Bacterial Genome Containing Chimeric DNA–RNA Sequences

Angad P. Mehta,†,# Yiyang Wang,†,# Sean A. Reed,‡ Lubica Supekova,† Tsotne Javahishvili,‡ John C. Chaput,§⊥⊥ and Peter G. Schultz*,†

†The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States
‡Bay Area Innovation Center, Corteva Agriscience, 4010 Point Eden Way, Hayward, California 94545, United States
§Department of Pharmaceutical Sciences, ‡Department of Chemistry, and ⊥⊥Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3958, United States

Supporting Information

ABSTRACT: Almost five decades ago Crick, Orgel, and others proposed the RNA world hypothesis. Subsequent studies have raised the possibility that RNA might be able to support both genotype and phenotype, and the function of RNA templates has been studied in terms of evolution, replication, and catalysis. Recently, we engineered strains of E. coli in which a large fraction of 2′-deoxycytidine in the genome is substituted with the modified base 5-hydroxymethyl-2′-deoxycytidine. We now report the generation of mutant strains derived from these engineered bacteria that show significant (~40–50%) ribonucleotide content in their genome. We have begun to characterize the properties of these chimeric genomes and the corresponding strains to determine the circumstances under which E. coli can incorporate ribonucleotides into its genome and herein report our initial observations.

INTRODUCTION

In all forms of life, genomic DNA predominantly contains A, C, G, and T. Modified bases are commonly observed in the genomes of bacteriophages, prokaryotes, and eukaryotes and play an important role in the regulation of transcription and restriction modification systems.1–3 In addition ribonucleotides have been found in the genomes of both prokaryotes and eukaryotes at low levels4–7 and there are viruses, bacteriophages, and viroids that possess RNA-only genomes, some of which replicate through reverse transcribed DNA and others that replicate with RNA-dependent RNA polymerases.4–6 Recently, we engineered strains of E. coli in which a large fraction of 2′-deoxycytidine in the genome is substituted with the modified base 5-hydroxymethyl-2′-deoxycytidine (5hmC). We now report the generation of mutant strains derived from these engineered bacteria that show significant (~40–50%) ribonucleotide content in their genome. Herein, we report our initial investigation into the presence and consequences of ribonucleotide content in their genome. We have begun to characterize the properties of these chimeric genomes and the corresponding strains to determine the circumstances under which E. coli can incorporate ribonucleotides into its genome and herein report our initial observations.

RESULTS

Pathway Engineering Coupled with Mutagenesis Results in E. coli Strains with rNMPs in the Genome

We previously used metabolic engineering to incorporate the modified base 5-hydroxymethyl-2′-deoxycytidine into the E. coli genome (Scheme 1AB).13,14 We manipulated the E. coli DNA pyrimidine biosynthetic pathways to generate large pools of ShmdCTP followed by systematic lowering of cytoplasmic dCTP pools. This was accomplished by introducing the following set of T4 bacteriophage genes into E. coli DH10B: gene 56, the phosphatase that removes dCTP and dCDP, as well as dUTP and dUDP, from the precursor pool; gene 42, dCMP hydroxymethylase; gene 1, the deoxyribonucleoside monophosphate kinase that catalyzes phosphorylation of dTMP, dGMP, and ShmdCMP (but not dCMP and dAMP);15 and gene cd, deoxyriboctyldate deaminase. Expression of these genes from plasmids pAM38 and pAM39, together with E. coli ndk (nucleoside diphosphate kinase), resulted in both the biosynthesis of ShmdCTP and a decrease in the

Received: July 4, 2018
Published: August 30, 2018
dCTP pool (Scheme 1B). As a consequence, approximately 60% of the genome had 5hmC substituted for 2′-deoxycytidine (dC) (a representative strain is designated as hmC-3 in this study).

