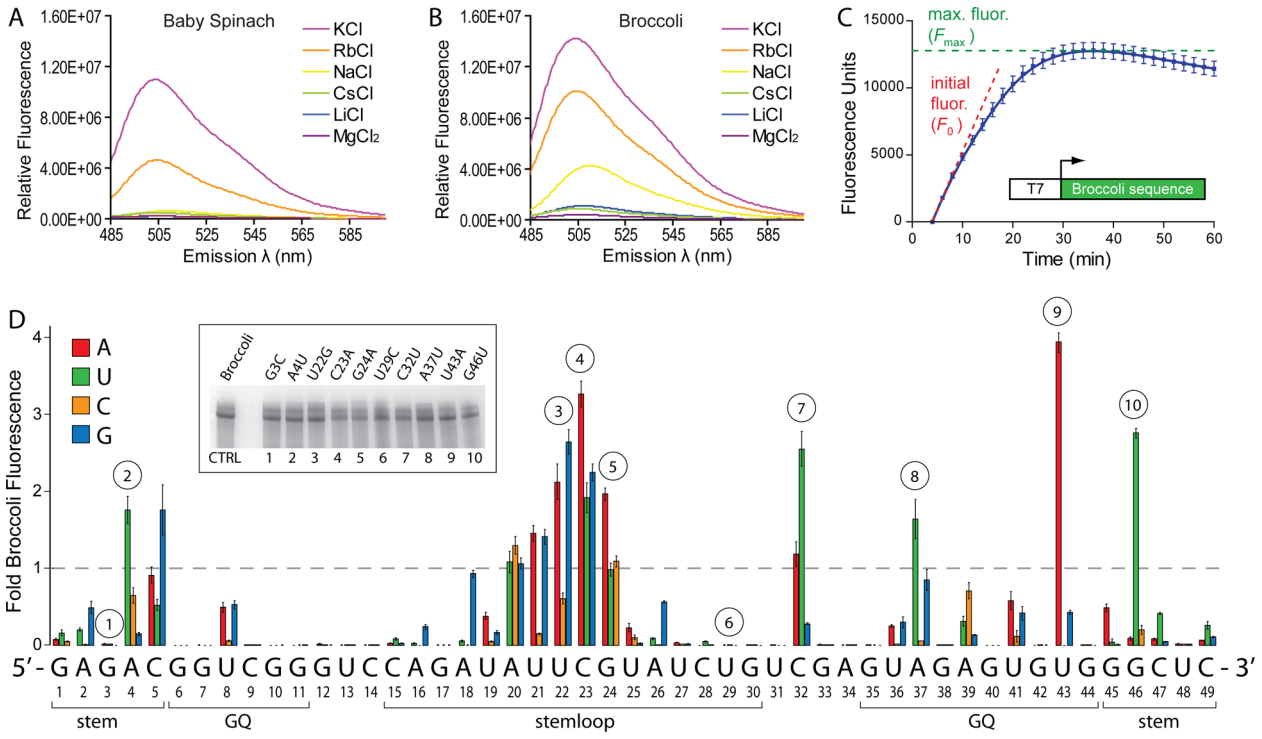


Figure 2. Metal-ion-dependent fluorescence and mutagenesis suggest the presence of a G-quadruplex and stem structures within broccoli RNA. Fluorescence emission is plotted for baby spinach (A) and broccoli (B) RNA aptamers in buffered solution containing DFHBI-1T and the indicated salts. (C) Example data for broccoli RNA using a transcription-based mutagenesis screen. T7 promoter DNA sequence is fused upstream of the broccoli sequence and used for *in vitro* transcription reactions containing the DFHBI-1T fluorophore. Initial fluorescence (F_0) and maximum fluorescence (F_{max}) are taken to approximate the relative folding and fluorescence activity of broccoli RNA concurrently with RNA synthesis. (D) Fluorescence results from systematic mutagenesis of broccoli RNA. Every nucleotide was individually substituted with one of three other nucleotides and fluorescence activity measured. Normalized activity of control broccoli RNA fluorescence is indicated by a dashed light gray line. Only F_{max} is plotted. Broccoli RNA sequence and nucleotide numbering are indicated below. Putative structures are indicated below and based upon the spinach crystal structure[16] and sequence alignment between broccoli and baby spinach. Mutant broccoli RNA reactions chosen for subsequent gel analysis (inset) by denaturing gel electrophoresis are indicated by circled numbers. Error bars are standard deviation from three separate analyses.

and in particular that it contains a G-quadruplex, considering that (1) it was selected using essentially the same fluorophore molecule that spinach activates, (2) spinach contains a G-quadruplex structure that binds DFHBI, (3) another fluorescent RNA aptamer that binds an HBI derivative was shown to have nuclear magnetic resonance chemical shifts indicative of G-quadruplexes, (4) a dependence for potassium ions was previously reported for spinach RNA fluorescence, and (5) broccoli contains a disproportionately high guanine to cytosine nucleotide ratio of 2:1.[16,18,19] In addition, simple alignment of the broccoli sequence with baby spinach sequence revealed high similarity (Figure 1B). Using the crystal structure of spinach as a guide,[16] the nucleotides that form its quadruplex core are nearly a perfect consensus between baby spinach and broccoli, with only a single A to U conversion at one position. Thus, we hypothesized that differences between broccoli and baby spinach fluorescence, magnesium preference, and stability likely arise from structural differences outside a quadruplex core.

To experimentally probe the structure of baby spinach and broccoli, we performed limited cleavage with lead acetate.[20] Lead is a divalent metal ion that accelerates phosphodiester bond cleavage. Cleavage correlates with regions of flexibility and accessibility within an RNA structure.[21,22] Both baby spinach and broccoli RNAs were refolded in the absence or presence of DFHBI-1T and either cesium chloride or potassium chloride. Potassium is known to stabilize G-quadruplexes by specific binding while cesium does not. We found no significant differences in the lead cleavage pattern of baby spinach or broccoli across these conditions (Figure 1C,D). However, overall lead cleavage patterns appeared quite similar between baby spinach and broccoli. These results suggest that large structural changes do not occur in the presence of potassium or DFHBI-1T and that baby spinach and broccoli possess similar structures.

Lead acetate is known to cause very strong phosphodiester bond cleavage at divalent metal ion binding sites within RNA structures,[21−23] thus acting as a specific probe for metal ion coordination. In agreement with this phenomenon, we observed excessive cleavage at the 3′ phosphodiester bond of uridine 32 (U32) in baby spinach, which corresponds to the same uridine in the spinach crystal structure that makes direct polar contact with a magnesium-coordinated water molecule via its 2′ hydroxyl (Figure 1E). This magnesium ion binds tightly and also helps stabilize the binding of DFHBI through polar contacts.[16] U31 of broccoli, which corresponds to U32 in baby spinach based on sequence alignment, also underwent very high levels of cleavage, suggesting that the same magnesium-dependent structure for binding DFHBI is also formed in the broccoli aptamer. Lead-induced cleavage was lower for U31 of broccoli, suggesting a less dynamic and more stable structure for broccoli around this nucleotide. By identifying the consensus sequence of spinach, baby spinach, and broccoli, we extrapolated the majority of the baby spinach aptamer

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

structure from the spinach crystal structure, except for part of the internal stem loop which is not present in the spinach crystal.[16,19] Overlaying the consensus and variable nucleotides of broccoli and baby spinach onto our extrapolated structure revealed where the differences between baby spinach and broccoli would likely lie within a structural context (Figure 1F). The only nucleotide that differed within the quadruplex core was A9 of baby spinach, corresponding to U8 of broccoli. Interestingly, A9 is extruded out of the spinach quadruplex core, is variable (replaced by a uridine in broccoli), and showed relatively strong lead-induced cleavage for broccoli. The accessibility to cleavage of U8 is in agreement with a predicted position on the periphery of the quadruplex core. Aside from A9/U8, all other variable nucleotides were in putative stem or stem-loop structures outside of the quadruplex-forming sequence. Together, these results indicate that broccoli and baby spinach are likely to possess very similar structures, including nearly identical quadruplex cores, and suggest that functional differences arise from quadruplex-flanking stem structures.

G-quadruplex structural stability is significantly enhanced in the presence of alkali metal ions, specifically potassium.[12] To further support the presence of a quadruplex structure in broccoli RNA and investigate the metal ion preferences of baby spinach and broccoli, we measured fluorescence activation of DFHBI-1T by each aptamer in the presence of 10 mM $MgCl_2$ or 50 mM of an alkali metal salt series: potassium (K^+), rubidium (Rb^+), sodium (Na^+), cesium (Cs^+), or lithium (Li^+) chloride. As controls, we used magnesium chloride alone and also tested the malachite green aptamer, which does not possess a quadruplex structure,[24] in the same salts and with its respective malachite green dye. Baby spinach showed strong fluorescence emission in the presence of K^+ ions but had a > 2-fold reduction in peak fluorescence with Rb^+ ions (Figure 2A). Virtually no fluorescence emission was observed in other salts. Broccoli fluorescence emission was higher than that of baby spinach in the presence of K^+ while Rb^+ also supported fluorescence quite well (Figure 2B). Sodium ions were also able to provide some degree of fluorescence emission, albeit at ~1/3 that of K^+ ions. Cesium, lithium, and magnesium alone supported little or no fluorescence of broccoli. Malachite green aptamer exhibited similar fluorescence emission spectra irrespective of the metal ions tested (Figure S1A). We performed titrations of KCl, NaCl, and LiCl and plotted the peak emission wavelength (507 nm) against salt concentration (Figure S1B,C). Baby spinach was found to respond to KCl with a 50% maximum fluorescence emission at about 10 mM (Figure S1B). However, no significant emission was detected in LiCl or NaCl up to 100 mM. In contrast, broccoli was nearly 2-fold more efficient in using K^+ ions with a KCl concentration of approximately 6 mM being sufficient to provide 50% maximum fluorescence emission (Figure S1C). Broccoli can also use sodium ions, but the concentration required for 50% maximum fluorescence was calculated to be over 200 mM.

G-quadruplexes uniquely coordinate potassium ions between stacked planar layers of guanine bases, which are formed by a combination of Watson−Crick and Hoogsteen base interactions.[25] Other ions of similar size and charge, such as rubidium and sodium, can sometimes also support quadruplex structure to a lesser extent.[12] In contrast, canonical Watson−Crick duplex structures usually have little preference for metal ion, which primarily act to neutralize negative phosphate backbone repulsions.[9,26] Likewise, cesium and lithium ions can sufficiently stabilize canonical nucleic acid helices by phosphate neutralization but do not stabilize G-quadruplexes very well.[12,27−29] Therefore, our results with various metal ions suggest that broccoli RNA possesses a G-quadruplex structure. The ability to activate fluorescence with sodium, however, demonstrates that broccoli's quadruplex is distinct from that of baby spinach. Because the predicted quadruplex-forming sequence is virtually unchanged between these two aptamers, flanking stem structures most likely influence alkali metal preferences by altering the local stability around the quadruplex.

To further correlate structure with function for the broccoli RNA aptamer and map critical nucleotides, we designed a systematic mutagenesis screen. DNA oligonucleotides containing a double-stranded T7 promoter sequence followed by single-stranded broccoli aptamer sequence were synthesized. This DNA template configuration is known to support efficient *in vitro* transcription by bacteriophage T7 RNA polymerase.[30] Every nucleotide position in the broccoli sequence was then systematically substituted with one of the three other nucleotide residues, generating almost 150 site-specific broccoli mutants. Rather than synthesize, purify, and test each broccoli mutant individually, we directly detected broccoli mutant fluorescence in real-time by including DFHBI-1T in transcription reactions and reading fluorescence over time in a quantitative PCR (qPCR) instrument.[31] This screening approach provided time-resolved readout of broccoli RNA mutant fluorescence as RNA molecules were synthesized (Figure 2C). Fluorescence can be directly correlated to broccoli RNA concentration by fitting relative fluorescence units to a standard curve (Figure S2A). The initial fluorescence from 4 to 8 min (F_0) and the maximum fluorescence values (F_{max}) observed over the course of the reaction were used to analyze the effect of each mutant on broccoli fluorescence activity. F_0 can be used as an approximate indicator of relative RNA folding, with presumably the fastest folding aptamers producing the greatest initial fluorescence. Once transcription has completed, broccoli RNA molecules with the greatest activation of fluorescence should provide the greatest F_{max} values, whereas those with substitutions at critical nucleotides would show reduced fluorescence.

Using this mutagenesis strategy, we screened all broccoli point mutants (Figure 2D). Aligning F_{max} values for all mutants together revealed several key features of broccoli RNA structure−function requirements. The initial discovery of broccoli suggested formation of a terminal stem and an internal stem loop structure but did not predict a quadruplex core nor how it would fold.[18] Our mutagenesis also supports the presence of the predicted stem structures. The first and last four nucleotides of broccoli could not be substituted successfully unless G:U wobble pairs were introduced (at positions 2, 4, 46, and 47), which then seemed to support a moderate degree of fluorescence. Positions 4 and 46 in particular responded well to mutation to a uridine, which would introduce a G:U or A-U pair, respectively, instead of an A:G purine−purine mismatch, to presumably stabilize the predicted terminal stem. The predicted internal stem also permitted moderate activity with similar mutations that converted A-U pairs to G:U pairs (at positions 16, 18, and 26), suggesting the presence of a Watson−Crick duplex. Near the center of the internal stem is a predicted loop with a consensus UUCG sequence shared between baby spinach and broccoli. Interestingly, UUCG is known to be a stabilizing sequence for loops,[32,33] which likely

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

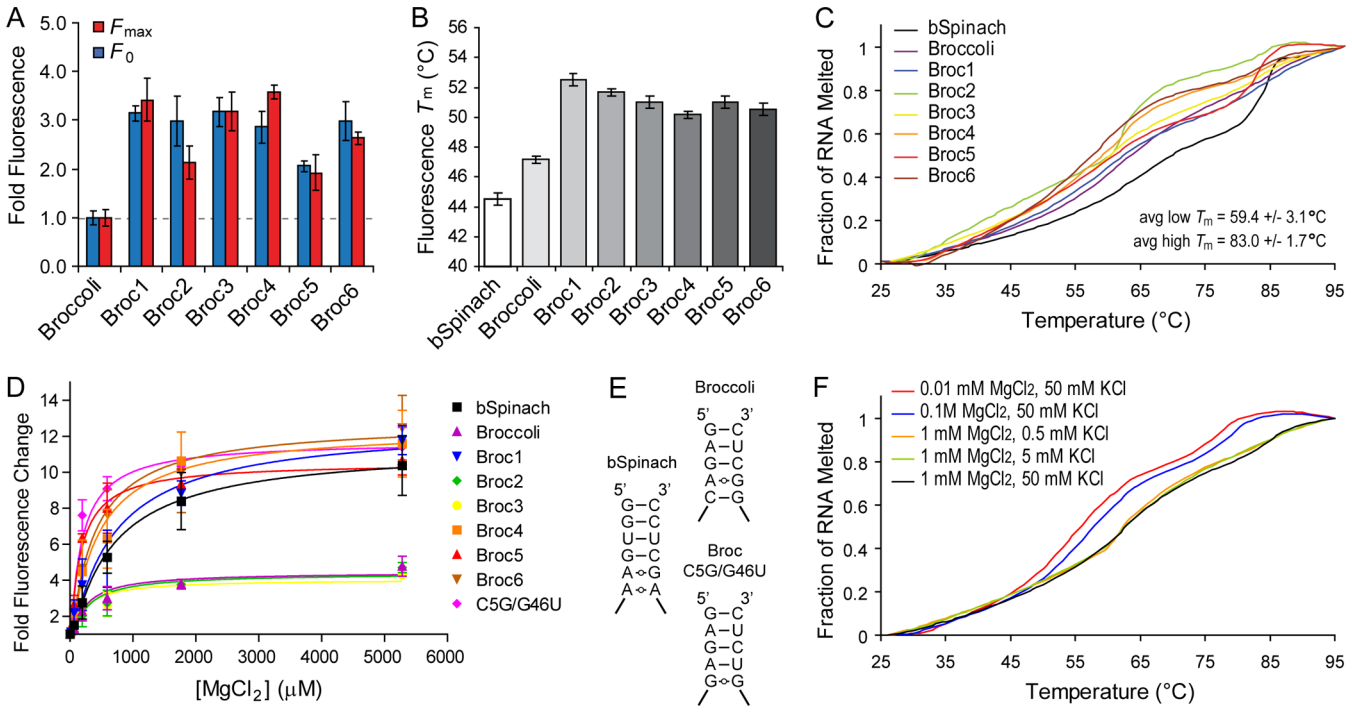


Figure 3. Sequence and structural variations in predicted quadruplex-flanking stem structures of broccoli RNA affect fluorescence, stability and magnesium-dependence. (A) Fluorescence activation of DFHBI-1T during *in vitro* transcription of broccoli mutants, referred to as Broc1 through Broc6. Normalized activity of control broccoli RNA fluorescence is indicated by a dashed gray line. Error is standard deviation. (B) Temperature-dependent fluorescence, defined as the temperature at 50% fluorescence activity, for Broc mutants. bspinach is baby spinach. Error bars are standard deviation. (C) Thermal denaturation of Broc mutants monitored by UV absorbance at 260 nm. Average melting temperature (T_m) values of the low or high temperature melt transitions are indicated. (D) Fluorescence of baby spinach (bspinach), broccoli, and Broc mutants during titration of $MgCl_2$. Peak emission (507 nm) is plotted against $MgCl_2$ concentration. C5G/G46U is a mutant of broccoli with sequence changes at base positions 5 and 46. Error bars are standard deviation. (E) Secondary structure illustrations of the terminal stem-to-quadruplex transitions of baby spinach, broccoli, and the C5G/G46U mutant. (F) Thermal denaturation of broccoli in buffered solutions with the indicated salts as monitored by UV absorbance at 260 nm.

explains its selection in the spinach aptamers.[15,16,18] However, this loop sequence was readily mutated while maintaining fluorescence that was similar to or better than regular broccoli. A recent systematic mutagenesis of spinach aptamer also found that mutation of the UUCG loop led to modulation of fluorescence and aptamer stability.[34] These results further support the formation of an internal stem loop.

Between the putative helical stem regions, numerous guanine residues exist that cannot be substituted without abolishing activity. These invariant guanine nucleotides suggest a direct role in quadruplex formation since no clear base-pairing interactions to cytosine or uridine nucleotides are apparent. Interspersed between invariant guanines, especially at the 3′ end, are several other residues that can be readily exchanged for other bases without suffering a severe loss of fluorescence. These residues may represent nucleotides bulged out of the quadruplex structure or otherwise involved in less critical hydrogen bonding or base stacking. These include nucleotides at positions 8, 32, 37, 39, 41, and 43. Based on the crystal structure of spinach and consensus sequence comparisons, all of these nucleotides except position 43 are extruded to the periphery of the quadruplex core and would be predicted to be less critical for quadruplex formation.[16,19] Certain mutations at some of these positions instead increased fluorescence activity, possibly due to improved stability through stacking or minimization of entropically costly solvent interactions.

Our mutagenesis screen also identified several mutants with substantially higher F_{max} values than the parent broccoli RNA. Arbitrary selection of 10 RNA transcription reactions from our mutagenesis screen followed by resolution on a denaturing polyacrylamide gel (Figure 2D) revealed similar levels of total RNA with no correlation to fluorescence activity, demonstrating that changes in fluorescence were not an artifact of different transcription efficiencies. Most of the enhanced fluorescence mutants were centered at the loop region of the putative stem loop. To investigate the general importance of the putative stem loop, we shortened the stem, deleted the loop, or deleted the entire stem loop (Figure S2B). All stem loop deletions significantly impaired or abolished fluorescence activity. Thus, broccoli fluorescence is very sensitive to changes in this stem loop region, in agreement with our site-specific mutations as well as previous mutagenesis by Filonov and co-workers.[18] These results indicate that the putative internal stem loop indeed plays an important role in mediating fluorophore activation.

On the basis of our initial mutagenesis, we generated a variety of combination mutants to further interrogate broccoli RNA structure and fluorescence. In our second round of mutagenesis, we combined several mutants, mostly in pairs (Figure S2C). Mutant names are simply listed as the position and the new nucleotide substituted at that position. Several mutations which seemed to enhance fluorescence were found to be incompatible with one another. These included combinatorial mutants containing 4U with 5G, 4U with 46U, and 43A with 46U, which presumably perturbed terminal stem formation or the transition from a duplex to quadruplex

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

structure when combined. Most combination mutants were designed to test compatibility of fluorescence-enhancing activity and revealed activity similar to or better than that of regular broccoli. The highest fluorescing mutants were again centered around the loop of the predicted internal stem loop. A third round of combinatorial mutagenesis identified several mutants with consistently higher fluorescence activity than regular broccoli (Figure S2D).

We selected four mutants and designed two new mutants to understand the basis of the apparent enhanced fluorescence. These six RNAs were named Broc1, Broc2, Broc3, Broc4, Broc5, and Broc6 (Figure S2E). The broccoli mutants exhibited 2−4-fold increases in F_0 or F_{max} during *in vitro* transcription in the presence of DFHBI-1T (Figure 3A). Thermal stability was considered a key limiting factor in the performance of the original spinach aptamer, which spurred mutagenesis to create spinach2.[7] To determine if thermal stability played a role in the enhanced fluorescence of our six selected broccoli mutants, we purified these RNAs and monitored fluorescence as temperature was increased. Taking the first derivative of the decrease in fluorescence, we estimated fluorescence-based melting temperature (T_m) values (Figure 3B). Baby spinach T_m was 44.5 °C, and broccoli T_m was approximately 47 °C, whereas Broc1−6 mutants ranged from 51 to 53 °C. Thus, all six Broc mutants exhibited a 4−6 °C increase in fluorescence-based thermal stability, which would be expected to improve fluorescence activity.

To obtain a closer look at thermal stability and RNA folding, we performed UV-monitored thermal denaturation studies with all six purified mutants in comparison to baby spinach and broccoli (Figure 3C). Three key points can be drawn from UV melt analyses. First, biphasic melt curves were observed to varying degrees for all aptamers. These results suggest at least a two-state folding model consistent with baby spinach and broccoli aptamers being composed of two relatively independent folding structures, which would include the predicted quadruplex and helical stem structures. Second, these aptamers displayed large variations in cooperative melting for the two structural states, indicating differences in the folding of the individual structures as well as the entire RNA. For example, Broc2 seems to fold the lower T_m structure with a more cooperative transition than the higher T_m structure. In contrast, Broc5 has the opposite profile where the higher T_m structure folds more cooperatively. Third, despite substantial changes in overall cooperativity of folding, the high and low T_m values calculated for all RNAs were similar, with ranges of ±3.1 °C at the lower T_m (59.4 °C average) and ±1.7 °C at the higher T_m (83.0 °C average; Figure S3A). These results suggest that the global structures of baby spinach, broccoli, and the Broc mutants are similar, which support out initial lead cleavage data, but that individually folding units with similar stabilities but different folding interactions exist. Melt profiles that are relatively linear, such as for regular broccoli, suggest a well-formed global structure with putative stem-to-quadruplex structures that fold and melt together. These thermal denaturation studies indicate that small but significant changes in the folding and stability of Broc1−6 may be associated with their apparent enhanced fluorescence.

To determine if metal ion preferences correlated with broccoli mutant fluorescence, we collected emission spectra in 50 mM of three different alkali metal ions, sodium (NaCl), rubidium (RbCl), or potassium (KCl). Fold change in fluorescence at the peak emission (507 nm) was plotted for each mutant in each salt condition (Figure S3B). Each mutant seemed to have similar fold increases in fluorescence when in the presence of potassium, and all were similar to or modestly better than broccoli. Broc2 and Broc3 were consistently similar to or better than broccoli for all metal ions tested.

Large enhancements in broccoli fluorescence from our mutagenesis screen were puzzling because broccoli was previously reported to be very efficient at fluorescence activation.[18] To rationalize our results, we considered that conditions in our transcription-based mutagenesis screen may be influencing fluorescence. Indeed, *in vitro* transcription reactions contained 40 mM $MgCl_2$, which would make it difficult to distinguish aptamer mutants with a magnesium dependence. A recent systematic mutagenesis to search for optimized spinach aptamers for *in vitro* applications also found enhanced aptamers in transcription reactions containing ∼20 mM $MgCl_2$.[35] However, the dependence of magnesium on aptamer function was not addressed. We considered that our mutagenesis screen results might help determine the basis for magnesium dependence among certain spinach family aptamers. We performed titrations of $MgCl_2$ with baby spinach, broccoli, and our six select Broc mutants (Figure 3D). We found a greater than 10-fold increase in fluorescence for baby spinach in the presence of high magnesium, while broccoli fluorescence increased less than 4-fold.

Among the Broc mutants, only Broc2 and Broc3 retained a low magnesium requirement like broccoli. Mutants Broc1, Broc4, Broc5, and Broc6 all exhibited a significant dependence on magnesium, similar to or greater than baby spinach. Upon inspection of the six Broc mutant sequences, we found that only Broc2 and Broc3 did not have a C5G and G46U mutation. When these two specific mutations were incorporated into parent broccoli to make a C5G/G46U mutant, magnesium-dependent fluorescence increased to over 10-fold mimicking the magnesium-dependence of baby spinach. This particular mutant had been screened during our combinatorial rounds of mutagenesis, and it showed increased fluorescence over broccoli (Figure S2C). In our initial screen, position C5 of broccoli could be converted to any base and still retained activity (Figure 2D). The proposed terminal stem of baby spinach and broccoli, based on the spinach crystal structure, shows noncanonical purine−purine pairs that are immediately adjacent to the quadruplex core (Figure 3E). These nucleotides and their pairing appear to directly influence magnesium binding or counterion effects and are likely to play a role in the stability of the stem-to-quadruplex transition. ispinach, a version of spinach recently optimized for *in vitro* applications, also retained these purine−purine pairs.[35] Thus, enhanced fluorescence of many of our mutants, in particular Broc1, -4, -5, and -6, could be relegated in part to a reliance on high magnesium for fluorescence. Importantly, all of the mutations for our six select Broc mutants were outside of the quadruplex region except for the A37U mutation, which is predicted to be extruded out of the quadruplex forming core based on our sequence and structural comparisons with spinach and baby spinach.

In support of magnesium-dependent stabilization of stem-to-quadruplex transitions, we performed UV thermal denaturation of broccoli in increasing magnesium or potassium concentrations. T_m values for both folding transitions increased as magnesium increased from 10 μM to 1 mM and the cooperativity of each transition became less pronounced. These results suggest that stem and quadruplex structures

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

fold independently without substantial magnesium, but in the presence of 1 mM magnesium their folding is more uniform, acting more like a single structure. Melting of broccoli in a constant 1 mM $MgCl_2$ but increasing KCl from 0.5 or 5 mM up to 50 mM revealed a change in the folding cooperativity of the more stable structure. In low KCl, the lower T_m structure did not change significantly, but a T_m value for the more stable structure could not be determined due to the lack of a clear transition. These results suggest that magnesium has a global effect on broccoli, most likely altering local stability and folding to improve transitions from stem to quadruplex structures. On the basis of these KCl-dependent effects, the more stable structure with a higher T_m value may correspond to a quadruplex while the less stable (lower T_m) structure may correspond to predicted stem structures.

Broccoli aptamers with mutations in quadruplex-flanking regions might translate to brighter or more stable fluorescence inside of cells. We cloned broccoli and the six select Broc mutants into bacterial expression vectors. Broccoli sequences were flanked by a T7 promoter and a T7 terminator without a scaffold sequence (ie, no tRNA scaffold) and transformed into Rosetta (DE3) *E. coli* cells. After induction of broccoli RNA expression, cells were incubated with DFHBI-1T and fluorescence measured. Compared to broccoli, Broc3 had the highest fluorescence by about 1.6 fold (Figure S3D). Comparing broccoli and Broc3 binding affinity to DFHBI-1T did not reveal any significant differences in the calculated K_d values (Figure S3E). On the basis of the small increase in potassium utilization, a T_m increase of a few degrees, similar magnesium dependence, and similar DFHBI-1T binding affinity, the modest improvement of Broc3 fluorescence over broccoli in *E. coli* seems reasonable. However, a direct correlation between *in vivo* and *in vitro* fluorescence for the tested Broc mutants was not clear, indicating that multiple factors contribute to effective fluorescence inside of cells and highlighting the need to optimize RNA scaffolds for *in vivo* applications.[36]

Effectively developing, optimizing, and using fluorescent RNA aptamers requires understanding of the principles of RNA structure that support successful fluorophore binding or activation. Such studies are expected to eventually contribute to the development of reliable, multicolored fluorescent RNA tool kits for molecular and chemical biology applications. Our results indicate that broccoli is a G-quadruplex-containing RNA with a structure that is very similar to that of baby spinach and spinach. Duplex stem structures that flank the central quadruplex structure alter the stability, metal ion preferences and fluorescence activity of baby spinach and broccoli. In particular, the transitions from stem to quadruplex are proposed to play a significant role in magnesium dependence of these GFP-mimic RNAs and is an important target for optimizing their stability and metal ion preferences. The role of magnesium in stabilizing tertiary RNA structures around transitions of independently folding units has been recognized previously, such as for pseudoknots, three-helix junctions, and tRNA folding.[37−39] Magnesium appears to aid in formation of a more uniform aptamer by improving the cooperativity of folding between stem and quadruplex structures in broccoli. Therefore, spinach derivatives likely exhibit diverse requirements for magnesium depending on their ability to achieve sufficiently favorable stem-to-quadruplex transitions. A systematic mutagenesis of the first generation spinach aptamer recently reported that purine−purine pairing at the terminal stem-to-quadruplex transition had important effects on thermodynamic stability and fluorescence activity.[34] These are the same purine−purine pairs that we have shown to modulate magnesium dependence. Analyses of DNA duplex-to-quadruplex transitions have also highlighted the importance of noncanonical pairing to accommodate stable transitions.[40] Thus, designing and optimizing better aptamers, such as RNA mimics of GFP, should take into consideration the interactions and transitions between independently folding structural elements and the role of metal ions in stabilizing them.

METHODS

RNA Synthesis. RNA was synthesized by T7 *in vitro* transcription starting with DNA templates purchased from Integrated DNA Technologies (IDT). Single-stranded DNA templates were annealed to a T7 promoter oligo (TAATACGACTCACTATA) to generate double-stranded promoter regions, which support *in vitro* transcription by T7 RNA polymerase.[30] T7-DNA template and synthesized RNA sequences for all mutants are listed in Table S1. *In vitro* transcription was performed by standard protocols. Briefly, reactions contained purified T7 RNA polymerase, 30 mM Tris (at pH 7.9), 12.5 mM NaCl, 40 mM $MgCl_2$, 2% PEG_{8000}, 0.05% Triton-X 100, 2 mM spermidine, and 2.5 μM T7-DNA template. Afterward, the DNA template was degraded by the addition of 1 unit of DNase I for every 20 μL of reaction and incubated at 37 °C for 15 min. Reactions were phenol-chloroform extracted and gel-purified from denaturing polyacrylamide gels. Purified RNA was quantified by measuring absorbance at 260 nm and calculated extinction coefficients using nearest neighbor approximations and Beer's Law.

RNA Structure Probing by Lead Acetate Cleavage. Lead cleavage mapping was performed similar to published protocols.[20] The 5′ phosphates of 100 pmol of *in vitro* transcribed baby spinach and broccoli RNA were removed by treatment with alkaline phosphatase. Dephosphorylated RNA was radiolabeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase then gel-purified. A total of 60 000 cpm of radiolabeled RNA (~2 pmol) was refolded by heating to 95 °C and cooling to 23 °C in 20 mM Tris (at pH 7.2), 5 mM $MgCl_2$, 0.2% DMSO, 1 mg mL^{-1} tRNA, 50 mM CsCl or KCl, and with or without 20 μM DFHBI-1T (Lucerna Technologies). Freshly prepared lead acetate was then added to a final of 2 mM, and the reaction was incubated at 23 °C for 20 min before being precipitated by the addition of 10 vols of 2% $LiClO_4$ in acetone. Alkaline hydrolysis ladders were prepared by incubation of radiolabeled RNA at 90 °C for 8 min in 20 mM $NaHCO_3$ (pH 10), 1 mM EDTA, and 1 mg mL^{-1} tRNA then precipitated with 2% $LiClO_4$ in acetone. T1 RNase ladders were prepared by incubation of radiolabeled RNA with 1 U of T1 RNase at 23 °C for 10 min in 25 mM phosphate buffer (pH 7.0), 150 mM NaCl, 1 mM $MgCl_2$, 0.2 mM EDTA, and 1 mg mL^{-1} tRNA then precipitated with 2% $LiClO_4$ in acetone. Precipitated reactions were washed with acetone, boiled in 90% formamide, 1× TBE, then resolved on a 15% denaturing polyacrylamide sequencing gel. Gels were dried and exposed to phosphorimager to visualize radioactive RNA bands.

Fluorescence Spectra, Salt Titrations, and DFHBI-1T Binding Affinity. Purified RNA was refolded by heating to 95 °C for 3 min then slow cooling to RT at a rate of approximately 1 °C/min in the following reaction: 20 mM cacodylate (pH 7.2), 1 mM $MgCl_2$, 0.05% DMSO, and 5 μM DFHBI-1T. RNA was included at 0.5 μM and XCl (X = K, Rb, Na, Li, or Cs) at 50 mM. For salt titrations, small volumes of 2 M salt solutions were added to 400 μL reactions and incubated at RT for 3 min, then emission spectra were collected. Fluorescence values were corrected for increased volume of each reaction over the course of titration. Fold fluorescence change was calculated by dividing the fluorescence value after each salt addition by the initial value before the addition of the titrated salt. For determining the binding affinity of broccoli and Broc3 RNA for DFHBI-1T, increasing concentrations of each RNA were refolded in the above-noted reaction buffer except

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

with 100 nM of DFHBI-1T. Excitation and emission spectra for these assays were collected on a Horiba Fluorolog-3 spectrofluorometer. Peak emission at 507 nm was plotted against salt concentration. All experiments were performed in duplicate or triplicate and curves fit by nonlinear regression in Prism (Graphpad).

Systematic Broccoli Aptamer Mutagenesis Monitored by Fluorescence. To perform systematic mutagenesis of broccoli RNA, we combined *in vitro* transcription with real time fluorescence readout using a BioRad CFX96 quantitative PCR machine. This approach was inspired by a similar technique described by Jaschke and co-workers.[31] Single-stranded broccoli mutant T7-DNA templates were purchased from IDT in 96-well plates normalized to 2 nmol of DNA. These single DNA strands represented the antisense sequence. To each well, 40 μL of resuspension buffer was added (5 mM Tris, pH 7.0, 0.1 mM EDTA) and 40 μL of T7 promoter oligo (sense sequence) at 50 μM in water, for a final of 25 μM template. To ensure efficient annealing of the T7 promoter oligo, plates were sealed and incubated at 80 °C for 10 min, then slow cooled to RT, spun down, and gently vortexed. For transcription reactions, 1 μL of each template was spotted into new 96-well plates followed by the addition of 9 μL of ice-cold transcription reaction mix on ice. The standard transcription reaction described above was used; however KCl was added to a final of 3 mM and DFHBI-1T to 60 μM. Reactions also contained 12.5 mM NaCl from the T7 RNA polymerase storage buffer. To initiate reactions, plates were placed into a BioRad CFX96 block set at 4 °C. The block was then cycled to 37 °C and held for 1 h. The plate was read using all channels (which includes the FAM/SYBR filter) once at 4 °C then every 1 min and 48 s (reading takes 12 s) at 37 °C such that fluorescence reading was at 2 min intervals. Data were exported to Excel for analysis. Maximum fluorescence values over a 1 h time course were used for comparing mutants. Initial fluorescence, defined as the change in fluorescence over the change in time from 4 to 8 min, was also used for comparing mutant performance. Four or more separate experiments were performed for each screen, and error was calculated by standard deviation.

