This is the published version of a paper published in *PLOS Genetics*.

Citation for the original published paper (version of record):

http://dx.doi.org/10.1371/journal.pgen.1004846

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-96881
Increased and Imbalanced dNTP Pools Symmetrically Promote Both Leading and Lagging Strand Replication Infidelity

Robert J. Buckland¹², Danielle L. Watt³, Balasubramanyam Chittoor¹, Anna Karin Nilsson¹, Thomas A. Kunkel³, Andrei Chabes¹²*¹

¹ Department of Medical Biochemistry and Biophysics, Umeå University, Umeå, Sweden, ² Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden, ³ Laboratory of Molecular Genetics and Laboratory of Structural Biology, National Institute of Environmental Health Sciences, NIH, DHHS, Research Triangle Park, North Carolina, United States of America

Abstract
The fidelity of DNA replication requires an appropriate balance of dNTPs, yet the nascent leading and lagging strands of the nuclear genome are primarily synthesized by replicases that differ in subunit composition, protein partnerships and biochemical properties, including fidelity. These facts pose the question of whether imbalanced dNTP pools differentially influence leading and lagging strand replication fidelity. Here we test this possibility by examining strand-specific replication infidelity driven by a mutation in yeast ribonucleotide reductase, rnr1-Y285A, that leads to elevated dTTP and dCTP concentrations. The results for the CAN1 mutational reporter gene present in opposite orientations in the genome reveal that the rates, and surprisingly even the sequence contexts, of replication errors are remarkably similar for leading and lagging strand synthesis. Moreover, while many mismatches driven by the dNTP pool imbalance are efficiently corrected by mismatch repair, others are repaired less efficiently, especially those in sequence contexts suggesting reduced proofreading due to increased mismatch extension driven by the high dTTP and dCTP concentrations. Thus the two DNA strands of the nuclear genome are at similar risk of mutations resulting from this dNTP pool imbalance, and this risk is not completely suppressed even when both major replication error correction mechanisms are genetically intact.


Editor: Gregory P. Copenhaver, The University of North Carolina at Chapel Hill, United States of America

Received September 6, 2014; Accepted October 23, 2014; Published December 4, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Project Z01 ES065070 to TAK from the Division of Intramural Research of the NIH, NIEHS, and by The Knut and Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research and the Swedish Cancer Society to AC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: andrei.chabes@medchem.umu.se

Introduction
The integrity of an organism’s genome is vital to its continued survival, whether unicellular microbe or complex large mammal [1]. Therefore, there are highly conserved mechanisms involved in regulating and protecting genetic material both during and post DNA replication. One of the first safety systems for DNA replication is the stringent control of dNTP synthesis by ribonucleotide reductase (RNR), which maintains concentrations of the individual dNTPs at different levels [1,2]. RNR catalyses the rate-limiting step in the production of all four dNTPs for the synthesis of nuclear and mitochondrial DNA [3,4]. In yeast, RNR is a multi-subunit complex comprised of a large subunit, which exist as a homodimer of Rnl1 proteins or a heterodimer of Rnl1/Rnl3 proteins, and a small subunit comprised of Rnl2/Rnl4 proteins. The large subunits contain allosteric specificity sites that modulate enzyme activity and control the balance of the four dNTPs by influencing the specific ribonucleoside monophosphate reduction reaction within the catalytic sites [5]. A highly conserved loop of 13 amino acid residues (Loop 2) connects the allosteric specificity and catalytic sites and is crucial for the correct allosteric regulation of the enzyme [6,7].