Subsequent mutagenesis of clones transformed with pAM38 and pAM39 using N-methyl-N′-nitro-N-nitrosoguanidine (NTG) led to the observation of a morphologically distinct colony (designated as HR-1 strain). Genomic DNA was isolated from this colony, and 16S rDNA (rsrH) was PCR amplified using universal bacterial rDNA primers (8F/1492R targeting locus)16 and sequenced, confirming the identity of E. coli and absence of detectable contamination. Similarly, an individual HR-1 colony was picked after replating and confirmed to be E. coli. Microscopic images of HR-1 (Figure 1A) showed spherical shaped cells instead of canonical rod-shaped E. coli, whereas hmC-3 cultures appear to be a mixture of spherical and rod-shaped cells. Previously, spherical shaped E. coli (E. coli C) have been reported that show a similar clustering behavior to the HR-1 strain, but were not well characterized.17 Surprisingly, when the genomic DNA of HR-1 colonies was analyzed, we observed the presence of ribonucleotides in the isolated genome. Moreover, upon a second generation of HR-1 growth in the presence of antibiotics, we observed the complete absence of 5hmC and a further increase in ribonucleotides in the isolated genome. Genomic DNA was isolated by proteinase K-mediated cell lysis and RNase A treatment, followed by DNA affinity column based genomic DNA purification (Purelink Mini kit, Life Technologies). The yield was approximately 5- to 10-fold lower than that for the E. coli DH10B strain, likely due to the observed sensitivity of these genomes to degradation at basic pH (pH 8.5 at 55 °C for 45 min) and RNase A treatment during isolation conditions. The genomic DNA was digested into single nucleosides (using phosphodiesterase, Benzonase, and alkaline phosphatase)14 followed by liquid chromatography mass spectrometry (LCMS) analysis, revealing signifi-
significant levels of ribonucleosides: $rC/(rC + dC) \approx 4 \pm 2\%$, $rG/(rG + dG) \approx 53 \pm 5\%$, and $rA/(rA + dA) \approx 57 \pm 4\%$ (Figure 1B, red chromatogram). Intriguingly, we did not detect any uridine in the genomic digests above the limit of detection for LCMS (2% of total sample) (Figure S3), suggesting that the isolated genomes were not contaminated with native *E. coli* RNA.

To rule out the possibility of RNA contamination, we spiked total RNA (isolated total RNA from *E. coli* DH10B using Qiagen RNeasy mini kit) into the *E. coli* lysates that were used to isolate genomic DNA and did not detect ribonucleosides in these genomes (Figure 1C, green chromatogram). Similarly, we spiked both ribonucleosides and ribonucleotides into the genomic lysates and after isolation detected only trace levels of RNA in the isolates for the 2 $\mu$mol spike (Figure 1C, orange and blue chromatogram). These experiments confirmed that ribonucleotide detection in the genomic DNA digests of the HR-1 strain is not an artifact of genomic DNA isolation.

To determine if this trait is heritable, the HR-1 culture was grown on LB agar plates with antibiotics (tetracycline and spectinomycin, corresponding to the pAM38 and pAM39 markers, respectively), and individual colonies were picked. Genomic DNA isolation and analysis showed similar levels of ribonucleotides in the isolated genome (Figure S4), suggesting that the ribonucleotide-bearing strain is not a contaminant of a DNA-only genome strain. However, a significantly slower growth rate was observed for the HR-1 strain as compared to the parent *E. coli* DH10B strains (Figure S8).

**LCMS Analysis Demonstrates the Presence of Chimeric Oligonucleotides.** To determine whether the genome of HR-1 is a covalent DNA–RNA chimera rather than a DNA–RNA hybrid, we sought to establish a chemical linkage between a deoxyribonucleotide and ribonucleotide in the isolated genome. Genomic DNA isolated from *E. coli* DH10B and HR-1 strains was digested with alkaline phosphatase (CIP) and Benzonase, and the resulting oligonucleotides (dinucleotides and trinucleotides) were analyzed by LCMS. Canonical dinucleotides (e.g., dC-dG, dG-dG, T-dG, dC-T, dC-dC, T-T) were detected for *E. coli* DH10B. Genomes were then isolated from HR-1 (20-fold larger culture volumes were required due to the low yields) and digested, and the resulting oligonucleotides were HPLC purified and concentrated (Figure 2A and Figure S5A). This sample showed masses corresponding to canonical dinucleotides, as well as chimeric dinucleotides, i.e., rG-dG and dC-rG (Figure 2B and 2C). The oligonucleotides were then analyzed by extracted ion chromatograms followed by comparison with authentic standards generated from synthetic chimeric oligonucleotides: the retention times of the chimeric standards matched those of the dinucleotides generated from HR-1 (Figure S5C,D). Similarly, several chimeric trinucleotide signals were detected in the genomes isolated from HR-1 strains (Figure 2D). Additionally, chimeric di- and trinucleotides were also found in *E. coli* strain hmc-ntg-31 (Figure 2E and 2F), which was generated independently and also contains high genomic RNA content (*vide infra*). These experiments indicate...
the formation of a chimeric DNA–RNA sequence rather than just noncovalent hybrids.