Fluorescence-Based Determination of RNA Melting Temperature. To determine temperature dependence of fluorescence, purified broccoli RNAs at 10 μM were refolded in 20 mM cacodylate (pH 7.2), 50 mM KCl, 1 mM $MgCl_2$, and 20 μM DFHBI-1T by heating to 95 °C for 3 min then slow cooling to RT. Reactions were placed in 96-well plates in a BioRad CFX96 and a typical qPCR melt analysis was performed. Reactions were heated from 25 to 95 °C at 0.5 °C increments. Reactions were held at each temperature for 30 s before plate reading. BioRad software on the CFX96 automatically determined the derivative of the melt curve and autocalculated T_m values.

Thermal Denaturation of RNA Monitored by UV Absorbance. RNA reactions were prepared as described for fluorescence-based T_m determination, but only 1 μM RNA and 5 μM DFHBI-1T were used. Degassed water was used to prepare samples. Data were collected in a Cary 400 UV/vis spectrophotometer. Reactions were heated to 95 °C then slow cooled at 1 °C/min. Samples were then heated and cooled from 25 to 95 °C three times at a ramp rate of 1 °C/min, while UV absorbance was collected every 1 min. Melting temperatures were determined using Van't Hoff calculations on up ramps only (dissociation curves) and error determined by standard deviation.

Expression and Fluorescence Measurement of Broccoli RNA in *E. coli*. Double-stranded DNA sequences encoding broccoli and broccoli mutants were synthesized by IDT and cloned into a pIDT-SMART vector (ampicillin resistant) containing a custom multiple cloning site. The broccoli sequences were inserted 30 bases downstream from a T7 promoter and 10 bases upstream of a T7 terminator. After sequencing to confirm cloning, plasmids were transformed into Rosetta (DE3) cells and grown at 37 °C with shaking. Cultures were measured by absorbance at 600 nm and normalized to an A_{600} of 0.5. IPTG was then added to a final of 0.5 mM and DFHBI-1T to a final of 10 μM, and shaking continued at 37 °C for 4 h. Cells were harvested, spun down, and resuspended in 400 μL of PBS and their absorbance at 600 nm checked again and adjusted to match, then DFHBI-1T was added to 10 μM and incubated on ice for 30 min. A total of 100 μL of each sample was spotted into four wells of a black 96-well fluorescence plate. Reactions were then read in a Promega GloMax plate reader three times using a green fluorescence filter set. Raw fluorescence was normalized and plotted. Error was calculated as standard deviation from three independent experiments.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.6b00047.

Fluorescence of malachite green, baby spinach, and broccoli RNA aptamers in alkali metal salts; fluorescence of broccoli aptamer stem-loop mutants and rounds 2−4 of site-specific mutagenesis; UV melts of broccoli and Broc1−6 mutants, fluorescence of Broc1−6 mutants in *E. coli*, binding affinity of broccoli and Broc3 to DFHBI-1T (PDF)

Table S1: Oligonucleotide Sequences (XLS)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ktgagnon@siu.edu.

Author Contributions

§These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank M. McCarroll for access to a Horiba Fluorolog spectrofluorometer and the R. Gupta laboratory for insightful discussions. This work was supported by a Team Development Grant from Southern Illinois University to K.T.G. and an ALS Association grant to K.T.G (16-IIP-260).

■ REFERENCES

(1) Feigon, J., Dieckmann, T., and Smith, F. W. (1996) Aptamer structures from A to zeta. *Chem. Biol. 3*, 611−617.

(2) Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science 249*, 505−510.

(3) You, M., Litke, J. L., and Jaffrey, S. R. (2015) Imaging metabolite dynamics in living cells using a spinach-based riboswitch. *Proc. Natl. Acad. Sci. U. S. A. 112*, E2756−2765.

(4) You, M., and Jaffrey, S. R. (2015) Designing optogenetically controlled RNA for regulating biological systems. *Ann. N. Y. Acad. Sci. 1352*, 13−19.

(5) Rogers, T. A., Andrews, G. E., Jaeger, L., and Grabow, W. W. (2015) Fluorescent monitoring of RNA assembly and processing using the split-spinach aptamer. *ACS Synth. Biol. 4*, 162−166.

(6) Arora, A., Sunbul, M., and Jaschke, A. (2015) Dual-colour imaging of RNAs using quencher- and fluorophore-binding aptamers. *Nucleic Acids Res. 43*, e144.

(7) Strack, R. L., Disney, M. D., and Jaffrey, S. R. (2013) A superfolding spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods 10*, 1219−1224.

(8) Dolgosheina, E. V., Jeng, S. C., Panchapakesan, S. S., Cojocaru, R., Chen, P. S., Wilson, P. D., Hawkins, N., Wiggins, P. A., and Unrau, P. J. (2014) RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. *ACS Chem. Biol. 9*, 2412−2420.

(9) Tan, Z. J., and Chen, S. J. (2011) Importance of diffuse metal ion binding to RNA. *Met. Ions Life Sci. 9*, 101−124.

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

(10) Pyle, A. M. (2002) Metal ions in the structure and function of RNA. *JBIC, J. Biol. Inorg. Chem. 7*, 679−690.
(11) Auffinger, P., D'Ascenzo, L., and Ennifar, E. (2016) Sodium and Potassium Interactions with Nucleic Acids. *Met. Ions Life Sci. 16*, 167−201.
(12) Largy, E., Mergny, J. L., and Gabelica, V. (2016) Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *Met. Ions Life Sci. 16*, 203−258.
(13) You, M., and Jaffrey, S. R. (2015) Structure and Mechanism of RNA Mimics of Green Fluorescent Protein. *Annu. Rev. Biophys. 44*, 187−206.
(14) Babendure, J. R., Adams, S. R., and Tsien, R. Y. (2003) Aptamers switch on fluorescence of triphenylmethane dyes. *J. Am. Chem. Soc. 125*, 14716−14717.
(15) Paige, J. S., Wu, K. Y., and Jaffrey, S. R. (2011) RNA mimics of green fluorescent protein. *Science 333*, 642−646.
(16) Warner, K. D., Chen, M. C., Song, W., Strack, R. L., Thorn, A., Jaffrey, S. R., and Ferre-D'Amare, A. R. (2014) Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat. Struct. Mol. Biol. 21*, 658−663.
(17) Song, W., Strack, R. L., Svensen, N., and Jaffrey, S. R. (2014) Plug-and-play fluorophores extend the spectral properties of spinach. *J. Am. Chem. Soc. 136*, 1198−1201.
(18) Filonov, G. S., Moon, J. D., Svensen, N., and Jaffrey, S. R. (2014) broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc. 136*, 16299−16308.
(19) Huang, H., Suslov, N. B., Li, N. S., Shelke, S. A., Evans, M. E., Koldobskaya, Y., Rice, P. A., and Piccirilli, J. A. (2014) A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol. 10*, 686−691.
(20) Gagnon, K. T., and Maxwell, E. S. (2015) Assessing intermolecular RNA:RNA interactions within a ribonucleoprotein complex using heavy metal cleavage mapping. *Methods Mol. Biol. 1240*, 125−134.
(21) Brown, R. S., Dewan, J. C., and Klug, A. (1985) Crystallographic and biochemical investigation of the lead(II)-catalyzed hydrolysis of yeast phenylalanine tRNA. *Biochemistry 24*, 4785−4801.
(22) Pan, T. (2000) Probing RNA structure by lead cleavage, *Curr. Protoc. Nucleic Acid Chem.* Chapter 6, Unit 6 3, p 6.3.1, John Wiley & Sons, Inc., Hoboken, NJ, DOI: 10.1002/0471142700.nc0603s00.
(23) Markley, J. C., Godde, F., and Sigurdsson, S. T. (2001) Identification and characterization of a divalent metal ion-dependent cleavage site in the hammerhead ribozyme. *Biochemistry 40*, 13849−13856.
(24) Baugh, C., Grate, D., and Wilson, C. (2000) 2.8 A crystal structure of the malachite green aptamer. *J. Mol. Biol. 301*, 117−128.
(25) Campbell, N. H., and Neidle, S. (2012) G-quadruplexes and metal ions. *Met. Ions Life Sci. 10*, 119−134.
(26) Dammertz, K., Hengesbach, M., Helm, M., Nienhaus, G. U., and Kobitski, A. Y. (2011) Single-molecule FRET studies of counterion effects on the free energy landscape of human mitochondrial lysine tRNA. *Biochemistry 50*, 3107−3115.
(27) Kim, B. G., Evans, H. M., Dubins, D. N., and Chalikian, T. V. (2015) Effects of Salt on the Stability of a G-Quadruplex from the Human c-MYC Promoter. *Biochemistry 54*, 3420−3430.
(28) Marincola, F. C., Virno, A., Randazzo, A., and Lai, A. (2007) Effect of rubidium and cesium ions on the dimeric quaduplex formed by the Oxytricha nova telomeric repeat oligonucleotide d-(GGGGTTTTGGGG). *Nucleosides, Nucleotides Nucleic Acids 26*, 1129−1132.
(29) Creze, C., Rinaldi, B., Haser, R., Bouvet, P., and Gouet, P. (2007) Structure of a d(TGGGGT) quadruplex crystallized in the presence of Li+ ions. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 63*, 682−688.
(30) Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res. 15*, 8783−8798.
(31) Hofer, K., Langejurgen, L. V., and Jaschke, A. (2013) Universal aptamer-based real-time monitoring of enzymatic RNA synthesis. *J. Am. Chem. Soc. 135*, 13692−13694.
(32) Heus, H. A., and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science 253*, 191−194.
(33) Dichtl, B., Pan, T., DiRenzo, A. B., and Uhlenbeck, O. C. (1993) Replacement of RNA hairpins by in vitro selected tetranucleotides. *Nucleic Acids Res. 21*, 531−535.
(34) Ketterer, S., Fuchs, D., Weber, W., and Meier, M. (2015) Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer spinach. *Nucleic Acids Res. 43*, 9564−9572.
(35) Autour, A., Westhof, E., and Ryckelynck, M. (2016) ispinach: a fluorogenic RNA aptamer optimized for in vitro applications. *Nucleic Acids Res. 44*, 2491−2500.
(36) Filonov, G. S., Kam, C. W., Song, W., and Jaffrey, S. R. (2015) In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chem. Biol. 22*, 649−660.
(37) Quigley, G. J., Teeter, M. M., and Rich, A. (1978) Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. *Proc. Natl. Acad. Sci. U. S. A. 75*, 64−68.
(38) Soto, A. M., Misra, V., and Draper, D. E. (2007) Tertiary structure of an RNA pseudoknot is stabilized by "diffuse" Mg2+ ions. *Biochemistry 46*, 2973−2983.
(39) Noeske, J., Schwalbe, H., and Wohnert, J. (2007) Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. *Nucleic Acids Res. 35*, 5262−5273.
(40) Lim, K. W., and Phan, A. T. (2013) Structural basis of DNA quadruplex-duplex junction formation. *Angew. Chem., Int. Ed. 52*, 8566−8569.

DOI: 10.1021/acschembio.6b00047
ACS Chem. Biol. XXXX, XXX, XXX−XXX

Rohilla and Gagnon *Acta Neuropathologica Communications* (2017) 5:63
DOI 10.1186/s40478-017-0468-y

Acta Neuropathologica Communications

REVIEW Open Access

RNA biology of disease-associated microsatellite repeat expansions

Kushal J. Rohilla[1] and Keith T. Gagnon[1,2*]

Abstract

Microsatellites, or simple tandem repeat sequences, occur naturally in the human genome and have important roles in genome evolution and function. However, the expansion of microsatellites is associated with over two dozen neurological diseases. A common denominator among the majority of these disorders is the expression of expanded tandem repeat-containing RNA, referred to as xtrRNA in this review, which can mediate molecular disease pathology in multiple ways. This review focuses on the potential impact that simple tandem repeat expansions can have on the biology and metabolism of RNA that contain them and underscores important gaps in understanding. Merging the molecular biology of repeat expansion disorders with the current understanding of RNA biology, including splicing, transcription, transport, turnover and translation, will help clarify mechanisms of disease and improve therapeutic development.

Keywords: Repeat expansion disease, Microsatellite, Tandem repeats, RNA, Splicing, Transcription, Transport, Export, Turnover, Translation, Mechanism, Therapeutics, Myotonic dystrophy, DM1, DM2, Huntington's disease, HD, C9ORF72, C9FTD/ALS, Amyotrophic lateral sclerosis, Spinocerebellar ataxia, SCA, Fragile X, SBMA, FXTAS

Introduction

In 1991 it was discovered that a microsatellite sequence expansion is the cause of two distinct neurological disorders, Fragile X syndrome (FXS) [303] and spinal bulbar muscular atrophy (SBMA), or Kennedy's disease [167]. Since then, simple repeat sequence expansions have been associated with over twenty more neurological disorders [166, 300, 333] (Table 1). What has been learned is that microsatellite expansions may cause disease in multiple ways. For nearly all of these neurological disorders, however, disease includes production of RNA that contains the aberrant repeat expansion sequence. Accordingly, the leading disease mechanisms involve repeat expansion RNA-mediated sequestration of critical RNA-binding proteins and translation of repeat expansion RNA into toxic repetitive polypeptides.

Tremendous progress has been made in understanding the metabolism of expanded tandem repeat-containing RNA (xtrRNA). Nonetheless, various gaps in our understanding of the underlying molecular biology and pathology remain, which in turn limits identification of promising therapeutic approaches. The goal of this review is to help address these gaps by discussing the potential impact of xtrRNA on cellular RNA metabolism. We begin with an overview that covers microsatellite origin, evolution, and expansion. We then follow xtrRNA through its life cycle, beginning with transcription and continuing through splicing, folding, protein interactions, localization, turnover, and translation. We rationalize the logic of current molecular disease models, note where important mechanistic information is lacking, and emphasize new pathways to consider for mechanistic insight. We use this discussion to also highlight areas where therapeutic intervention may be useful.

Origin and expansion of microsatellites in human disease

Simple tandem repeat sequences in the human genome

Microsatellite sequences comprise approximately 3% of the human genome, about twice as much as protein coding sequence [1, 171]. Microsatellites, interchangeably known as simple or short tandem repeats (STRs), are

* Correspondence: ktgagnon@siu.edu
[1]Department of Biochemistry and Molecular Biology, Southern Illinois University, School of Medicine, Carbondale, Illinois, USA
[2]Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, Illinois, USA

Table 1 Microsatellite repeat expansion disorders

Disorder	Repeating unit	Genomic location	Gene name	Normal length	Pathogenic length	Expanded repeats result in:				Repeat discovery & references
						Gene silencing	xtrRNA transcription	xtrRNA proteins	xtrRNA foci	
FXS/FRAXA†	CGG	5′UTR	*FMR1*	6-55	200+	Yes*	No	No	No	[114, 161, 230, 303]
SBMA†	CAG	Coding	*AR*	9-36	38-62	No	Yes	Yes*	L .D.	[167]
DM1	CTG	3′UTR	*DMPK*	5-37	50-10000	No	Yes*	Yes	Yes*	[193]
HD	CAG	Coding	*HTT*	10-35	35+	No	Yes	Yes*	L.D.	[192]
SCA1	CAG	Coding	*ATXN1*	6-35	49-88	No	Yes	Yes*	L.D.	[232]
FRAXE	CCG	5'UTR	*AFF2*	4-39	200-900	Yes*	No	No	No	[152]
DRPLA	CAG	Coding	*ATN1*	6-35	49-88	No	Yes	Yes*	L.D.	[155]
SCA3	CAG	Coding	*ATXN3*	12-40	55-86	No	Yes	Yes*	Yes	[140]
SCA2	CAG	Coding	*ATXN2*	14-32	33-77	No	Yes	Yes*	L.D.	[127, 253]
FRDA	GAA	Intron	*FXN*	8-33	90+	Yes*	Yes / No	No	No	[30]
SCA6	CAG	Coding	*CACNA1A*	4-18	21-30	No	Yes	Yes*	L.D.	[337]
EPM1	CCCCGC CCCGCG	Promoter	*CSTB*	2-3	30-80	Yes*	Yes / No	No	No	[170]
SCA7	CAG	Coding	*ATXN7*	7-17	38-120	No	Yes	Yes*	L.D.	[57]
OPMD	GCG	Coding	*PABPN1*	6-10	12-17	No	Yes	Yes*	No	[23]
SCA8	CTG	3′UTR	*ATXN8*	16-34	74+	No	Yes	Yes*	Yes*	[156]
SCA12	CAG	5′UTR	*PPP2R2B*	7-28	66-78	No	Yes*	No	No	[121]
SCA10	ATTCT	Intron	*ATXN10*	10-20	500-4500	No	Yes	?	Yes*	[205]
SCA17	CAG	Coding	*TBP*	25-42	47-63	No	Yes	Yes*	L.D.	[224]
DM2	CCTG	Intron	*CNBP*	10-26	75-11000	No	Yes	?	Yes*	[184]
FXTAS/FXPOI	CGG	5′UTR	*FMR1*	6-55	55-200	No	Yes	Yes*	Yes*	[143, 287]
HDL2	CTG/CAG	3'UTR/antisense	*JPH3*	<50	50+	No	Yes	?	Yes*	[120]
SCA31	TGGAA	Intron	*TK2/BEAN*	0	45+	No	Yes	Yes*	Yes*	[256]
SCA36	GGCCTG	Intron	*Nop56*	3-14	650+	No	Yes	?	Yes*	[153]
C9FTD/ALS	GGGGCC	Intron	*C9ORF72*	2-25	25+	No	Yes	Yes*	Yes*	[59, 248]
FRA7A	CGG	Intron	*ZFN713*	5-22	85+	No	Yes* / No	No	No	[209]
FRA2A	CGG	Intron	*AFF3*	8-17	300+	No	Yes* / No	No	No	[210]

Disorders are listed in order of the year they were discovered, with the appropriate references relating to their discovery. This table highlights known RNA biology for each disease with respect to xtrRNA transcription, translation, and formation of nuclear focal aggregates

Dagger symbol (†) indicates that athough the CAG repeat for SBMA was discovered first, the CGG repeat for FXS was published first. Asterisk (*) indicates the most likely repeat-associated disease mechanism(s) for that disorder. *L.D.* length-dependent, *SBMA* Spinal-Bulbar Muscular Atrophy, *EPM1* Progressive Myoclonus Epilepsy 1 (Unverricht–Lundborg Disease), *FXS/FRAXA* Fragile X Syndrome, *DM* Myotonic Dystrophy, *HD* Huntington's Disease, *SCA* Spinocerebellar Ataxia, *FRAXE* Fragile X E Syndrome, *DRPLA* Dentatorubral-Pallidoluysian Atrophy, *FRDA* Friedreich Ataxia, *OPMD* Oculopharyngeal Muscular Dystrophy, *FXTAS* Fragile X–Associated Tremor/Ataxia Syndrome, *FXPOI* Fragile X-Associated Primary Ovarian Insufficiency, *HDL2* Huntington's Disease-Like 2, *C9FTD/ALS C9ORF72*-Associated Frontotemporal Dementia and Amyotrophic Lateral Sclerosis, *FRA7A* CGG Expansion at Fragile Site 7A, *FRA2A* CGG Expansion at Fragile Site 2A

usually defined as simple sequence motifs of one to six nucleotides that are contiguously repeated at least a few times [24, 69]. Microsatellites occur throughout the genome, but are predominantly found in noncoding promoters, introns, 5' and 3' untranslated regions (UTRs), and intergenic regions [236, 257, 291]. Intergenic microsatellites seem to fit neutral evolution models, although not perfectly [69], and are among the most variable genomic sequences [32]. Therefore, they serve as the basis for forensic DNA analyses and as markers for population genetics studies [24, 61, 65, 236].

The origin and evolution of microsatellites is incompletely understood. They may have derived from simple and repetitive sequence motifs found in mobile genetic elements, such as non-LTR (long terminal repeat) retrotransposons like *Alu* and L1 [6, 51, 95]. Transposable elements have colonized the human genome extensively and their remains have undergone mutation and replication, providing the starting material for simple tandem repeats [69]. The dense repetitive sequence of centromeres and telomeres is proposed to originate from incorporation of mobile genetic elements during early

eukaryotic evolution [87, 304]. Small sequence duplications of these simple sequences can further produce microsatellites with multiple repeats. The STRs that are expanded in Friedreich ataxia (FRDA) and myotonic dystrophy type 2 (DM2), for example, have been traced to an *Alu* element origin [41, 163]. In contrast, *de novo* genesis by events like random mutation, replication slippage, and duplication of unique sequence may also account for the birth of microsatellite sequences [28, 69]. STRs have been shown to have positive roles in evolution, such as bacterial resistance to antibiotics and circadian clock adaptation to the environment in *Neurospora crassa* and *Drosophila melanogaster* [83, 133, 211, 258, 279, 307]. The placement of microsatellite sequences in and near regulatory and coding regions of the genome also implicate them in control of gene expression and genetic interactions [133, 331].

Expansion of simple tandem repeats

There are several distinct mechanisms that can contribute to the expansion of naturally occurring microsatellites. In this section we provide a brief overview of these mechanisms. Many excellent reviews on this topic are cited in this section and recommended for further reading (see [84, 130, 149, 206, 212, 260, 297, 331, 333]).

A major source of microsatellite expansion in dividing cells is DNA replication, although mitotic recombination is also recognized as a contributing factor [84, 149, 212, 242]. During replication, repetitive sequences can cause problems at the replication fork and result in fork reversals or template switching, which can insert extra repeats [76, 144, 149, 212]. At the strand level, polymerase slipping can cause expansions in the leading or lagging strand [84, 137, 144, 149]. Repeats may also induce imperfect Okazaki fragment ligation and add repeats in the lagging strand [81, 93, 278]. The pathway followed for expansion of a repeat has been proposed to be a balancing act between several factors [84, 149, 212]. These include relative repeat length, the stability and types of non-canonical structures the repeat sequence can form, and nearby flanking sequences. After repeat sequences are added to one or both strands, the daughter strands reanneal. Misalignment and slippage will occur and extra sequences will bulge out to form non-canonical (non-B-form) structures like hairpins or quadruplexes [237, 331]. If these structures persist to the next round of replication, or if they undergo flawed repair, they can result in permanent expansions [130, 149, 212, 260, 297]. During DNA recombination, which repairs single-end or double-strand breaks, unequal crossing over or template switching can cause misalignments and introduction of additional repeats [208, 242, 306].

Repeat expansion events are intimately tied to the repair of non-canonical DNA structures and DNA damage. Multiple DNA damage control pathways have been implicated, including mechanisms that replace DNA bases, like base excision repair (BER) or nucleotide excision repair (NER), especially as sources for repeat expansion in non-dividing cells [206]. However, mismatch repair (MMR) has been argued to be a primary driver of repeat expansion [75, 106, 130, 260, 271]. MMR expands repeats through recognition and processing of unusual DNA structures, such as small bulges and hairpins [260], via the enzyme MutSβ (MSH2-MSH3 complex) [130, 260, 334]. The processing and damage rectification steps are carried out by MutSβ and associated proteins, including the MutLα (MLH1-PMS2 complex) or MutLγ (MLH1-MLH3 complex) endonucleases that help remove DNA lesions [106, 130, 241]. Polymerases like Polβ are then recruited, which can insert extra repeats due to flawed priming or templating [33, 190].

An important question is how repeats are able to expand out of control, sometimes into the hundreds or thousands of perfect tandem copies, without accumulating significant interruptions? Microsatellites that are evolutionarily neutral, typically in intergenic regions, become highly mutable when they exceed thresholds above just a few tandem repeats [68, 95, 320]. Therefore, the likelihood of remaining as a perfect tandem repeat without interruption is expected to decrease with tandem repeat length. This suggests that accumulation of large expansions must either occur quickly, before mutations can accumulate, or their disruption must be guarded against [320]. Genic regions of the genome, where all currently known disease-associated repeat expansions occur [31, 236] (Table 1), seem to enjoy special favor through positive evolutionary selection processes that protect sequence fidelity [191, 236, 284]. However, it seems unlikely that this would contribute significantly to large repeat expansions. For example, non-repetitive codons would presumably be preferred and selected over unstable repeat codons.

Mechanisms have been proposed that could provide large expansions in a single step, including template switching replication models where repeats are already sufficiently large enough [225, 266] and out-of-register synthesis during homologous recombination-based repair of double-strand breaks (DSBs) [212, 242, 249, 250, 283]. One intriguing mechanism for rapid and large repeat accumulation is break-induced replication (BIR) [148, 176]. BIR is a homologous recombination pathway that can rescue collapsed or broken replication forks [195]. It is induced when a replisome collides with a broken single-end DSB [189]. BIR is also believed to be selective for structure-prone or GC-rich repeats that are long enough to form stable structures [148]. In this mechanism of expansion, stable structures would cause fork reversals. Resolution of these four-way junction

structures would result in a one-ended DSB. To restart the fork, the one-ended DSB invades the sister chromatid to form a D-loop, but likely does so out-of-register because of the repetitive sequence, thus leading to expansion. While this BIR study was performed in yeast, the results are expected to translate to human cells [176].

Incremental expansions, such as those caused by MMR, are typically on the order of 1-3 repeats at a time [260]. Could these events generate large uninterrupted expansions? Rapid accumulation of expansions via MMR or other DNA damage repair pathways might be facilitated by transcription across the repeat. It has been shown that transcription is required for expansion of the CGG repeat in a mouse model of FXS [2, 333]. Several studies have shown that transcription at repeat expansions is associated with repeat instability, possibly via formation of DNA-RNA hybrids, or R-loops [180, 181, 183, 195, 223, 246, 333]. It is possible that these events could allow DNA damage. One report has shown a correlation between R-loop formation and replication fork stalling, offering a familiar mechanism for repeat expansion through DNA replication [86, 109]. Alternative mechanisms might involve oxidation of free DNA strands, or simple misalignment upon strand reannealing, to signal DNA damage repair [180]. The latter model suggests that GC-rich or structure-prone repeats would be more susceptible to expansion during transcription, which might explain why transcription levels alone are not predictive of expansion [333]. Thus, cycles of transcription and R-loop formation might accelerate repeat expansion for structure-prone repeats via ongoing DNA damage repair [180]. A cell-based model where transcription levels or R-loop formation could be controlled, repeat sequence and size altered, expansions monitored, and DNA damage repair mechanisms systematically tested (perhaps building on HeLa cell models recently described [187]) might allow more direct testing of these ideas.

A common theme among sources of repeat instability and expansion is DNA metabolism associated with strand separation and reannealing at microsatellite sequences. These events can lead to formation of non-canonical structures and recruitment of DNA damage responses that ultimately and inadvertently add more repeats. Thus, mechanisms meant to maintain and protect the genome can also lead to large tandem repeat expansions and cause human disease [11, 130, 333].

Microsatellite repeat expansion disorders

Since it was first discovered that microsatellite expansions can cause disease, at least two dozen microsatellite repeat expansion disorders have been subsequently reported (Table 1). The latest discoveries are autism spectrum disorders caused by expansions in fragile 7A (FRA7A) and fragile 2A (FRA2A) fragile site loci [209, 210]. Comparing and contrasting these disorders can highlight several trends. Almost half of the microsatellite expansion disorders result from CAG trinucleotide expansions, mostly occurring in coding exons. All STRs for known repeat expansion disorders are GC-rich except for the trinucleotide GAA repeat of FRDA and the ATTCT and TGGAA pentanucleotide repeats of spinocerebellar ataxia 10 (SCA10) and 31 (SCA31), respectively. In this review we focus on large microsatellite repeat expansions that are transcribed into RNA, a feature that is shared by nearly all repeat expansion disorders (Table 1).

Microsatellite expansions cause disease through two broad molecular mechanisms (Fig. 1): loss-of-function for the associated gene or gain-of-function for the repeat expansion sequence. In loss of function mechanisms, gene expression can be silenced at the transcriptional level, such as by epigenetic modification, resulting in the complete loss of that gene's normal functions [70, 112]. Alternatively, the affected gene may lose function at the protein level by the introduction of unusually long polypeptide tracts in the translated protein product (Fig. 1) [168, 268]. In gain-of-function mechanisms the repetitive polypeptide can take on new roles, such as protein aggregation. Many of these mutant misfolded proteins cannot be degraded efficiently and will accumulate in cellular aggregates or inclusions [48, 168, 332]. Aggregation also tends to sequester proteins and critical cellular components and is taxing on cellular proteostasis [48]. The xtrRNA can also acquire gain of function mechanisms, primarily through interaction with nucleic acid-binding proteins (Fig. 1). The repetitive xtrRNA forms length-dependent focal aggregates in cell nuclei in several diseases [35, 59, 196, 262, 311]. Loss-of-function and gain-of-function mechanisms can result in complicated molecular disease pathologies and some disorders can simultaneously exhibit multiple mechanisms (Table 1).

Transcription and splicing at simple tandem repeat expansions

Transcribing repeat expansion sequences

Repeat expansion sequences are known to inhibit or impede RNA Polymerase II (Pol II) initiation or elongation either directly or via induction of a repressed chromatin state [100]. Expansions like the GAA repeat in FRDA [19, 94, 97, 162, 231], the CTG repeat in myotonic dystrophy type 1 (DM1) [25], the GGGGCC repeat in *C9ORF72*-associated frontotemporal dementia and amyotrophic lateral sclerosis (C9FTD/ALS) [108], and the CGG repeat in FXS (also known as FRAXA) [44, 285] have all been implicated in reduced or silenced transcription. For FXS [230, 298], Fragile XE (FRAXE) [18], FRDA [97], FRA2A [210] and FRA7A [209], transcription

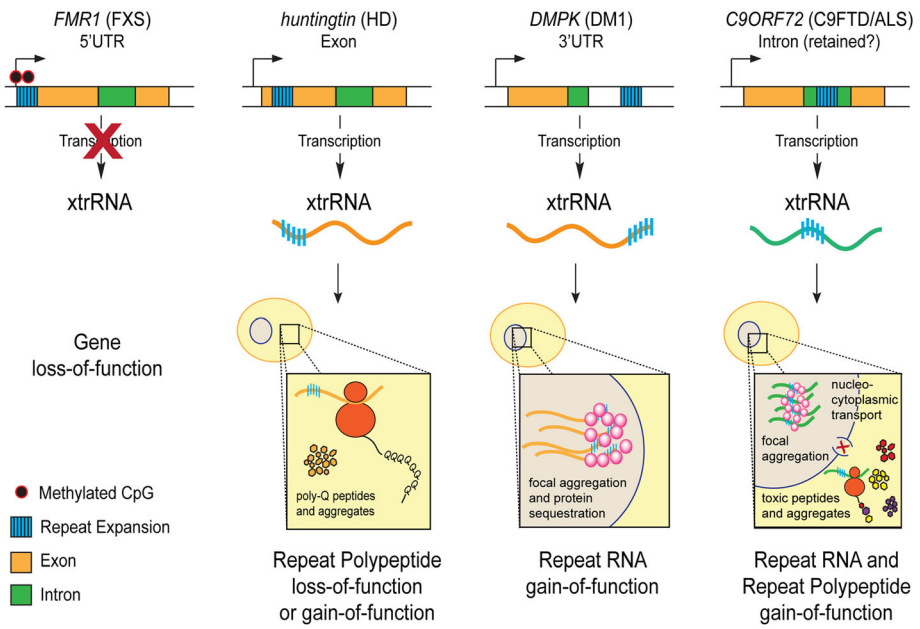


Fig. 1 Distinct loss-of-function and gain-of-function mechanisms of disease for various repeat expansion disorders. Repeat expansions can occur in 5′ or 3′ UTRs, exons, or introns. Expanded tandem repeat-containing RNA (xtrRNA) may not be transcribed due to epigenetic silencing, thereby causing loss of gene function. If transcribed, xtrRNA may become trapped in the cell nucleus where it can form focal aggregates and functionally deplete important RNA binding proteins. The xtrRNA may also be exported to the cytoplasm where it can undergo translation to produce repeat-containing polypeptides that disrupt cellular processes. In some cases, xtrRNA can form focal nuclear aggregates and also be translated into repeat-containing polypeptides. Repeat-containing polypeptides can be toxic in multiple ways, including insoluble aggregation, blocking normal host protein function, inhibiting nucleocytoplasmic transport, and disrupting other critical cellular functions

appears to be blocked or significantly reduced by DNA methylation of the repeat expansion or nearby CpG islands. However, although transcription may be well below basal levels, it is possible that xtrRNA can still contribute to disease in some cases [44].

Slowed or stalled transcription across repeat expansions may lead to R-loops, which further slow transcription [123] and inadvertently contribute to deposition of repressive chromatin marks and silence transcription (Fig. 2) [44, 99, 316, 317]. R-loops play important roles in biology, such as immunoglobulin class switching [323], keeping CpG islands unmethylated [91, 254], and defining transcription termination signals [254, 270]. R-loop formation is common in transcription of C-rich template sequences [324], which most disease-associated repeat expansion genomic loci possess. The impact of R-loop formation on disease at repeat expansions is still unclear. Whether R-loop formation will trigger DNA methylation, transcriptional silencing, or other events may be dependent upon a number of factors specific to the affected gene or locus.

Bidirectional transcription of repeat expansions

Bidirectional transcription has been reported to occur in DM1, C9FTD/ALS, Huntington's disease (HD), spinocerebellar ataxia 8 (SCA8), and Huntington's disease-like 2 (HDL2), among other diseases [26, 40, 126, 312]. Slowed transcription across a repeat may also be able to induce antisense transcription of the non-template DNA strand via R-loop formation [270]. For example, FRDA-associated GAA repeat expansion sequences were shown to initiate transcription and act as promoters in yeast [330]. However, many genes exhibit bidirectional transcription [293] and in microsatellite diseases bidirectional transcription typically initiates outside of the repeat (Fig. 2) [26, 113]. Bidirectional transcription across repeats can also result in double R-loops that amplify repeat instability and accelerate methylation and transcriptional silencing [181, 183, 223]. Antisense transcription can often interfere with transcription of the coding gene [145]. Most relevant to this review is the production of two xtrRNAs from bidirectional transcription and the potential to synthesize repetitive polypeptides from both xtrRNA. For example, in C9FTD/ALS both xtrRNAs form nuclear foci [59, 90, 248, 340] that sequester RNA-binding proteins [47, 175, 217] and are translated into repetitive polypeptides [9, 216], highlighting the importance of bidirectional transcription to molecular disease pathology.