The DNA polymerase selectivity, proofreading and mismatch repair are subsequent safety systems that determine the fidelity of DNA replication. The DNA polymerase selectivity ensures insertion of the correct nucleotide during DNA synthesis. Although the major replicative polymerases alpha (Pol α), delta (Pol δ), and epsilon (Pol ε) are high fidelity enzymes, their accuracy is dependent upon the supply of dNTPs [8]. The second mechanism is proofreading in which errors are removed from primer termini during replication by a 3’–5’ exonuclease activity. Errors that escape proofreading can still be repaired post-replication, through the mismatch repair system (MMR) [reviewed in [9]]. The major components of MMR are the homologs of the bacterial MutS proteins, a heterodimer of either Msh2-Msh6 or Msh2-Msh3 that recognise and bind to the mismatch. Msh2-Msh6 is mainly responsible for repairing single base-base mismatches, short insertions and deletions (indels) and small loops, whereas Msh2-Msh3 is involved in larger loop repair. Therefore, Msh2 is essential for MMR [10] and loss of this activity elevates mutation rates [11]. Mutation or loss of Msh2 in humans is associated with microsatellite instability and hereditary nonpolyposis colorectal cancer (HNPCC) [12] and gall bladder cancer [13].
The building blocks of DNA, dNTPs, are vital to life, and thus their production is carefully controlled within each cell. Under certain conditions, such as cancer, infection, or drugs, the overall dNTP level or dNTP balance can change. Using yeast genetics, we manipulated the dNTP pool balance in unicellular baker’s yeast and analysed the effects upon fidelity of DNA replication. We also disrupted mismatch repair, an internal safety system that corrects replication errors and is mutated in many cancers. By sequencing DNA from yeast cells with these alterations, we gained insights into the mechanisms of mutation formation that contribute to genome instability. We find that the leading and lagging strand replication fidelity is affected similarly by the dNTP pool imbalance and that the mismatch repair machinery corrects replication errors driven by a dNTP pool imbalance with highly variable efficiencies.

Figure 1. dNTP pools of the strains with the imbalanced dNTP pools. Numbers above columns show the factor increase over wt. Error bars show Standard Error of Mean. doi:10.1371/journal.pgen.1004846.g001

**Results**

dNTP pools of the *mrnl-Y285A* and *msh2Δ* strains

To examine potential differences in mutational specificity between the major replicative polymerases in the presence of the imbalanced dNTP pools, we reversed the orientation of the CAN1 gene (*CAN1-OR2*). To investigate the effect of this dNTP pool imbalance in the absence of MMR, we deleted MSW2 in the *rnl1-Y285A CAN-OR1* strain. The *msh2Δ* single mutant strain had normal dNTP pools (Fig. 1). The dNTP pools in the double mutant had the same imbalance as in the single *rnl1-Y285A* [17], with approximately 26- and 14-fold higher concentrations of dCTP and dTTP, respectively, compared to wild type (wt) whilst the concentration of dATP was about double and dGTP was normal (Fig. 1).

**CAN1 spontaneous mutation rates and types**

The spontaneous CAN1 mutation rate in the *mrnl-Y285A CAN1-OR2* was 13-fold higher than wt (Table 1), which was similar to the *CAN1-OR1* previously published [OR1] [17]. The *msh2Δ* mutant had a 15-fold higher mutation rate compared to wt, however, the *rnl1-Y285A msh2Δ* strain mutation rate was over 500-fold greater than that of wt and over 30-fold either of the relative single mutants. Indels were the major mutation type observed in all four mutant strains whereas it was single base substitutions in wt (Fig. 2). The *mrnl-Y285A*, with CAN1 in both orientations, and *msh2Δ* strains had an average increase in the indel rate of more than 40-fold the wt strain (0.5 x 10^-6 for wt versus 33 x 10^-6 for OR1 [17], 37 x 10^-6 for OR2, and 42 x 10^-6 for *msh2Δ*). However, the double mutant indel rate was increased the most at more than 2000-fold over wt. In addition to single base indels, base substitutions were also significantly increased in the mutants to over 8-fold higher in the single mutants and 350-fold higher in *mrnl-Y285A msh2Δ* compared to wt. Complex mutations, defined as mutations involving insertions or deletions of multiple bases, were much more common in the double mutant, occurring at a rate over 30 times higher than that in wild type.