**Metabolomic Analysis of dNTP and rNTP Pools.** One possible explanation for a high genomic ribonucleotide content is the modulation of dNTP pools by engineering dCMP metabolism. A decrease in dNTP levels, coupled with mutations in the DNA repair machinery, may result in the incorporation of rNMPs into the genome. Indeed, previous *in vitro* studies showed that some mutants of *E. coli* DNA polymerase I (PolA) incorporate ribonucleotides into DNA templates.18,19 Similarly, *in vitro* ribonucleotide incorporation into DNA templates by yeast replicative DNA polymerase and human DNA polymerase delta has been studied *in vitro*, and certain mutants of archaeal DNA polymerases (QGLK mutants) have been shown to efficiently incorporate rNTPs in a primer extension assay.22–25

To explore the possibility that a decrease in dNTP pools is responsible for the observed chimeric genome, we first analyzed the cellular free deoxyribonucleoside and ribonucleoside levels by lysing *E. coli* DH10B and HR-1 cells, passing the soluble cell lysate through a 3 kDa cut-off filter, treating the small molecule pool with alkaline phosphatase, and determining the relative levels of ribonucleosides and deoxyribonucleosides by LC-MS analysis (Figure 3A–D, Figure S6A–C). A significant drop was observed in the levels of dC, dA, and dG in *E. coli* HR-1 relative to *E. coli* DH10B (Figure 3B–D), and 5hmC levels were below the LCMS detection limit by extracted ion chromatogram. However, the levels of T were relatively constant (the ratio of T:U was difficult to quantify due to interfering signals). These intracellular deoxynucleotide levels are consistent with our genomic DNA analysis in which we detected rC, rG, and rA, but not U in the HR-1 genome. At present it is unclear as to why the levels of 5hmC drop in the HR-1 strains.

**In Vitro Primer Extension Assays with DNA Polymerases Demonstrate Incorporation of rNMPs in Product Oligonucleotides.** To further explore the notion that a decrease in dNTP pools leads to DNA polymerase mediated incorporation of RNA in the genome, we investigated whether bacterial or archaeal polymerases incorporate increased rNMPs into replicative templates as the ratio of rNTPs:dNTPs is increased. We initially used the readily available Klone fragment of *E. coli* DNA polymerase I and 3P rNTPs in an *in vitro* assay with the DNA template sequences shown in Figure 4A and 4C. In the absence of rNTPs, we observed the incorporation of dNTPs into templates as expected. Further, as the ratio of rNTP:dNTP increases, we detect higher incorporation of rNMPs in the template (~35% to 40% as the ratio of rNTP:dNTP reaches 1:1000, Figure 4B and 4D). This result is consistent with our hypothesis that an increase in the rNTP:dNTP ratio leads to incorporation of rNMPs into the genome. This result is also consistent with literature studies on nucleotide pool imbalance at replication forks.26

We also overexpressed and isolated the archaeal DNA polymerase mutant previously described, *Thermococcus* sp. 9N exo- (9n) polymerase (QGLK mutant), and performed primer extension assays with the DNA templates (described in the Supporting Information (SI)) in the presence of either only rNTPs or only dNTPs.27 We isolated the products of the