The role of Supt4h in xtrRNA transcription

Transcribing microsatellite expansions into xtrRNA requires processivity across repetitive sequence tracts that

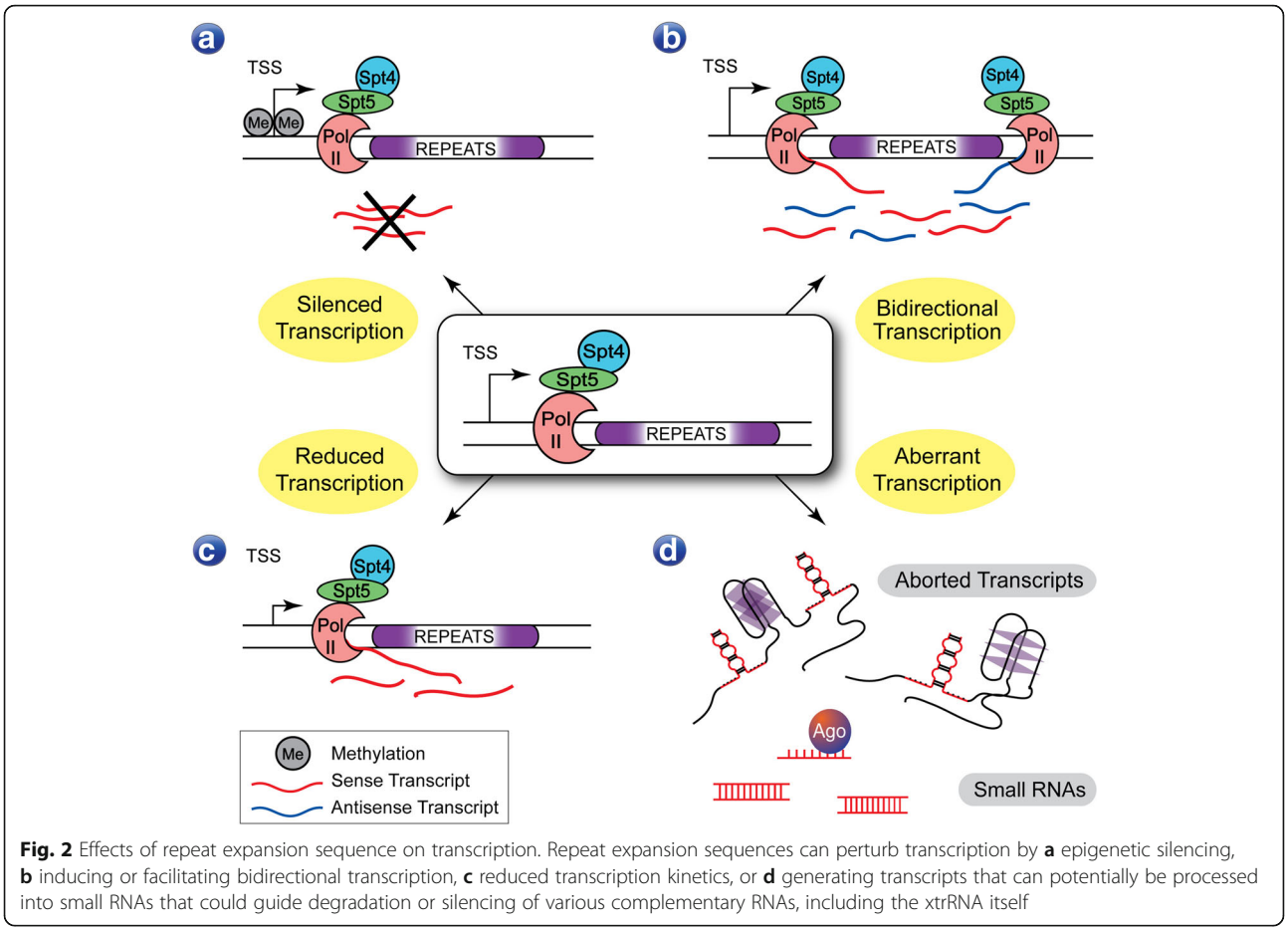


Fig. 2 Effects of repeat expansion sequence on transcription. Repeat expansion sequences can perturb transcription by **a** epigenetic silencing, **b** inducing or facilitating bidirectional transcription, **c** reduced transcription kinetics, or **d** generating transcripts that can potentially be processed into small RNAs that could guide degradation or silencing of various complementary RNAs, including the xtrRNA itself

can have very high GC content. The 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor (DSIF), composed of Supt4h and Supt5h proteins (Spt4 and Spt5 in yeast), aids RNA Polymerase II (Pol II) in transcription elongation and transcription rate [305, 308]. The DSIF complex is important for traversing sequences that elicit pausing of RNA Pol II [305] and has been identified as a factor involved in the transcription of RNA containing large simple repeat sequences. For example, transcription of repeat-containing RNA from the *huntingtin* and *C9ORF72* genes significantly decreases when Supt4h is deleted or knocked-down [159, 186]. Supt5h is a conserved transcription factor with a homolog known as NusG in bacteria that is important for elongation and processivity [185, 202]. Supt5h binds directly to the clamp coiled-coil domain of RNA Pol II while Supt4h interacts through contact with Supt5h [17, 119, 201]. Together, the DSIF complex interacts with the DNA template outside of the transcription bubble [17, 50, 151, 185]. Supt4h has a zinc-finger domain that may be important for modulating DNA interactions of DSIF [308], and thereby improve processivity by maintaining RNA Pol II template interaction during periods of extended pausing [50, 151, 309]. Long repetitive sequences prone to formation of secondary structure in the transcription bubble, such as repeat-induced hairpin or R-loop structures, may represent prime sites for pausing or backtracking [251, 260, 333].

DSIF is also used by RNA Pol I to presumably ensure robust transcription of abundant and repetitive ribosomal RNA [122, 309]. It is worth noting that repeat expansions might occur in ribosomal RNA genes but they have either not been characterized or have not been associated with disease [122]. In contrast, RNA Pol III, which only transcribes relatively small noncoding RNA genes, does not interact with the DSIF complex [309]. Thus, transcription is unlikely to be successful if large microsatellite expansions occur in the small RNA genes transcribed by RNA Pol III. These observations may lend some rationale as to why all disease-associated repeat expansions to date are associated with Pol II-transcribed genic regions [7, 31, 236].

Splicing of xtrRNA

Splicing involves several regulated steps, many accessory factors and the spliceosome, a complex multi-component

enzyme. There is currently a lack of mechanistic insight regarding how the splicing apparatus reacts when encountering pre-mRNA containing large repetitive sequence tracts [14]. Since introns can be excessively large while still allowing productive and accurate splicing [263], the size of the repeat expansion itself is not expected to significantly impede splicing. However, transcription rates across microsatellite expansions can be reduced, which can influence alternative splicing [58, 270], and stem loop structures in large pre-mRNA introns have been predicted to affect splicing [263].

Examples of microsatellite repeat expansions modulating splicing include the GAA repeat expansion associated with FRDA. When placed near reporter gene exons or in the first intron of a *frataxin* minigene system, the GAA repeat caused complex splicing defects and accumulation of aberrant splice products [15]. The mechanism proposed involved binding of various splicing factors to the GAA repeat-containing transcripts [15]. In C9FTD/ALS, the intronic GGGGCC repeat has been implicated in splicing by favoring retention of the intron-containing repeat, suggesting a mechanism by which *C9ORF72* xtrRNA can escape to the cytoplasm for translation [227]. Expanded CAG repeats of HD are also linked to production of short alternatively spliced forms of the *huntingtin* mRNA that contain the CAG repeat expansion and add to the production of toxic polyglutamine protein [255].

Potential impact on splicing factors

If repeat expansion sequences can mimic the binding motif of splicing regulators, they could recruit splicing factors and affect splice site selection. In DM1 the MBNL family of splicing factors and CUG binding proteins (CUGBPs) have an affinity for repetitive CUG and CAG sequence. Although the splicing of DM1 protein kinase (DMPK) mRNA does not appear to be affected by the CUG repeat expansion that it contains, the splicing pattern of an antisense transcript across the *DMPK* repeat, which contains a CAG repeat, appears to be altered by the expansion [105]. In HD the expanded CAG repeats have been proposed to interact with the splicing factor SRSF6, which is believed to contribute to altered splicing to generate truncated repeat-containing huntingtin mRNA [255].

Repeat expansion sequences in xtrRNA could also alter splicing by recruiting factors that are not typically involved in splicing. These factors might modulate splice site selection or spliceosome activity by changing local ribonucleoprotein (RNP) structure or access to splice signals [67]. The repetitive structural nature of repeat expansions could also sterically hinder access to splice signals, depending on their proximity to splice enhancer or silencer elements. Alternative splicing is a complicated interplay of modular protein and RNA interactions that are difficult to predict at present and local sequence and context will likely be important for understanding the impact of expanded repeats on splicing [14].

Therapeutic approaches to control xtrRNA transcription and splicing

Characterizing the effect of microsatellite expansions on transcription and splicing will directly benefit therapeutic approaches for repeat expansion disorders. Proof-of-principle methods to locally disrupt the interactions of xtrRNA at repeat expansion loci, such as R-loops, have been demonstrated for FXS and FRDA using small molecules and nucleic acids [44, 177]. Disrupting the interaction of Spt4 and Spt5, or modulating Spt4 function, could provide a therapeutic avenue for a number of repeat expansion disorders by reducing xtrRNA expression. This has been demonstrated for CAG and GGGGCC repeat expansions [159, 186] and might be particularly valuable in disorders exhibiting bidirectional transcription across the repeat expansion. For splicing-based therapeutics, blocking inclusion of repeat expansion-containing introns, such as with splice-modulating antisense oligonucleotides or small RNAs, could prove to be useful for disorders like FRDA and C9FTD/ALS.

With the emergence of gene editing technologies, the direct removal of repeat expansions from the genome may also be possible. Removal of genomic repeat expansions could eliminate the possibility of xtrRNA expression or reverse repressive epigenetic states. Careful SNP selection followed by targeting with CRISPR-Cas9 has been shown to block promoter function and silence the mutant expanded allele in HD [215] or completely delete large portions of the mutant HD allele [265]. Targeting CRISPR-Cas9 to sequences flanking the CTG repeat in DM1 also caused large repeat deletions [299]. In model cells of DM1, CRISPR-Cas9 was used to introduce a poly-A signal upstream of the CTG expansion in the DMPK gene to prevent CUG repeat transcription, which led to a reversal of molecular disease [318]. While potential CRISPR-based therapeutics are exciting, precautions must be taken to address potential pitfalls and challenges like off-target effects, delivery, and cell-type specific mechanisms of DNA damage repair [16, 54, 71, 85, 229, 238, 240, 322].

Structure, protein interactions, and localization of xtrRNA

Structure of xtrRNAs and targeting with small molecules

During and after the synthesis and processing of xtrRNA, the repetitive RNA will fold into repetitive and unique structures and interact with proteins that have an affinity for its sequence or structure. Watson-Crick

pairing apparently dominates folding since all atomic resolution investigations of disease-associated xtrRNA to date, including CAG, CUG, CCG, CGG, CCUG, AUUCU, and CCCCGG, are imperfect A-form-like duplexes [39, 62, 147, 234, 341]. These structures possess repeating units of Watson-Crick and mismatch paired nucleotides [147, 341]. While some studies have identified G-quadruplexes or tetraplexes [45, 79, 194, 247], other reports suggest that xtrRNA either do not form quadruplexes or are transient and interconvert readily with Watson-Crick paired conformations, especially as the number of repeating units increase [62, 108, 281, 329, 341]. Some reports of tetraplex structure may be the result of unusual interactions like dimerization between imperfect repeat duplex RNA, as was observed for CGG repeat RNAs [103]. Convincing evidence for the presence or biological significance of RNA G-quadruplexes inside human cells is still lacking [20, 107, 164, 194], therefore direct roles for quadruplex RNA in repeat expansion disease remain unclear.

Available structures of short repeat RNAs reveal A-form-like conformations with unique mismatches that may be targeted with artificial molecules to selectively bind repeat expansion RNA structure. Small molecule screening and structure-guided synthesis have experimentally identified a variety of small molecules that can bind xtrRNA, such as the CUG, CCUG, CGG, and GGGGCC repeat RNAs associated with DM1, DM2, FXS or FXTAS (Fragile X-associated tremor/ataxia syndrome), and C9FTD/ALS, respectively [38, 39, 226, 281, 292, 314, 325]. These molecules have been shown to stabilize repetitive structure or disrupt protein binding, which can correct molecular disease markers like nuclear RNA foci and repetitive polypeptide translation, or improve pathology in cells and animal models. Although promising, their eventual therapeutic application will need to demonstrate exquisite specificity for the RNA target, minimal non-specific interactions, and pharmacologic safety and efficacy [102, 252].

Protein interactions and localization of xtrRNA

Both sequence specific and structure specific interactions likely underlie protein binding to xtrRNA. The repetitive nature of xtrRNA can result in multiple tandem binding sites for proteins. In DM1, the disease-associated xtrRNA contains hundreds or thousands of CUG repeats that bind and recruit possibly as many copies of MBNL-1 protein and potentially other CUG-binding proteins [173, 197, 235]. MBNL-1 recognizes CG dinucleotides separated by 1-17 nucleotides [92], which include motifs in pre-mRNA where MBNL-1 helps to regulate splicing [245]. Examples include the pre-mRNA of sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase 1 (SERCA1), which contains several YGCU(U/G)Y motifs downstream from exon 22. MBNL-1 usually interacts with these motifs to cause inclusion of exon 22 but in DM1 exon 22 is excluded during splicing [118, 34]. Blocking the interaction of proteins with DM1 xtrRNA by using morpholino oligonucleotides rescued splicing defects and molecular pathology [310]. Thus, a major contributor to disease mechanism in DM1 is the sequestration of splicing factors, particularly MBNL proteins.

A number of diseases are characterized by binding of specific proteins to xtrRNA or colocalization of proteins with xtrRNA focal aggregates (Table 1) [327]. These include proteins like MBNL-1 in DM1, DM2, HD, spinocerebellar ataxia 3 (SCA3), SCA8 and HDL2 [197, 282, 327], hnRNP K in SCA10 and C9FTD/ALS [46, 311], Pur-α, hnRNP F and SRSF2 in C9FTD/ALS [47, 108, 319], and Sam68 and hnRNP A2/B1 in FXTAS [262, 276]. As such, protein interactions with xtrRNA play key roles in disease mechanism and are expected to be important mediators of aberrant xtrRNA localization and aggregation [214]. Foci containing xtrRNA are believed to be the result of RNA-binding protein sequestration that can functionally deplete those proteins and partially protect the xtrRNA from degradation [214, 327].

Sequence specific interactions may not entirely explain xtrRNA localization or foci formation. While certain proteins that prefer to bind G-rich sequence, like hnRNP H/F, have been found to associate strongly with the GGGGCC repeats of C9FTD/ALS, other interacting proteins do not appear to have strong GGGGCC sequence-binding specificity, such as ALY/REF, SC-35, SF2, and nucleolin [47, 108, 175]. Imperfect A-form-like duplexes, or duplexes inter-converting with tetraplex conformations, may attract proteins that recognize the unique structures of xtrRNA rather than the specific sequence. Glycine-arginine-rich (GAR) proteins containing RGG/RG motifs, for example, are believed to recognize the structure of their nucleic acid partners rather than sequence [289]. The GAR domain-containing proteins FUS (fused in sarcoma), FMRP, and hnRNP U all recognize structured guanine-rich RNA sequences with an apparent preference for transitions between canonical duplexes and non-canonical structures like quadruplexes [233]. One explanation for foci is that proteins bind specifically to repeat sequence or structural elements of xtrRNA and then seed aggregation that recruits additional secondary interacting factors. Thus, xtrRNA may form foci by either merging with existing nuclear bodies or else establishing their own novel versions of RNA granules. While focal aggregation of xtrRNA can be detrimental to sequestered protein function, it may also protect the cell by preventing nuclear escape and translation of repeat RNAs [150].

xtrRNA localization and membrane-free cellular organelles

Whether there is a specific localization pattern of xtrRNA inside cell nuclei is not entirely clear. Foci might be expected to nucleate at the site of transcription. DMPK mRNA usually localizes to SC-35 splicing speckles after transcription. However, when containing CUG repeat expansions, the DMPK mRNA has been shown to localize peri-transcriptionally outside of SC-35 splicing speckles [274]. RNA containing CAG, CUG and GGGGCC repeats were also shown to localize to SC-35 splicing speckles and nuclear speckles [132, 295]. However, in other studies the xtrRNAs, specifically CUG and CGG RNAs, appeared to form foci stochastically [243, 280]. Live cell imaging of Spinach2 aptamer-tagged CGG repeat xtrRNA revealed rapid aggregation and formation of very stable foci [280]. CGG xtrRNA foci were additionally found to be mobile and dynamic and colocalized with Sam68 protein. They migrated around the nucleus over time and could be seen to merge into larger foci or disaggregate into smaller foci. Live cell imaging of CUG repeat xtrRNA tagged with the MS2-GFP system found similar effects for aggregation, foci formation and dynamics [243]. CUG repeat RNA foci formation depended on the presence of MBNL-1 protein. In live-cell experimental approaches the xtrRNA is likely to be over-expressed from an artificial genetic context and may not represent the true dynamics or localization of endogenous repeat expansions. Nonetheless, live and fixed cell imaging have revealed that xtrRNA foci are dynamic, stable aggregates that likely depend on protein interactions and may co-localize with known nuclear bodies.

Nuclear bodies can be built around RNA and the molecular forces that govern nuclear body formation may help explain xtrRNA foci formation and localization. For example, nuclear paraspeckles depend on the long noncoding RNA NEAT1 (nuclear paraspeckle assembly transcript 1) [321]. Nuclear bodies are essentially membrane-free organelles that are held together by transient or dynamic protein-protein and protein-RNA interactions. These interactions collectively provide a type of phase separation to organize and compartmentalize cellular processes [336]. It was recently demonstrated that CAG, CUG and GGGGCC repeat containing RNAs form soluble aggregates with sol-gel phase separation properties and behave similar to liquid-like droplets [132]. These properties were dependent on the repeat expansion length and base-pairing interactions. In contrast, CCCCGG repeats did not form phase transitions, suggesting that not all xtrRNA will possess these properties. Interestingly, guanine-rich nucleic acids are less soluble than other nucleic acids and appear to be intrinsically aggregate-prone apart from protein, especially when packing into quartets or higher-order quadruplex structures [21, 89, 179]. The disruption of membrane-free organelles, which are abundant in the nucleus, is linked to disease [198, 228, 272]. In fact, the disruption of membrane-free organelle assembly and dynamics by repetitive poly-glycine-arginine (poly-GR) and poly-proline-arginine (poly-PR) translation products has emerged as a leading molecular disease mechanism for C9FTD/ALS [165, 174, 182]. Association of certain proteins with xtrRNA, dependent upon RNA sequence and structure, may strongly influence the subsequent localization of xtrRNA with membrane-free cellular compartments.

Abundance and turnover of xtrRNA

Abundance of foci-forming xtrRNA

Understanding the biology of an RNA includes knowing the effective concentration or abundance of that RNA and its turnover and decay pathways. Three current studies highlight the importance of characterizing cellular xtrRNA abundance. The cellular abundance of CUG repeat-containing transcripts was recently measured using transgenes and endogenous DMPK RNA in mouse models of DM1 and human tissues from DM1 patients [104]. Surprisingly, a large 1000-fold discrepancy for transcript number was discovered across mouse models. In human samples only a few dozen DMPK mRNA molecules were detected per cell, with only half of those expected to contain the repeat expansion. In a similar study looking at the abundance and processing of an antisense transcript across the *DMPK* repeat expansion, only a handful of repeat containing antisense transcripts were quantified per cell [105]. Quantification of the repeat-containing intron of C9ORF72 in C9FTD/ALS patient cells found only a few copies per cell, concluding that each foci might be composed of as few as one xtrRNA transcript [188]. Therefore, one or a few copies of xtrRNA may be enough to generate focal aggregates. Importantly, the stochastic nature of foci formation, where many cells contain no foci but some contain several, suggests that there may be a disproportionate contribution to disease for xtrRNA at the individual cell level [188]. These reports indicate that knowing the number and type of xtrRNA species inside of cells will be important for correct interpretation of data and for understanding the role of xtrRNA in disease.

Nuclear xtrRNA retention and surveillance mechanisms

The nuclease enzymes primarily responsible for degrading nuclear RNA are the exosome complex (3'-5' exoribonuclease activity) and 5'-3' exoribonuclease 2 (XRN2) [67]. These enzymes act as part of a nuclear RNA quality control and surveillance pathway that monitors transcription, splicing, and 3'-end formation of pre-mRNAs, as well as their packaging into mRNP particles (Fig. 3)

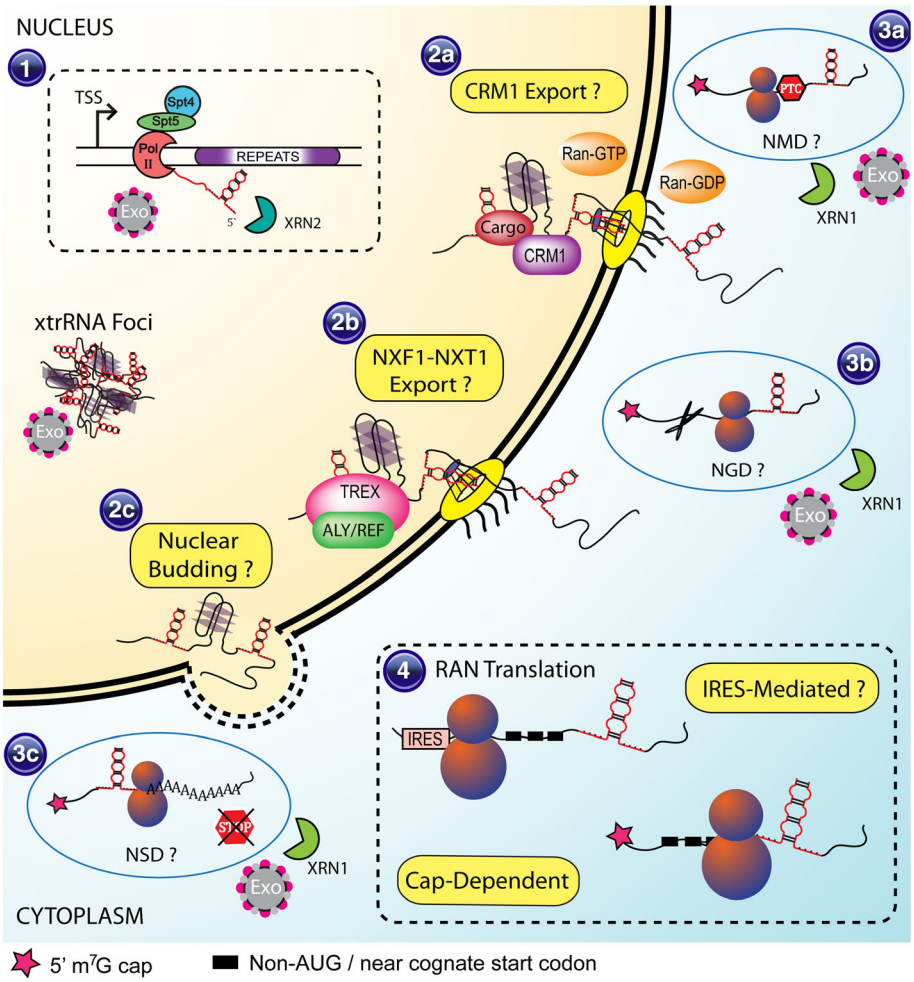


Fig. 3 Possible mechanisms of nuclear and cytoplasmic RNA surveillance, nuclear export, and translation of xtrRNA. RNA containing large repeat expansion sequences may be subject to nuclear RNA surveillance mechanisms, including degradation by the nuclear exosome (1) or the XRN2 5'-3' exoribonuclease (1). Export of xtrRNA likely involves bulk mRNA transport via NXF1 (2b), but may also include alternative mechanisms like CRM1-mediated export (2a) or possibly nuclear envelope budding (2c). Cytoplasmic RNA surveillance mechanisms that may control xtrRNA levels and translation include nonsense-mediated decay (NMD) (3a), no-go decay (NGD) (3b), or nonstop decay (NSD) (3c). Translation of xtrRNA is likely to follow canonical cap-dependent translation (4), especially when repeat expansions are embedded in normal coding regions of an mRNA, but may potentially involve internal ribosome entry site (IRES)-like mechanisms (4). RAN translation has been shown to be cap-dependent for some repeat expansions, but complete mechanistic details remain to be determined

[67, 146, 338]. Instead of degradation, these pathways can also signal for retention of aberrant transcripts in the nucleus, typically at the site of transcription [52, 220] but sometimes near nuclear pores [67]. Retention at the site of transcription is coupled to nuclear exosome activity, particularly the Rrp6p subunit [66, 220]. The TPR protein, a mammalian ortholog of yeast Mlp1/2p, is implicated in retention at nuclear pores for mRNAs with retained introns that normally exit the nucleus through the nuclear export factor 1 (NXF1) pathway [49]. Both of these mechanisms may be relevant to xtrRNA, especially when repeat expansions are found in retained introns [43, 110, 227].

Surveillance mechanisms are also related to transcription or splicing of xtrRNA since these might be expected to trigger degradation [67, 146]. However, the existence of foci and nuclear export of xtrRNA argue that surveillance mechanisms are incomplete or inefficient for xtrRNA removal. At present it is unknown how many molecules of any repeat expansion-containing RNA are synthesized versus how many survive to form foci or exit the nucleus for translation. It is likely that repeat expansion-containing RNAs survive due to protection by protein binding, such as hnRNP proteins [125, 327]. Alternatively, factors responsible for recruiting RNA to the nuclear exosome, such as the TRAMP (Trf4/Air2/

Mtr4p Polyadenylation) complex or NEXT (nuclear exosome targeting) complex, are unable to efficiently recognize and bind the xtrRNA [146, 259]. Thus, xtrRNA transcripts may escape degradation if they appear as "normal" mRNPs, having undergone proper RNA processing like capping, splicing and polyadenylation and are associated with appropriate post-processing factors.

Turnover and decay of xtrRNA

An unanswered question remains as to whether foci might contain partially degraded fragments of repeat RNA in addition to larger intact transcripts. A case in point is C9FTD/ALS where the microsatellite expansion occurs in an intron but nuclear RNA foci and cytoplasmic translation are both observed. When introns are spliced out of pre-mRNA transcripts they are typically destined for rapid degradation unless they contain a stably folding RNA element or recruit RNA binding proteins [115]. Examples include small nucleolar RNAs and microRNAs [56, 203]. It is possible that the structures that repeat expansion RNAs form, as well as the proteins that they bind, allow them to persist, accumulate, and aggregate as foci [3]. At present the exact type and number of xtrRNA species that are trapped in foci versus free and soluble for any disease is unknown. It is also not known whether partially degraded xtrRNA fragments are stable enough to accumulate. Furthermore, there is no distinction between what are the major species or which are nuclear versus cytoplasmic.

When mRNAs that contain microsatellite repeat expansions reach the cytoplasm they may encounter additional quality control mechanisms designed to eliminate aberrant mRNA and prevent its translation. These include nonsense-mediated decay (NMD), no-go decay (NGD), and non-stop decay (NSD) (Fig. 3) [267]. NMD recognizes premature stop codons through possibly multiple mechanisms. NMD can be triggered if exon-junction complexes (EJCs), which are deposited near splicing junctions, are encountered downstream of a stop codon [172, 199]. Another mechanism of NMD may involve the relative length of sequence 3' to the stop codon [4, 172]. Repeat expansions could conceivably alter the positioning of a stop codon or extend the 3' UTR region and trigger NMD. When ribosome translation is significantly slowed or stalled then NGD can be triggered. Stalling is thought to be initiated by unusually stable RNA structures or protein-binding motifs and result in endonucleolytic cleavage [63]. Since repeat expansions are believed to fold into stable hairpins or tetraplexes they could possibly trigger NGD. In the case of NSD, the ribosome can become stuck on mRNA that does not possess a stop codon. These complexes must be resolved by cleavage and degradation of the mRNA [82, 302]. NSD could become activated if repeat expansions alter the reading frame or cause loss of the stop codon, such as via mis-splicing.

For xtrRNA embedded in mRNAs, either as exons or introns, mRNA surveillance mechanisms should act to reduce translation. Activation of these pathways will lead to degradation by the cytoplasmic version of the exosome and XRN1, a 5'-3' exoribonuclease. Accessory factors that guide cleavage include the Ski complex (composed of Ski2, Ski3, and Ski8 proteins) and the Ski7 protein [5, 267]. However, similar to nuclear RNA surveillance mechanisms, the cytoplasmic pathways seem unable to detect or completely remove xtrRNA and prevent translation. Methods to enhance RNA surveillance mechanisms might represent reasonable targets for therapeutic intervention. For example ataluren (PTC124) is a small molecule drug that increases NMD and might sensitize mRNA surveillance to repeat expansions [73].

Some studies have uncovered potential RNA turnover mechanisms associated with repeat expansions. In a recent RNAi screen several RNA processing factors were identified as suppressors of toxicity in *C. elegans* expressing a $(CUG)_{123}$ repeat expansion in a 3'-UTR of GFP [88]. These factors included RNases, helicases and RNA binding proteins that, when knocked-down, caused increased toxicity and enhanced nuclear foci formation. A nuclear pore complex (NPC) protein, *npp-4*, was also a suppressor in this screen. Interestingly, *smg-2*, a conserved helicase and central component of the NMD pathway, was a strong suppressor. Knock-down of NMD resulted in a several-fold increase in GFP-$(CUG)_{123}$ RNA expression levels and increased GFP translation. The substantial increase in 3'-UTR GC-content was identified as the likely trigger for NMD of the GFP-$(CUG)_{123}$ RNA [88]. *Smg-2* was also identified as a suppressor of poly-glutamine aggregation previously, likely through NMD of aberrant repeat-containing HTT transcripts [328]. These results provide one example of the role of RNA turnover in controlling toxicity of xtrRNA in repeat expansion disorders.

Nuclear export and translation of xtrRNA

Canonical mRNA export pathways

RNA export from the nucleus involves several distinct pathways depending on the RNA and the various protein factors that constitute the RNP particle [273]. For nuclear mRNA there are two main export pathways: NXF1-mediated and chromosome region maintenance 1 (CRM1)-mediated, although NXF1 is the primary transport system for bulk mRNAs (Fig. 3) [60]. Both mechanisms rely on adapter proteins to specify the RNA cargo for export. Many of the factors required for successful export are deposited co-transcriptionally and post-transcriptionally during mRNP assembly [22]. The C-

terminal domain of RNA Pol II serves as a docking platform for a wide variety of mRNA processing and mRNP assembly proteins and plays critical roles in establishing mRNP composition [139]. Correct processing of mRNA, such as capping, splicing and 3'-end formation, determines the ability of these factors to bind mRNA and assemble export-competent mRNP particles [36, 60, 335].

The mRNA-associated adapter proteins and complexes represent a complicated matrix of possible interactions that dictate export efficiency [22, 60]. NXF1 can specifically bind the constitutive transport element found in some mRNAs and viral RNAs, like that of the Mason Pfizer monkey virus, to directly facilitate export [101, 178]. However, for bulk cellular mRNA transport NXF1 uses adapters like TREX (transcription export complex) [60, 111]. Although the NXF1 protein interacts loosely with RNA, TREX helps mediate specific binding through its subunit ALY/REF [124], a protein previously reported to interact with *C9ORF72* GGGGCC RNA repeats [47]. TREX associates with mRNA during synthesis and processing via mRNA capping and splicing events [138, 204] and appears to be primarily recruited to the 5' ends of mRNAs in human cells via interaction between ALY/REF and the cap binding complex component CBP80 [36]. In addition to ALY/REF, TREX is composed of the THO complex, CIP29, and UAP56, a component of the EJC [37, 53, 157]. For repeat expansion disorders, NXF1 seems to be the most likely pathway since disease-associated xtrRNA are transcribed from coding gene loci and TREX is deposited onto mRNAs early during transcription [204].

Nuclear export of xtrRNA

A recent study connected NXF1 transport of C9FTD/ALS intronic xtrRNA via interaction with the export adapter SR-rich splicing factor 1 (SRSF1) [110]. SRSF1 appeared to interact and colocalize with C9ORF72 xtrRNA. Depletion of SRSF1 prevented neurodegeneration in a fly model and suppressed cell death in patient-derived motor neurons and astrocytes. Depleting SRSF1 or preventing interaction with NXF1 inhibited nuclear export of repeat-containing *C9ORF72* transcripts and blocked RAN translation. Thus, SRSF1 might serve as a therapeutic target in C9FTD/ALS. This report highlights the value of understanding RNA biology in the context of repeat expansion disorders.

Most disease-associated xtrRNA is embedded in exonic or untranslated regions (Table 1) and therefore likely exits the nucleus via mRNA export pathways. CRM1 exports proteins and their associated RNAs via interaction with nuclear export signal sequences and Ran-GTP (Fig. 3) [60, 74, 77]. CRM1 interacts directly with the NPC at the nuclear periphery and commonly exports noncoding RNAs like spliceosomal RNA (snRNA) [10, 74, 131]. There is no reported RNA binding affinity of CRM1 so selective export of mRNAs depends on the RNA-binding properties of its cargo proteins [60]. Export of xtrRNA by CRM1 might only require that the repeat expansion sequence or structure somehow recruit a CRM1 cargo protein.

Export of intronic xtrRNA would be expected to require aberrant splicing that resulted in its retention in mRNA, as has been implicated for the intronic C9FTD/ALS repeat expansion [227]. Alternative export pathways exist but seem unlikely given their very specific nature. For example, transfer RNA (tRNA) undergoes multiple maturation phases that cumulatively result in two separate import and export steps [273]. These export pathways involve specific RNA-protein interactions, such as EXP-t and EXP5 [8, 29], that are unlikely to mediate xtrRNA export. For any export pathway through the NPC, xtrRNA must somehow establish RNP complexes that pass the requisite tests for licensing of export.

Nuclear exit of RNP granules, such as nuclear xtrRNA foci, might also be possible through nuclear envelope budding (Fig. 3). This mechanism involves TorsinA, nuclear lamina, and other uncharacterized factors. Nuclear budding was discovered as part of the nuclear egress mechanism of large nucleocapsid particles of Herpes viruses [55, 78, 200, 221, 277]. Nuclear envelope budding has been found to be a natural process for nuclear release of large RNP complexes during development of neuromuscular junctions in *Drosophila melanogaster* [135, 277]. However, knock-out of TorsinA in HeLa cells had little impact on Herpes virus production [294], suggesting alternative factors or mechanisms in human cells. If xtrRNA is exported by nuclear envelope budding it would have to mimic specific RNP granule formation that elicits nuclear envelope budding, which at present involves mechanisms that are largely uncharacterized [78].