**Mutation hotspots**

Replication of the CAN1 gene originates from ARS507 and travels through the gene towards the telomere [18–20] (Fig. 3A). Therefore, in *mrnl-Y285A CAN1-OR1* the leading strand polymerase, presumed to be Pol ε [21], uses the non-coding strand as the template while the coding strand is the template for lagging strand replication primarily by Pol δ and Pol α [14]. In CAN1-OR2, Pol ε now copies the coding strand and Pol δ/Pol α copy the non-coding strand (see Fig. 3A). An example is given in Fig. 3B, for the single base substitution at 648 bp. During leading strand synthesis in OR1, no mistake is made when copying template G due to high concentration of dCTP. However, during lagging strand replication dTTP is inserted opposite template C, as dTTP is at a much higher concentration than the dGTP required for correct pairing. As the succeeding incoming nucleotides are also at an increased concentration, rapid extension then follows stabilizing the C: dT mismatch. In the next round of replication, a C to A mutation arises. When the gene is reversed in *mrnl-Y285A CAN1-OR2*, Pol ε now copies the template C with low fidelity by misinserting dTTP, which ultimately results in a C to A mutation, and Pol δ/Pol α replication is error-free.
Hotspots, mutation sites where the rates were ≥10-fold greater than wt, were assigned to have occurred during leading or lagging strand synthesis by the nature of the mutation observed and dNTP pool imbalance. Simplified mutational spectra illustrating the hotspots in the CAN1 gene for each strain are shown in Fig. 4 with the full spectra in Figure S1-S3. The hotspot mutation rate was calculated by the equation [(frequency/total no. of samples) x CAN1 mutation rate]. The majority of hotspots in mrr1-Y285A CAN1-OR1 and OR2 show no leading – lagging strand bias and have similar mutation rates in both strains (Fig. 5). However, the major hotspot at position 425–427 bp was only seen in OR2 and had a mutation rate of 4.8 x 10^{-7} which was 15-fold higher than in OR1.

The mrr1-Y285A and msh2Δ strains had several shared major hotspots. Three single base deletions occurred in G: C homopolymeric runs at 757–760 bp, 795–797 bp, and 857–859 bp and two single base substitutions at 313 bp and 1379 bp (Fig. 4 and Table S1). Whilst the double mutant shared these five hotspots with both single mutants, there was a more than 100-fold increase in rates for base substitutions at 313 bp and 1379 bp (Fig. 6A and Table S1). The major hotspots in mrr1-Y285A msh2Δ were those seen in msh2Δ at 1118–1121 bp, 1392 bp, and especially the dominant deletion hotspot at 620–625 bp, 964–969 bp, and 1381–1386 bp. The site-specific mutation rates in the double mutant ranged from 4- to 800-fold the single mutants.

**MMR efficiency**

Analysis of the mutation spectra in the mrr1-Y285A and mrr1-Y285A msh2Δ strains showed that MMR had different efficiency at distinct mismatches. The ratio of mutation rates in the msh2Δ and MMR-proficient strains gave site-specific MMR correction efficiency (Fig. 6B and Table S1). The five hotspots (314, 718, 757, 795 and 857 bp), which include those with the highest mutation rates in the single mrr1-Y285A mutant, were the same sites that MMR was the least efficient at repairing errors. The correction factors were less than 30 and only around 10 in four of these sites (i.e., around 10% of errors at these sites will remain uncorrected by MMR). The majority of sites with the highest mutation rates in the double mutant (620, 964, 1118, 1381, and 1392 bp) were those that MMR was best at repairing, namely at T:A mononucleotide repeats. Loss of MSH2 increased the mutation rate at these sites by 260- to 800-fold.