---

**Figure 3.** Metabolomics analysis of *E. coli* DH10B and HR-1 strains. (A) HPLC traces (270 nm): maroon trace, deoxyribonucleoside standards; blue trace, ribonucleoside standards; black trace, *E. coli* DH10B metabolome treated with alkaline phosphatase (30 U); green trace, HR-1 metabolome treated with alkaline phosphatase (30 U); red trace, hmC-ntg-31 metabolome treated with alkaline phosphatase (30 U). (B) [Deoxadenosine]:[adenosine] ratios in the analyzed metabolomes. (C) [Deoxyguanosine]:[guanosine] ratios in the analyzed metabolomes. (D) [Deoxycytidine]:[cytosine] ratios in the analyzed metabolomes. Note: Due to several overlapping UV signals, all ratios are the ratios of the peak areas of the extracted ion chromatograms of deoxyribonucleosides:ribonucleosides. Ratios are normalized to the phosphatase-treated *E. coli* DH10B metabolome. Identity of the analyzed metabolites was confirmed by co-injection, coelution, MS, and MS/MS. Quantitation method described in Figure S6.

**Figure 4.** (A) Template for incorporation of rATP in the Klenow fragment primer extension assays. (B) Amount of rA incorporation in product DNA (template shown in A). (C) Template for incorporation of rCTP in the Klenow fragment assays. (D) Amount of rC incorporation in the product DNA (substrate shown in D). (E) HPLC (254 nm) traces for primer extension assays with 9N DNA polymerase when only dNTPs are used as substrates. (F) HPLC (254 nm) traces for primer extension assays with 9N DNA polymerase (QGLK mutant) when only rNTPs are used as substrates. (G) HPLC quantitation of the observed levels of rNTPs in the primer extension assay. Blue: % rA/(rA + dA), green: % rC/(rC + dC), yellow: % rG/(rG + dG), red: % U/(U + T).

---

DOI: 10.1021/jacs.8b07046

polymerase assays and digested them with phosphodiesterase I, Benzonase, and alkaline phosphatase to generate single nucleosides, which were analyzed by HPLC as before to quantitate the levels of ribonucleosides and deoxyribonucleosides. Consistent with the literature, we observed efficient incorporation of rNTPs in primer extension assays (Figure 4E–G).

Further, we investigated if similar primer extension assays could be performed on larger DNA templates. To this end, primers were annealed to the M13 ssDNA genome, and primer extension assays were performed in the presence of rNTPs alone. As shown in Figure 5A, we detect a product band consistent with the size of the full-length M13 genome. Further, we showed that the product is sensitive to RNase H, consistent with the formation of a RNA–DNA hybrid product (Figure 5B). We further analyzed the ribonucleotide content of the product by digestion to single nucleosides and HPLC quantitation, which confirmed the efficient incorporation of rNTPs in primer extension assays with mutant archaeal polymerases (i.e., a ribonucleoside:deoxyribonucleoside ratio of ∼50:50 in isolated products, consistent with expected ratio for complete primer extension). In the absence of rNTPs, we observed the incorporation of dNTPs into templates as expected. Further, as the ratio of rNTP:dNTP increases, we detect efficient rNMP incorporation into the product template (∼75% to 100% of the expected value, depending on the nucleotide, as the ratio of rNTP:dNTP reaches 1:1000, Figure 5C). Again, this result is consistent with our hypothesis that an increase in the rNTP:dNTP ratio could lead to the incorporation of rNMPs into the genome. Collectively, these assays suggest that a combination of deoxynucleotide pool imbalance along with mutations in polymerase can result in the incorporation of rNTPs during genome replication. These experiments also suggest that it may be possible to rationally engineer or evolve an archaeal or C-family of bacterial polymerases that incorporate rNTPs into plasmids or genomes in vivo. Such defined systems will allow us to further probe those factors that affect genomic ribonucleotide content.

