PhD Project Proposal

Funder details

Studentship funded by: MRC-DTP

Project details

Typos in the genome: Defining the molecular mechanism of replication slippage

Supervisory team

Primary Supervisor: Dr. Gideon Coster

Associate Supervisor(s):

Secondary Supervisor: Prof. Jessica Downs

Divisional affiliation

Primary Division: Cell and Molecular Biology

Primary Team: Genome Replication

Site: Chelsea

Project background

Inactivation of mismatch repair (MMR) causes inherited and sporadic cancers[1]. These tumours exhibit microsatellite instability (MSI) - length alterations within short repeats. MSI is a diagnostic and prognostic tool and plays a direct causative role. The most common MSI events are single-nucleotide deletions within poly(dA:dT) stretches[2]. Inactivation of proofreading by replicative polymerases drives similar hypermutated cancers, but with single-nucleotide insertions. Surprisingly, longer homopolymers (n>15 bp) exhibit larger deletions. Although these events are all thought to be driven by replication slippage[3], current models fail to explain these different mutational outcomes.

To efficiently stratify and target MSI cancers, we must understand what drives these distinct phenotypes. We will therefore define the molecular mechanism of slippage. Key questions we will address are whether slippage occurs differently on leading/lagging strands, whether replicative polymerases exhibit distinct slippage behaviours and if other replisome components modulate slippage.

We have recently established a system to study the mechanism of replication slippage in nucleotide resolution using in vitro reconstituted replisomes in conjunction with templates that harbour different lengths of poly(dA:dT) stretches. The system currently employs yeast replisomes but will soon also employ reconstituted human replisomes[4].

Briefly, in vitro replicated products are cleaved and the insert separated on sequencing gels. End-labelled inserts serve as references. Using radiolabelled dATP or dTTP allows us to discriminate between nascent leading or lagging strands. By comparing the length of the original template with the replicated products,

we can determine if slippage occurs on the leading or lagging strand and whether it is more likely to occurs with poly(dA) or poly(dT) templates. The power of a defined in vitro system means we can directly determine the role that different replisome components play in replication slippage

Project aims

- **Determine the mechanism of replication slippage using reconstituted replisomes**
- **Directly sequence *in vitro* replicated products using single molecule PacBio sequencing**
- **Establish how local DNA context affects replication slippage**
- **Develop a high-throughput single-molecule approach to study replication fidelity in vitro**

Research proposal

Given the broad role that replication plays in human health, defining how replisomes cope with challenging templates is of fundamental importance. While replication stress can be induced by chemicals or radiation, DNA itself has been implicated as an endogenous source of stress. Specifically, sequences that adopt various secondary structures[5] (e.g., hairpins, cruciforms, G-quadruplexes (G4s), intercalated motifs (i-motifs), and triplexes) are hotspots for mutations and translocations across various cancers[6,7]. However, such sequences play important biological roles and mechanisms must be in place to preserve them. A complete molecular understanding of how these sequences are preserved while maintaining genome stability is lacking. Furthermore, the causal relationship between perturbed replication dynamics, altered fidelity, and loss of genome integrity is unclear.

This proposal aligns well with the MRC DTP theme of “Genome Stability and DDR” and will impact a broad range of disciplines and fields, including nucleic acid chemistry and structure, cancer biology, genome evolution and repeat expansion disorders.

We have recently discovered that structure-forming sequences stall reconstituted budding yeast replisomes[8,9]. This highlights the DNA template as a direct source of replication stress. To understand the relationship between altered replication dynamics and fidelity, we must first define the mechanisms that impact replication fidelity within repetitive sequences. This is the major goal of this proposed project.

Specific aims:

- Determine the mechanism of replication slippage using reconstituted replisomes

  Here, the PhD candidate will used a well-established approach in the lab which allows us to reconstitute DNA replication in a test tube using purified budding yeast proteins. This powerful approach generates mechanistic insight that cannot be obtained in other ways. A similar system has recently been developed for human replisomes, which will also be established in the lab. To obtain nucleotide-resolution results, the PhD candidate will employ denaturing sequencing gels. Comparing the length and distribution of replicated products from different templates and with different reaction conditions will allow us to determine fundamental aspects of replication slippage.

- Directly sequence in vitro replicated products using single molecule PacBio sequencing

  We have an onoging collaboration with the lab of Prof. Vincent Dion (University of Cardiff). The Dion lab have developed an experimental and bioinformatic pipeline for single-molecule sequencing of repetitive sequences[10]. Identifying rare mutations within repeats using conventional

sequencing is very difficult as repeat are often filtered away or generate too much noise. However, these challenged can be resolved using single-molecule sequencing, such as that obtained by the Pacific Biosciences (PacBio) platform. The Dion lab have an in-house Sequel IIe platform and have developed an algorithm to identify mutations within complex repetitive sequences.

Here, the PhD candidate will collaborate with the Dion lab to directly sequence in-vitro replication products. We have devised an approach to specifically sequence replicated material whilst also differentiating leading versus lagging strand products. This will provide direct evidence of replication slippage and will also pinpoint exactly which mutations are produced, where and at what frequency.