Translation of xtrRNA

If xtrRNA can successfully exit the nucleus it is a potential candidate for translation. However, mRNAs that contain expanded tandem repeats are possibly the only practical source for translation of repeat expansion polypeptides since they contain the prerequisite sequence elements and protein factors to mediate canonical cap-dependent translation. These typically include 5' cap structures, bound eukaryotic initiation factors (eIFs), a poly-A tail, and appropriate mRNP complexes like the EJC [116, 301, 315]. Translation of xtrRNA sequence embedded within the coding exon of a gene, such as is found in SBMA, HD, DRPLA (dentatorubral-pallidoluysian atrophy), OPMD (oculopharyngeal muscular dystrophy) and several of the SCA

disorders (Table 1), are translated by canonical mechanisms. Most repeat expansions form stable secondary structures that have been shown to reduce the amount of overall translation by presumably inducing stalling, frame-shifting or abortive translation [72, 222, 244, 313]. In contrast, the specific binding of MID1 protein to huntingtin mRNA, which contains CAG repeat expansions, has been reported to enhance translation and lead to greater levels of aberrant protein [160]. This mechanism has also been proposed to enhance translation of other CAG repeat expansion-containing genes that cause disease [98]. Canonical translation of repeat expansions that are found in-frame in coding sequences is expected to generate otherwise normal protein that simply contain long tracts of repetitive polypeptide [296].

Repeat-associated non-AUG translation

The translation of noncoding xtrRNA irrespective of a canonical start codon was recently discovered and termed repeat-associated non-AUG (RAN) translation [42, 96, 290, 339]. Repeat expansion diseases where this mechanism has been observed now include SCA2, SCA8, SCA31, HD, FXTAS/FXPOI, and C9FTD/ALS [9, 13, 27, 96, 129, 218, 261, 290, 339, 340]. RAN translation of xtrRNA sequence can occur in many contexts, including repeat expansions found in untranslated regions, retained introns, and even those embedded in coding exons [96]. The mechanisms of RAN translation remain poorly understood and could involve several scenarios, possibly even internal ribosome entry site (IRES)-like mechanisms (Fig. 3) [96, 339]. For the CGG repeats of FMR1 that cause FXTAS a more straightforward mechanism is emerging. In this case, RAN translation is m^7G cap-dependent where a pre-initiation complex scans the RNA looking for a start codon [96, 141]. When the CGG repeats are present and stable structures are presumably encountered then stalling occurs and significantly enhances the ability of the ribosome to select a near-cognate start codon, or possibly any codon, to initiate translation [141, 154, 158, 339]. A similar mechanism is favored for the CAG and CUG repeats of sense and antisense transcripts in SCA8 [339]. This mechanism is proposed to allow translation initiation upstream of a repeat expansion in multiple reading frames [42, 96]. The sequence context, such as the leader sequence during scanning, the types of potential near-cognate start codons, and the repeat expansion sequence and size all appear to modulate the degree of RAN translation [12, 141, 261, 339].

The mechanism of RAN translation may be related to translation of upstream open reading frames (uORFs), a widespread phenomena revealed through high-throughput ribosomal footprint profiling [128]. RAN translation could even represent a specialized form of uORF translation that is triggered by stable xtrRNA structures. Both mechanisms can initiate at near-cognate start codons (although RAN translation may use other codons or other mechanisms, like frame-shifting) and are influenced by surrounding sequence context that might impact RNA folding or protein interactions [96, 117].

Recent investigations have demonstrated that certain RAN translation products of C9FTD/ALS disrupt the function of membrane-free cellular organelles, such as stress granules, Cajal bodies and the nucleolus [174, 182]. These polypeptides seem to block the formation or critical interaction dynamics of membrane-free organelles and RNA granules, which are important for neuronal cell signaling and health [269, 288]. Transport of macromolecules through the nuclear pore complex depends on interactions that resemble membrane-free organelle structure [207, 219]. They are organized by dynamic protein interactions of low complexity domain proteins, including phenylalanine-glycine (FG) repeats, which may explain why certain C9FTD/ALS RAN translation products are reported to disrupt nucleocytoplasmic transport [80, 136, 264, 326]. RAN translation products can also aggregate and are implicated in the disruption of a variety of other pathways [12, 42, 96, 142, 286, 339].

Several important questions remain concerning the mechanisms of RAN translation. For example, how similar are the mechanisms of RAN translation across diverse repeat expansion and sequence contexts [42, 96]? RAN translation maybe a spectrum of related mechanisms based upon modulation of ribosomal scanning, translation initiation, and translation elongation [301]. RAN translation can initiate just upstream from the repeat expansion, but how often can RAN translation initiate within the repeat sequence itself [339]? *In vitro* and cell-based model systems suggest that RAN translation can proceed uninterrupted through an entire repeat expansion [141, 213, 339, 340]. Yet some expansions are massive in size. Therefore, how often do repeat expansions induce frame-shifting or possibly even early translation termination [313]? Also, what factors are unique to RAN translation? Finding answers to these mechanistic questions may be critical for developing future therapeutic molecules that can target and selectively block xtrRNA translation.

Conclusion

RNA species that contain simple tandem repeat sequences occupy an underexplored world of RNA biology. Recent studies have begun to revisit the transcription and translation of repeat expansions. However, significant gaps remain for processes like cellular

transport and turnover of xtrRNA. Placing repeat expansion disease mechanism studies in the context of current RNA biology will help reveal a better understanding of how the cell deals with xtrRNA and identify mechanisms unique to repeat expansions.

Investigations into the biology of xtrRNA promise to unlock new approaches to therapeutics. Transcription across repeat expansions has opportunities for therapeutic development, such as modulating the function of Supt4h. Likewise, translation of repeat expansions, especially RAN translation, may become more targetable as molecular mechanisms become better characterized and specific factors identified. Selectively blocking both the synthesis of xtrRNA or its translation are attractive therapeutic approaches since they could extrapolate to multiple repeat expansion disorders. Turnover of xtrRNA should become increasingly important since several potential therapeutic strategies employ targeted and selective degradation of repeat expansion-containing RNA, such as antisense oligonucleotides and small interfering RNAs [64, 134, 169, 239, 275].

Funding
This work was supported by a United States Department of Defense Amyotrophic Lateral Sclerosis Research Program (ALSRP) Grant and a Judith and Jean Pape Adams Charitable Foundation Grant to K.T.G.

Authors' contributions
KJR performed an initial literature survey and draft of the manuscript and prepared tables and figures. KTG performed an in-depth literature survey, finalized tables and figures, and wrote the final draft of the manuscript. Both authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Received: 28 June 2017 Accepted: 22 August 2017
Published online: 29 August 2017

References
1. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65. doi:10.1038/nature11632
2. Adihe Lokanga R, Zhao XN, Entezam A, Usdin K (2014) X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the fragile X-related disorders: implications for the mechanism of repeat expansion. Hum Mol Genet 23:4985–4994. doi:10.1093/hmg/ddu213
3. Akiyama BM, Eiler D, Kieft JS (2016) Structured RNAs that evade or confound exonucleases: function follows form. Curr Opin Struct Biol 36:40–47. doi:10.1016/j.sbi.2015.12.006
4. Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature 432:112–118. doi:10.1038/nature03060
5. Anderson JS, Parker RP (1998) The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J 17:1497–1506. doi:10.1093/emboj/17.5.1497
6. Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995) Alu repeats: a source for the genesis of primate microsatellites. Genomics 29:136–144. doi:10.1006/geno.1995.1224
7. Arimbasseri AG, Rijal K, Maraia RJ (2014) Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation. Transcription 5:e27639. doi:10.4161/trns.27369
8. Arts GJ, Kuersten S, Romby P, Ehresmann B, Mattaj IW (1998) The role of exportin-t in selective nuclear export of mature tRNAs. EMBO J 17:7430–7441. doi:10.1093/emboj/17.24.7430
9. Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW, 3rd, Rademakers Ret al (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron 77: 639-646. doi:10.1016/j.neuron.2013.02.004
10. Bai B, Moore HM, Laiho M (2013) CRM1 and its ribosome export adaptor NMD3 localize to the nucleolus and affect rRNA synthesis. Nucleus 4:315–325. doi:10.4161/nucl.25342
11. Bak ST, Sakellariou D, Pena-Diaz J (2014) The dual nature of mismatch repair as antimutator and mutator: for better or for worse. Front Genet 5:287. doi:10.3389/fgene.2014.00287
12. Banez-Coronel M, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, Pletnikova O, Borchelt DR, Ross CA, Margolis RL et al (2015) RAN Translation in Huntington Disease. Neuron 88:667–677. doi:10.1016/j.neuron.2015.10.038
13. Banez-Coronel M, Porta S, Kagerbauer B, Mateu-Huertas E, Pantano L, Ferrer I, Guzman M, Estivill X, Marti E (2012) A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity. PLoS Genet 8:e1002481. doi:10.1371/journal.pgen.1002481
14. Baralle D, Buratti E (2017) RNA splicing in human disease and in the clinic. Clin Sci (Lond) 131:355–368. doi:10.1042/CS20160211
15. Baralle M, Pastor T, Bussani E, Pagani F (2008) Influence of Friedreich ataxia GAA noncoding repeat expansions on pre-mRNA processing. Am J Hum Genet 83:77–88. doi:10.1016/j.ajhg.2008.06.018
16. Beaudet AL, Meng L (2016) Gene-targeting pharmaceuticals for single-gene disorders. Hum Mol Genet 25:R18–R26. doi:10.1093/hmg/ddv476
17. Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P (2016) Structure of transcribing mammalian RNA polymerase II. Nature 529:551–554. doi:10.1038/nature16482
18. Biancalana V, Taine L, Bouix JC, Finck S, Chauvin A, De Verneuil H, Knight SJ, Stoll C, Lacombe D, Mandel JL (1996) Expansion and methylation status at FRAXE can be detected on EcoRI blots used for FRAXA diagnosis: analysis of four FRAXE families with mild mental retardation in males. Am J Hum Genet 59:847–854
19. Bidichandani SI, Ashizawa T, Patel PI (1998) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am J Hum Genet 62:111–121. doi:10.1086/301680
20. Biffi G, Di Antonio M, Tannahill D, Balasubramanian S (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat Chem 6:75–80. doi:10.1038/nchem.1805
21. Biyani M, Nishigaki K (2005) Structural characterization of ultra-stable higher-ordered aggregates generated by novel guanine-rich DNA sequences. Gene 364:130–138. doi:10.1016/j.gene.2005.05.041
22. Bjork P, Wieslander L (2014) Mechanisms of mRNA export. Semin Cell Dev Biol 32:47–54. doi:10.1016/j.semcdb.2014.04.027
23. Brais B, Bouchard JP, Xie YG, Rochefort DL, Chretien N, Tome FM, Lafreniere RG, Rommens JM, Uyama E, Nohira Oet al (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nat Genet 18:164–167. doi:10.1038/ng0298-164
24. Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415. doi:10.1086/301869
25. Brouwer JR, Huguet A, Nicole A, Munnich A, Gourdon G (2013) Transcriptionally Repressive Chromatin Remodelling and CpG Methylation in the Presence of Expanded CTG-Repeats at the DM1 Locus. J Nucleic Acids 2013:567435. doi:10.1155/2013/567435
26. Budworth H, McMurray CT (2013) Bidirectional transcription of trinucleotide repeats: roles for excision repair. DNA repair 12:672–684. doi:10.1016/j.dnarep.2013.04.019
27. Buijsen RA, Visser JA, Kramer P, Severijnen EA, Gearing M, Charlet-Berguerand N, Sherman SL, Berman RF, Willemsen R, Hukema RK (2016) Presence of inclusions positive for polyglycine containing protein,

FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. Hum Reprod 31:158–168. doi:10.1093/humrep/dev280
28. Buschiazzo E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryotic genomes. BioEssays : news and reviews in molecular, cellular and developmental biology 28:1040–1050. doi:10.1002/bies.20470
29. Calado A, Treichel N, Muller EC, Otto A, Kutay U (2002) Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. EMBO J 21:6216–6224
30. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A et al (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271:1423–1427
31. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M et al (2015) Resolving the complexity of the human genome using single-molecule sequencing. Nature 517:608–611. doi:10.1038/nature13907
32. Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci U S A 94:1041–1046
33. Chan NL, Guo J, Zhang T, Mao G, Hou C, Yuan F, Huang J, Zhang Y, Wu J, Gu L et al (2013) Coordinated processing of 3' slipped (CAG)n/(CTG)n hairpins by DNA polymerases beta and delta preferentially induces repeat expansions. J Biol Chem 288:15015–15022. doi:10.1074/jbc.M113.464370
34. Charlet BN, Savkur RS, Singh G, Philips AV, Grice EA, Cooper TA (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. Mol Cell 10:45–53
35. Chen IC, Lin HY, Lee GC, Kao SH, Chen CM, Wu YR, Hsieh-Li HM, Su MT, Lee-Chen GJ (2009) Spinocerebellar ataxia type 8 larger triplet expansion alters histone modification and induces RNA foci. BMC Mol Biol 10:9. doi:10.1186/1471-2199-10-9
36. Cheng H, Dufu K, Lee CS, Hsu JL, Dias A, Reed R (2006) Human mRNA export machinery recruited to the 5' end of mRNA. Cell 127:1389–1400. doi:10.1016/j.cell.2006.10.044
37. Chi B, Wang Q, Wu G, Tan M, Wang L, Shi M, Chang X, Cheng H (2013) Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. Nucleic Acids Res 41:1294–1306. doi:10.1093/nar/gks1188
38. Childs-Disney JL, Hoskins J, Rzuczek SG, Thornton CA, Disney MD (2012) Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. Acs Chem Biol 7:856–862. doi:10.1021/cb200408a
39. Childs-Disney JL, Yildirim I, Park H, Lohman JR, Guan L, Tran T, Sarkar P, Schatz GC, Disney MD (2014) Structure of the myotonic dystrophy type 2 RNA and designed small molecules that reduce toxicity. Acs Chem Biol 9:538–550. doi:10.1021/cb4007387
40. Chung DW, Rudnicki DD, Yu L, Margolis RL (2011) A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. Hum Mol Genet 20:3467–3477. doi:10.1093/hmg/ddr263
41. Clark RM, Dalgliesh GL, Endres D, Gomez M, Taylor J, Bidichandani SI (2004) Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. Genomics 83:373–383. doi:10.1016/j.ygeno.2003.09.001
42. Cleary JD, Ranum LP (2014) Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. Curr Opin Genet Dev 26C:6–15. doi:10.1016/j.gde.2014.03.002
43. Cohen-Hadad Y, Altarescu G, Eldar-Geva T, Levi-Lahad E, Zhang M, Rogaeva E, Gotkine M, Bartok O, Ashwal-Fluss R, Kadener S et al (2016) Marked Differences in C9orf72 Methylation Status and Isoform Expression between C9/ALS Human Embryonic and Induced Pluripotent Stem Cells. Stem Cell Rep 7:927–940. doi:10.1016/j.stemcr.2016.09.011
44. Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, Disney MD, Jaffrey SR (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science 343:1002–1005. doi:10.1126/science.1245831
45. Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, Manley JL (2016) The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. eLife 5. doi:10.7554/eLife.17820
46. Cooper-Knock J, Higginbottom A, Stopford MJ, Highley JR, Ince PG, Wharton SB, Pickering-Brown S, Kirby J, Hautbergue GM, Shaw PJ (2015) Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. Acta Neuropathol 130:63–75. doi:10.1007/s00401-015-1429-9
47. Cooper-Knock J, Walsh MJ, Higginbottom A, Robin Highley J, Dickman MJ, Edbauer D, Ince PG, Wharton SB, Wilson SA, Kirby J et al (2014) Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain 137:2040-2051. doi:10.1093/brain/awu120
48. Cortes CJ, La Spada AR (2015) Autophagy in polyglutamine disease: Imposing order on disorder or contributing to the chaos? Mol Cell Neurosci 66:53–61. doi:10.1016/j.mcn.2015.03.010
49. Coyle JH, Bor YC, Rekosh D, Hammarskjold ML (2011) The Tpr protein regulates export of mRNAs with retained introns that traffic through the Nxf1 pathway. RNA 17:1344–1356. doi:10.1261/rna.2616111
50. Crickard JB, Fu J, Reese JC (2016) Biochemical Analysis of Yeast Suppressor of Ty 4/5 (Spt4/5) Reveals the Importance of Nucleic Acid Interactions in the Prevention of RNA Polymerase II Arrest. J Biol Chem 291:9853–9870. doi:10.1074/jbc.M116.716001
51. Crouau-Roy B, Clisson I (2000) Evolution of an Alu DNA element of type Sx in the lineage of primates and the origin of an associated tetranucleotide microsatellite. Genome 43:642–648
52. Custodio N, Carmo-Fonseca M, Geraghty F, Pereira HS, Grosveld F, Antoniou M (1999) Inefficient processing impairs release of RNA from the site of transcription. EMBO J 18:2855–2866. doi:10.1093/emboj/18.10.2855
53. Custodio N, Carvalho C, Condado I, Antoniou M, Blencowe BJ, Carmo-Fonseca M (2004) In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. RNA 10:622–633
54. Dai WJ, Zhu LY, Yan ZY, Xu Y, Wang QL, Lu XJ (2016) CRISPR-Cas9 for in vivo Gene Therapy: Promise and Hurdles. Mol Ther Nucleic Acids 5:e349. doi:10.1038/mtna.2016.58
55. Darlington RW, Moss LH 3rd (1968) Herpesvirus envelopment. J Virol 2:48–55
56. Daugaard I, Hansen TB (2017) Biogenesis and Function of Ago-Associated RNAs. Trends Genet. doi:10.1016/j.tig.2017.01.003
57. David G, Abbas N, Stevanin G, Durr A, Yvert G, Cancel G, Weber C, Imbert G, Saudou F, Antoniou Eet al (1997) Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. Nat Genet 17:65–70. doi:10.1038/ng0997-65
58. de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell 12:525–532
59. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J et al (2011) Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron 72:245–256. doi:10.1016/j.neuron.2011.09.011
60. Delaleau M, Borden KL (2015) Multiple Export Mechanisms for mRNAs. Cells 4:452–473. doi:10.3390/cells4030452
61. Diegoli TM (2015) Forensic typing of short tandem repeat markers on the X and Y chromosomes. Forensic Sci Int Genet 18:140–151. doi:10.1016/j.fsigen.2015.03.013
62. Dodd DW, Tomchick DR, Corey DR, Gagnon KT (2016) Pathogenic C9ORF72 Antisense Repeat RNA Forms a Double Helix with Tandem C:C Mismatches. Biochemistry 55:1283–1286. doi:10.1021/acs.biochem.6b00136
63. Doma MK, Parker R (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature 440:561–564. doi:10.1038/nature04530
64. Donnelly CJ, Zhang PW, Pham JT, Haeusler AR, Mistry NA, Vidensky S, Daley EL, Poth EM, Hoover B, Fines DM et al (2013) RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron 80:415–428. doi:10.1016/j.neuron.2013.10.015
65. Dumache R, Ciocan V, Muresan C, Enache A (2016) Molecular DNA Analysis in Forensic Identification. Clin Lab 62:245–248
66. Eberle AB, Hessle V, Helbig R, Dantoft W, Gimber N, Visa N (2010) Splice-site mutations cause Rrp6-mediated nuclear retention of the unspliced RNAs and transcriptional down-regulation of the splicing-defective genes. PLoS One 5:e11540. doi:10.1371/journal.pone.0011540
67. Eberle AB, Visa N (2014) Quality control of mRNP biogenesis: networking at the transcription site. Semin Cell Dev Biol 32:37–46. doi:10.1016/j.semcdb.2014.03.033
68. Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. Nat Genet 24:400–402. doi:10.1038/74249
69. Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445. doi:10.1038/nrg1348

70. Evans-Galea MV, Hannan AJ, Carrodus N, Delatycki MB, Saffery R (2013) Epigenetic modifications in trinucleotide repeat diseases. Trends Mol Med 19:655–663. doi:10.1016/j.molmed.2013.07.007
71. Fellmann C, Gowen BG, Lin PC, Doudna JA, Corn JE (2017) Cornerstones of CRISPR-Cas in drug discovery and therapy. Nat Rev Drug Discov 16:89–100. doi:10.1038/nrd.2016.238
72. Feng Y, Zhang F, Lokey LK, Chastain JL, Lakkis L, Eberhart D, Warren ST (1995) Translational suppression by trinucleotide repeat expansion at FMR1. Science 268:731–734
73. Finkel RS (2010) Read-through strategies for suppression of nonsense mutations in Duchenne/ Becker muscular dystrophy: aminoglycosides and ataluren (PTC124). J Child Neurol 25:1158–1164. doi:10.1177/0883073810371129
74. Floer M, Blobel G (1999) Putative reaction intermediates in Crm1-mediated nuclear protein export. J Biol Chem 274:16279–16286
75. Foiry L, Dong L, Savouret C, Hubert L, te Riele H, Junien C, Gourdon G (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. Hum Genet 119:520–526. doi:10.1007/s00439-006-0164-7
76. Follonier C, Oehler J, Herrador R, Lopes M (2013) Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions. Nat Struct Mol Biol 20:486–494. doi:10.1038/nsmb.2520
77. Fornerod M, Ohno M, Yoshida M, Mattaj IW (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. Cell 90:1051–1060
78. Fradkin LG, Budnik V (2016) This bud's for you: mechanisms of cellular nucleocytoplasmic trafficking via nuclear envelope budding. Curr Opin Cell Biol 41:125–131. doi:10.1016/j.ceb.2016.05.001
79. Fratta P, Mizielinska S, Nicoll AJ, Zloh M, Fisher EMC, Parkinson G, Isaacs AM (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci Rep-Uk 2:1016. doi:10.1038/Srep01016
80. Freibaum BD, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee KH, Badders N, Valentine M, Miller BL, Wong PC et al (2015) GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. Nature 525:129–133. doi:10.1038/nature14974
81. Freudenreich CH, Kantrow SM, Zakian VA (1998) Expansion and length-dependent fragility of CTG repeats in yeast. Science 279:853–856
82. Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science 295:2258–2261. doi:10.1126/science.1067338
83. Froehlich AC, Liu Y, Loros JJ, Dunlap JC (2002) White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. Science 297:815–819. doi:10.1126/science.1073681
84. Gadgil R, Barthelemy J, Lewis T, Leffak M (2017) Replication stalling and DNA microsatellite instability. Biophys Chem 225:38–48. doi:10.1016/j.bpc.2016.11.007
85. Gagnon KT, Corey DR (2015) Stepping toward therapeutic CRISPR. Proc Natl Acad Sci U S A 112:15536–15537. doi:10.1073/pnas.1521670112
86. Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X (2011) R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev 25:2041–2056. doi:10.1101/gad.17010011
87. Garavis M, Gonzalez C, Villasante A (2013) On the origin of the eukaryotic chromosome: the role of noncanonical DNA structures in telomere evolution. Genome Biol Evol 5:1142–1150. doi:10.1093/gbe/evt079
88. Garcia SM, Tabach Y, Lourenco GF, Armakola M, Ruvkun G (2014) Identification of genes in toxicity pathways of trinucleotide-repeat RNA in C. elegans. Nat Struct Mol Biol 21:712–720. doi:10.1038/nsmb.2858
89. Gellert M, Lipsett MN, Davies DR (1962) Helix formation by guanylic acid. Proc Natl Acad Sci U S A 48:2013–2018
90. Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE, Caulfield T, Daughrity L, Dunmore JH, Castanedes-Casey M, Chew J et al (2013) Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol 126:829–844. doi:10.1007/s00401-013-1192-8
91. Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell 45:814–825. doi:10.1016/j.molcel.2012.01.017
92. Goers ES, Purcell J, Voelker RB, Gates DP, Berglund JA (2010) MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. Nucleic Acids Res 38:2467–2484. doi:10.1093/nar/gkp1209
93. Gordenin DA, Kunkel TA, Resnick MA (1997) Repeat expansion–all in a flap? Nat Genet 16:116–118. doi:10.1038/ng0697-116
94. Grabczyk E, Usdin K (2000) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic Acids Res 28:2815–2822
95. Grandi FC, An W (2013) Non-LTR retrotransposons and microsatellites: Partners in genomic variation. Mobile Genet Elem 3:e25674. doi:10.4161/mge.25674
96. Green KM, Linsalata AE, Todd PK (2016) RAN translation-What makes it run? Brain Res. doi:10.1016/j.brainres.2016.04.003
97. Greene E, Mahishi L, Entezam A, Kumari D, Usdin K (2007) Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. Nucleic Acids Res 35:3383–3390. doi:10.1093/nar/gkm271
98. Griesche N, Schilling J, Weber S, Rohm M, Pesch V, Matthes F, Auburger G, Krauss S (2016) Regulation of mRNA Translation by MID1: A Common Mechanism of Expanded CAG Repeat RNAs. Front Cell Neurosci 10:226. doi:10.3389/fncel.2016.00226
99. Groh M, Lufino MM, Wade-Martins R, Gromak N (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet 10:e1004318. doi:10.1371/journal.pgen.1004318
100. Groh M, Silva LM, Gromak N (2014) Mechanisms of transcriptional dysregulation in repeat expansion disorders. Biochem Soc Trans 42:1123–1128. doi:10.1042/BST20140049
101. Gruter P, Tabernero C, von Kobbe C, Schmitt C, Saavedra C, Bachi A, Wilm M, Felber BK, Izaurralde E (1998) TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. Mol Cell 1:649–659
102. Guan L, Disney MD (2012) Recent advances in developing small molecules targeting RNA. Acs Chem Biol 7:73–86. doi:10.1021/cb200447r
103. Gudanis D, Popenda L, Szpotkowski K, Kierzek R, Gdaniec Z (2016) Structural characterization of a dimer of RNA duplexes composed of 8-bromoguanosine modified CGG trinucleotide repeats: a novel architecture of RNA quadruplexes. Nucleic Acids Res 44:2409–2416. doi:10.1093/nar/gkv1534
104. Gudde AE, Gonzalez-Barriga A, van den Broek WJ, Wieringa B, Wansink DG (2016) A low absolute number of expanded transcripts is involved in myotonic dystrophy type 1 manifestation in muscle. Hum Mol Genet 25:1648–1662. doi:10.1093/hmg/ddw042
105. Gudde AE, van Heeringen SJ, de Oude AI, van Kessel ID, Estabrook J, Wang ET, Wieringa B, Wansink DG (2017) Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)n repeat. RNA Biol 0. doi:10.1080/15476286.2017.1279787
106. Guo J, Gu L, Leffak M, Li GM (2016) MutSbeta promotes trinucleotide repeat expansion by recruiting DNA polymerase beta to nascent (CAG)n or (CTG)n hairpins for error-prone DNA synthesis. Cell Res 26:775–786. doi:10.1038/cr.2016.66
107. Guo JU, Bartel DP (2016) RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science 353. doi:10.1126/science.aaf5371
108. Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ, Troncoso JC, Pandey A, Sattler Ret al (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature 507:195–200. doi:10.1038/nature13124
109. Hamperl S, Cimprich KA (2016) Conflict Resolution in the Genome: How Transcription and Replication Make It Work. Cell 167:1455–1467. doi:10.1016/j.cell.2016.09.053
110. Hautbergue GM, Castelli LM, Ferraiuolo L, Sanchez-Martinez A, Cooper-Knock J, Higginbottom A, Lin YH, Bauer CS, Dodd JE, Myszczynska MA et al (2017) SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. Nat Commun 8:16063. doi:10.1038/ncomms16063
111. Hautbergue GM, Hung ML, Golovanov AP, Lian LY, Wilson SA (2008) Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. Proc Natl Acad Sci U S A 105:5154–5159. doi:10.1073/pnas.0709167105
112. He F, Todd PK (2011) Epigenetics in nucleotide repeat expansion disorders. Semin Neurol 31:470–483. doi:10.1055/s-0031-1299786
113. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW (2008) The antisense transcriptomes of human cells. Science 322:1855–1857. doi:10.1126/science.1163853
114. Heitz D, Rousseau F, Devys D, Saccone S, Abderrahim H, Le Paslier D, Cohen D, Vincent A, Toniolo D, Della Valle G et al (1991) Isolation of

sequences that span the fragile X and identification of a fragile X-related CpG island. Science 251:1236–1239
115. Hesselberth JR (2013) Lives that introns lead after splicing. Wiley Interdiscip Rev RNA 4:677–691. doi:10.1002/wrna.1187
116. Hinnebusch AG (2014) The scanning mechanism of eukaryotic translation initiation. Annu Rev Biochem 83:779–812. doi:10.1146/annurev-biochem-060713-035802
117. Hinnebusch AG, Ivanov IP, Sonenberg N (2016) Translational control by 5'-untranslated regions of eukaryotic mRNAs. Science 352:1413–1416. doi:10.1126/science.aad9868
118. Hino S, Kondo S, Sekiya H, Saito A, Kanemoto S, Murakami T, Chihara K, Aoki Y, Nakamori M, Takahashi MP et al (2007) Molecular mechanisms responsible for aberrant splicing of SERCA1 in myotonic dystrophy type 1. Hum Mol Genet 16:2834–2843 doi:10.1093/hmg/ddm239
119. Hirtreiter A, Damsma GE, Cheung AC, Klose D, Grohmann D, Vojnic E, Martin AC, Cramer P, Werner F (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. Nucleic Acids Res 38:4040–4051. doi:10.1093/nar/gkq135
120. Holmes SE, O'Hearn E, Rosenblatt A, Callahan C, Hwang HS, Ingersoll-Ashworth RG, Fleisher A, Stevanin G, Brice A, Potter NT et al (2001) A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. Nat Genet 29:377–378. doi:10.1038/ng760
121. Holmes SE, O'Hearn EE, McInnis MG, Gorelick-Feldman DA, Kleiderlein JJ, Callahan C, Kwak NG, Ingersoll-Ashworth RG, Sherr M, Sumner AJ et al (1999) Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. Nat Genet 23:391–392. doi:10.1038/70493
122. Houseley J, Tollervey D (2011) Repeat expansion in the budding yeast ribosomal DNA can occur independently of the canonical homologous recombination machinery. Nucleic Acids Res 39:8778–8791. doi:10.1093/nar/gkr589
123. Huertas P, Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12:711–721
124. Hung ML, Hautbergue GM, Snijders AP, Dickman MJ, Wilson SA (2010) Arginine methylation of REF/ALY promotes efficient handover of mRNA to TAP/NXF1. Nucleic Acids Res 38:3351–3361. doi:10.1093/nar/gkq033
125. Iglesias N, Stutz F (2008) Regulation of mRNP dynamics along the export pathway. FEBS letters 582:1987–1996. doi:10.1016/j.febslet.2008.03.038
126. Ikeda Y, Daughters RS, Ranum LP (2008) Bidirectional expression of the SCA8 expansion mutation: one mutation, two genes. Cerebellum 7:150–158. doi:10.1007/s12311-008-0010-7
127. Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier JM, Weber C, Mandel JL, Cancel G, Abbas N et al (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nat Genet 14:285–291. doi:10.1038/ng1196-285
128. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223. doi:10.1126/science.1168978
129. Ishiguro T, Sato N, Ueyama M, Fujikake N, Sellier C, Kanegami A, Tokuda E, Zamiri B, Gall-Duncan T, Mirceta M et al (2017) Regulatory Role of RNA Chaperone TDP-43 for RNA Misfolding and Repeat-Associated Translation in SCA31. Neuron. doi:10.1016/j.neuron.2017.02.046
130. Iyer RR, Pluciennik A, Napierala M, Wells RD (2015) DNA Triplet Repeat Expansion and Mismatch Repair. Annu Rev Biochem. doi:10.1146/annurev-biochem-060614-034010
131. Izumi H, McCloskey A, Shinmyozu K, Ohno M (2014) p54nrb/NonO and PSF promote U snRNA nuclear export by accelerating its export complex assembly. Nucleic Acids Res 42:3998–4007. doi:10.1093/nar/gkt1365
132. Jain A, Vale RD (2017) RNA phase transitions in repeat expansion disorders. Nature 546:243–247. doi:10.1038/nature22386
133. Jansen A, Gemayel R, Verstrepen KJ (2012) Unstable microsatellite repeats facilitate rapid evolution of coding and regulatory sequences. Genome Dyn 7:108–125. doi:10.1159/000337121
134. Jiang J, Zhu Q, Gendron TF, Saberi S, McAlonis-Downes M, Seelman A, Stauffer JE, Jafar-Nejad P, Drenner K, Schulte D et al (2016) Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. Neuron 90:535–550. doi:10.1016/j.neuron.2016.04.006
135. Jokhi V, Ashley J, Nunnari J, Noma A, Ito N, Wakabayashi-Ito N, Moore MJ, Budnik V (2013) Torsin mediates primary envelopment of large ribonucleoprotein granules at the nuclear envelope. Cell Rep 3:988–995. doi:10.1016/j.celrep.2013.03.015
136. Jovicic A, Mertens J, Boeynaems S, Bogaert E, Chai N, Yamada SB, Paul JW, 3rd, Sun S, Herdy JR, Bieri Get al (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. Nat Neurosci 18: 1226-1229 Doi doi:10.1038/nn.4085
137. Kang S, Jaworski A, Ohshima K, Wells RD (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. Nat Genet 10:213–218. doi:10.1038/ng0695-213
138. Katahira J (2012) mRNA export and the TREX complex. Biochimica Et Biophysica Acta 1819:507–513. doi:10.1016/j.bbagrm.2011.12.001
139. Katahira J (2015) Nuclear export of messenger RNA. Genes 6:163–184. doi:10.3390/genes6020163
140. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I et al (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1 Nat Genet 8:221–228. doi:10.1038/ng1194-221
141. Kearse MG, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC, Todd PK (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. Mol Cell 62:314–322. doi:10.1016/j.molcel.2016.02.034
142. Kearse MG, Todd PK (2014) Repeat-associated non-AUG translation and its impact in neurodegenerative disease. Neurotherapeutics 11:721–731. doi:10.1007/s13311-014-0292-z
143. Kenneson A, Zhang F, Hagedorn CH, Warren ST (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. Hum Mol Genet 10:1449–1454
144. Kerrest A, Anand RP, Sundararajan R, Bermejo R, Liberi G, Dujon B, Freudenreich CH, Richard GF (2009) SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. Nat Struct Mol Biol 16:159–167. doi:10.1038/nsmb.1544
145. Khorkova O, Myers AJ, Hsiao J, Wahlestedt C (2014) Natural antisense transcripts. Hum Mol Genet 23:R54–R63. doi:10.1093/hmg/ddu207
146. Kilchert C, Wittmann S, Vasiljeva L (2016) The regulation and functions of the nuclear RNA exosome complex. Nat Rev Mol Cell Biol 17:227–239. doi:10.1038/nrm.2015.15
147. Kiliszek A, Rypniewski W (2014) Structural studies of CNG repeats. Nucleic Acids Res 42:8189–8199
148. Kim JC, Harris ST, Dinter T, Shah KA, Mirkin SM (2017) The role of break-induced replication in large-scale expansions of (CAG)n/(CTG)n repeats. Nat Struct Mol Biol 24:55–60. doi:10.1038/nsmb.3334
149. Kim JC, Mirkin SM (2013) The balancing act of DNA repeat expansions. Curr Opin Genet Dev 23:280–288. doi:10.1016/j.gde.2013.04.009
150. Kino Y, Washizu C, Kurosawa M, Oma Y, Hattori N, Ishiura S, Nukina N (2015) Nuclear localization of MBNL1: splicing-mediated autoregulation and repression of repeat-derived aberrant proteins. Hum Mol Genet 24:740–756. doi:10.1093/hmg/ddu492
151. Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. Proc Natl Acad Sci U S A 108:546–550. doi:10.1073/pnas.1013828108
152. Knight SJ, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembrey ME et al (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell 74:127–134
153. Kobayashi H, Abe K, Matsuura T, Ikeda Y, Hitomi T, Akechi Y, Habu T, Liu W, Okuda H, Koizumi A (2011) Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am J Hum Genet 89:121–130. doi:10.1016/j.ajhg.2011.05.015
154. Kochetov AV, Palyanov A, Titov II, Grigorovich D, Sarai A, Kolchanov NA (2007) AUG_hairpin: prediction of a downstream secondary structure influencing the recognition of a translation start site. BMC Bioinformatics 8:318. doi:10.1186/1471-2105-8-318
155. Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). Nat Genet 6:9–13. doi:10.1038/ng0194-9
156. Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LP (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet 21:379–384. doi:10.1038/7710