**Possibility of Multiple Genomes in the HR-1 Strain.** It is possible that multiple full or partial genomes are present in the HR-1 strain with a DNA-only genome being the functionally competent genome. It has been previously shown that with certain replication defects E. coli can accumulate more than one genome or generate so-called anucleate minicells with no genome at all.26 To investigate this possibility, E. coli DH10B and HR-1 were grown in LB medium in the presence of 32P-labeled phosphate salts (250 μCi, K2HPO4), and the genome was isolated using agarose plugs instead of column-based isolation methods (see SI for this experimental method). Gel-based calculations (Typhoon Scanner) were performed to calculate the amount of radioactivity in each of the genomes, and these levels were then correlated to the colony forming units (cfu) of E. coli DH10B and HR-1. Under these isolation conditions we observed approximately a 1:1 ratio (DH10B:HR-1 = 0.8 ± 0.3) of the 32P-labeled genome for E. coli DH10B and HR-1 strains (Figure S7). The results of these experiments indicate the presence of multiple genomes in HR-1 is unlikely.

**Alternate Protocol to Generate E. coli Strains with Chimeric DNA–RNA Genomes.** We next investigated alternate ways to generate strains with a similar phenotype and genotype to HR-1. To this end, we first created a random mutant library of E. coli DH10B using NTG and transformed these cells with plasmids pAM38 and pAM39.14 We observed both regular colonies, which showed only deoxyribonucleotides (including 5hmC) in the genome, and colonies with altered morphologies (hmC-ntg-2, hmC-ntg-31, hmC-ntg-33), which had rNMPs in their genomic digests. The levels of rNMPs varied from strain to strain (Figure 6A, rC/(rC + dC + 5hmC) ≈ 0.2–0.5, rG/(rG + dG) ≈ 0.3–0.6, rA/(rA + dA) ≈ 0.3–0.65). We chose one of the variants, hmC-ntg-31, to repeat the key experiments performed on the HR-1 strain. We verified that this colony was E. coli by sequencing 16S rDNA (rrsH) and also observed the presence of chimeric dinucleotides and trinucleotides during our mutation/tranformation protocol, no strains containing ribonucleotides were observed (Figure 2E and 2F). Consistent with the HPLC analysis of HR-1 and hmC-ntg-31 genomes, U-containing dinucleotides and trinucleotides are observed only in the case of hmC-ntg-31 and not in the case of HR-1 genomic digests.

To probe the relevance of the 5hmC pathway for RNA incorporation in the genome, we deleted the gene encoding T4 hydroxymethylase (T4 gp 42) from plasmid pAM38 (to obtain pAM148). When plasmids pAM148 and pAM39 were used (instead of pAM38 and pAM39) during the mutagenesis/ transformation protocol, no strains containing ribonucleotides were observed, demonstrating that alterations in the nucleotide pools resulting from dCMP metabolic engineering play a key role in the incorporation of rNMPs in the genome. However, the exact role of the hydroxymethylase in modulating dNTP levels remains to be determined. Further, the hmC-ntg-31 strain differs from HR-1 in the presence of significan levels of U in the genome, which may occur if there are additional defects in the uracil DNA glycosylase (UDG) or thymidylate synthase genes or gene expression in the hmC-ntg-31 strain. We PCR amplified and sequenced these genes in both the HR-1 and hmC-ntg-31 strains but observed no mutations in the coding region when compared to wild-type E. coli DH10B.
Pulsed Field Gel Electrophoresis (PFGE) to Analyze RNase Sensitivity of Genomes. To further verify that the genomes of the HR-1 and hmC-ntg-31 strains are chimeric in nature, we isolated large fragments of genomic DNA and analyzed their sensitivity to RNase digestion. Any DNA-only genome should resist RNase digestion. To this end, we made agarose plugs of *E. coli* DH10B and HR-1 cultures, lysed the cells in plugs using proteinase K at 55 °C, washed the plugs, and analyzed them by PFGE. An intense band was seen for the *E. coli* DH10B genome above 300 kb (Figure 7A). On the other hand, a significantly fragmented genome was observed for the HR-1 and hmC-ntg-31 strains (Figure 7A). This observation is consistent with the presence of ribonucleotides in the genome of HR-1 and hmC-ntg-31, as these chimeric sequences are known to be less stable than DNA sequences and may partially fragment under cell lysis conditions at 55 °C.