**Discussion**

Increased dCTP and dTTP drive different polymerases to make similar errors

Despite the inherent differences in complexity of continuous (leading strand) and non-continuous (lagging strand) synthesis, the increased dCTP and dTTP drive the same kind of mutations at identical sequences regardless of the replicative DNA polymerase. Most of the mutations occurred at a G:C base pair in which the cytosine served as the template for synthesis and was often flanked by a 5’-A or a tract of purines as exemplified in Fig. 3. With the concentration of dGTP being the lowest and dCTP and dTTP the highest, the deletion of a G:C base pair in a mononucleotide repeat is stabilized by the rapid incorporation of the next incoming nucleotide (dTTP opposite the template A), as described in detail in our previous report [17]. This dNTP imbalance and sequence context also explains the G to T: A base substitutions at 1118–1121 bp, 1392 bp, and especially the dominant deletion hotspots at 620–625 bp, 964–969 bp, and 1381–1386 bp. The site-specific mutation rates in the double mutant ranged from 4- to 800-fold the single mutants.

**Table 1. CAN1 mutation rates and observed events.**

<table>
<thead>
<tr>
<th>Strain</th>
<th>wt</th>
<th>mrr1-Y285A</th>
<th>mrr1-Y285A CAN1 OR2</th>
<th>msh2Δ</th>
<th>mrr1-Y285A msh2Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutation rate (x10^{-7})</td>
<td>4.2</td>
<td>57</td>
<td>60</td>
<td>66</td>
<td>2236</td>
</tr>
<tr>
<td>95% CI</td>
<td>1.6–4.4</td>
<td>43–103</td>
<td>40–108</td>
<td>53.4–90</td>
<td>1883–3272</td>
</tr>
<tr>
<td>can1 mutants sequenced</td>
<td>93</td>
<td>173</td>
<td>170</td>
<td>164</td>
<td>259</td>
</tr>
<tr>
<td>Base substitutions</td>
<td>65</td>
<td>72</td>
<td>80</td>
<td>62</td>
<td>131</td>
</tr>
<tr>
<td>Single base indels</td>
<td>11</td>
<td>101</td>
<td>104</td>
<td>106</td>
<td>130</td>
</tr>
<tr>
<td>Others</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Total mutations</td>
<td>93</td>
<td>173</td>
<td>184</td>
<td>169</td>
<td>264</td>
</tr>
</tbody>
</table>

Data previously published in [17].

Figure 2. Mutations rates by class.
doi:10.1371/journal.pgen.1004846.g002

Figure 3. Mutations rates by class.
doi:10.1371/journal.pgen.1004846.g002

**Figure 2. Mutations rates by class.**

doi:10.1371/journal.pgen.1004846.g002
dGTP was larger, from ~4:1 in the wt strain to ~38:1 in the rnr1-Y285A strains, which may explain the prevalence of the G: C to T: A transversions. Furthermore, the lack of T: A to G: C transversions may be due to the intrinsic difference in the rates at which the two errors are generated. Recent genome-wide studies in S. cerevisiae have reported that G: C to T: A transversions were observed at a higher rate than T: A to G: C in strains with normal dNTP pools [22,23]. The three major replicative polymerases were more prone to generate G: C to T: A transversions but very rarely generated T: A to G: C transversions [23]. In addition, tumours with somatic mutations in the exonuclease domain of Pol ε have a higher prevalence of C to A mutations [24–28].

Figure 3. Strand Assignment Model. A. Cartoon representation of CAN1 orientation in rnr1-Y285A strains. B. Model showing strand assignment in the two rnr1-Y285A mutants (OR1 and OR2) using the hotspot at site 648 bp as an example. MI = Misinsertion. Red characters represent the mutational event and green characters represent bases where the dNTP is at an excessively high concentration. doi:10.1371/journal.pgen.1004846.g003
Figure 4. Simplified *CAN1* mutation spectra showing hotpots (where mutation rate is greater than 10-fold that in wt) for *mr1-Y285A CAN1-OR1* (n = 173), *mr1-Y285A CAN1-OR2* (n = 170), *mr1-Y285A msh2Δ* (n = 259), and *msh2Δ* (n = 164) where n = number of individual colonies sequenced. Symbols indicate the following: plus - additions, triangles - deletions, squares - transversions, circles - transitions, red - occur during leading strand synthesis, blue - occur during lagging strand synthesis, black - mutation cannot be assigned to a strand. * Reanalysed from the dataset published in [17].