157. Kota KP, Wagner SR, Huerta E, Underwood JM, Nickerson JA (2008) Binding of ATP to UAP56 is necessary for mRNA export. J Cell Sci 121:1526–1537. doi:10.1242/jcs.021055
158. Kozak M (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Natl Acad Sci U S A 87:8301–8305
159. Kramer NJ, Carlomagno Y, Zhang YJ, Almeida S, Cook CN, Gendron TF, Prudencio M, Van Blitterswijk M, Belzil V, Couthouis J et al (2016) Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts. Science 353:708–712. doi:10.1126/science.aaf7791
160. Krauss S, Griesche N, Jastrzebska E, Chen C, Rutschow D, Achmuller C, Dorn S, Boesch SM, Lalowski M, Wanker E et al (2013) Translation of HTT mRNA with expanded CAG repeats is regulated by the MID1-PP2A protein complex. Nat Commun 4:1511. doi:10.1038/ncomms2514
161. Kremer EJ, Yu S, Pritchard M, Nagaraja R, Heitz D, Lynch M, Baker E, Hyland VJ, Little RD, Wada Met al (1991) Isolation of a human DNA sequence which spans the fragile X. Am J Hum Genet 49:656–661
162. Kumari D, Biacsi RE, Usdin K (2011) Repeat expansion affects both transcription initiation and elongation in friedreich ataxia cells. J Biol Chem 286:4209–4215. doi:10.1074/jbc.M110.194035
163. Kurosaki T, Ueda S, Ishida T, Abe K, Ohno K, Matsuura T (2012) The unstable CCTG repeat responsible for myotonic dystrophy type 2 originates from an AluSx element insertion into an early primate genome. PLoS One 7:e38379. doi:10.1371/journal.pone.0038379
164. Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods 13:841–844. doi:10.1038/nmeth.3965
165. Kwon I, Xiang S, Kato M, Wu L, Theodoropoulous P, Wang T, Kim J, Yun J, Xie Y, McKnight SL (2014) Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science 345:1139–1145
166. La Spada AR, Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat Rev Genet 11:247–258. doi:10.1038/nrg2748
167. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77–79. doi:10.1038/352077a0
168. Labbadia J, Morimoto RI (2013) Huntington's disease: underlying molecular mechanisms and emerging concepts. Trends Biochem Sci 38:378–385. doi:10.1016/j.tibs.2013.05.003
169. Lagier-Tourenne C, Baughn M, Rigo F, Sun S, Liu P, Li HR, Jiang J, Watt AT, Chun S, Katz M et al (2013) Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc Natl Acad Sci U S A 110:E4530–E4539. doi:10.1073/pnas.1318835110
170. Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, Malafosse A, Antonarakis SE (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature 386:847–851. doi:10.1038/386847a0
171. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. doi:10.1038/35057062
172. Le Hir H, Izaurralde E, Maquat LE, Moore MJ (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J 19:6860–6869. doi:10.1093/emboj/19.24.6860
173. Lee JE, Cooper TA (2009) Pathogenic mechanisms of myotonic dystrophy. Biochem Soc Trans 37:1281–1286. doi:10.1042/BST0371281
174. Lee KH, Zhang P, Kim HJ, Mitrea DM, Sarkar M, Freibaum BD, Cika J, Coughlin M, Messing J, Molliex A et al (2016) C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell 167(774-788):e717. doi:10.1016/j.cell.2016.10.002
175. Lee YB, Chen HJ, Peres JN, Gomez-Deza J, Attig J, Stalekar M, Troakes C, Nishimura AL, Scotter EL, Vance C et al (2013) Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic. Cell Rep 5:1178–1186. doi:10.1016/j.celrep.2013.10.049
176. Leffak M (2017) Break-induced replication links microsatellite expansion to complex genome rearrangements. Bioessays 39. doi:10.1002/bies.201700025
177. Li L, Matsui M, Corey DR (2016) Activating frataxin expression by repeat-targeted nucleic acids. Nat Commun 7:10606. doi:10.1038/ncomms10606
178. Li Y, Bor YC, Misawa Y, Xue Y, Rekosh D, Hammarskjold ML (2006) An intron with a constitutive transport element is retained in a Tap messenger RNA. Nature 443:234–237. doi:10.1038/nature05107
179. Li YY, Abu-Ghazalah R, Zamiri B, Macgregor RB Jr (2016) Concentration-dependent conformational changes in GQ-forming ODNs. Biophys Chem 211:70–75. doi:10.1016/j.bpc.2016.02.002
180. Lin Y, Hubert L Jr, Wilson JH (2009) Transcription destabilizes triplet repeats. Mol Carcinog 48:350–361. doi:10.1002/mc.20488
181. Lin Y, Leng M, Wan M, Wilson JH (2010) Convergent transcription through a long CAG tract destabilizes repeats and induces apoptosis. Mol Cell Biol 30:4435–4451. doi:10.1128/MCB.00332-10
182. Lin Y, Mori E, Kato M, Xiang S, Wu L, Kwon I, McKnight SL (2016) Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. Cell 167:789–802 e712. doi:10.1016/j.cell.2016.10.003
183. Lin Y, Wilson JH (2012) Nucleotide excision repair, mismatch repair, and R-loops modulate convergent transcription-induced cell death and repeat instability. PLoS One 7:e46807. doi:10.1371/journal.pone.0046807
184. Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science 293:864–867. doi:10.1126/science.1062125
185. Liu B, Steitz TA (2017) Structural insights into NusG regulating transcription elongation. Nucleic Acids Res 45:968–974. doi:10.1093/nar/gkw1159
186. Liu CR, Chang CR, Chern Y, Wang TH, Hsieh WC, Shen WC, Chang CY, Chu IC, Deng N, Cohen SN et al (2012) Spt4 is selectively required for transcription of extended trinucleotide repeats. Cell 148:690–701. doi:10.1016/j.cell.2011.12.032
187. Liu G, Chen X, Bissler JJ, Sinden RR, Leffak M (2010) Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells. Nat Chem Biol 6:652–659. doi:10.1038/nchembio.416
188. Liu J, Hu J, Ludlow AT, Pham JT, Shay JW, Rothstein JD, Corey DR (2017) c9orf72 Disease-Related Foci Are Each Composed of One Mutant Expanded Repeat RNA. Cell Chem Biol 24:141–148. doi:10.1016/j.chembiol.2016.12.018
189. Llorente B, Smith CE, Symington LS (2008) Break-induced replication: what is it and what is it for? Cell Cycle 7:859–864. doi:10.4161/cc.7.7.5613
190. Lokanga RA, Senejani AG, Sweasy JB, Usdin K (2015) Heterozygosity for a hypomorphic Polbeta mutation reduces the expansion frequency in a mouse model of the Fragile X-related disorders. PLoS Genet 11:e1005181. doi:10.1371/journal.pgen.1005181
191. Lujan SA, Clausen AR, Clark AB, MacAlpine HK, MacAlpine DM, Malc EP, Mieczkowski PA, Burkholder AB, Fargo DC, Gordenin DA et al (2014) Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. Genome Res 24:1751–1764. doi:10.1101/gr.178335.114
192. MacDonald ME, Barnes G, Srinidhi J, Duyao MP, Ambrose CM, Myers RH, Gray J, Conneally PM, Young A, Penney J et al (1993) Gametic but not somatic instability of CAG repeat length in Huntington's disease. J Med Genet 30:982–986
193. Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K et al (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science 255:1253–1255
194. Malgowska M, Gudanis D, Kierzek R, Wyszko E, Gabelica V, Gdaniec Z (2014) Distinctive structural motifs of RNA G-quadruplexes composed of AGG, CGG and UGG trinucleotide repeats. Nucleic Acids Res 42:10196–10207. doi:10.1093/nar/gku710
195. Malkova A, Haber JE (2012) Mutations arising during repair of chromosome breaks. Annu Rev Genet 46:455–473. doi:10.1146/annurev-genet-110711-155547
196. Mankodi A, Logigian E, Callahan L, McClain C, White R, Henderson D, Krym M, Thornton CA (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. Science 289:1769–1773
197. Mankodi A, Urbinati CR, Yuan QP, Moxley RT, Sansone V, Krym M, Henderson D, Schalling M, Swanson MS, Thornton CA (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. Hum Mol Genet 10:2165–2170
198. Mao YS, Zhang B, Spector DL (2011) Biogenesis and function of nuclear bodies. Trends Genet 27:295–306. doi:10.1016/j.tig.2011.05.006
199. Maquat LE, Li X (2001) Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. RNA 7:445–456
200. Maric M, Shao J, Ryan RJ, Wong CS, Gonzalez-Alegre P, Roller RJ (2011) A functional role for TorsinA in herpes simplex virus 1 nuclear egress. J Virol 85:9667–9679. doi:10.1128/JVI.05314-11

201. Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J 30:1302–1310. doi:10.1038/emboj.2011.64
202. Mason SW, Greenblatt J (1991) Assembly of transcription elongation complexes containing the N protein of phage lambda and the Escherichia coli elongation factors NusA, NusB, NusG, and S10. Genes Dev 5:1504–1512
203. Massenet S, Bertrand E, Verheggen C (2016) Assembly and trafficking of box C/D and H/ACA snoRNPs. RNA Biol 1–13. doi:10.1080/15476286.2016.1243646
204. Masuda S, Das R, Cheng H, Hurt E, Dorman N, Reed R (2005) Recruitment of the human TREX complex to mRNA during splicing. Genes Dev 19:1512–1517. doi:10.1101/gad.1302205
205. Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, Khajavi M, McCall AE, Davis CF, Zu L et al (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet 26:191–194. doi:10.1038/79911
206. McMurray CT (2010) Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet 11:786–799. doi:10.1038/nrg2828
207. Meng F, Na I, Kurgan L, Uversky VN (2015) Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments. Int J Mol Sci 17. doi:10.3390/ijms17010024
208. Meservy JL, Sargent RG, Iyer RR, Chan F, McKenzie GJ, Wells RD, Wilson JH (2003) Long CTG tracts from the myotonic dystrophy gene induce deletions and rearrangements during recombination at the APRT locus in CHO cells. Mol Cell Biol 23:3152–3162
209. Metsu S, Rainger JK, Debacker K, Bernhard B, Rooms L, Grafodatskaya D, Weksberg R, Fombonne E, Taylor MS, Scherer SW et al (2014) A CGG-repeat expansion mutation in ZNF713 causes FRA7A: association with autistic spectrum disorder in two families. Hum Mutat 35:1295–1300. doi:10.1002/humu.22683
210. Metsu S, Rooms L, Rainger J, Taylor MS, Bengani H, Wilson DI, Chilamakuri CS, Morrison H, Vandeweyer G, Reyniers E et al (2014) FRA2A is a CGG repeat expansion associated with silencing of AFF3. PLoS Genet 10:e1004242. doi:10.1371/journal.pgen.1004242
211. Michael TP, Park S, Kim TS, Booth J, Byer A, Sun Q, Chory J, Lee K (2007) Simple sequence repeats provide a substrate for phenotypic variation in the Neurospora crassa circadian clock. PLoS One 2:e795. doi:10.1371/journal.pone.0000795
212. Mirkin SM (2007) Expandable DNA repeats and human disease. Nature 447:932–940. doi:10.1038/nature05977
213. Mizielinska S, Gronke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, Moens T, Norona FE, Woollacott IO, Pietrzyk J et al (2014) C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science 345:1192–1194. doi:10.1126/science.1256800
214. Mohan A, Goodwin M, Swanson MS (2014) RNA-protein interactions in unstable microsatellite diseases. Brain Res 1584:3–14. doi:10.1016/j.brainres.2014.03.039
215. Monteys AM, Ebanks SA, Keiser MS, Davidson BL (2017) CRISPR/Cas9 Editing of the Mutant Huntingtin Allele In Vitro and In Vivo. Mol Ther 25:12–23. doi:10.1016/j.ymthe.2016.11.010
216. Mori K, Arzberger T, Grasser FA, Gijselinck I, May S, Rentzsch K, Weng SM, Schludi MH, van der Zee J, Cruts M et al (2013) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. Acta Neuropathol 126:881–893. doi:10.1007/s00401-013-1189-3
217. Mori K, Lammich S, Mackenzie IR, Forne I, Zilow S, Kretzschmar H, Edbauer D, Janssens J, Kleinberger G, Cruts M et al (2013) hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol 125:413–423. doi:10.1007/s00401-013-1088-7
218. Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Van Broeckhoven C et al (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science 339:1335–1338. doi:10.1126/science.1232927
219. Moussavi-Baygi R, Mofrad MR (2016) Rapid Brownian Motion Primes Ultrafast Reconstruction of Intrinsically Disordered Phe-Gly Repeats Inside the Nuclear Pore Complex. Sci Rep 6:29991. doi:10.1038/srep29991
220. Muhlemann O, Mock-Casagrande CS, Wang J, Li S, Custodio N, Carmo-Fonseca M, Wilkinson MF, Moore MJ (2001) Precursor RNAs harboring nonsense codons accumulate near the site of transcription. Mol Cell 8:33–43
221. Muranyi W, Haas J, Wagner M, Krohne G, Koszinowski UH (2002) Cytomegalovirus recruitment of cellular kinases to dissolve the nuclear lamina. Science 297:854–857. doi:10.1126/science.1071506
222. Myers SJ, Huang Y, Genetta T, Dingledine R (2004) Inhibition of glutamate receptor 2 translation by a polymorphic repeat sequence in the 5'-untranslated leaders. J Neurosci 24:3489–3499. doi:10.1523/JNEUROSCI.4127-03.2004
223. Nakamori M, Pearson CE, Thornton CA (2011) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. Hum Mol Genet 20:580–588. doi:10.1093/hmg/ddq501
224. Nakamura K, Jeong SY, Uchihara T, Anno M, Nagashima K, Nagashima T, Ikeda S, Tsuji S, Kanazawa I (2001) SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. Hum Mol Genet 10:1441–1448
225. Neil AJ, Kim JC, Mirkin SM (2017) Precarious maintenance of simple DNA repeats in eukaryotes. BioEssays. doi:10.1002/bies.201700077
226. Nguyen L, Luu LM, Peng S, Serrano JF, Chan HY, Zimmerman SC (2015) Rationally Designed Small Molecules That Target Both the DNA and RNA Causing Myotonic Dystrophy Type 1. J Am Chem Soc 137:14180–14189. doi:10.1021/jacs.5b09266
227. Niblock M, Smith BN, Lee YB, Sardone V, Topp S, Troakes C, Al-Sarraj S, Leblond CS, Dion PA, Rouleau GA et al (2016) Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. Acta neuropathol Commun 4:18. doi:10.1186/s40478-016-0289-4
228. Nunes VS, Moretti NS (2017) Nuclear subcompartments: an overview. Cell Biol Int 41:2–7. doi:10.1002/cbin.10703
229. O'Geen H, Yu AS, Segal DJ (2015) How specific is CRISPR/Cas9 really? Curr Opin Chem Biol 29:72–78. doi:10.1016/j.cbpa.2015.10.001
230. Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boue J, Bertheas MF, Mandel JL (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252:1097–1102
231. Ohshima K, Montermini L, Wells RD, Pandolfo M (1998) Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J Biol Chem 273:14588–14595
232. Orr HT, Chung MY, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LP, Zoghbi HY (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet 4:221–226. doi:10.1038/ng0793-221
233. Ozdilek BA, Thompson VF, Ahmed NS, White CI, Batey RT, Schwartz JC (2017) Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. Nucleic Acids Res. doi:10.1093/nar/gkx460
234. Park H, Gonzalez AL, Yildirim I, Tran T, Lohman JR, Fang P, Guo M, Disney MD (2015) Crystallographic and Computational Analyses of AUUCU Repeating RNA That Causes Spinocerebellar Ataxia Type 10 (SCA10). Biochemistry 54:3851–3859. doi:10.1021/acs.biochem.5b00551
235. Paul S, Dansithong W, Jog SP, Holt I, Mittal S, Brook JD, Morris GE, Comai L, Reddy S (2011) Expanded CUG repeats Dysregulate RNA splicing by altering the stoichiometry of the muscleblind 1 complex. J Biol Chem 286:38427–38438. doi:10.1074/jbc.M111.255224
236. Payseur BA, Jing P, Haasl RJ (2011) A genomic portrait of human microsatellite variation. Mol Biol Evol 28:303–312. doi:10.1093/molbev/msq198
237. Pearson CE, Sinden RR (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. Biochemistry 35:5041–5053. doi:10.1021/bi9601013
238. Peng R, Lin G, Li J (2015) Potential Pitfalls of CRISPR/Cas9-mediated Genome Editing. The FEBS J. doi:10.1111/febs.13586
239. Pfister EL, Kennington L, Straubhaar J, Wagh S, Liu W, DiFiglia M, Landwehrmeyer B, Vonsattel JP, Zamore PD, Aronin N (2009) Five siRNAs targeting three SNPs may provide therapy for three-quarters of Huntington's disease patients. Curr Biol 19:774–778. doi:10.1016/j.cub.2009.03.030
240. Pineda M, Moghadam F, Ebrahimkhani MR, Kiani S (2017) Engineered CRISPR Systems for Next Generation Gene Therapies. ACS Synthetic Biol. doi:10.1021/acssynbio.7b00011
241. Pluciennik A, Burdett V, Baitinger C, Iyer RR, Shi K, Modrich P (2013) Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLalpha endonuclease activation. Proc Natl Acad Sci U S A 110:12277–12282. doi:10.1073/pnas.1311325110
242. Polleys EJ, NCM H, Freudenreich CH (2017) Role of recombination and replication fork restart in repeat instability. DNA Repair 56:156–165. doi:10.1016/j.dnarep.2017.06.018

243. Querido E, Gallardo F, Beaudoin M, Menard C, Chartrand P (2011) Stochastic and reversible aggregation of mRNA with expanded CUG-triplet repeats. J Cell Sci 124:1703–1714 Doi 2461242/jcs.073270
244. Raca G, Siyanova EY, McMurray CT, Mirkin SM (2000) Expansion of the (CTG)(n) repeat in the 5'-UTR of a reporter gene impedes translation. Nucleic Acids Res 28:3943–3949
245. Ranum LP, Cooper TA (2006) RNA-mediated neuromuscular disorders. Annu Rev Neurosci 29:259–277. doi:10.1146/annurev.neuro.29.051605.113014
246. Reddy K, Schmidt MH, Geist JM, Thakkar NP, Panigrahi GB, Wang YH, Pearson CE (2014) Processing of double-R-loops in (CAG).(CTG) and C9orf72 (GGGGCC).(GGCCCC) repeats causes instability. Nucleic Acids Res 42:10473–10487. doi:10.1093/nar/gku658
247. Reddy K, Zamiri B, Stanley SYR, Macgregor RB, Pearson CE (2013) The Disease-associated r(GGGGCC)(n) Repeat from the C9orf72 Gene Forms Tract Length-dependent Uni- and Multimolecular RNA G-quadruplex Structures. J Biol Chem 288:9860–9866. doi:10.1074/jbc.C113.452532
248. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L et al (2011) A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. Neuron 72:257–268. doi:10.1016/j.neuron.2011.09.010
249. Richard GF, Cyncynatus C, Dujon B (2003) Contractions and expansions of CAG/CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a MUS81-independent mechanism. J Mol Biol 326:769–782
250. Richard GF, Goellner GM, McMurray CT, Haber JE (2000) Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11-RAD50-XRS2 complex. EMBO J 19:2381–2390. doi:10.1093/emboj/19.10.2381
251. Richard P, Manley JL (2016) R Loops and Links to Human Disease. J Mol Biol. doi:10.1016/j.jmb.2016.08.031
252. Rzuczek SG, Colgan LA, Nakai Y, Cameron MD, Furling D, Yasuda R, Disney MD (2017) Precise small-molecule recognition of a toxic CUG RNA repeat expansion. Nat Chem Biol 13:188–193. doi:10.1038/nchembio.2251
253. Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, Wakisaka A, Tashiro K, Ishida Y, Ikeuchi T et al (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet 14:277–284. doi:10.1038/ng1196-277
254. Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, Chedin F (2016) Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell 63:167–178. doi:10.1016/j.molcel.2016.05.032
255. Sathasivam K, Neueder A, Gipson TA, Landles C, Benjamin AC, Bondulich MK, Smith DL, Faull RL, Roos RA, Howland D et al (2013) Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. Proc Natl Acad Sci U S A 110:2366–2370. doi:10.1073/pnas.1221891110
256. Sato N, Amino T, Kobayashi K, Asakawa S, Ishiguro T, Tsunemi T, Takahashi M, Matsuura T, Flanigan KM, Iwasaki S et al (2009) Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. Am J Hum Genet 85:544–557. doi:10.1016/j.ajhg.2009.09.019
257. Sawaya S, Bagshaw A, Buschiazzo E, Kumar P, Chowdhury S, Black MA, Gemmell N (2013) Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. PLoS One 8:e54710. doi:10.1371/journal.pone.0054710
258. Sawyer LA, Hennessy JM, Peixoto AA, Rosato E, Parkinson H, Costa R, Kyriacou CP (1997) Natural variation in a Drosophila clock gene and temperature compensation. Science 278:2117–2120
259. Schmidt K, Butler JS (2013) Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity. Wiley Interdiscip Rev RNA 4:217–231. doi:10.1002/wrna.1155
260. Schmidt MH, Pearson CE (2016) Disease-associated repeat instability and mismatch repair. DNA Repair 38:117–126. doi:10.1016/j.dnarep.2015.11.008
261. Scoles DR, Ho MH, Dansithong W, Pflieger LT, Petersen LW, Thai KK, Pulst SM (2015) Repeat Associated Non-AUG Translation (RAN Translation) Dependent on Sequence Downstream of the ATXN2 CAG Repeat. PLoS One 10:e0128769. doi:10.1371/journal.pone.0128769
262. Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ et al (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J 29:1248–1261. doi:10.1038/emboj.2010.21
263. Shepard S, McCreary M, Fedorov A (2009) The peculiarities of large intron splicing in animals. PLoS One 4:e7853. doi:10.1371/journal.pone.0007853
264. Shi KY, Mori E, Nizami ZF, Lin Y, Kato M, Xiang S, Wu LC, Ding M, Yu Y, Gall JG et al (2017) Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. Proc Natl Acad Sci U S A 114:E1111–E1117. doi:10.1073/pnas.1620293114
265. Shin JW, Kim KH, Chao MJ, Atwal RS, Gillis T, MacDonald ME, Gusella JF, Lee JM (2016) Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. Hum Mol Genet 25:4566–4576. doi:10.1093/hmg/ddw286
266. Shishkin AA, Voineagu I, Matera R, Cherng N, Chernet BT, Krasilnikova MM, Narayanan V, Lobachev KS, Mirkin SM (2009) Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell 35:82–92. doi:10.1016/j.molcel.2009.06.017
267. Shoemaker CJ, Green R (2012) Translation drives mRNA quality control. Nat Struct Mol Biol 19:594–601. doi:10.1038/nsmb.2301
268. Shoubridge C, Gecz J (2012) Polyalanine tract disorders and neurocognitive phenotypes. Adv Exp Med Biol 769:185–203
269. Shukla S, Parker R (2016) Hypo- and Hyper-Assembly Diseases of RNA-Protein Complexes. Trends Mol Med 22:615–628. doi:10.1016/j.molmed.2016.05.005
270. Skourti-Stathaki K, Proudfoot NJ, Gromak N (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell 42:794–805. doi:10.1016/j.molcel.2011.04.026
271. Slean MM, Panigrahi GB, Castel AL, Pearson AB, Tomkinson AE, Pearson CE (2016) Absence of MutSbeta leads to the formation of slipped-DNA for CTG/CAG contractions at primate replication forks. DNA repair 42:107–118. doi:10.1016/j.dnarep.2016.04.002
272. Sleeman JE, Trinkle-Mulcahy L (2014) Nuclear bodies: new insights into assembly/dynamics and disease relevance. Curr Opin Cell Biol 28:76–83. doi:10.1016/j.ceb.2014.03.004
273. Sloan KE, Gleizes PE, Bohnsack MT (2016) Nucleocytoplasmic Transport of RNAs and RNA-Protein Complexes. J Mol Biol 428:2040–2059. doi:10.1016/j.jmb.2015.09.023
274. Smith KP, Byron M, Johnson C, Xing Y, Lawrence JB (2007) Defining early steps in mRNA transport: mutant mRNA in myotonic dystrophy type I is blocked at entry into SC-35 domains. J Cell Biol 178:951–964. doi:10.1083/jcb.200706048
275. Sobczak K, Wheeler TM, Wang W, Thornton CA (2013) RNA interference targeting CUG repeats in a mouse model of myotonic dystrophy. Mol Ther 21:380–387. doi:10.1038/mt.2012.222
276. Sofola OA, Jin P, Qin Y, Duan R, Liu H, de Haro M, Nelson DL, Botas J (2007) RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. Neuron 55:565–571. doi:10.1016/j.neuron.2007.07.021
277. Speese SD, Ashley J, Jokhi V, Nunnari J, Barria R, Li Y, Ataman B, Koon A, Chang YT, Li Q et al (2012) Nuclear envelope budding enables large ribonucleoprotein particle export during synaptic Wnt signaling. Cell 149:832–846. doi:10.1016/j.cell.2012.03.032
278. Spiro C, Pelletier R, Rolfsmeier ML, Dixon MJ, Lahue RS, Gupta G, Park MS, Chen X, Mariappan SV, McMurray CT (1999) Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. Mol Cell 4:1079–1085
279. Stern A, Brown M, Nickel P, Meyer TF (1986) Opacity genes in Neisseria gonorrhoeae: control of phase and antigenic variation. Cell 47:61–71
280. Strack RL, Disney MD, Jaffrey SR (2013) A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. Nat Methods 10:1219–1224. doi:10.1038/Nmeth.2701
281. Su Z, Zhang Y, Gendron TF, Bauer PO, Chew J, Yang WY, Fostvedt E, Jansen-West K, Belzil VV, Desaro P et al (2014) Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. Neuron 83:1043–1050. doi:10.1016/j.neuron.2014.07.041
282. Sun X, Li PP, Zhu S, Cohen R, Marque LO, Ross CA, Pulst SM, Chan HY, Margolis RL, Rudnicki DD (2015) Nuclear retention of full-length HTT RNA is mediated by splicing factors MBNL1 and U2AF65. Sci Rep 5:12521. doi:10.1038/srep12521
283. Sundararajan R, Gellon L, Zunder RM, Freudenreich CH (2010) Double-strand break repair pathways protect against CAG/CTG repeat expansions, contractions and repeat-mediated chromosomal fragility in Saccharomyces cerevisiae. Genetics 184:65–77. doi:10.1534/genetics.109.111039
284. Supek F, Lehner B (2015) Differential DNA mismatch repair underlies mutation rate variation across the human genome. Nature 521:81–84. doi:10.1038/nature14173
285. Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, Warren ST (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol Genet 1:397–400

286. Tao Z, Wang H, Xia Q, Li K, Jiang X, Xu G, Wang G, Ying Z (2015) Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. Hum Mol Genet. doi:10.1093/hmg/ddv005
287. Tassone F, Hagerman RJ, Taylor AK, Hagerman PJ (2001) A majority of fragile X males with methylated, full mutation alleles have significant levels of FMR1 messenger RNA. J Med Genet 38:453–456
288. Taylor JP, Brown RH Jr, Cleveland DW (2016) Decoding ALS: from genes to mechanism. Nature 539:197–206. doi:10.1038/nature20413
289. Thandapani P, O'Connor TR, Bailey TL, Richard S (2013) Defining the RGG/RG motif. Mol Cell 50:613–623. doi:10.1016/j.molcel.2013.05.021
290. Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, Renoux AJ, Chen KC, Scaglione KM, Basrur V et al (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron 78:440–455. doi:10.1016/j.neuron.2013.03.026
291. Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10:967–981
292. Tran T, Childs-Disney JL, Liu B, Guan L, Rzuczek S, Disney MD (2014) Targeting the r(CGG) repeats that cause FXTAS with modularly assembled small molecules and oligonucleotides. Acs Chem Biol 9:904–912. doi:10.1021/cb400875u
293. Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM (2004) An abundance of bidirectional promoters in the human genome. Genome Res 14:62–66. doi:10.1101/gr.1982804
294. Turner EM, Brown RS, Laudermilch E, Tsai PL, Schlieker C (2015) The Torsin Activator LULL1 Is Required for Efficient Growth of Herpes Simplex Virus 1. J Virol 89:8444–8452. doi:10.1128/JVI.01143-15
295. Urbanek MO, Jazurek M, Switonski PM, Figura G, Krzyzosiak WJ (2016) Nuclear speckles are detention centers for transcripts containing expanded CAG repeats. Biochim Biophys Acta 1862:1513–1520. doi:10.1016/j.bbadis.2016.05.015
296. Usdin K (2008) The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. Genome Res 18:1011–1019. doi:10.1101/gr.070409.107 18/7/1011
297. Usdin K, House NC, Freudenreich CH (2015) Repeat instability during DNA repair: Insights from model systems. Crit Rev Biochem Mol Biol 50:142–167. doi:10.3109/10409238.2014.999192
298. Usdin K, Kumari D (2015) Repeat-mediated epigenetic dysregulation of the FMR1 gene in the fragile X-related disorders. Front Genet 6:192. doi:10.3389/fgene.2015.00192
299. van Agtmaal EL, Andre LM, Willemse M, Cumming SA, van Kessel IDG, van den Broek W, Gourdon G, Furling D, Mouly V, Monckton DG et al (2017) CRISPR/Cas9-Induced (CTGCAG)n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing. Mol Ther 25:24–43. doi:10.1016/j.ymthe.2016.10.014
300. van Blitterswijk M, DeJesus-Hernandez M, Rademakers R (2012) How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? Curr Opin Neurol 25:689–700
301. Van Der Kelen K, Beyaert R, Inze D, De Veylder L (2009) Translational control of eukaryotic gene expression. Crit Rev Biochem Mol Biol 44:143–168. doi:10.1080/10409230902882090
302. van Hoof A, Frischmeyer PA, Dietz HC, Parker R (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science 295:2262–2264. doi:10.1126/science.1067272
303. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914
304. Villasante A, Abad JP, Mendez-Lago M (2007) Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. Proc Natl Acad Sci U S A 104:10542–10547. doi:10.1073/pnas.0703808104
305. Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F et al (1998) DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev 12:343–356
306. Warren ST (1997) Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. Science 275:408–409
307. Weiser JN, Love JM, Moxon ER (1989) The molecular mechanism of phase variation of H. influenzae lipopolysaccharide. Cell 59:657–665
308. Wenzel S, Martins BM, Rosch P, Wohrl BM (2009) Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. Biochem J 425:373–380. doi:10.1042/BJ20091422
309. Werner F (2012) A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. J Mol Biol 417:13–27. doi:10.1016/j.jmb.2012.01.031
310. Wheeler TM, Sobczak K, Lueck JD, Osborne RJ, Lin X, Dirksen RT, Thornton CA (2009) Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. Science 325:336–339
311. White MC, Gao R, Xu W, Mandal SM, Lim JG, Hazra TK, Wakamiya M, Edwards SF, Raskin S, Teive HA et al (2010) Inactivation of hnRNP K by expanded intronic AUUCU repeat induces apoptosis via translocation of PKCdelta to mitochondria in spinocerebellar ataxia 10. PLoS Genet 6:e1000984. doi:10.1371/journal.pgen.1000984
312. Wilburn B, Rudnicki DD, Zhao J, Weitz TM, Cheng Y, Gu X, Greiner E, Park CS, Wang N, Sopher BL et al (2011) An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice. Neuron 70:427–440. doi:10.1016/j.neuron.2011.03.021
313. Wojciechowska M, Olejniczak M, Galka-Marciniak P, Jazurek M, Krzyzosiak WJ (2014) RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. Nucleic Acids Res 42:11849–11864. doi:10.1093/nar/gku794
314. Wong CH, Fu Y, Ramisetty SR, Baranger AM, Zimmerman SC (2011) Selective inhibition of MBNL1-CCUG interaction by small molecules toward potential therapeutic agents for myotonic dystrophy type 2 (DM2). Nucleic Acids Res 39:8881–8890. doi:10.1093/nar/gkr415
315. Woodward LA, Mabin JW, Gangras P, Singh G (2016) The exon junction complex: a lifelong guardian of mRNA fate. Wiley Int Rev RNA. doi:10.1002/wrna.1411
316. Xi Z, Zhang M, Bruni AC, Maletta RG, Colao R, Fratta P, Polke JM, Sweeney MG, Mudanohwo E, Nacmias B et al (2015) The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. Acta Neuropathol 129:715–727. doi:10.1007/s00401-015-1401-8
317. Xi Z, Zinman L, Moreno D, Schymick J, Liang Y, Sato C, Zheng Y, Ghani M, Dib S, Keith J et al (2013) Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. Am J Hum Genet 92:981–989. doi:10.1016/j.ajhg.2013.04.017
318. Xia G, Gao Y, Jin S, Subramony SH, Terada N, Ranum LP, Swanson MS, Ashizawa T (2015) Genome modification leads to phenotype reversal in human myotonic dystrophy type 1 induced pluripotent stem cell-derived neural stem cells. Stem Cells 33:1829–1838. doi:10.1002/stem.1970
319. Xu Z, Poidevin M, Li X, Li Y, Shu L, Nelson DL, Li H, Hales CM, Gearing M, Wingo TS et al (2013) Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proc Natl Acad Sci U S A 110:7778–7783. doi:10.1073/pnas.1219643110
320. Yamada NA, Smith GA, Castro A, Roques CN, Boyer JC, Farber RA (2002) Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells. Mutat Res 499:213–225
321. Yamazaki T, Hirose T (2015) The building process of the functional paraspeckle with long non-coding RNAs. Front Biosci (Elite Ed) 7:1–41
322. Yang W, Tu Z, Sun Q, Li XJ (2016) CRISPR/Cas9: Implications for Modeling and Therapy of Neurodegenerative Diseases. Front Mol Neurosci 9:30. doi:10.3389/fnmol.2016.00030
323. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol 4:442–451. doi:10.1038/ni919
324. Yu K, Roy D, Huang FT, Lieber MR (2006) Detection and structural analysis of R-loops. Methods Enzymol 409:316–329. doi:10.1016/S0076-6879(05)09018-X
325. Zamiri B, Reddy K, Macgregor RB Jr, Pearson CE (2014) TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. J Biol Chem 289:4653–4659. doi:10.1074/jbc.C113.502336
326. Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, Daley EL, Miller SJ, Cunningham KM, Vidensky S et al (2015) The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature 525:56–61. doi:10.1038/nature14973
327. Zhang N, Ashizawa T (2017) RNA toxicity and foci formation in microsatellite expansion diseases. Curr Opin Genet Dev 44:17–29. doi:10.1016/j.gde.2017.01.005

328. Zhang S, Binari R, Zhou R, Perrimon N (2010) A genomewide RNA interference screen for modifiers of aggregates formation by mutant Huntingtin in Drosophila. Genetics 184:1165–1179. doi:10.1534/genetics.109.112516
329. Zhang Y, Roland C, Sagui C (2016) Structure and Dynamics of DNA and RNA Double Helices Obtained from the GGGGCC and CCCCGG Hexanucleotide Repeats That Are the Hallmark of C9FTD/ALS Diseases. ACS Chem Neurosci. doi:10.1021/acschemneuro.6b00348
330. Zhang Y, Shishkin AA, Nishida Y, Marcinkowski-Desmond D, Saini N, Volkov KV, Mirkin SM, Lobachev KS (2012) Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. Mol Cell 48:254–265. doi:10.1016/j.molcel.2012.08.002
331. Zhao J, Bacolla A, Wang G, Vasquez KM (2010) Non-B DNA structure-induced genetic instability and evolution. Cell Mol Life Sci 67:43–62. doi:10.1007/s00018-009-0131-2
332. Zhao T, Hong Y, Li XJ, Li SH (2016) Subcellular Clearance and Accumulation of Huntington Disease Protein: A Mini-Review. Front Mol Neurosci 9:27. doi:10.3389/fnmol.2016.00027
333. Zhao X, Usdin K (2015) The Repeat Expansion Diseases: The dark side of DNA repair. DNA Repair 32:96–105
334. Zhao XN, Lokanga R, Allette K, Gazy I, Wu D, Usdin K (2016) A MutSbeta-Dependent Contribution of MutSalpha to Repeat Expansions in Fragile X Premutation Mice? PLoS Genet 12:e1006190. doi:10.1371/journal.pgen.1006190
335. Zhou Z, Luo MJ, Straesser K, Katahira J, Hurt E, Reed R (2000) The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. Nature 407:401–405. doi:10.1038/35030160
336. Zhu L, Brangwynne CP (2015) Nuclear bodies: the emerging biophysics of nucleoplasmic phases. Curr Opin Cell Biol 34:23–30. doi:10.1016/j.ceb.2015.04.003
337. Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, Subramony SH, Zoghbi HY, Lee CC (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. Nat Genet 15:62–69. doi:10.1038/ng0197-62
338. Zinder JC, Lima CD (2017) Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. Genes Dev 31:88–100. doi:10.1101/gad.294769.116
339. Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MA et al (2011) Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A 108:260–265. doi:10.1073/pnas.1013343108
340. Zu T, Liu Y, Banez-Coronel M, Reid T, Pletnikova O, Lewis J, Miller TM, Harms MB, Falchook AE, Subramony SH et al (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci U S A 110:E4968–E4977. doi:10.1073/pnas.1315438110
341. Zumwalt M, Ludwig A, Hagerman PJ, Dieckmann T (2007) Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. RNA Biol 4:93–100

Biochemistry

Communication

Cite This: Biochemistry 2018, 57, 3027−3031

pubs.acs.org/biochemistry

Chimeric Guides Probe and Enhance Cas9 Biochemical Activity

Zachary J. Kartje,[†] Christopher L. Barkau,[‡] Kushal J. Rohilla,[‡] Eman A. Ageely,[†] and Keith T. Gagnon*[,†,‡]

[†]Department of Chemistry and Biochemistry, Southern Illinois University, Carbondale, Illinois 62901, United States

[‡]Department of Biochemistry and Molecular Biology, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901, United States

Supporting Information

ABSTRACT: DNA substitutions in RNA can probe the importance of A-form structure, 2′-hydroxyl contacts, and conformational constraints within RNA-guided enzymes. Using this approach, we found that Cas9 biochemical activity tolerated significant substitution with DNA nucleotides in the clustered regularly interspaced short palindromic repeat RNA (crRNA). Only minimal RNA content was needed in or near the seed region. Simultaneous substitution at all positions with predicted crRNA−Cas9 2′-hydroxyl contacts had no effect on enzyme activity. The *trans*-activating crRNA (tracrRNA) also tolerated >50% substitution with DNA. DNA substitutions in the tracrRNA-pairing region of crRNA consistently enhanced cleavage activity while maintaining or improving target specificity. Together, results point to a prominent role for guide:target A-form-like helical structure and a possible regulatory role for the crRNA−tracrRNA pairing motif. A model chimeric crRNA with high activity did not significantly alter RNP assembly or target binding but did reduce Cas9 ribonucleoprotein stability, suggesting effects through conformation or dynamics. Cas9 directed by chimeric RNA−DNA guides may represent a cost-effective synthetic or molecular biology tool for robust and specific DNA cleavage.