More revealing was the sensitivity of the genomes to RNases. As a positive control, we treated the agarose plugs containing *E. coli* DH10B, HR-1, and hmC-ntg-31 genomes with Benzonase (which hydrolyzes DNA and RNA templates) and observed complete fragmentation of all three genomes (Figure S10). Next, we tested the effect of RNases (RNase H and RNase I) on the genomes of *E. coli* DH10B, HR-1, and hmC-ntg-31 (Figure 7B and Figure 7C). Upon treatment of HR-1 and hmC-ntg-31 agarose plugs with RNase I (cleaves all RNA dinucleotide bonds and has a preference for ssRNA over dsRNA), almost all of the high molecular weight genomic nucleic acid was digested, whereas the wild-type *E. coli* DH10B genome remained intact (Figure 7B). When agarose plugs were treated with RNase H (hydrolyzes the phosphodiester bonds of RNA, which is hybridized to DNA and does not digest ssDNA or dsDNA), partial fragmentation of high molecular weight genomic nucleic acids was observed for the HR-1 and hmC-ntg-31 strains, whereas the wild-type *E. coli* DH10B genome again remained intact (Figure 7C). These results are consistent with the presence of a chimeric genome in which ribonucleotide tracts are frequently base-paired with deoxyribonucleotide tracts. Finally, we partially degraded the genomes isolated from the HR-1 and hmC-ntg-31 strains to...
determine the length of single-stranded deoxyoligonucleotides generated (ssDNA) after complete degradation of the ribonucleotide linkages. Since unlike deoxyribonucleotides, ribonucleotide linkages are sensitive to alkaline conditions, *E. coli* DH10B, hmC-3, HR-1, and hmC-ntg-31 genomic templates were treated with pH 10 buffer (10 mM Tris-HCl, 10 mM MgCl₂, pH 10) at 55 °C for 3 h, neutralized, and analyzed on 6% TBE-urea gels. For the HR-1 and hmC-ntg-31 strains we observed two sets of oligonucleotides: one between 300 nucleotides and 500 nucleotides and the other around 10 nucleotides and 20 nucleotides (Figure S9), suggesting the distribution of ribonucleotides in the genomic templates is relatively uniform. No such fragmentation was observed for *E. coli* DH10B genomic DNA.

**Sequencing Attempts and Proteomics Analysis.**

Attempts to transform the HR-1 or hmC-ntg strains by a variety of methods and plasmids were unsuccessful, which impeded our efforts to reverse engineer the HR-1 and hmC-ntg strains. However, we were able to amplify and sequence some of the polymerases, ribonucleotide reductases, RNases, and strains. However, we were able to amplify and sequence some of the polymerases, ribonucleotide reductases, RNases, and repair enzymes (polA, nrdE, nrdA, rnhA, rnhB, unuC) which showed no mutations. A caveat in this PCR-based sequencing (or any primer-based amplification techniques such as RT-PCR) is that there may be a heterogeneous population of chimeric genomes isolated from a pool of cells and our PCR probes target only those sequences (could be DNA-only sequences) that are similar or identical to wild-type *E. coli* DH10B sequences. Attempts to sequence the whole genome (Illumina sequencing) failed, likely due to the presence of a DH10B genomic DNA.

Representative overproduced protein bands were isolated for LCMS-based proteomic analysis. As expected, peptides corresponding to the wild-type *E. coli* proteins were observed from the gel bands isolated from the *E. coli* DH10B and hmC-3 proteomes. On the other hand, we could only match some of the peptides corresponding to the parent *E. coli* proteome from the gel bands isolated from the HR-1 and hmC-ntg-31 proteomes. Interestingly, the majority of these latter peptides showed one to three mutations (point mutations, insertions, and deletions were observed) when matched to the parent *E. coli* proteome (SI Excel sheet, for peptide analysis data). This level of mutation rate is unprecedented for NTG-mediated chemical mutagenesis. These results raise the possibility of a correlation between RNA incorporation in the genomic templates and a high mutation rate during replication or, based on our DNA sequencing results, more likely a decreased transcriptional fidelity of DNA-dependent RNA polymerases while transcribing chimeric DNA–RNA templates.