doi:10.1371/journal.pgen.1004846.g004
MMR repairs replication errors driven by a dNTP pool imbalance with highly variable efficiencies

MMR efficiency was dependent upon the site and mismatch generated from the dNTP pool imbalance. The increase in indels in the msh2Δ strains was not surprising as MMR is known to be highly active at repairing mistakes at mononucleotide repeats [29,30]. The indels were almost entirely unique to sequences with ≥3 mononucleotide repeats in the double mutant (99.2%, 127 of 128) compared to 91% in the msh2Δ mutant and most frequently occurred in A: T runs. This can be predicted as A-T mononucleotide repeats are often the site of indels in MMR deficient strains [31] and are by far the most common in the CAN1 gene sequence (Figure S4). Indeed, it appears that the relationship between mutation rate and mononucleotide repeat length is exponential as others have found across the whole yeast genome [22].

The MMR correction factor for all indels in the rnr1-Y285A background was 32, which means that on average, 31 of 32 indels are corrected by MMR (compare Fig. 2 rates). Nevertheless, this is ~3-fold lower than that in the wt RNR background suggesting that some indels driven by this dNTP pool imbalance escape MMR. In addition, there were several major indel hotspots, mainly at G: C base pairs in mononucleotide repeats, with correction factors of 10 compared to the indels at A: T repeats which ranged from 200 to 800. This is a huge variation in the vital post-replication repair machinery that supports the notion of MMR efficiency being dependent on the dNTP pool imbalance, sequence context, and identity of the mismatch.

There are several possibilities as to why MMR is not efficient at these sites in the rnr1-Y285A strain. First, there could be a saturation of MMR due to the volume of errors induced by the

---

**Figure 5. Comparison of CAN1 mutation rates at hotspots (predominant mutation at site) in rnr1-Y285A strains with natural (OR1) and reversed (OR2) orientation of the CAN1 gene.** <Denotes that no events were detected.
doi:10.1371/journal.pgen.1004846.g005

**Figure 6. Mismatch repair efficiency.** A. Comparison of site mutation rates at major hotspots in rnr1-Y285A with (left axis) and without MMR (right axis). B. MMR correction factor at these sites in the presence of an rnr1-Y285A dNTP pool imbalance. <Denotes no events observed.
doi:10.1371/journal.pgen.1004846.g006
pool imbalance that are not corrected by proofreading [32]. Consider the hotspots at 795 and 857 bp which dominated the spectra in \textit{rnr1-12854}. The correction factor for the wt RNR strain was more than 2- and 5-fold higher than for the \textit{rnr1-12854} mutant at the 795 and 857 hotspots, respectively (S1 Table). Therefore, MMR was more accurate at repairing deletions at these sites in the wt RNR strain with normal dNTP pools. Second, MMR itself may require a natural dNTP pool balance in order to correctly repair mistakes. If MMR complexes recognize the mismatches generated but recruit an error-prone or even high fidelity polymerase, the imbalanced dNTP concentration may result in the same mismatch; thus, the mutation is maintained. Finally, some mismatches may not be subjected to MMR if they are damaged or generated outside of DNA replication [33–37].

Materials and Methods

Yeast strains

The \textit{CAN1} gene was replaced with \textit{URA3} from the pUG72 plasmid [38] (primers “\textit{CAN1 Del Ura3}” in Materials and Methods S1) in the RNR mutated strain (\textit{rnr1-12854}) previously described [16]. PCR amplified \textit{WT CAN1} in reversed orientation (primers “\textit{CAN1 orientation}” in Materials and Methods S1) was then transformed into the \textit{can1:: URA3} strain to give \textit{rnr1-12854 CAN1 OR2}. 5-FOA selection allowed the elimination of any \textit{can1:: URA3} cells and the \textit{CAN1} reverse orientation was confirmed by sequencing (“\textit{can1 ori scr}” primers in Materials and Methods S1).