RNA-guided enzymes are evolutionarily ancient ribonucleoproteins (RNPs) that act upon nucleic acid substrates using associated RNA guides for specificity.[1,2] Understanding how an RNA guide and its associated protein cooperate is important for characterizing enzyme mechanisms and engineering RNA-guided enzymes. One step toward understanding the functional marriage between an RNA and protein is probing with chemically modified residues.[3−5] DNA nucleotides, simply lacking a 2′-hydroxyl group, are effective probes.[6] They are affordable to synthesize and have well-characterized properties. DNA typically forms a B-form double helix when paired to another DNA strand but can readily undergo a transition to an A-form helix depending on the local environment.[7] Chimeric RNA−DNA oligonucleotides prefer to form A-form-like helices due to RNA dominance.[8−10] This is attributed to the relative flexibility of the 2′-deoxyribose sugar when incorporated into nucleic acids, as opposed to more sterically constrained ribonucleotides.[11] Inspiration for incorporating DNA nucleotides also arises from naturally occurring DNA-guided enzymes like Argonaute.[12,13]

Clustered regularly interspaced short palindromic repeat (CRISPR) RNAs (crRNAs) and the factors that associate with them, Cas proteins, are recently discovered RNA-guided proteins.[1,14] Enzymatic CRISPR−Cas complexes can recognize and degrade foreign DNA in bacteria and Archaea.[14] Enzymes like Cas9 from *Streptococcus pyogenes* have been heavily co-opted for genome editing and synthetic biology. They are straightforward to program and can elicit sequence-specific DNA cleavage.[15−17] Their continued development will depend on characterization of mechanism and their tolerance to modifications and unnatural conditions.

We chose the CRISPR−Cas9 complex from *S. pyogenes* (*Sp*Cas9) as a model system to investigate. Because *Sp*Cas9 evolved to use two separate guide RNAs,[14,16] the crRNA and *trans*-activating crRNA (tracrRNA), and to simplify chemical synthesis of RNA−DNA chimeras, we chose to use the more natural dual RNA-guided CRISPR−Cas9 complex (Figure 1A). This is in contrast to the more commonly used artificial single-guide RNA (sgRNA) that fuses a crRNA and tracrRNA together.[16] We assembled *Sp*Cas9 RNP complexes and assessed *in vitro* DNA cleavage activity with purified components (Figure S1A). Using two substrates (Figures S1B and S2), a linearized EGFP plasmid and a 60 bp fluorescently labeled DNA duplex (FAM duplex), we systematically tested DNA substitution of the crRNA. The FAM duplex targets facilitated later biochemical experiments and extended our general conclusions. We found that cleavage activity trends were quite similar for both types of target DNA substrate.

DNA nucleotides were initially incorporated at the 5′ end of the guide. Activity against a single sequence (E2 target site) steadily decreased as the DNA composition expanded beyond nine residues and into the seed region of the guide (Figure 1B and Figure S1C). The seed sequence is the portion of the guide region that nucleates target hybridization.[16] These results suggested a need for RNA in the seed sequence. To test this hypothesis, we substituted eight contiguous seed residues with DNA (crE2-I) and surprisingly observed robust activity. We suspected that the guide may instead require a minimum amount of RNA to maintain A-form-like helical structure upon hybridization to target DNA. Upon complete substitution with

Received: January 30, 2018
Revised: April 19, 2018
Published: May 10, 2018

DOI: 10.1021/acs.biochem.8b00107
Biochemistry 2018, 57, 3027−3031

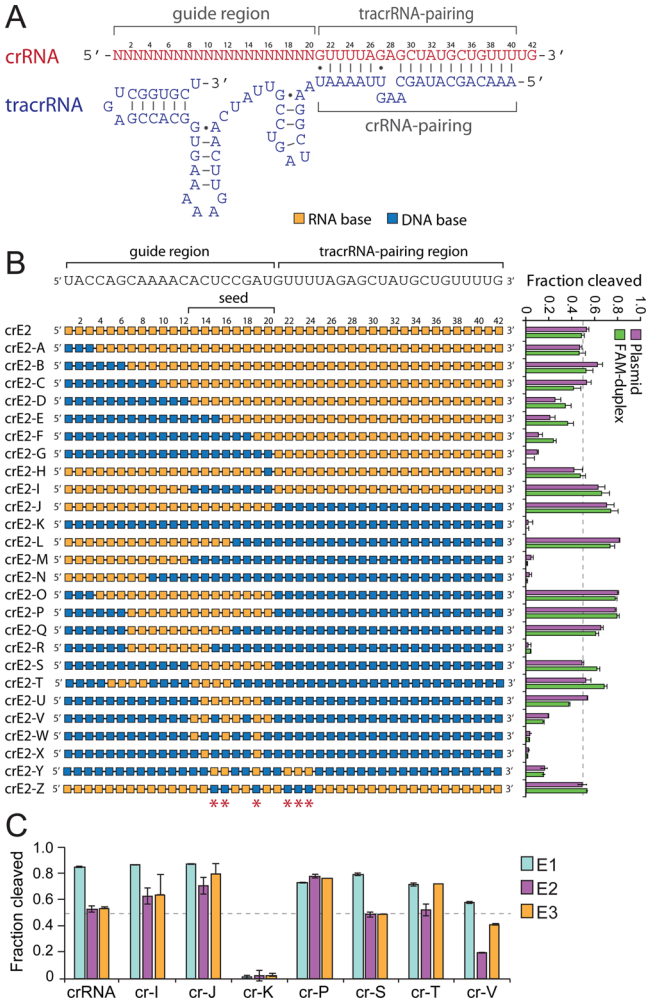


Figure 1. Chimeric RNA–DNA crRNAs are tolerated and can enhance the cleavage activity of *Sp*Cas9. (A) Sequence and secondary structure of a dual-RNA guide for *Sp*Cas9. (B) DNA substitution patterns for crRNA and subsequent *in vitro* cleavage activity for *Sp*Cas9. The DNA (blue) and RNA (orange) composition is schematically shown with the corresponding catalytic activity to the right. Cleavage of the plasmid (purple) or short FAM duplex DNA (green) is shown. Critical contacts between Cas9 and 2′-hydroxyls of crRNA are denoted with red asterisks. Error bars represent the standard error of the mean (SEM). (C) Cleavage of plasmid by *Sp*Cas9 RNPs assembled with chimeric crRNAs targeting different sequences. E2 cleavage data are taken from panel B for comparison. crRNA naming follows the substitution scheme and convention in panel B. Error bars are SEM.

DNA at the 3′ end that pairs with the tracrRNA, we observed significantly enhanced cleavage activity (crE2-J) (Table S1). Chimeras with complete DNA substitution in this region, including crE2-L, crE2-O, and crE2-P, also exhibited enhanced cleavage. Cleavage activity was, however, compromised when too many DNA residues were substituted into the guide region, in particular the seed region (i.e., crE2-M, crE2-N, and crE2-R). Minimal RNA content was achieved with as few as six RNA residues in the seed (crE2-U and crE2-V), although activity was reduced. These results suggest that a minimum number of RNA residues might be required in the guide, preferably in or near the seed region, for catalysis. The six critical 2′-hydroxyl contacts predicted between crRNA and Cas9,[18] three of which are in the seed region, are lost to varying degrees in several of these chimeras. To directly test the role of specific 2′-hydroxyl contacts, we generated an all-DNA crRNA with all six positions converted to RNA (crE2-Y). This resulted in activity similar to that of another crRNA with only six RNA residues, crE2-V. In contrast, substituting an all-RNA crRNA with DNA at these same six residues (crE2-Z), thereby eliminating all predicted critical 2′-hydroxyls between crRNA and Cas9, did not significantly alter cleavage activity (Table S1). These results demonstrate that predicted 2′-hydroxyl contacts are dispensable for biochemical activity[19] and indicate that RNA content is important for maintaining an A-form conformation of the crRNA.

To test whether enhanced cleavage was independent of guide region sequence, we targeted two additional sites of the EGFP plasmid, E1 and E3 (Figure 1C and Figure S1D,E). The tested crRNA substitution patterns gave similar results for most targets. Although the E1 target was more efficiently cleaved overall, configurations like cr-J and cr-P consistently conferred high activity.

Other CRISPR–Cas systems may also tolerate DNA substitution. To evaluate this possibility, we assembled dual RNA-guided *Staphylococcus aureus* Cas9 (*Sa*Cas9) RNPs with DNA-substituted crRNAs against the E2 target site. DNA substitution patterns showed high activity for some configurations (crE2-I, crE2-J, and crE2-P) (Figure S3A). However, others that had previously induced efficient cleavage for *Sp*Cas9 provided little activity. These results suggest that chimeric RNA–DNA guides will generally be compatible with other CRISPR–Cas systems, although substitution patterns tolerated in the guide region may vary in an enzyme-specific manner.

We reasoned that higher activity might be the result of nonspecific cleavage. To evaluate this possibility, we resolved radiolabeled short duplex cleavage products on a denaturing polyacrylamide sequencing gel (Figure 2A). We found that

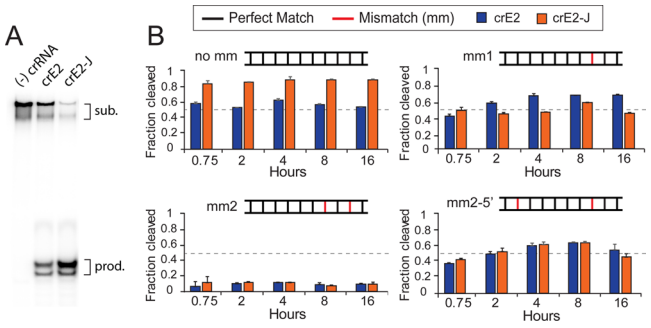


Figure 2. Chimeric crRNA maintains Cas9 specificity. (A) Resolution of radiolabeled *Sp*Cas9 cleavage products on a sequencing gel. Two radiolabeled products that differ by a single nucleotide can be resolved. (B) Cleavage activity using mismatched FAM duplex targets. *Sp*Cas9 RNPs were incubated with the targets indicated. Cleavage activity was measured after the time specified. Mismatch target sequences are listed in Table S2. Error bars are SEM.

*Sp*Cas9 guided by native crE2 and a model chimeric crE2-J catalyzed cleavage of the same phosphodiester bonds, a major and minor product as expected from previous reports.[15,16] Therefore, cleavage remained site-specific. We also tested a FAM duplex target with an NTT instead of an NGG protospacer adjacent motif (PAM) sequence. The PAM sequence is an additional sequence requirement for each Cas protein, with NGG being specific for *Sp*Cas9.[16,20,21] Little or no

DOI: 10.1021/acs.biochem.8b00107
Biochemistry 2018, 57, 3027–3031

cleavage was observed for crE2 or crE2-J with an NTT PAM (Figure S3B).

To further address specificity, we synthesized three FAM duplex targets containing nucleobase changes that would result in mismatches with the crRNA. These substrates can test the substrate specificity of Cas9 assembled with chimeric crRNA. A single seed mismatch at position 16 (mm1) should only moderately impact activity, while two seed mismatches at positions 16 and 12 (mm2) should abrogate cleavage. In contrast, combining a mismatch outside of the seed (position 6) with a position 16 mismatch in the seed (mm2-5′) should be tolerated like a single mismatch in the seed.[22−24] Activity of crE2 was compared to that of substituted crE2-J for all targets (Figure 2B, Figure S3C,D, and Table S2). The level of cleavage of the mm1 target by crE2 increased from ∼40% to almost 70% over time, while the level of cleavage by crE2-J stayed nearly constant at ∼50%. This represents a significant discrimination against the mm1 target at all time points by crE2-J compared to crE2 (Table S1). Cleavage of mm2 and mm2-5′ was similar for both crE2 and crE2-J. These targets represent a small fraction of possible mismatches. Nonetheless, these results together demonstrate that chimeric crE2-J can maintain or improve Cas9 specificity.

Having established tolerance for DNA in the crRNA, we also probed the tracrRNA with DNA substitution (Figure 3).

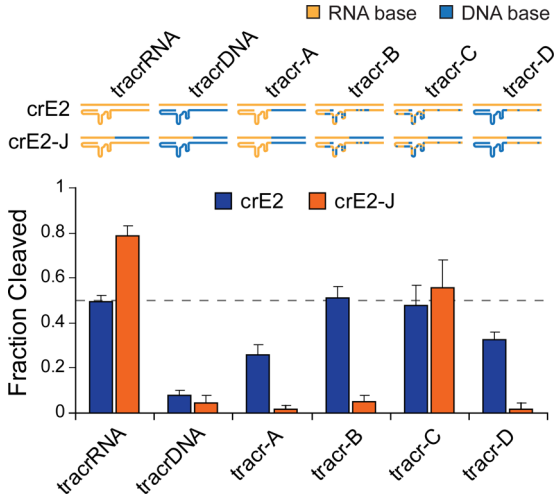


Figure 3. DNA substitutions in tracrRNA are tolerated. Cleavage activity of *Sp*Cas9 RNP assembled with crE2 and crE2-J and chimeric tracrRNA. Substitution illustrations are shown above the graph. Full substitution patterns are listed in Table S1. Error bars are SEM.

Complete conversion of tracrRNA to DNA (tracrDNA) supported little or no activity. Substitution of DNA in the crRNA-pairing region of the tracrRNA (tracr-A) supported approximately half of the usual cleavage activity for crE2 but none for crE2-J. More selective replacement of RNA with DNA in certain parts of the tracrRNA, such as stem structures (tracr-B), improved crE2 activity to that of unsubstituted tracrRNA but did not yield activity for crE2-J. Therefore, we used a crystal structure of an *Sp*Cas9 ternary complex to guide additional substitutions and potentially conserve critical Cas9 contacts or RNA structure.[18] We substituted 34 of 66 (∼50%) RNA nucleotides with DNA (tracr-C) and observed restored activity for both crE2 and crE2-J that was similar to that of an unsubstituted tracrRNA. As an additional control, we synthesized a tracrRNA containing DNA everywhere except for 16 bases in the crRNA-pairing region (tracr-D). Surprisingly, this design (∼75% DNA) supported approximately two-thirds of the activity of a normal tracrRNA for crE2. These results demonstrate that the entire 3′ portion of tracrRNA that anchors Cas9 binding can be completely substituted with DNA and still provide moderate enzyme activity. Thus, binding of tracrRNA to Cas9 may involve an induced fit of the tracrRNA. Combined with substitutions in the crRNA that enhanced activity, these results indicate that elements for regulating Cas9 activity may reside in the crRNA−tracrRNA pairing motif.

To better understand enhanced activity with chimeric crRNAs, we performed time-course cleavage, RNP assembly, target binding, and RNP stability assays. In time-course experiments, we observed rapid cleavage of the linearized plasmid by *Sp*Cas9 when guided by native crE2, confirming previous reports of fast kinetics[15,16] (Figure 4A). Cleavage was largely complete within 30 s. The crE2-J chimera likewise showed rapid kinetics but a greater overall level of cleavage, supporting our usual end-point assays. These data suggest that crE2-J may force assembly of Cas9 into a conformation that is more poised for cleavage. To determine whether DNA substitutions affected RNP assembly or global conformation, we performed electrophoretic mobility shift assays (EMSAs), or gel shifts. A catalytically inactive "dead" *Sp*Cas9 (dCas9) was titrated with radiolabeled tracrRNA−crRNA and resolved on a native polyacrylamide gel (Figure 4B and Figure S4A−C). Either crE2, crE2-J, or crE2-K as a control was tested for RNP assembly. The gel shift patterns and calculated apparent binding affinities from replicate gels were nearly the same for all three crRNAs irrespective of DNA content (Table S1). Because crE2-K Cas9 RNPs are catalytically inactive, we instead

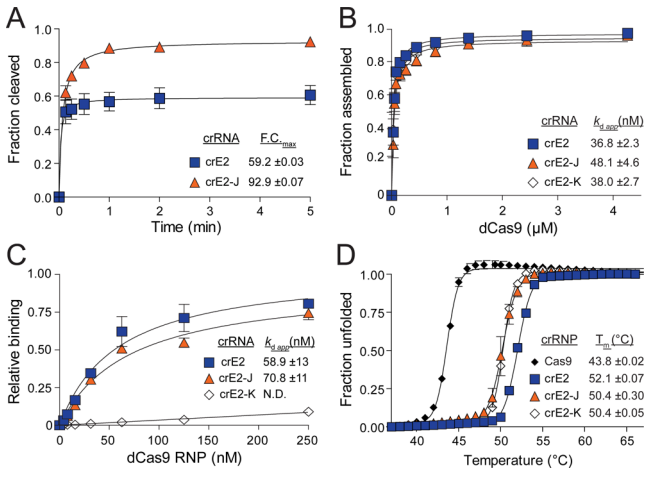


Figure 4. Model chimeric crRNA maintains RNP assembly and target engagement but reduces RNP stability. (A) *Sp*Cas9 RNP complexes were assayed for plasmid cleavage over time. $F.C._{max}$ is the maximum fraction cleaved. Error bars are SEM. (B) Quantification of binding of dCas9 to radiolabeled tracrRNA and indicated crRNAs by an EMSA. Error bars and the error of calculated apparent dissociation constants ($k_{d\ app}$) are reported as SEM. (C) Quantification of preassembled dCas9 RNP binding to radiolabeled duplex DNA as determined by dot-blot filter binding. Error bars and the error of calculated apparent dissociation constants ($k_{d\ app}$) are reported as SEM. (D) Thermal stability of *Sp*Cas9 alone or assembled into RNPs as determined by ultraviolet (UV) melts (280 nm). UV melt profiles are the average of two replicates, and the error is reported as SEM for calculated T_m values.

DOI: 10.1021/acs.biochem.8b00107
Biochemistry 2018, 57, 3027−3031

tested their ability to engage a double-stranded DNA target. We assembled RNP complexes with dCas9 and titrated them with radiolabeled duplex DNA targets. RNP−target complexes were separated by filter binding. No significant difference in binding between crE2 and crE2-J was observed (Figure 4C and Table S1). In contrast, target binding by crE2-K RNPs resulted in a nearly complete loss of target binding. Finally, to assess RNP stability, we assembled *Sp*Cas9 RNP complexes and monitored their melting at 280 nm (Figure 4D). The *Sp*Cas9 protein alone had a T_m of 43.8 °C. For assembled RNP, we observed a small but significant decrease in the thermal stability with crE2-J and crE2-K (T_m of 50.4 °C each) versus crE2 (T_m of 52.1 °C) (Table S1).

In this study, we used DNA substitutions to probe the structure−function relationship between guide RNA and Cas9 protein. Our results emphasize a requirement for A-form-like structure for target binding and cleavage rather than a conservation of 2′-hydroxyl contacts between Cas9 and crRNA, in contrast to a previous report.[19] Substantial DNA substitution could be tolerated, with as few as six to eight RNA residues in or near the seed region, and removal of all 2′-hydroxyls at predicted Cas9 contacts had no effect on activity. A fully substituted, all-DNA guide was able to assemble stable RNP complexes but unable to efficiently engage target DNA. It is likely that a B-form DNA−DNA guide:target duplex cannot fit into the Cas9 RNP. Structures of the CRISPR−Cas9 complex bound to target DNA show that the guide:target duplex assumes an A-form-like architecture.[18,23,25] It is possible that DNA−DNA hybridization thermodynamics could also reduce the level of Cas9 RNP target binding and therefore reduce activity. However, comparing the activity of other crRNAs, like crE2-M and crE2-S, reveals that DNA placement in the guide is the determining factor rather than the overall DNA content. We also performed a thermal denaturation of crRNA crE2, crE2-J, and crE2-K when annealed to complementary target DNA and found very similar T_m values (Figure S4D). Thus, our results argue for structural constraints when accommodating an R-loop structure and fitting substrate DNA into the Cas9 RNP.[23]

DNA substitutions of tracrRNA showed that the 3′ hairpins and nexus could be completely replaced by DNA and still support moderate cleavage. Cas9 may be able to remodel the DNA and induce it to fit into the RuvC and PI domains.[18,25,26] Many induced fit interactions have been described for RNA−protein complexes and usually involve conformational plasticity of the RNA.[27−29] The flexibility of DNA makes it a useful probe for studying induced fit interactions.[6] Previous studies have noted that high-affinity interactions between Cas9 and truncated sgRNAs can still occur, although they do not support catalysis.[26,30]

Substitution of DNA in the tracrRNA-pairing region of crRNAs consistently enhanced Cas9 biochemical activity. Cleavage activity was site-specific, even showing specificity greater than that of a native crRNA for a single seed mismatch. This same substitution pattern had no impact on RNP assembly but did reduce RNP stability. Effects on enzyme conformation or dynamics are possible explanations for these observations. DNA nucleotides could destabilize inactive conformations or allow more frequent sampling of catalytically competent conformations.[25,30,31]

The duplex between crRNA and tracrRNA may be sensitive to DNA substitutions as it forms a unique bulged motif due to imperfect pairing.[18] This motif is partially buried in the α-helical REC lobe of *Sp*Cas9.[18,32] It makes significant contacts with two short coiled regions of the REC lobe (amino acids 103−126) and the NUC lobe (amino acids 1109−1135) that lack strong α-helical or β structure. These two regions are completely unresolved in the apo *Sp*Cas9 structure,[32] indicating that RNA binding may play a role in stabilizing them through an induced fit of the protein.[27] The paired crRNA−tracrRNA motif might bridge Cas9 domain communication through these two regions, suggesting a possible allosteric role in regulation.[32] For example, large REC lobe movements during guide RNA binding[30] are necessary to maintain an open structure for engaging and processing target DNA.[23,25,30]

Our results indicate a need for conserving A-form-like helical structure in the crRNA of the CRISPR−Cas9 complex and suggest a potential new regulatory role for the crRNA−tracrRNA pairing motif. DNA-substituted crRNAs may find utility in molecular or synthetic biology applications as affordable alternatives to native RNA guides while maintaining predictably robust and specific DNA cleavage. Our results also contribute to the rational design of chemically modified CRISPR systems.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b00107.

Experimental details and supporting data (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ktgagnon@siu.edu.

ORCID

Keith T. Gagnon: 0000-0002-5868-675X

Funding

This work was supported by startup funding provided to K.T.G. by the Southern Illinois University School of Medicine and a U.S. Department of Defense Amyotrophic Lateral Sclerosis Research Program (ALSRP) award to K.T.G.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the M. J. Damha laboratory (McGill University, Montreal, QC) and H. Yin (Massachusetts Institute of Technology, Cambridge, MA) for helpful discussions.

■ REFERENCES

(1) Koonin, E. V. (2017) Evolution of RNA- and DNA-guided antivirus defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. *Biol. Direct 12*, 5.

(2) Huttenhofer, A., and Schattner, P. (2006) The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nat. Rev. Genet. 7*, 475−482.

(3) Kubota, M., Tran, C., and Spitale, R. C. (2015) Progress and challenges for chemical probing of RNA structure inside living cells. *Nat. Chem. Biol. 11*, 933−941.

(4) Nilsen, T. W. (2015) Nucleotide analog interference mapping. *Cold Spring Harbor Protocols 2015*, 604−608.

(5) Peacock, H., Kannan, A., Beal, P. A., and Burrows, C. J. (2011) Chemical modification of siRNA bases to probe and enhance RNA interference. *J. Org. Chem. 76*, 7295−7300.

DOI: 10.1021/acs.biochem.8b00107
Biochemistry 2018, 57, 3027−3031

(6) Ui-Tei, K., Naito, Y., Zenno, S., Nishi, K., Yamato, K., Takahashi, F., Juni, A., and Saigo, K. (2008) Functional dissection of siRNA sequence by systematic DNA substitution: modified siRNA with a DNA seed arm is a powerful tool for mammalian gene silencing with significantly reduced off-target effect. *Nucleic Acids Res. 36*, 2136−2151.
(7) Kulkarni, M., and Mukherjee, A. (2017) Understanding B-DNA to A-DNA transition in the right-handed DNA helix: Perspective from a local to global transition. *Prog. Biophys. Mol. Biol. 128*, 63−73.
(8) Wahl, M. C., and Sundaralingam, M. (2000) B-form to A-form conversion by a 3′-terminal ribose: crystal structure of the chimera d(CCACTAGTG)r(G). *Nucleic Acids Res. 28*, 4356−4363.
(9) Fedoroff, O., Salazar, M., and Reid, B. R. (1993) Structure of a DNA: RNA hybrid duplex. Why RNase H does not cleave pure RNA. *J. Mol. Biol. 233*, 509−523.
(10) Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S., and Reid, B. R. (1993) The DNA strand in DNA.RNA hybrid duplexes is neither B-form nor A-form in solution. *Biochemistry 32*, 4207−4215.
(11) Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag Inc., New York.
(12) Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Snijders, A. P., Wang, Y., Patel, D. J., Berenguer, J., Brouns, S. J., and van der Oost, J. (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature 507*, 258−261.
(13) Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T., and Patel, D. J. (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell 19*, 405−419.
(14) Mojica, F. J., and Rodriguez-Valera, F. (2016) The discovery of CRISPR in archaea and bacteria. *FEBS J. 283*, 3162−3169.
(15) Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A. 109*, E2579−2586.
(16) Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science 337*, 816−821.
(17) Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-guided human genome engineering via Cas9. *Science 339*, 823−826.
(18) Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell 156*, 935−949.
(19) Rueda, F. O., Bista, M., Newton, M. D., Goeppert, A. U., Cuomo, M. E., Gordon, E., Kroner, F., Read, J. A., Wrigley, J. D., Rueda, D., and Taylor, B. J. M. (2017) Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nat. Commun. 8*, 1610.
(20) Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science 315*, 1709−1712.
(21) Shah, S. A., Erdmann, S., Mojica, F. J., and Garrett, R. A. (2013) Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol. 10*, 891−899.
(22) Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., Cheng, A. W., Trevino, A. E., Konermann, S., Chen, S., Jaenisch, R., Zhang, F., and Sharp, P. A. (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol. 32*, 670−676.
(23) Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., and Doudna, J. A. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science 351*, 867−871.
(24) Hua Fu, B. X., Hansen, L. L., Artiles, K. L., Nonet, M. L., and Fire, A. Z. (2014) Landscape of target:guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Res. 42*, 13778−13787.
(25) Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J. A. (2015) A Cas9-guide RNA complex preorganized for target DNA recognition. *Science 348*, 1477−1481.
(26) Wright, A. V., Sternberg, S. H., Taylor, D. W., Staahl, B. T., Bardales, J. A., Kornfeld, J. E., and Doudna, J. A. (2015) Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci. U. S. A. 112*, 2984−2989.
(27) Williamson, J. R. (2000) Induced fit in RNA-protein recognition. *Nat. Struct. Biol. 7*, 834−837.
(28) Turner, B., Melcher, S. E., Wilson, T. J., Norman, D. G., and Lilley, D. M. (2005) Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. *RNA 11*, 1192−1200.
(29) Qin, F., Chen, Y., Wu, M., Li, Y., Zhang, J., and Chen, H. F. (2010) Induced fit or conformational selection for RNA/U1A folding. *RNA 16*, 1053−1061.
(30) Sternberg, S. H., LaFrance, B., Kaplan, M., and Doudna, J. A. (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature 527*, 110−113.
(31) Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature 507*, 62−67.
(32) Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., and Doudna, J. A. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science 343*, 1247997.

DOI: 10.1021/acs.biochem.8b00107
Biochemistry 2018, 57, 3027−3031

Nucleic Acids Research, 2018 **1**
doi: 10.1093/nar/gky1214

Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity

Daniel O'Reilly[1,†], Zachary J. Kartje[2,†], Eman A. Ageely[2], Elise Malek-Adamian[1], Maryam Habibian[1], Annabelle Schofield[1], Christopher L. Barkau[3], Kushal J. Rohilla[3], Lauren B. DeRossett[2], Austin T. Weigle[2], Masad J. Damha[1,*] and Keith T. Gagnon[2,3,*]

[1]Department of Chemistry, McGill University, Montreal, Quebec H3A 0B8, Canada, [2]Department of Chemistry & Biochemistry, Southern Illinois University, Carbondale, Illinois 62901, USA and [3]Department of Biochemistry & Molecular Biology, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901, USA

Received August 14, 2018; Revised November 20, 2018; Editorial Decision November 21, 2018; Accepted November 29, 2018

ABSTRACT

CRISPR (clustered regularly interspaced short palindromic repeat) endonucleases are at the forefront of biotechnology, synthetic biology and gene editing. Methods for controlling enzyme properties promise to improve existing applications and enable new technologies. CRISPR enzymes rely on RNA cofactors to guide catalysis. Therefore, chemical modification of the guide RNA can be used to characterize structure-activity relationships within CRISPR ribonucleoprotein (RNP) enzymes and identify compatible chemistries for controlling activity. Here, we introduce chemical modifications to the sugar–phosphate backbone of *Streptococcus pyogenes* Cas9 CRISPR RNA (crRNA) to probe chemical and structural requirements. Ribose sugars that promoted or accommodated A-form helical architecture in and around the crRNA 'seed' region were tolerated best. A wider range of modifications were acceptable outside of the seed, especially *D*-2′-deoxyribose, and we exploited this property to facilitate exploration of greater chemical diversity within the seed. 2′-fluoro was the most compatible modification whereas bulkier *O*-methyl sugar modifications were less tolerated. Activity trends could be rationalized for selected crRNAs using RNP stability and DNA target binding experiments. Cas9 activity *in vitro* tolerated most chemical modifications at predicted 2′-hydroxyl contact positions, whereas editing activity in cells was much less tolerant. The biochemical principles of chemical modification identified here will guide CRISPR-Cas9 engineering and enable new or improved applications.

INTRODUCTION

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) is an adaptive immune defence system that evolved to combat foreign invading nucleic acids in bacteria and archaea (1–6). Their core enzymatic component typically comprises a CRISPR-associated (Cas) endonuclease bound to CRISPR RNA (crRNA) cofactors that guide sequence-specific binding and subsequent phosphodiester bond cleavage of double-stranded DNA (7,8). A well-characterized and prototypical system is CRISPR-Cas9 from *Streptococcus pyogenes* (8,9). The *S. pyogenes* crRNA is complementary to target DNA at its 5′ guide and base-pairs with a *trans*-activating crRNA (tracrRNA) via its 3′ repeat sequence. The tracrRNA uses 3′ stem-loops to bind Cas9 while the crRNA makes several direct Cas9 contacts, including six predicted 2′-hydroxyl contacts (Figure 1A) (7,10,11). The *S. pyogenes* Cas9 can also use an artificial single-guide RNA (sgRNA) where the crRNA and tracrRNA are artificially fused together (8,12,13).

