Restarted replication forks drive CAG repeat instability

Michaela A. Gold¹, Jenna M. Whalen¹, Karine Freon², Zixin Hong¹, Ismail Iraqui², Sarah A. E. Lambert² and Catherine H. Freudenreich¹

¹Department of Biology, Tufts University, Medford, MA USA
²Institut Curie, CNRS, Orsay, France
Repeat Expansion Diseases

- CGG, CAG
  - Fragile X (A)
  - Fragile X (E)
  - SCA12
- CAG → Glu
- GCN → Ala
- GAC → Asp
- GAA
  - Friedreich’s Ataxia
  - SCA10
  - ALS
- ATTCT
- G4C2
- CTG
  - Myotonic Dystrophy
  - SCA8

C/G 12mer
- EPM1

Huntingtons
- SCA1,2,3,6,7
- SBMA
- DRPLA
- TBP

intron
Trinucleotide Repeat Expansions are **Dynamic Mutations**

**Myotonic Dystrophy (CTG)**
- Full mutation: 200–2000 repeats
- Premutation: 50–180 repeats
- Normal: 5–37 repeats

**Huntington’s Disease (CAG)**
- Full mutation: 36–121 repeats
- Normal: 6–35 repeats

Expansions occur Intergenerationally and Somatically

1. 6-34 REPEATS (STABLE LENGTH) → 35-38 REPEATS (THRESHOLD)
2. EXPANSIONS → 40-150 REPEATS (HD)
   → 40-2000 REPEATS (DM)
When do Expansions Occur?

• In the Germline
  - Paternal (HD, SCA) and Maternal expansion biases (FRAX, DM1) exist
  - Paternal age expansion bias- in dividing premeiotic and postmeiotic spermatogonia.
  - Paternal contractions happen in proliferating germ stem cells that lack methylation (FRAXA, FRDA, SCA8)
  - Expansions in maternal germline in Fragile X and myotonic dystrophy
  - Germline expansions are important in disease inheritance (intergenerational)

• In Somatic tissue
  - Early embryogenesis – cells dividing quickly
  - Mature somatic tissues – brain
  - Somatic expansions have an important contribution to disease progression
How do Expansions Occur?
-during either replication or repair (requires DNA synthesis)

CTG and CAG repeats form hairpin structures that interfere with DNA replication and DNA repair

Tm CTG > Tm CAG
Tract-length changes can occur during Replication

- Direction of replication determines the expansion-contraction bias (bacteria, yeast, human cells)

- Disease loci tend to be in the expansion-prone orientation
  - CAG on the lagging strand template at the DM1 locus (Cleary..Pearson, 2010)

- Mutation or inhibition of replication proteins increases expansion and contraction frequency
Structure-forming Trinucleotide Repeats Interfere with DNA Replication

Repeat names indicate the sequence on the lagging strand template.
Question: What stages of replication are most prone to repeat instability?

- Slippage or template switch during normal replication progression through the repeat?
- Slippage or misalignment at a replication stall caused by the repeat?
- During subsequent replication of the repeat by a restarted fork that has altered properties?
- After breakage at the repeat during replication: expansion or contraction occurring during BIR?

Figure from Polleys, House, Freudenreich 2017

Figure from Khristich & Mirkin, 2020
Question: What stages of replication are most prone to repeat instability?

- Slippage or template switch during normal replication progression through the repeat?
- Slippage or misalignment at a replication stall caused by the repeat?
- **During subsequent replication of the repeat by a restarted fork that has altered properties?**
- After breakage at the repeat during replication: expansion or contraction occurring during BIR?