- Establish how local DNA context affects replication slippage

  There is evidence that the local DNA sequence around a repetitive sequence can affect the mutational outcomes of replication slippage[11]. Why or how this is the case is unclear.
  To determine the effect of local DNA context, the PhD candidate will clone and test the effect of different flanking 5' and 3' sequences derived from the human genome and whether this is modulated by the replicative polymerase used.

- Develop a high-throughput single-molecule approach to study replication fidelity in vitro

  Since there is limited scope in studying a small subset of repeats and sequence contexts, an ideal approach would be a high-throughput assay where a library of repeats is replicated and then sequenced using single-molecule PacBio platforms. Thus, the PhD candidate will develop a method for replicating and sequencing a library of barcoded repeats with a large variety of flanking sequence contexts. This will provide a broad understanding of the sequence determinants that drive replication slippage.

Altogether, this project will provide a mechanism for replication slippage while establishing state-of-the-art biochemical approaches. It will identify which replisome components and DNA sequences play a role. The results are relevant not only for all MSI cancers but also beyond, impacting the fields of genome stability and cancer evolution.

Literature references

1. Lynch, H.T., Snyder, C.L., Shaw, T.G., Heinen, C.D., and Hitchins, M.P., Milestones of Lynch syndrome: 1895-2015*. Nat Rev Cancer*, 2015. **15**(3): p. 181-194.
2. Chung, J., Maruvka, Y.E., Sudhaman, S., Kelly, J., Haradhvala, N.J., Bianchi, V., Edwards, M., Forster, V.J.*, et al.*, DNA Polymerase and Mismatch Repair Exert Distinct Microsatellite Instability Signatures in Normal and Malignant Human Cells*. Cancer Discovery*, 2021. **11**(5): p. 1176-1191.
3. Streisinger, G. and Owen, J.E., Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4*. Genetics*, 1985. **109**(4): p. 633-659.
4. Baris, Y., Taylor, M.R.G., Aria, V., and Yeeles, J.T.P., Fast and efficient DNA replication with purified human proteins*. Nature*, 2022. **606**(7912): p. 204-210.
5. Khristich, A.N. and Mirkin, S.M., On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability*. Journal of Biological Chemistry*, 2020. **295**(13): p. 4134-4170.
6. Bacolla, A., Tainer, J.A., Vasquez, K.M., and Cooper, D.N., Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences*. Nucleic Acids Research*, 2016. **44**(12): p. 5673-5688.
7. Georgakopoulos-Soares, I., Morganella, S., Jain, N., Hemberg, M., and Nik-Zainal, S., Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis*. Genome Research*, 2018. **28**(9): p. 1264-1271.

8. Casas-Delucchi, C.S., Daza-Martin, M., Williams, S.L., and Coster, G., The mechanism of replication stalling and recovery within repetitive DNA*. Nature Communications*, 2022. **13**(1): p. 3953.
9. Williams, S.L., Casas-Delucchi, C.S., Raguseo, F., Guneri, D., Li, Y., Minamino, M., Fletcher, E.E., Yeeles, J.T.P., et al., Replication-induced DNA secondary structures drive fork uncoupling and breakage. *EMBO J*, 2023. **42**: e114334 https://doi.org/10.15252/embj.202311433410.
10. Taylor, A.S., Barros, D., Gobet, N., Schuepbach, T., McAllister, B., Aeschbach, L., Randall, E.L., Trofimenko, E., et al., Repeat Detector: versatile sizing of expanded tandem repeats and identification of interrupted alleles from targeted DNA sequencing. *NAR genomics and bioinformatics*, 2022. **4**(4): p. lqac089.
11. Herzog, M., Alonso-Perez, E., Salguero, I., Warringer, J., Adams, David J., Jackson, S.P., and Puddu, F., Mutagenic mechanisms of cancer-associated DNA polymerase ϵ alleles. *Nucleic Acids Research*, 2021. **49**(7): p. 3919-3931

Candidate profile

Note: the ICR's standard minimum entry requirement is a relevant undergraduate Honours degree (First or 2:1).

Pre-requisite qualifications of applicants:

B.Sc. or M.Sc. (First or 2:1, or equivalent) in any of the following disciplines:

>> Life Sciences / Biology
>> Biochemistry
>> Biomedical Sciences
>> Molecular Biology
>> Genetics

Intended learning outcomes:

During the course of this studentship, the PhD candidate will be fully supported and encouraged to develop multiple skills and become proficient in the following:

>> Molecular biology techniques (cloning, PCR, western blotting etc…)
>> Protein expression and purification using a variety of systems (bacteria / yeast / baculovirus)
>> Experimental design and interpretation
>> Biochemical analysis of complex reactions
>> Reading and critical analysis of published journal articles
>> Writing and publishing original research papers
>> Presentation of scientific projects to an expert audience

Advertising details

Project suitable for a student with a background in:

☒ Biological Sciences
☐ Physics or Engineering

☒ Chemistry

☐ Maths, Statistics or Epidemiology

☐ Computer Science