CRISPR systems are being actively co-opted for biotechnology, therapeutics and synthetic biology (14,15) and further development will depend on a more detailed understanding of mechanistic principles (7,16). Chemical modification of RNA is an effective method to probe mechanism and RNA–protein relationships within ribonucleoprotein (RNP) enzymes (17). Ribose modifications can characterize the contributions of RNA to helical structure, sterics, flexibility, sugar pucker and 2′-hydroxyl requirements (18). In addition, chemical modifications can impart nuclease resistance, improved bioavailability and increased target

*To whom correspondence should be addressed. Tel: +1 618 453 3895; Fax: +1 618 453 6440; Email: ktgagnon@siu.edu
Correspondence may also be addressed to Masad J. Damha. Tel: +1 514 398 7552; Fax: +1 514 398 3797; Email: masad.damha@mcgill.ca
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

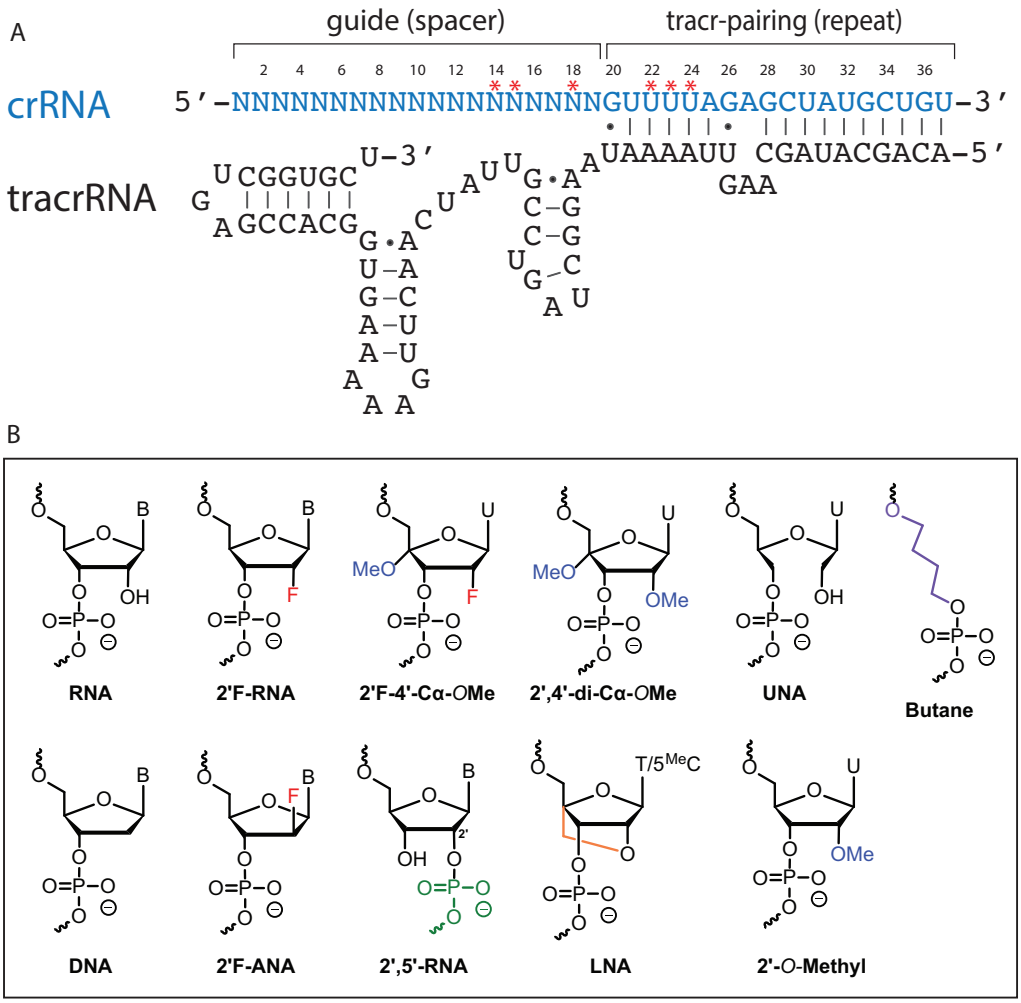


Figure 1. Dual RNA-guided Cas9 and sugar–phosphate chemical modifications. (**A**) Sequence and secondary structure of a dual RNA guide for *Streptococcus pyogenes* Cas9. Asterisks indicate structure-predicted 2′-OH contacts with Cas9. (**B**) Chemically modified nucleotides and linkers used in this study.

affinity (19,20). Ribose modification has been the subject of efforts to enhance the binding, stability and function of nucleic acids in antisense oligonucleotide (AON), RNA interference (RNAi) and aptamer technologies. Chemical modifications at the 2′-position of the sugar ring have been indispensable for transitioning nucleic acids into FDA-approved drugs (21,22). It is likely that CRISPR-based technologies will also benefit from chemical modification of RNA (14,23–25).

Several studies have focused on chemical modification of crRNA, tracrRNA and sgRNA for their compatibility with gene editing applications (25–30). However, these studies primarily relied on a limited set of commercially available modifications to maintain gene editing, which involves multiple steps and poorly understood mechanisms inside of cells. The application-driven nature of these studies has made it difficult to extract guidelines based on enzyme structure or intrinsic biochemical mechanisms. For example, it has been concluded by several studies that RNA cannot be removed at certain positions, particularly the 'seed' region or those that make 2′-hydroxyl contacts with Cas9 (27–29,31). However, we previously found that Cas9 biochemical cleavage activity was unaffected when all of the RNA residues at these critical positions were simultaneously substituted with DNA (32). To better understand the underlying rules governing crRNA modification, we incorporated a broader and more diverse set of 2′-modified DNA and RNA analogues in crRNA and characterized their impact on Cas9's cleavage activity. We sought to specifically address the biochemical requirement of RNA by investigating the role of A-form structure, conformational flexibility and steric constraints.

Here, we identify guidelines for modification of crRNA that are compatible with Cas9 biochemical activity. We propose that the seed region of crRNA guides prefer an A-form-like architecture whereas the tracrRNA-pairing (tracr-pairing) repeat region is compatible with a broader array of modifications, including DNA, 2′F-ANA and 2′,5′-RNA. Despite strong A-form conformational preference, relatively bulky modifications such as C2′- or C4′-*O*Me groups were only tolerated in the seed when minimal replacements were made. The RNA analogue 2′F-RNA was

the most compatible modification in most positions. The high compatibility of DNA nucleotides outside of the seed region suggests that flexibility may be an important factor in conformational transitions of Cas9 (7,33,34). Chemical modification at positions of predicted Cas9–2′-hydroxyl contacts reduced efficient gene editing in cells despite efficient cleavage activity *in vitro*. These results help elucidate the functional role of RNA and provide more rational guidelines for chemical modification of CRISPR-Cas9 for synthetic biology, biotechnology and future therapeutic applications.

MATERIALS AND METHODS

Chemically modified oligonucleotide synthesis

Standard phosphoramidite solid-phase synthesis conditions were used for the synthesis of all modified and unmodified oligonucleotides. Syntheses were performed on an Applied Biosystems 3400 or Expedite DNA Synthesizer at a 1 micromole scale using Unylink CPG support (ChemGenes). All phosphoramidites were prepared as 0.15 M solutions in acetonitrile (ACN), except DNA, which was prepared as 0.1 M solutions. 5-Ethylthiotetrazole (0.25 M in ACN) was used to activate phosphoramidites for coupling. Detritylations were accomplished with 3% trichloroacetic acid in CH_2Cl_2 for 110 s. Capping of failure sequences was achieved with acetic anhydride in tetrahydrofuran (THF) and 16% N-methylimidazole in THF. Oxidation was done using 0.1M I_2 in 1:2:10 pyridine:water:THF. Coupling times were 10–15 min for RNA, 2′F-ANA, 2′F-RNA, 2′F,4′*O*Me-RNA, 2′,4′-di*O*Me-RNA and LNA phosphoramidites. Mixed sugar modifications were prepared using premixed 1:1 equivalents of RNA 2′-amidite with RNA 3′-amidite, or 1:1 equivalents of RNA 2′-amidite or DNA 3′-amidite, or 1:1 equivalents of RNA 2′-amidite with 2′F-RNA 3′-amidite. This 1:1 ratio results in ∼0.77:1 incorporation of 2′-amidite to 3′-amidite. Deprotection and cleavage from the solid support was accomplished with either 3:1 NH_4OH:EtOH for 48 h at room temperature (RT), or at 55°C for 16 h. Oligonucleotides containing RNA were synthesized with standard 2′-TBDMS phosphoramidites, and desilylation was achieved with either neat triethylamine trihydrofluoride for 48 h at RT, or with triethylamine trihydrofluoride/N-methyl pyrrolidone/triethylamine (1.5:0.75:1 by volume) for 2.5 h at 65°C. Crude oligonucleotides were purified by anion exchange HPLC on an Agilent 1200 Series Instrument using a Protein-Pak DEAE 5PW column (7.5 × 75 mm) at a flow rate of 1 ml/min. The gradient was 0–24% of 1 M $LiClO_4$ over 30 min at 60°C. Samples were desalted on NAP-25 desalting columns according to manufacturer protocol. Modified crRNAs were prepared for RNP assembly by heating to 95°C then placing on ice to prevent formation of stable secondary structures.

RNA and RNA–DNA oligonucleotide synthesis

DNA oligonucleotides and DNA–RNA chimeric oligonucleotides were synthesized by Integrated DNA Technologies (IDT). Chimeric crRNAs were prepared for RNP assembly by heating to 95°C then placing on ice to prevent formation of stable secondary structures. Cas9 tracrRNAs were prepared by T7 RNA polymerase *in vitro* transcription with DNA templates synthesized by IDT. Single-stranded DNA templates were annealed to T7 promoter oligo to generate double-stranded promoter regions, which support *in vitro* transcription by T7 RNA polymerase. *In vitro* transcriptions were performed by standard protocols for 2 h. Briefly, reactions contained purified T7 RNA polymerase (4 μM), 30 mM Tris (at pH 7.9), 12.5 mM NaCl, 40 mM MgCl2, 2% PEG 8000, 0.05% Triton X-100, 2 mM spermidine and 2.5 μM T7-DNA template. Afterward, the DNA template was degraded by DNase I treatment. Reactions were phenol–chloroform extracted and gel-purified from denaturing polyacrylamide gels. Purified RNA was refolded by heating to 95°C for 5 min in a heating block followed by slow cooling the block to RT (∼40 min). RNA was quantified by measuring absorbance at 260 nm and calculated extinction coefficients using nearest neighbor approximations and Beer's Law.

Preparation of Cas9 enzymes

Plasmid encoding an *Sp*Cas9 (simply referred to as Cas9) with a C-terminal fusion of a nuclear localization signal (NLS) and a 6x-Histidine tag (pET-Cas9-NLS-6xHis) was obtained from Addgene (62933). A dead Cas9 (dCas9) version was prepared by performing site-directed mutagenesis on this plasmid to generate H840A and D10A mutations (pET-dCas9-NLS-6xHis). Cas9 proteins were prepared similar to that previously described (35). Briefly, protein expression was induced in Rosetta (DE3) cells with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18°C for 16 h. Cell pellets were resuspended in 12 ml of chilled binding buffer (20 mM Tris–HCl, pH 8.0, 250 mM NaCl, 1 mM PMSF, 5 mM imidazole) per 0.5 l of culture pellet. Resuspended cells were sonicated and clarified by centrifugation. Supernatant was purified by His-Pur Cobalt-CMA resin (Thermo Scientific) by sequentially increasing concentrations of NaCl wash buffer (Tris–HCl, pH 8, 0.25/0.5/0.75/1.0 M NaCl, 10 mM imidazole). Protein was eluted with 130 mM imidazole wash buffer. The eluent was concentrated and exchanged into 2x Cas9 storage buffer (40 mM HEPES-KOH, pH 7.5, 300 mM KCl, 2 mM ethylenediaminetetraacetic acid (EDTA), 2 mM DTT) then one volume of glycerol added. Concentration was determined by UV absorbance at 280 nm using a calculated extinction coefficient (120 450 M^{-1} cm^{-1}) and Beer's law.

In vitro Cas9 cleavage activity assays

In vitro cleavage assays were performed as previously described (32). Linearized plasmid target DNA (100 ng) harboring the tetracycline receptor (TR) gene or EGFP (EG) were combined with the Cas9 RNP complex (0.75 μM Cas9, 0.25 μM tracrRNA, 0.3 μM crRNA final concentration) in a 1× cleavage buffer (20 mM Tris–HCl, pH 7.5, 100 mM KCl, 5% glycerol, 1 mM DTT, 0.5 mM EDTA, 2 mM $MgCl_2$) supplemented with 0.1 mg/ml of purified yeast tRNA. The concentration of tracrRNA was purposely set as the limiting component of the RNP complex and used to predict final RNP concentration. Molar excess of Cas9 and

crRNA will ensure complete assembly of tracrRNA into RNP complexes and does not result in aggregation at concentrations used in our assays (32). Standard reaction conditions were 37°C for 2 h in a final reaction volume of 40 μl. The mixture was treated with 10 μg of RNase A (Thermo Scientific) for 15 min followed by 20 μg of Proteinase K (Thermo Scientific) for 15 min at room temperature. The DNA products were precipitated in 10 volumes of acetone with 2% $LiClO_4$ at −20°C for >1 h. DNA cleavage products were resolved by agarose electrophoresis and visualized using ethidium bromide staining. The fraction of target plasmid cleaved was quantified using ImageJ software. The band intensity for the cleavage product band was divided by the combined intensity of the largest cleavage product and uncut substrate plasmid bands and reported as fraction cleaved (i.e. 'cut'/'cut + uncut'). Error bars for all quantified data represent experimental replicates, not technical replicates. Sample size was selected based on the expectation that three or more replicates will be representative of typical *in vitro* assay conditions.

Radiolabeling of DNA target

A total of 100 pmols of antisense DNA target strand was radiolabeled with [γ- 32P]-ATP using T4 polynucleotide kinase following the manufacturer's recommended enzyme protocol (Thermo Fisher). Reactions were phenol-chloroform extracted and radiolabeled DNA was gel-purified on 15% denaturing polyacrylamide gels by the crush-and-soak method. Gel-purified radiolabeled RNA and DNA was quantified by scintillation counting.

Dot-blot filter binding assays for duplex target binding

For target binding by Cas9 RNP complexes, radiolabeled duplex target DNA (1000 cpm/reaction) was combined with increasing concentrations of a pre-assembled dCas9–tracrRNA–crRNA complex in a final reaction of 20 μl 1× cleavage buffer and 0.1 mg/ml tRNA. After incubation at 37°C for 15 min, reactions were vacuum filtered over nitrocellulose membrane (Protran Premium NC, Amersham) using a 96-well dot blot apparatus. Wells were washed twice with 200 μl of 1× cleavage buffer. Membrane was then removed and washed with 1× PBS solution three times and air dried at RT. Binding of radiolabeled DNA was then visualized by GE Typhoon phosphorimager. Spots were quantified with ImageQuant software, values plotted in Prism (GraphPad) and data fit to a one-site binding hyperbola equation.

Thermal denaturation monitored by UV absorbance

Thermal denaturation was as previously described (32). Cas9 alone or complexed to tracrRNA and crRNA at 1 μM final concentration (equimolar concentrations of all components) was incubated at room temperature for 10 min in degassed 1× UV melt buffer (20 mM Cacodylate, pH 7.5, 150 mM KCl, 1 mM $MgCl_2$). Samples were melted in a Cary 400 UV/Vis spectrophotometer at a ramp rate of 1°C/min while UV absorbance at 280 nm was collected every 1 min. Experiments were repeated in duplicate. Melting temperatures were determined using Van't Hoff calculations and error determined by standard error of the mean using two experimental replicates for each sample. Melt data was plotted using Prism (GraphPad) software.

Cell-based editing measured by flow cytometry

HEK 293T cells expressing EGFP and *Sp*Cas9 were a kind gift from Wen Xue (UMass Medical Center) (29). Cells were grown in Dulbecco's modified eagle's medium (DMEM) with 1× non-essential amino acids (NEAA), 5% Cosmic calf serum (CCS) and 2.5% fetal bovine serum (FBS) without antibiotics. Cells were reverse transfected (40–50 000 cells) in four experimental replicates in 96-well plates with 20 pmols of crRNA:tracrRNA complex in a final of 200 μl using RNAiMAX (Invitrogen) following the manufacturer's recommended protocol. After 12 h, one volume of media containing 15% FBS and 1× Penicillin-Streptomycin solution was added to the Opti-MEM and cells incubated for an additional 12 h. Media was then replaced with full media and cells grown for an additional 4 days.

For flow cytometry, cells were washed with PBS, trypsinized, washed again and then fixed with 1% paraformaldehyde in PBS for 8 min. Cells were washed again and counted in an Accuri C6 Flow Cytometer. EGFP was detected using the blue laser at excitation 488 nm; emission detection 530 ± 15 nm (FL1 channel). At least 20 000 events were collected and analyzed by Accuri CFlow Plus software. The cells were first gated based on forward and side scattering (FSC-A/SSC-A) to remove cell debris, then gated to select single cells (FSC-H/FSC-A). At last, cells were gated to select EGFP positive cells. The quadrant gate was established using the signal from non-EGFP expressing control cells. Untreated HEK 293T cells expressing EGFP and Cas9 contained ∼6% non-fluorescent cells. The average from four replicates was used for background subtraction to determine the extent of cell-based editing after treatment.

RESULTS

We used a dual RNA-guided (separate crRNA and tracrRNA) *S. pyogenes* CRISPR-Cas9 to probe the role of the guide RNA in enzyme activity (Figure 1A). As opposed to an artificial single-guide RNA (sgRNA) (8), dual RNA-guided Cas9 reflects the naturally occurring enzyme (36,37). Given their shorter length, dual guides are also more amenable to chemical modification (28). To facilitate more efficient crRNA chemical synthesis, we truncated a crRNA by 2 nt at the 5′ end and 4 nt at the 3′ end to yield a 36 nt crRNA (28). For this study, we have selected several modified nucleotides with properties that mimic or antagonize the properties of RNA (Figure 1B). To isolate and compare the effects of these modifications, we initially used a guide sequence targeting the tetracycline receptor (TR) gene and quantified site-specific *in vitro* cleavage of a linearized plasmid (Supplementary Figure S1) (32).

Complete crRNA modification can support target DNA cleavage activity

To establish a baseline for modification effects, we completely modified a native crRNA (crTR1) to 2′F-RNA (crTR2), a modification that strongly mimics RNA sugar pucker but cannot donate hydrogen bonds (38) (Figure 2). Enzyme activity was very high, comparable to crTR1, suggesting that strong A-form conformation and small chemical changes are preferred whereas hydrogen bonding may be less important. We also synthesized a control crRNA composed entirely of DNA (crTR3). Consistent with our earlier report, crTR3 did not support target DNA cleavage (32). A DNA–DNA hybrid with the target would adopt a B-form helical structure (39) not easily accommodated by Cas9 (10,11).

2′,5′-RNA is an RNA mimic where 3′,5′-phosphodiester linkages are replaced by the regioisomeric 2′,5′ linkage; it binds to complementary native RNA with good affinity, and incorporation of a few 2′,5′ linkages into an otherwise unmodified RNA strand has a minor impact on duplex stability (40). The 2′,5′-RNA modification is known to induce 'A-like' structure on oligonucleotides and sequences containing this modification prefer to hybridize to RNA over single-stranded DNA (ssDNA) (41,42). Thus, 2′,5′ linkages should serve as an alternative A-form structure probe. Complete conversion of a crRNA to 2′,5′-RNA abrogated enzyme activity (crTR4), possibly reflecting its poor affinity towards single-stranded DNA. To mitigate overpowering effects of 2′,5′ modifications at individual positions and generate a more moderate probe, we incorporated a mixture of 2′,5′ and 3′,5′ linkages at each position. Recent high-resolution crystallographic data on mixed-backbone RNA duplexes show that RNA duplexes containing a few 2′,5′ linkages share the same global A-like structure as the native duplex (41,43). Remarkably, RNA helical structures are well retained with 40% backbone heterogeneity (43). The result is a mixed population of crRNA regioisomers with the same base sequence. The resulting crRNA crTR5 consisted of a statistical mixture of 2^{35} (or 3.44×10^{10}) isomers. A similar isomeric effect is observed with phosphorothioate linkages, which has not been a major hurdle for their use in development of FDA-approved drugs like Nusinersen (44). Incorporating 3′,5′ linkages into the inactive 2′,5′-crRNA strand significantly restored activity (crTR4 versus crTR5), albeit not to the level of a native 3′,5′ crRNA (crTR1). Combining 2′,5′-RNA with 2′F-RNA in a mixed configuration (crTR6) also provided substantial activity, also supporting the A-form-like structural accommodation of 2′,5′ linkages.

We and others have previously found that Cas9 enzymes can accept partial RNA–DNA chimeric crRNAs (29,31,32). In particular, converting the entire tracr-pairing region to DNA significantly enhanced cleavage activity and maintained or enhanced specificity (32). Thus, we rationalized that synergy might be achieved by taking into consideration the modular nature of the crRNA. We modified the guide and tracr-pairing regions independently and confirmed that DNA in the tracr-pairing region was highly compatible with our target sequence (crTR7) (32). DNA in this region will pair with RNA in the tracrRNA and likely maintain an A-form architecture while increasing conformational freedom (32). While keeping DNA in the tracr-pairing region, we then substituted 2′F-RNA (crTR8), 2′,5′-RNA (crTR9), RNA/2′,5′-RNA mixed nucleotide modification (crTR10) and 2′F-ANA (crTR11) into the guide region. The 2′F-RNA configuration was very compatible, supporting high activity. The 2′,5′-RNA and RNA/2′,5′-RNA mixed nucleotides exhibited little or no activity.

The guide/seed region strongly prefers A-form helical structure

2′F-ANA (2′-deoxy-2′-fluoro-arabinonucleic acid), the 2′ epimer of 2′F-RNA, is known to acquire a C2′/*O4′-endo* sugar pucker and prefer a B-form helix (45,46), which results in more DNA-like properties but with less flexibility. The 2′F-ANA configuration (crTR11) was inactive. These results suggested that the guide region is more reliant on A-form helical structure than the tracr-pairing region. This is not unexpected since it will pair with a DNA target and must assume an A-form-like conformation for efficient catalysis (10,11). We therefore tested B-form conformational preference in the tracr-pairing region by substituting in 2′F-ANA (crTR12). This modification configuration was also completely inactive. Since 2′F-ANA lacks the conformational flexibility of DNA (45), it is likely that 2′F-ANA in certain tracr-pairing positions is incompatible due to inflexibility. In support of this possibility, we found that a crRNA with a 2′F-RNA guide and a 2′F-ANA tracr-pairing region could support high cleavage activity when the first 4 nt of the tracr-pairing region were DNA residues (crTR13).

Within the crRNA guide region is a 'seed' sequence where crRNA pairing to a target DNA is nucleated (7). Previous investigations reported that seed sequence residues cannot be fully modified and specific positions require RNA (27–29,31). Likewise, crystal structures have predicted crRNA 2′-hydroxyl contacts with Cas9 (10,11). Arguing against a strict requirement for 2′-hydroxyl contacts, we previously showed that all predicted contacts could be simultaneously converted to DNA in an otherwise native crRNA with no effect on biochemical activity (32). To probe the seed sequence, we substituted eight contiguous seed residues while maintaining RNA in the rest of the crRNA. DNA (crTR14), as well as a mixed RNA/2′,5′-RNA strand (crTR15), provided high activity, indicating that predicted 2′ polar contacts are of limited importance. However, complete conversion of the seed to 2′,5′-RNA (crTR16) or 2′F-ANA (crTR17) resulted in little or no activity. To determine the importance of flanking nucleotide conformation on the seed region, we started with 2′,5′-RNA or 2′F-ANA in the seed and then substituted flanking regions with various modifications (crTR18-crTR24). None of these combinations supported activity, suggesting an inability to strongly influence seed structure. Therefore, the seed region appears to strongly prefer A-form helical structure and is not significantly impacted by flanking modifications. An exception may be observed for conformationally flexible nucleotides like DNA (crTR14).

Based on the above results, we tested a variety of 'altimers' where two or more modifications are alternated in a pattern (47). Altimers can exploit unique synergy between

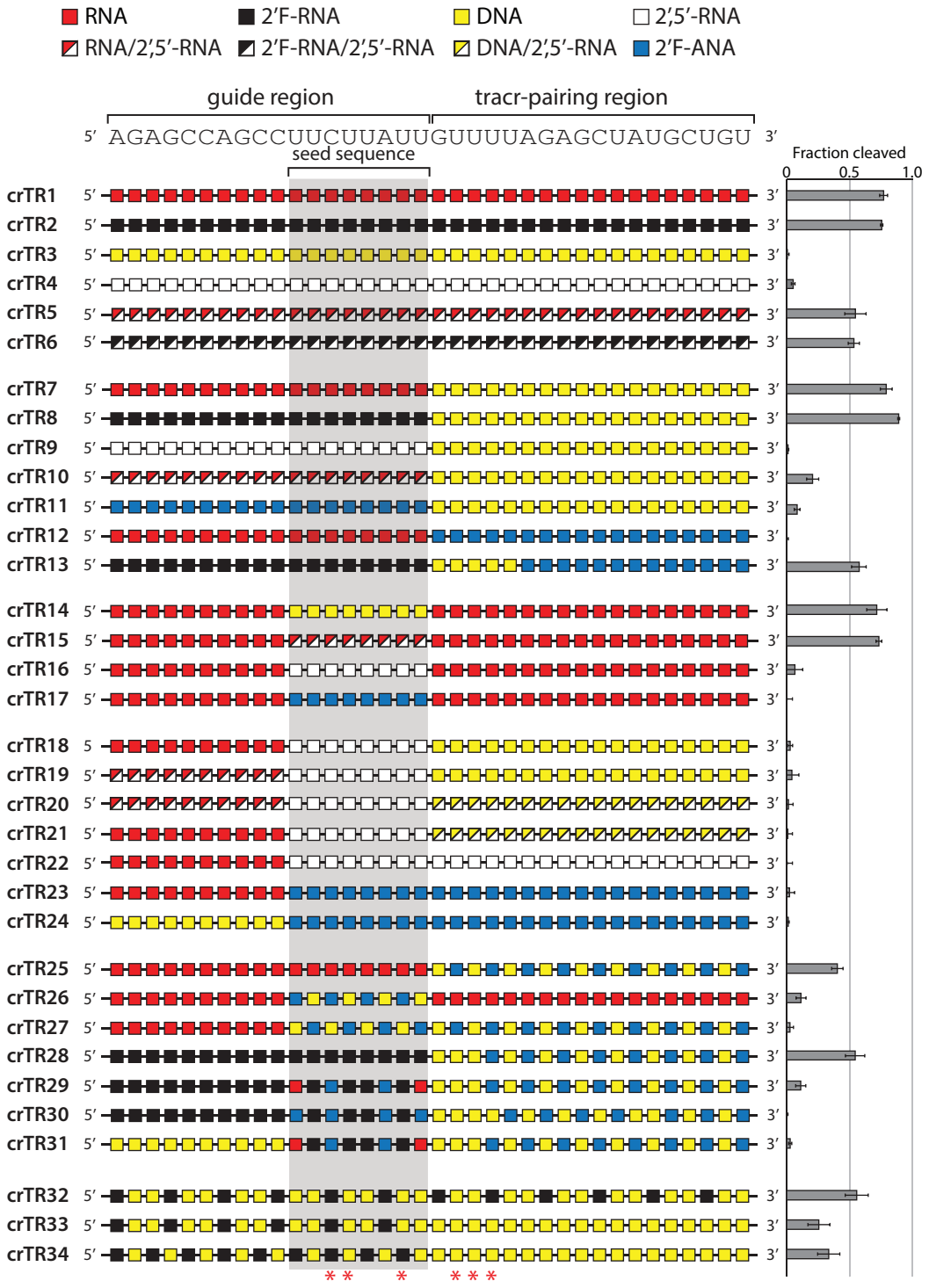


Figure 2. Modification schemes and the corresponding Cas9 enzyme activity for chemically modified crRNAs. Entire crRNAs or large sequence tracts in the guide, seed or tracrRNA-pairing region are modified. crRNA sequence is shown above and structure-predicted 2′-OH contacts with Cas9 are indicated with asterisks below. Enzyme activity is reported to the right. Error is reported as standard error of the mean (s.e.m.) for three or more replicates.

adjacent modifications and modulate effects. Indeed, altimers of DNA and 2′F-ANA in the tracr-pairing region supported high activity when coupled with RNA (crTR25) or 2′F-RNA (crTR28) in the guide region. The seed region was not compatible with altimer configurations using 2′F-ANA (crTR29-crTR31). However, altimers of 2′F-RNA and DNA across the seed region were generally tolerated (crTR32-crTR34), further supporting a predominant role for A-form conformational effects and the benefit of conformational flexibility (32).

Our previous study characterizing RNA–DNA chimeric crRNAs found that RNA residues in the seed region alone, with DNA in all other positions, was sufficient to support robust *in vitro* enzyme activity by Cas9 (32). We confirmed this effect (crTR35) and used the high compatibility of DNA outside of the seed region to further probe the structural and conformational requirements of the seed region (Figure 3). In this configuration, there is a lack of strong interference from flanking sequence because of the relative conformational neutrality of DNA. Once again, we found that 2′F-RNA was a suitable replacement for RNA (crTR36). As well, 2′,5′-RNA (crTR38) or 2′-FANA (crTR39) in the seed region showed little activity. Mixed linkages (2′,5′/3′,5′-RNA) in the seed were more efficient (crTR37) than 2′,5′-RNA, reinforcing our earlier observation that RNA can help mitigate the structural effects of 2′,5′-linkages.

We wanted to explore more diverse modifications, some of which are only available as uridines. Therefore, we modified the six uridines in the 8-nt seed region while keeping all other positions DNA. With only six uridine nucleotides to modify, we explored replacement of RNA (crTR40) with 2′F-RNA (crTR41), 2′,5′-RNA (crTR42), 2′F-ANA (crTR43), DNA (crTR44) (to distinguish thymine versus uracil effects), 2′,4′-di-Cα-*O*Me RNA (crTR45) and 2′F-4′-Cα-*O*Me RNA (crTR46) (48–51). Other than 2′F-RNA, and to a small degree 2′,5′-RNA, none of these modifications supported activity. This was surprising given the RNA-like (C3′-*endo*) sugar pucker previously reported for 2′,4′-di-Cα-*O*Me and 2′F-4′-Cα-*O*Me (48,50,51). Multiple 4′-*O*-methyl groups may introduce unfavorable steric constraints due to their bulkiness relative to a hydrogen atom.

Nearest neighbor effects can also modulate the conformation or properties of adjacent nucleotides. For example, DNA is easily induced into A-form-like conformation by neighboring RNA nucleotides (52,53). Likewise, juxtaposing 2′F-ANA and 2′F-RNA next to each other can relax the strong RNA-like properties of 2′F-RNA (54). To determine whether nearest neighbor effects could make modifications work synergistically, we combined RNA with 2′F-ANA in the seed (crTR47 and crTR48). RNA was unable to improve activity for 2′F-ANA in the seed. Combining RNA with 2′,5′-RNA (crTR49), 2′,4′-di-Cα-*O*Me (crTR50), or 2′F-4′-Cα-*O*Me (crTR51) also resulted in no activity. Thus, RNA was unable to offset the unfavorable effects of these modifications. In contrast, combining RNA with 2′F-RNA (crTR52 and crTR53) and LNA (crTR54), a 2′-4′ bridging modification that locks C3′-*endo* sugar pucker (55), resulted in robust enzyme activity. Combining 2′F-RNA (with and without RNA) with 2′F-ANA was unable to restore activity (crTR55 and crTR56). Attempting to increase conformational freedom by placing a UNA nucleotide (56) (crTR57) or butane linker (crTR58) at the very 3′ nucleotide of the seed region did not restore activity.

2′F-ANA was well tolerated in the tracr-pairing region, as long as a few of the first 5′ nucleotides were DNA (crTR13 and crTR59). However, placement of 2′F-ANA at the second position (along with the fifth and eleventh positions) in the tracr-pairing region (crTR60) abrogated activity. Combined with the activity of 2′F-ANA-containing crRNAs crTR12, crTR13, crTR25, crTR27 and crTR28, this result suggests that one or more positions closest to the seed must assume A-form structure, either through conformational predisposition or flexibility. DNA may temper the formation of A/B junctions or rigid versus flexible A-form transitions between the seed and tracr-pairing region. Placement of 2′F-ANA outside of the seed but still in the guide region (crTR61 and crTR62) was also detrimental to activity. Conversely, placing 2′F-RNA in the same guide and tracr-pairing positions resulted in robust activity (crTR63 and crTR64). These results reinforce a strong preference for A-form helical structure in the guide region, especially the seed sequence and a greater tolerance for modification in most of the tracr-pairing region. This difference may derive from hybridization of the crRNA 5′ guide region to a DNA target versus pairing of the 3′ region to a tracrRNA partner.

To further probe the role of RNA in the seed region, we replaced two uridines that are predicted to mediate 2′-hydroxyl contacts (10,11) while maintaining RNA in all other positions (crTR65-crTR70). Replacement with 2′F-RNA was the most compatible (crTR65), while DNA (crTR66), 2′F-4′-Cα-*O*Me (crTR67), 2′,4′-di-Cα-*O*Me (crTR68) and even 2′F-ANA (crTR69) were well-tolerated. The incorporation of 2′-*O*-methyl, a known RNA conformational mimic (57), was only moderately effective, reinforcing steric issues with bulky modifications at the 2′ position. We selected only uridines for replacement, so sequence or base composition may influence modification effects. However, these results show that judicious placement in a few seed positions can make modifications tolerable, even in sensitive 2′-hydroxyl contact positions.