The Idea: separate the repeat-mediated stall from the restarted fork

- Induce restart with a protein-mediated stall
- Restarted fork travels through a CAG repeat

Figure from Sarah Lambert
Creating a single and local replication stress site with the power of yeast genetics

**The site specific RTS1-RFB**

- Inducible polar fork arrest mediated by the binding of Rtf1 to RTS1 sequence
- Blockade at the same, unique locus at Chromosome III in > 90% of cells
- Blocked forks are restarted by Recombination-Dependent Replication (RDR), a form of non-canonical DNA synthesis

*Schizosaccharomyces pombe* fission yeast

Recombination-Dependent Replication (RDR) as a fork-restart pathway

Lambert, Murray, Whitby and Carr labs

- DNA-bound protein = Rft1

**Fork-resection** (MRN-Ctp1-Ku: initial; Exo1: Long-range; Fft3$^{\text{Fun30/SMARCAD1}}$)

**Strand Invasion** (Dependent on Rad52 & Rad51 strand exchange activity)

**Restarted fork (in ~ 20min):**
- Semi-conservative DNA synthesis (≠ BIR, = Canonical fork)
- Pol δ/ Pol δ DNA synthesis (≠ Canonical fork)
- Error-prone, liable to slippage mutations and template switch
- Prone to GCRs (Gross Chromosomal Rearrangements)
The System:

Put a (CAG)70 repeat tract after an inducible replication fork barrier (RFB)

Does replication through the repeat by the restarted fork cause repeat expansions or contractions?
Analysis of Replication Intermediates in the Region

B CAG-70 + RTS1 CAG-70 No RTS1
OFF (Rtf1 repressed)
ON (Rtf1 expressed)

CAG-70-containing fragment (EcoRV digest, Kan probe)

C
No CAG repeat +RTS1 CAG-70 + RTS1 CAG-70 No RTS1
OFF (Rtf1 repressed)
ON (Rtf1 expressed)

RTS1-containing fragment (Ase1 digest, Ura4 probe)
Both CAG Expansions and Contractions increase after induction of a nearby Replication Fork Barrier (RFB)

Conclusions:
1. The RFB increases CAG expansions
   Surprisingly, the effect is more dramatic for a WEAK RFB and increases further from the RFB
PCR Assay to detect both CAG Expansions and Contractions
Distribution of CAG Expansion and Contraction Sizes

Expansions ranged from +5 to +50 repeats
Contractions ranged from -5 to -65 repeats

N= 109-487 per group

CAG-70 1.9 kb

CAG-70 6.7 kb

N= 109-487 per group
Mechanisms of replication fork rescue

DNA-bound protein

Fork-arrest

Fork-restart

RFB system is **Rad52-dependent**
(Re-priming is another option)

Rescue by converging fork
(Firing of dormant origins)
Not **Rad52-dependent**
Rad52-dependent fork restart is required to cause RFB-dependent CAG expansions

CAG-70, rad52Δ (raw data)

CAG-70, Expansions (fold over wild-type)

CAG expansions are happening during Rad52-dependent fork restart

Contractions may be occurring by another mechanism
How are expansions occurring during fork restart (by what mechanism)?

*Figure from Bonner & Zhao, 2016*

*S. pombe* Rad8  
*S. cerevisiae* Rad5  
human HTLF needed for template switch
A Rad8 (scRad5, hHTLF)-dependent process (template switch) causes some RFB-dependent CAG expansions and contractions.

Some CAG expansions and contractions happening during template switch.
A Rad8 (scRad5, hHTLF)-dependent process (template switch) causes RFB-dependent CAG expansions and contractions.
Expansions and Contractions are not occurring during DSB repair
-Not dependent on single-strand annealing (Swi10) or resection (Exo1)
Expansions and contractions are not occurring during DSB repair - not dependent on SSA (Swi10) or resection (Exo1)
Why does the RFB cause such a high frequency of contractions? Most are not dependent on Rad52-dependent restart.

![Graph showing Instability frequency (%)](chart.png)

- **CAG-70, Wild-type**
- **Instability frequency (%)**
- **Expansions**
  - 1.9 kb: No RFB, RFB Off (Weak RFB), RFB On (Strong RFB)
  - 6.7 kb: No RFB, RFB Off (Weak RFB), RFB On (Strong RFB)
- **Contractions**
  - 1.9 kb: No RFB, RFB Off (Weak RFB), RFB On (Strong RFB)
  - 6.7 kb: No RFB, RFB Off (Weak RFB), RFB On (Strong RFB)

* denotes statistical significance.
Mechanisms of replication fork rescue

- DNA-bound protein
  - Fork-arrest
  - Fork-restart
  - RFB system is Rad52-dependent