An MMR deficient strain was created by deleting \textit{MSH2} in the AC402 wt (to give \textit{msh2D}) using the pAG32 plasmid and transfection technique [39] and the primers \textit{msh2_hphMX4}, shown in Materials and Methods S1. The deletion was confirmed using primers flanking \textit{MSH2}. This strain was then crossed with \textit{rnr1-12854} [16], sporulated and dissected spores selected on Hygromycin and –Trp plates for the double mutant \textit{rnr1-12854 msh2D}.

Culture conditions and Canavanine resistance assay

All culturing was carried out at 30°C in YPAD (1% yeast extract, 2% bacto-peptone, 20 mg/l adenine, 2% agar for plates) liquid cultures in a shaking incubator at 160rpm. The Canavanine extract, 2% bacto-peptone, 20 mg/l adenine, 2% agar for plates) were used to grow yeast cultures and the Canavanine resistance assay was used to calculate mutation rates as previously described [16]. The \textit{Can1} colonies were picked and the resistance assay was used to calculate mutation rates as previously described [16], [40,41]. The Canavanine resistance assay was used to calculate mutation rates as previously described [16,40,41]. The Can1 colonies were picked and the resistance assay was used to calculate mutation rates as previously described [16], [40,41].

dNTP pool measurements

dNTP pools were measured in asynchronous cultures as described previously [16] with minor changes as described in [42]. Briefly, cells were harvested by filtration at a density of 0.4×10^7 to 0.5×10^7 cells/ml and NTPs and dNTPs were extracted in trichloroacetic acid and MgCl₂ followed by a Fecon-trioctylamine mix. dNTPs were separated using boronate columns (Affigel 601, BioRad) and analysed by HPLC on a LaChrom Elite UV detector (Hitachi) with a Partisphere SAX column (Hichrom, UK).

Supporting Information

Figure S1 Full \textit{CAN1} mutation spectrum of \textit{rnr1-12854 CAN1-OR2} strain, showing individual mutations.

Figure S2 Full \textit{CAN1} mutation spectrum of \textit{msh2D} strain, showing individual mutations.

Figure S3 Full \textit{CAN1} mutation spectrum of \textit{rnr1-12854} \textit{msh2D} strain, showing individual mutations.

Figure S4 Base composition and mutation rates of \textit{CAN1} gene. A. Comparison of hotspot mutation rate and mononucleotide repeat length (all bases) in \textit{msh2D} strain. B. Base composition and mononucleotide repeat frequency of wild type \textit{CAN1} gene sequence.

Table S1 \textit{CAN1} hotspots observed in strains used in this study arranged into speculative classification as follows: (1) are \textit{msh2D} hotspots that are enhanced by high dCTP and dTTP levels, (2) are \textit{rnr1-12854} hotspots that are enhanced by the loss of MMR, and (3) are hotspots that exist in both single mutants and are enhanced in combination. Classification of hotspots observed in the \textit{rnr1-12854 msh2D} double mutant strain. MMR correction factor for wt and \textit{rnr1-12854} strains. \textit{<c>rate calculation based on if one event was observed. Only one event was observed.}

Materials and Methods S1 Primers used in this study.

Acknowledgments

We thank Dr. Jörgen Wiberg for creating the \textit{msh2D} strain and Anders Clausen and Scott Lujan for thoughtful comments on the manuscript.

Author Contributions

Conceived and designed the experiments: RJB DLW TAK AC. Performed the experiments: RJB DLW BC AKN. Analyzed the data: RJB DLW TAK AC. Wrote the paper: RJB DLW TAK AC.

References