RNP stability, target DNA engagement and a second guide sequence support crRNA activity trends

To rationalize the effect of modifications, we assembled Cas9 RNP complexes and determined their thermal stability (Figure 4A) and their binding to a double-stranded DNA target (Supplementary Figure S2 and Figure 4B). Thermal stability was measured by absorbance at 280 nm as temperature was increased (32). Cas9 alone exhibited a melting temperature (T_m) of 44°C. Upon assembly with a tracrRNA, the T_m increased to 46.6°C. Higher order assembly with native crRNA (crTR1) resulted in an additional increase in the T_m, up to 49.5°C. We selected a handful of representative crRNAs to determine their impact on RNP stability. Two crRNAs that provided high enzyme activity, crTR2 and crTR35, fell into a high T_m regime like native crRNA. One exception was crTR3, composed entirely of DNA. This modified oligonucleotide seems to assemble stable RNP complexes but does not support catalysis (32). crRNAs that fell into an intermediate T_m regime generally

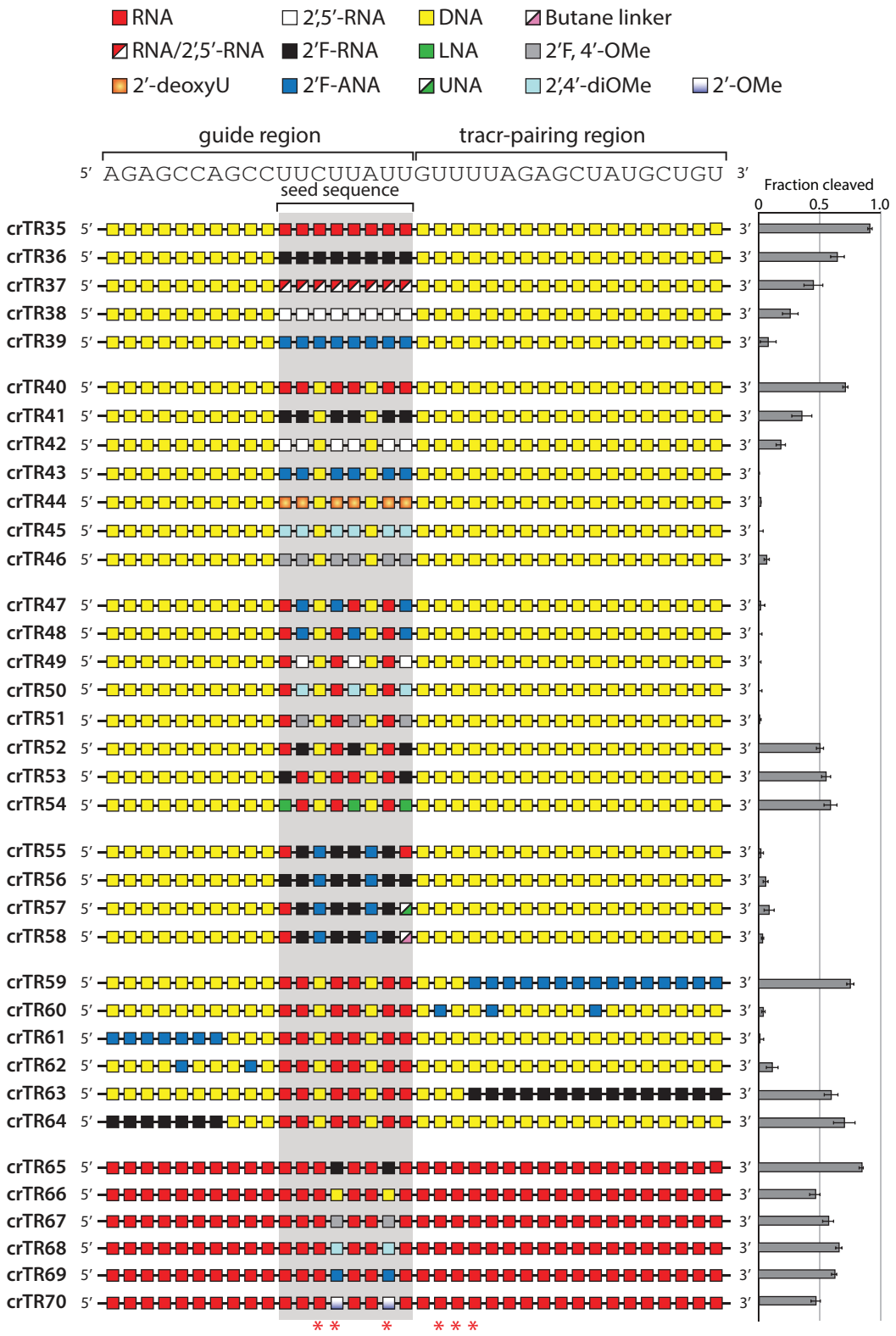


Figure 3. Modification schemes investigating effects on Cas9 enzyme activity primarily at the seed region. crRNA sequence is shown above and structure-predicted 2′-OH contacts with Cas9 are indicated with asterisks below. Enzyme activity is reported to the right. Error is reported as s.e.m for three or more replicates.

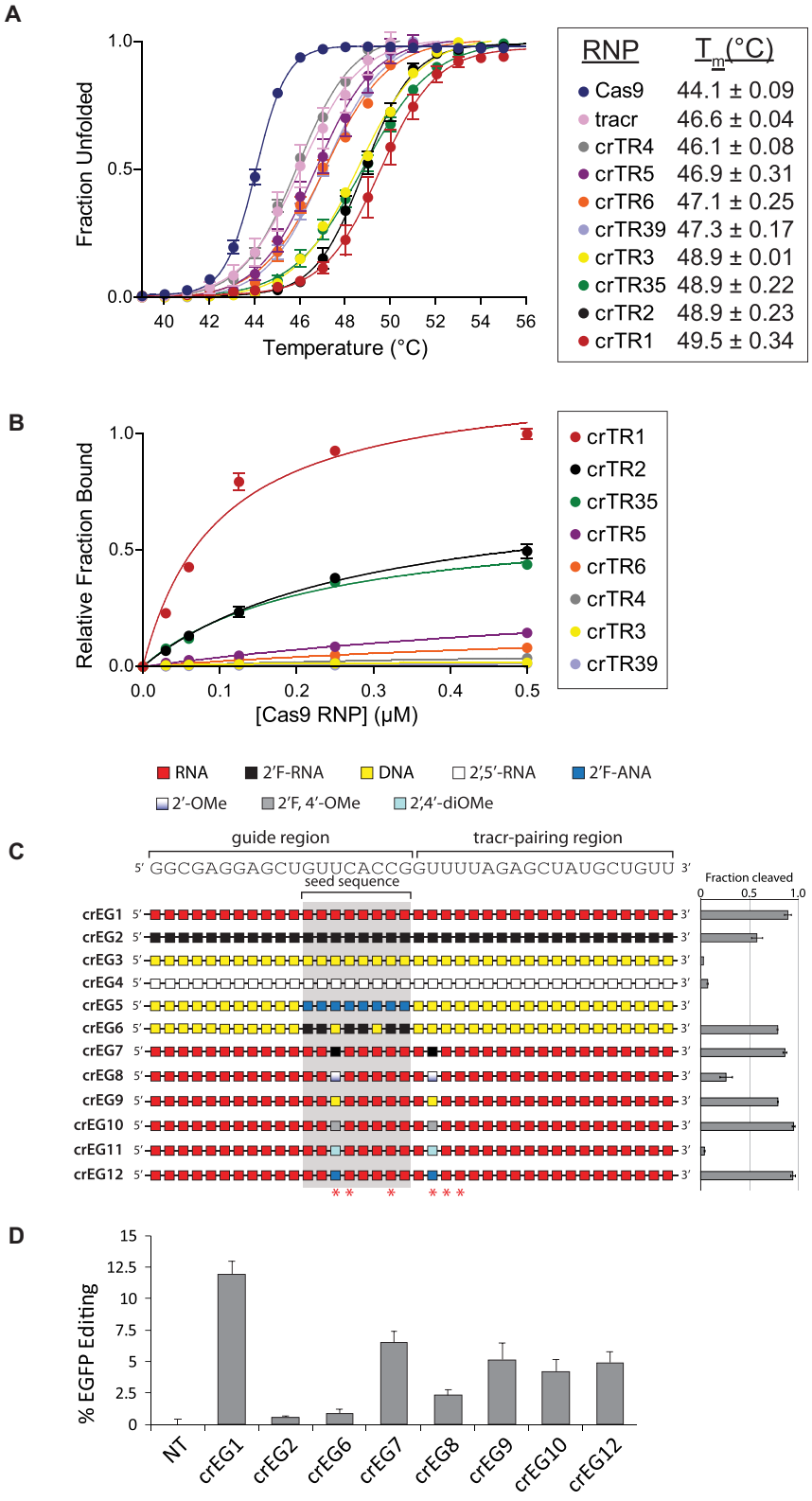


Figure 4. Thermal denaturation, target binding and a second guide sequence help rationalize crRNA activity trends. (**A**) RNP thermal denaturation monitored by UV absorbance at 280 nm. Error is reported as s.e.m. for two replicates. (**B**) Substrate engagement measured by dot-blot filter binding of radiolabeled target DNA. Error is reported as s.e.m. for three or more replicates. (**C**) *In vitro* cleavage activity of Cas9 RNPs assembled with chemically modified crRNAs targeting an EGFP gene. crRNA sequence is shown above and structure-predicted 2′-OH contacts with Cas9 are indicated with asterisks below. Enzyme activity is reported to the right. Error is reported as s.e.m. for three or more replicates. (**D**) Cell-based editing activity of crRNAs co-transfected with tracrRNA into HEK 293T cells stably expressing EGFP and Cas9. Gene editing efficiency was measured as loss of EGFP fluorescence by flow cytometry 5 days post-transfection. Error is reported as s.e.m. for four replicate treatments.

supported moderate activity (crTR5 and crTR6). However, crTR39, with 2′F-ANA in the seed, also fell into the intermediate T_m range. This result and the T_m of crTR3 illustrate the modular nature of crRNAs where tracrRNA-pairing might support assembly, yet the guide cannot support catalysis. Only crTR4, composed entirely of 2′-5′ linkages, did not increase RNP T_m. This crRNA may be unable to efficiently assemble a stable RNP complex with Cas9-tracrRNA, explaining its lack of activity.

The ability of these chemically modified crRNAs to stably engage target DNA mirrored the observations from RNP stability (Figure 4B and Supplementary Figure S2). Native crRNA (crTR1) facilitated strong equilibrium binding of Cas9 RNP to target double-stranded DNA. Two crRNAs with high activity and stable RNP assembly, crTR2 and crTR35, exhibited moderate but stable target binding. crRNAs with reduced enzyme activity (crTR3 and crTR4) had lower target binding and those that possessed no cleavage activity did not bind the target. The all-DNA crTR3, while able to assemble a stable RNP complex, was unable to engage target DNA, as previously reported (32). Thus, the effects of chemical modification can be correlated with RNP assembly and target substrate engagement.

To determine if some of the general principles we have observed could be extended to another sequence, we selected a few modification schemes and a new guide sequence that targets an EGFP gene. We performed *in vitro* cleavage assays as before but using a linearized plasmid harboring the EGFP gene as substrate (Figure 4C). A native crRNA targeting EGFP (crEG1) supported about 90% cleavage of the target plasmid. Complete modification with 2′F-RNA (crEG2) resulted in substantial, albeit reduced, cleavage at about 55%. As expected, complete conversion to DNA (crEG3) or 2′,5′-RNA (crEG4) resulted in loss of activity. For the crTR guide we showed that DNA outside of the seed region was compatible with cleavage activity if the seed contained modified nucleotides with appropriate RNA-like properties. This was confirmed by testing either 2′F-ANA or 2′F-RNA in the seed. The 2′F-ANA seed did not support enzyme activity (crEG5) whereas 2′F-RNA in only six positions in the seed region was well-tolerated (crEG6). These results support the overall preference for A-form like structure, especially in the seed region and the compatibility of DNA in other regions of the crRNA.

To extend some of these results to Cas9 activity inside of cells, we co-transfected tracrRNA and crRNA modified with 2′F-RNA (crEG2) or 2′F-RNA and DNA (crEG6) into HEK 293T cells stably expressing EGFP and Cas9. These modified guides target EGFP and should result in loss of EGFP expression if gene editing was successful. Editing was measured by flow cytometry five days after transfection. Despite high activity *in vitro* and the presence of the strong RNA mimic 2′F-RNA, these guides failed to produce substantial editing activity (Figure 4D). A recent report suggested that modification at certain predicted 2′-hydroxyl contacts is a bottleneck for complete modification of editing-active crRNAs (27).

To focus on guides that might reveal the impact of chemical modifications at these positions, we generated a new set of modified crRNAs (crEG7-crEG12) against the same EGFP sequence. In this case, however, only two predicted 2′-hydroxyl contact positions were modified. These modified crRNAs are similar to a subset tested *in vitro* in Figure 3 (crTR65-crTR70), but the modifications were placed at two different positions. The positions chosen contained uridine nucleotides to allow testing of a more diverse variety of chemical modifications. These two positions were also shown to be sensitive to modification in a previous report (27).

In vitro cleavage revealed that these crRNAs either supported high activity (crEG7, 9, 10 and 12), moderate activity (crEG8), or no activity (crEG11) (Figure 4C). All except crEG11 were subsequently co-transfected with tracrRNA and their ability to support editing was measured by flow cytometry. These crRNAs showed editing efficiencies that were lower than an unmodified crRNA, but quite similar to the trends observed *in vitro* (Figure 4D). These results underscore the importance of predicted 2′-hydroxyl contacts with Cas9 for maintaining high editing efficiency in cells. They also support our *in vitro* observations regarding the general compatibility of chemical modifications that prefer or can assume A-form helical structure and are not sterically bulky.

DISCUSSION

The potential use of CRISPR-Cas9 for gene editing in patients, and its possible other applications in pharmacotherapy, will likely face key hurdles not unlike other oligonucleotide-based therapeutics (19–22). Very early work with AONs and siRNAs was undertaken with unmodified, natural molecules. However, it soon became clear that native oligonucleotides were subject to relatively rapid degradation and lacked sufficient drug-like properties, such as biodistribution or cellular uptake (19,58). Although CRISPR-Cas9 is a large RNP enzyme that possesses unique challenges for drug development (14,23,24), it is anticipated that chemical modification of the guide RNA will significantly benefit certain therapeutic applications (59). Previous investigations into CRISPR-Cas9 RNA chemical modification have generally focused on screens for guide RNAs that are tolerated in gene editing applications with an emphasis on nuclease resistance or off-target effects. These have included modification of sgRNA with 2′-*O*-methyl, 2′F-RNA, and phosphorothioate (PS) or thiophosphonoacetate linkages at the termini (26) or internal residues (30), partial substitution of crRNAs with DNA (29) or 2′F-RNA, 2′-*O*-methyl, 2′-4′ bridged nucleic acid and PS linkages (28), or site-specific incorporation of 2′-*O*-methyl-3′-phosphonoacetate (25) and 2′-4′ bridged nucleic acids (60). A recent study explored extensive structure-guided modification with 2′F-RNA, 2′-*O*-methyl and PS linkages in both the crRNA and tracrRNA and found that heavy modification could still support satisfactory gene editing (27). While informative, these studies did not fully establish clear mechanistic or rational rules for guide RNA modification or correlated gene editing results with a biochemical understanding of Cas9 RNP structure–function.

In this study, we sought to establish more fundamental rules for chemical modification that correlate with the intrinsic biochemical activity of Cas9. These guidelines are likely to impact a wide range of CRISPR applica-

tions. In a previous study of crRNA structure-function, we took advantage of the excellent probing properties of 2′-deoxynucleotides. Substituting RNA with DNA at specific residues provided characterization of 2′-hydroxyl contacts, steric flexibility and A-form helical structure. Systematic substitution suggested strong preferences for A-form helical structure in the crRNA, a lack of requirement for 2′-hydroxyl contacts with Cas9, and potential regulatory features of the crRNA that could enhance Cas9 activity (32). Here we more rigorously tested our hypothesis of A-form helical structure requirements for the crRNA while evaluating the impact of conformational flexibility and steric restrictions on Cas9 activity.

We used both commercially available and custom modified nucleotides that can mimic RNA or DNA nucleotides and better control crRNA conformation, flexibility, polar contacts and steric constraints. Among the chemically modified nucleotides we tested, those possessing conformational properties similar to RNA were 2′F-RNA, 2′,5′-RNA (31–33), LNA (44), 2′,4′-di-Cα-*O*Me RNA (51), 2′F-4′-Cα-*O*Me RNA (39,40) and 2′-*O*-methyl (57). Among these, 2′F-RNA is one of the most prominent RNA substitutions in nucleic acid-based applications because of its C3′-*endo* sugar pucker that increases binding affinity to target RNA, increased endonuclease resistance, and small substituent size (38,61,62). In contrast, 2′F-ANA mimics the DNA conformation, has minimal steric constraints, possesses hydrogen bond acceptor potential and has substantial nuclease resistance (63). We found that replacing RNA bases with 2′F-RNA generally maintained an excellent degree of activity. DNA could sometimes be partially replaced with 2′F-ANA, such as in the tracr-pairing repeat region. In contrast, placing 2′F-ANA in positions that require A-form helical architecture, specifically the seed region, severely reduced activity presumably due to its inflexibility. The 2′,5′-RNA nucleotides were largely detrimental when replacing RNA. However, when 2′,5′-RNA was used to replace RNA or 2′F-RNA nucleotides in a mixed nucleotide configuration (crTR5 and crTR6), substantial cleavage was often observed. Approximately 40% activity was retained when three of six RNA bases in the seed (in an otherwise all-DNA crRNA) were replaced with LNA bases (crTR54). Replacing too many RNA bases with ribonucleotides possessing 2′F,4′-*O*Me and 2′,4′-di-*O*Me moieties resulted in complete loss of activity. Instead, these modifications performed better when placed at only one or two select positions, as in the case of siRNAs (48). The effectiveness of DNA and 2′F-RNA suggests that guide flexibility and conformational preference is a key factor in Cas9′s ability to achieve catalytically competent conformational states.

Together, our results support a need for A-form structure throughout the crRNA. A requirement in the guide region, specifically the seed sequence, likely arises from its role in nucleation of target hybridization and the sensing of specific duplex shape by Cas9 in this region (64–66). The tracr-pairing repeat region is likely to tolerate broader modification because hybridization with tracrRNA should help induce an A-form duplex architecture. The tracr-pairing region also plays a poorly understood role in modulating Cas9 activity (32). A-form conformation and flexibility at or near the junction between the guide and tracr-pairing region may offer insight into this regulation. For example, DNA or 2′F-RNA are tolerated at or near this junction, but not 2′F-ANA or 2′,5′-RNA. Our results indicate that modified nucleotides in the crRNA guide region, especially the seed sequence, that mimic the properties of RNA will be the most successful. Overuse of modifications with bulky moieties should be avoided in the seed region. The use of multiple modification types is supported, which can take advantage of synergistic effects among modifications. DNA nucleotides, for example, can offset conformationally rigid modifications like 2′F-ANA, although both offer potentially advantageous properties when judiciously placed in the tracr-pairing region.

Previous reports have suggested a difficulty in replacing predicted Cas9–2′-hydroxyl contacts with chemically modified nucleotides when trying to preserve gene editing activity in cells (27,28,31). Our investigations support this phenomenon. Although two crRNAs lacking any RNA residues (crEG2 and crEG6) performed very well *in vitro*, they supported very little gene editing inside of cells. Specifically focusing chemical modification on two 2′-hydroxyl contact positions in the crRNA (crEG7-crEG12) revealed a clear reduction in editing efficiency for all modification types tested. Thus, our results further emphasize a disconnect between the intrinsic biochemical activity of Cas9 and its ability to induce gene editing in cells that centers on a handful of 2′-hydroxyl contacts with the guide RNA. While several mechanistic causes are possible, the difference in substrate complexity between *in vitro* (plasmid DNA) and cell-based editing (chromatinized DNA) experiments is evident. Characterizing the source of discrepancy between *in vitro* and cell-based activity should provide a path forward for the generation of CRISPR-Cas9 enzymes that can utilize completely modified guides with high efficiency for gene editing applications in the future.

CONCLUSION

This study establishes A-form helical structure and steric constraint as the primary determinants of the crRNA structure–function relationship with Cas9. Based on this observation, we identify rational guidelines for chemical modification of the crRNA sugar–phosphate backbone that are compatible with Cas9 enzyme activity. In summary, the guide region prefers an A-form-like architecture, primarily due to nucleation of target DNA pairing at the seed. The tracr-pairing region is compatible with a broader array of modifications due to hybridization with tracrRNA and possible roles in conformational regulation of Cas9 activity. Potentially bulky modifications should be avoided in the seed sequence while a broad variety of modifications are likely to be well-tolerated at crRNA 5′ and 3′ termini (26). The combination of multiple modification types can provide flexibility in tuning Cas9 activity by combining synergistic properties. The phenomenon of apparent 2′-hydroxyl requirement at certain guide positions, which are dispensable *in vitro*, pose restrictions to highly efficient gene editing by fully modified guides (27–29,31,32). The discovery of a clear mechanistic underpinning or the identification of modifications that can compensate for the positional 2′-hydroxyl requirement should unlock this restriction and make the

guidelines established here highly relevant for rational guide modification.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Science and Engineering Council of CANADA (NSERCC) (to M.J.D); NSERCC Create (to D.O.); Southern Illinois University (SIU) Startup Funding (to K.T.G.); SIU Chemistry Department Gower Fellowship (to Z.J.K.); SIU Graduate Fellowship (to C.L.B.); United States Department of Defense Amyotrophic Lateral Sclerosis Research Program (ALSRP) Grant (to K.T.G.). Funding for open access charge: Startup Funding (to K.T.G.).
Conflict of interest statement. None declared.

REFERENCES

1. Bolotin,A., Quinquis,B., Sorokin,A. and Ehrlich,S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**, 2551–2561.
2. Brouns,S.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J., Snijders,A.P., Dickman,M.J., Makarova,K.S., Koonin,E.V. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
3. Marraffini,L.A. and Sontheimer,E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, **322**, 1843–1845.
4. Mojica,F.J., Diez-Villasenor,C., Garcia-Martinez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
5. Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
6. Garneau,J.E., Dupuis,M.E., Villion,M., Romero,D.A., Barrangou,R., Boyaval,P., Fremaux,C., Horvath,P., Magadan,A.H. and Moineau,S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, **468**, 67–71.
7. Jiang,F. and Doudna,J.A. (2017) CRISPR-Cas9 structures and mechanisms. *Annu. Rev. Biophys.*, **46**, 505–529.
8. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
9. Shmakov,S., Smargon,A., Scott,D., Cox,D., Pyzocha,N., Yan,W., Abudayyeh,O.O., Gootenberg,J.S., Makarova,K.S., Wolf,Y.I *et al.*2017) Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **15**, 169–182.
10. Jiang,F., Taylor,D.W., Chen,J.S., Kornfeld,J.E., Zhou,K., Thompson,A.J., Nogales,E. and Doudna,J.A. (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, **351**, 867–871.
11. Nishimasu,H., Ran,F.A., Hsu,P.D., Konermann,S., Shehata,S.I., Dohmae,N., Ishitani,R., Zhang,F. and Nureki,O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, **156**, 935–949.
12. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
13. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
14. Fellmann,C., Gowen,B.G., Lin,P.C., Doudna,J.A. and Corn,J.E. (2017) Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat. Rev. Drug Discov.*, **16**, 89–100.
15. Wright,A.V., Nunez,J.K. and Doudna,J.A. (2016) Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell*, **164**, 29–44.
16. Bisaria,N., Jarmoskaite,I. and Herschlag,D. (2017) Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-Based genome editing. *Cell Syst.*, **4**, 21–29.
17. Nilsen,T.W. (2015) Nucleotide analog interference mapping. *Cold Spring. Harb. Protoc.*, **2015**, 604–608.
18. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag New York Inc., NY.
19. Deleavey,G.F. and Damha,M.J. (2012) Designing chemically modified oligonucleotides for targeted gene silencing. *Chem. Biol.*, **19**, 937–954.
20. Manoharan,M. (1999) 2′-carbohydrate modifications in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation. *Biochim. Biophys. Acta*, **1489**, 117–130.
21. Stein,C.A. and Castanotto,D. (2017) FDA-approved oligonucleotide therapies in 2017. *Mol. Ther.*, **25**, 1069–1075.
22. Shen,X. and Corey,D.R. (2018) Chemistry, mechanism and clinical status of antisense oligonucleotides and duplex RNAs. *Nucleic Acids Res.*, **46**, 1584–1600.
23. Dai,W.J., Zhu,L.Y., Yan,Z.Y., Xu,Y., Wang,Q.L. and Lu,X.J. (2016) CRISPR-Cas9 for in vivo Gene Therapy: Promise and Hurdles. *Mol. Ther. Nucleic Acids*, **5**, e349.
24. Wilson,R.C. and Gilbert,L.A. (2017) The promise and challenge of in vivo delivery for genome therapeutics. *ACS Chem. Biol.*, **13**, 376–382.
25. Ryan,D.E., Taussig,D., Steinfeld,I., Phadnis,S.M., Lunstad,B.D., Singh,M., Vuong,X., Okochi,K.D., McCaffrey,R., Olesiak,M. *et al.* (2018) Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res.*, **46**, 792–803.
26. Hendel,A., Bak,R.O., Clark,J.T., Kennedy,A.B., Ryan,D.E., Roy,S., Steinfeld,I., Lunstad,B.D., Kaiser,R.J., Wilkens,A.B. *et al.* (2015) Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.*, **33**, 985–989.
27. Mir,A., Alterman,J.F., Hassler,M.R., Debacker,A.J., Hudgens,E., Echeverria,D., Brodsky,M.H., Khvorova,A., Watts,J.K. and Sontheimer,E.J. (2018) Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing. *Nat. Commun.*, **9**, 2641.
28. Rahdar,M., McMahon,M.A., Prakash,T.P., Swayze,E.E., Bennett,C.F. and Cleveland,D.W. (2015) Synthetic CRISPR RNA-Cas9-guided genome editing in human cells. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E7110–E7117.
29. Yin,H., Song,C.Q., Suresh,S., Kwan,S.Y., Wu,Q., Walsh,S., Ding,J., Bogorad,R.L., Zhu,L.J., Wolfe,S.A. *et al.* (2018) Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat. Chem. Biol.*, **14**, 311–316.
30. Yin,H., Song,C.Q., Suresh,S., Wu,Q., Walsh,S., Rhym,L.H., Mintzer,E., Bolukbasi,M.F., Zhu,L.J., Kauffman,K. *et al.* (2017) Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.*, **35**, 1179–1187.
31. Rueda,F.O., Bista,M., Newton,M.D., Goeppert,A.U., Cuomo,M.E., Gordon,E., Kroner,F., Read,J.A., Wrigley,J.D., Rueda,D. *et al.* (2017) Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nat. Commun.*, **8**, 1610.
32. Kartje,Z.J., Barkau,C.L., Rohilla,K.J., Ageely,E.A. and Gagnon,K.T. (2018) Chimeric guides probe and enhance Cas9 biochemical activity. *Biochemistry*, **57**, 3027–3031.
33. Dagdas,Y.S., Chen,J.S., Sternberg,S.H., Doudna,J.A. and Yildiz,A. (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.*, **3**, eaao0027.
34. Sternberg,S.H., LaFrance,B., Kaplan,M. and Doudna,J.A. (2015) Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature*, **527**, 110–113.
35. Anders,C. and Jinek,M. (2014) In vitro enzymology of Cas9. *Methods Enzymol.*, **546**, 1–20.
36. Lim,Y., Bak,S.Y., Sung,K., Jeong,E., Lee,S.H., Kim,J.S., Bae,S. and Kim,S.K. (2016) Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.*, **7**, 13350.
37. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J., Charpentier,E., Haft,D.H. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.

38. Patra,A., Paolillo,M., Charisse,K., Manoharan,M., Rozners,E. and Egli,M. (2012) 2′-Fluoro RNA shows increased Watson-Crick H-bonding strength and stacking relative to RNA: evidence from NMR and thermodynamic data. *Angew. Chem. Int. Ed. Engl.*, **51**, 11863–11866.
39. Kulkarni,M. and Mukherjee,A. (2017) Understanding B-DNA to A-DNA transition in the right-handed DNA helix: perspective from a local to global transition. *Prog. Biophys. Mol. Biol.*, **128**, 63–73.
40. Giannaris,P.A. and Damha,M.J. (1993) Oligoribonucleotides containing 2′,5′-phosphodiester linkages exhibit binding selectivity for 3′,5′-RNA over 3′,5′-ssDNA. *Nucleic Acids Res.*, **21**, 4742–4749.
41. Sheng,J., Li,L., Engelhart,A.E., Gan,J., Wang,J. and Szostak,J.W. (2014) Structural insights into the effects of 2′-5′ linkages on the RNA duplex. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 3050–3055.
42. Wasner,M., Arion,D., Borkow,G., Noronha,A., Uddin,A.H., Parniak,M.A. and Damha,M.J. (1998) Physicochemical and biochemical properties of 2′,5′-linked RNA and 2′,5′-RNA:3′,5′-RNA "hybrid" duplexes. *Biochemistry*, **37**, 7478–7486.
43. Shen,F., Luo,Z., Liu,H., Wang,R., Zhang,S., Gan,J. and Sheng,J. (2017) Structural insights into RNA duplexes with multiple 2-5-linkages. *Nucleic Acids Res.*, **45**, 3537–3546.
44. Aartsma-Rus,A. (2017) FDA Approval of Nusinersen for Spinal Muscular Atrophy Makes 2016 the Year of Splice Modulating Oligonucleotides. *Nucleic Acid Ther.*, **27**, 67–69.
45. Mangos,M.M., Min,K.L., Viazovkina,E., Galarneau,A., Elzagheid,M.I., Parniak,M.A. and Damha,M.J. (2003) Efficient RNase H-directed cleavage of RNA promoted by antisense DNA or 2′F-ANA constructs containing acyclic nucleotide inserts. *J. Am. Chem. Soc.*, **125**, 654–661.
46. Martin-Pintado,N., Yahyaee-Anzahaee,M., Campos-Olivas,R., Noronha,A.M., Wilds,C.J., Damha,M.J. and Gonzalez,C. (2012) The solution structure of double helical arabino nucleic acids (ANA and 2′F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res.*, **40**, 9329–9339.
47. Min,K.L., Viazovkina,E., Galarneau,A., Parniak,M.A. and Damha,M.J. (2002) Oligonucleotides comprised of alternating 2′-deoxy-2′-fluoro-beta-D-arabinonucleosides and D-2′-deoxyribonucleosides (2′F-ANA/DNA 'altimers') induce efficient RNA cleavage mediated by RNase H. *Bioorg. Med. Chem. Lett.*, **12**, 2651–2654.
48. Malek-Adamian,E., Guenther,D.C., Matsuda,S., Martinez-Montero,S., Zlatev,I., Harp,J., Burai Patrascu,M., Foster,D.J., Fakhoury,J., Perkins,L. *et al.* (2017) 4′-C-Methoxy-2′-deoxy-2′-fluoro modified ribonucleotides improve metabolic stability and elicit efficient RNAi-Mediated gene silencing. *J. Am. Chem. Soc.*, **139**, 14542–14555.
49. Burai Patrascu,M., Malek-Adamian,E., Damha,M.J. and Moitessier,N. (2017) Accurately modeling the conformational preferences of nucleosides. *J. Am. Chem. Soc.*, **139**, 13620–13623.
50. Harp,J.M., Guenther,D.C., Bisbe,A., Perkins,L., Matsuda,S., Bommineni,G.R., Zlatev,I., Foster,D.J., Taneja,N., Charisse,K. *et al.* (2018) Structural basis for the synergy of 4′- and 2′-modifications on siRNA nuclease resistance, thermal stability and RNAi activity. *Nucleic Acids Res.*, **46**, 8090–8104.
51. Malek-Adamian,E., Patrascu,M.B., Jana,S.K., Martinez-Montero,S., Moitessier,N. and Damha,M.J. (2018) Adjusting the structure of 2′-Modified nucleosides and oligonucleotides via C4′-alpha-F or C4′-alpha-OMe Substitution: Synthesis and conformational analysis. *J. Org. Chem.*, **83**, 9839–9849.
52. Salazar,M., Fedoroff,O.Y., Miller,J.M., Ribeiro,N.S. and Reid,B.R. (1993) The DNA strand in DNA.RNA hybrid duplexes is neither B-form nor A-form in solution. *Biochemistry*, **32**, 4207–4215.
53. Wahl,M.C. and Sundaralingam,M. (2000) B-form to A-form conversion by a 3′-terminal ribose: crystal structure of the chimera d(CCACTAGTG)r(G). *Nucleic Acids Res.*, **28**, 4356–4363.
54. Martin-Pintado,N., Deleavey,G.F., Portella,G., Campos-Olivas,R., Orozco,M., Damha,M.J. and Gonzalez,C. (2013) Backbone FC-H…O hydrogen bonds in 2′F-substituted nucleic acids. *Angew. Chem. Int. Ed. Engl.*, **52**, 12065–12068.
55. Koshkin,A.A., singh,S.K., Nielsen,P., Rajwanshi,V.K., Kumar,R., Meldgaard,M., Olsen,C.E. and Wengel,J. (1998) LNA (Locked Nucleic Acids): synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron*, **54**, 3607–3630.
56. Langkjaer,N., Pasternak,A. and Wengel,J. (2009) UNA (unlocked nucleic acid): a flexible RNA mimic that allows engineering of nucleic acid duplex stability. *Bioorg. Med. Chem.*, **17**, 5420–5425.
57. Lubini,P., Zurcher,W. and Egli,M. (1994) Stabilizing effects of the RNA 2′-substituent: crystal structure of an oligodeoxynucleotide duplex containing 2′-O-methylated adenosines. *Chem. Biol.*, **1**, 39–45.
58. Uhlmann,E. and Peyman,A. (1990) Antisense oligonucleotides: a new therapeutic principle. *Chem. Rev.*, **90**, 543–584.
59. Gagnon,K.T. and Corey,D.R. (2015) Stepping toward therapeutic CRISPR. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 15536–15537.
60. Cromwell,C.R., Sung,K., Park,J., Krysler,A.R., Jovel,J., Kim,S.K. and Hubbard,B.P. (2018) Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat. Commun.*, **9**, 1448.
61. Kawasaki,A.M., Casper,M.D., Freier,S.M., Lesnik,E.A., Zounes,M.C., Cummins,L.L., Gonzalez,C. and Cook,P.D. (1993) Uniformly modified 2′-deoxy-2′-fluoro phosphorothioate oligonucleotides as nuclease-resistant antisense compounds with high affinity and specificity for RNA targets. *J. Med. Chem.*, **36**, 831–841.
62. Pallan,P.S., Greene,E.M., Jicman,P.A., Pandey,R.K., Manoharan,M., Rozners,E. and Egli,M. (2011) Unexpected origins of the enhanced pairing affinity of 2′-fluoro-modified RNA. *Nucleic Acids Res.*, **39**, 3482–3495.
63. Watts,J.K., Martin-Pintado,N., Gomez-Pinto,I., Schwartzentruber,J., Portella,G., Orozco,M., Gonzalez,C. and Damha,M.J. (2010) Differential stability of 2′F-ANA*RNA and ANA*RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility. *Nucleic Acids Res.*, **38**, 2498–2511.
64. Jiang,F., Zhou,K., Ma,L., Gressel,S. and Doudna,J.A. (2015) A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*, **348**, 1477–1481.
65. Sternberg,S.H., Redding,S., Jinek,M., Greene,E.C. and Doudna,J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62–67.
66. Semenova,E., Jore,M.M., Datsenko,K.A., Semenova,A., Westra,E.R., Wanner,B., van der Oost,J., Brouns,S.J. and Severinov,K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10098–10103.