CAG expansions are Rad52-dependent, therefore, happening during fork restart

Are CAG contractions happening during fork rescue?
Why does the RFB cause such a high frequency of contractions? 
-most are not dependent on Rad52-dependent restart 

Replication direction switch model: CTG forms more stable hairpins than CAG 
Therefore, CTG on the lagging strand template may form more structures, leading to contractions
CTG sequence on the lagging strand template causes a high frequency of contractions: consistent with replication direction switch model.
Testing the replication direction switch model: blocking the converging fork reduces CAG contractions

Contractions mostly occur due to a fork switch
Are expansions dependent on fork reversal or fork resection through the repeat?

Placing the CAG tract just upstream of the barrier does not cause RFB-dependent instability.
Are expansions occurring during the initial strand invasion phase of restart?

Placing the CAG tract just downstream of the barrier does not cause RFB-dependent instability
MutS causes CAG expansions in *S. pombe* (as in humans and mouse models), but this is independent of the RFB.

---

**CAG-70, Expansions**

- **Wild-type**
  - No RFB: 3.0%
  - Weak RFB: 2.5%
  - Strong RFB: 7.5%

- **msh2Δ**
  - No RFB: 1.5%
  - Weak RFB: 2.0%
  - Strong RFB: 5.0%

---

**CAG-70, Expansions**

- **msh2Δ**
  - No RFB: 1.0 times wild-type
  - Weak RFB: 1.5 times wild-type
  - Strong RFB: 2.5 times wild-type
MutS deletion causes CAG contractionss (as in humans and mouse models), but this is independent of the RFB
Model

A. CAG repeat instability occurring during Rad52 dependent fork restart

1. Replication of repeat by less processive δ-δ restarted fork
2. Strand slippage
3. Expansions and Contractions

B. CAG repeat contractions occurring due to replication by converging replication fork

1. Fork restart delayed or fails
2. Repeat replicated by incoming epsilon-delta fork
3. Contractions
Take Home Messages for Replication Fidelity:

• Restarted replication forks are especially prone to slippage through repetitive DNA tracts
• Increased template switching by the uncoupled fork can lead to repeat expansions
• Replication fork barriers in genomes can lead to a change in fork direction, and this has implications for the stability of structure-forming sequences
• This mechanism of repeat instability could be relevant to barriers caused either by the repeat itself, or neighboring barriers.
  • Cancer cells have altered replication programs and rely heavily on replication restart mechanisms
Implications for Repeat Expansion Diseases:

• Transient replication fork barriers can drive repeat instability – could this explain the restricted developmental time window of intergenerational expansions?
  • Cell type or timing-specific barriers could lead to changes in the replication profile of the repeat locus

• Many expandable CAG/CTG tracts are near CTCF (chromatin insulator) binding sites that are associated with Topologically Associated Domains (TADs)
  • CTCF may cause fork slowing through the Myotonic Dystrophy locus (Cleary..Pearson, 2010)

Thanks to Sarah Lambert for hosting me and being a terrific collaborator!

Freudenreich lab members:
Michaela Gold (undergrad at the time)
Zixin Hong (Tufts undergrad)
Jenna Whalen (Grad student at time)

Lambert lab members:
Karine Freon (Research Associate)
Ismail Iraqui (post-doc)

Funding:
**CHF**: NIH grant GM122880; Mayent Rothschild fellowship from Institut Curie
**SAL**: Ligue contre le cancer “Equipe Labellisée 2020 (EL2020LNCC/Sal)
**MAG**: Tufts Undergraduate Research Award; Tufts University Sandler International Research Program award to M.A.G., J.M.W and C.H.F.

Catherine and Sarah at the Recombination meeting in Alicante, Spain
Figure S3

A. Diagram showing different orientations of the t-ura4sd20-ori sequence.

B. Bar graph showing the frequency of Ura+ cells X 10^-5 for WT and rad8Δ strains.

C. Diagram illustrating the converging fork and arrested fork.

D. Bar graph showing the % of signal for converging and resected forks in WT and rad8Δ strains, with p=0.016.