**Analysis of RNase H Mutants of *E. coli.***

Previous studies demonstrated that the genomic DNA from RNase H2-deficient budding yeast has a few thousand ribonucleotides incorporated per genome per cell cycle; similarly, RNase HII-null *B. subtilis* also has ribonucleotides incorporated into its genome.26 We therefore obtained RNase H mutants of *E. coli* (∆rnhA *E. coli* and ∆rnhB *E. coli*) to determine if these mutants have ribonucleotides embedded in their genome. Initially, we isolated the genomes from ∆rnhA and ∆rnhB strains of *E. coli*, digested them to single nucleosides, and analyzed the pool by LCMS.

We observed the presence of similar levels of rNMPs in both strains: U/(U + T) ≈ 2 ± 2%, rC/(rC + dC) ≈ 3 ± 2%, rG/(rG + dG) ≈ 5 ± 2%, and rA/(rA + dA) ≈ 6 ± 2% (Figure 9). This lower content of rNMPs is consistent with rNMPs being selected against during replication.30,31 These results also support the hypothesis that rNMPs are selected against during replication.

**Figure 8.** 2D-gel electrophoresis for the proteomes isolated from the *E. coli* DH10B, hmC-ntg-17, HR-1, and hmC-ntg-31 strains. All gels are 10% Bis-Tris Protein gels (NuPAGE, Thermo Fisher) and are silver stained. The circled bands in each case were isolated for LCMS-based peptide analysis.

**Figure 9.** HPLC chromatograms at 270 nm: pink trace, deoxyribonucleoside standards and ribonucleoside standards; blue trace, genomic digestion of ∆rnhA *E. coli* strain; violet trace, genomic digestion of ∆rnhA *E. coli* with complementation of rnhA expression; red trace, genomic digestion of ∆rnhB *E. coli* strain; green trace, genomic digestion of ∆rnhB *E. coli* with complementation of rnhB expression.
low level of ribonucleotides in the genome is likely either due to incomplete Okazaki primer removal or due to promiscuous misincorporation of ribonucleotides at low levels. Analysis of the sensitivity of genomes from these strains to RNase treatment (RNase I and RNase H) by PFGE as described above did not result in significant genomic digestion (Figures S11 and S12). Further, induction of rnhA or rnhB expression (using constructs pAM145 and pAM147) in E. coli ΔrnhA or E. coli ΔrnhB, respectively, decreased ribonucleotides below 2% in the corresponding genomes (Figure 9). We also investigated if we could generate strains similar to HR-1 and hmC-ntg by transforming E. coli ΔrnhA or E. coli ΔrnhB with the ShmC-encoding plasmids (pAM38/39). Interestingly, we observed a huge drop in colony-forming units when either the E. coli ΔrnhA or E. coli ΔrnhB strains were transformed with pAM38/39 as compared to E. coli DH10B. One explanation for this observation is that the E. coli ΔrnhA or E. coli ΔrnhB strains of E. coli have a highly compromised RNA repair mechanism, which when combined with expressed genes affecting dNTP pools likely crosses the threshold of RNA the genome can tolerate.

■ DISCUSSION

Low levels of ribonucleotides have been observed in genomes. For example, inactivation of RNase H II in the mouse embryonic fibroblasts leads to accumulation of more than 1 million rNMPs (ribo nucleotide monophosphates) per genome. Moreover, there are preferred sites for these embedded rNMPs, although their exact role has not yet been elucidated. Ribonucleotides have also been found at specific loci in mouse and human mitochondrial genomes, and a decrease in dNTP pools was suggested to result in increased ribonucleotide levels in the mitochondrial genome. In addition, studies have shown that genomic DNA from RNase H2-deficient budding yeast has a few thousand rNMPs incorporated per genome per cell cycle, and similarly, RNase HII null B. subtilis has rNMPs incorporated in its genome. Because RNase H is usually responsible for removing Okazaki primers during DNA replication of the reverse strand, the low level of rNMP incorporation in RNase H-deficient strains is most likely due to incomplete primer removal.

During our efforts to maximize the levels of ShmC by metabolic engineering of the E. coli genome we observed the presence of high levels of ribonucleotides (in some cases up to ~50%) in individual RNP populations based on LCMS analysis in the isolated genomes from mutant strains. Furthermore, we showed that (1) these genomes are base and RNase labile; (2) the presence of ribonucleotides is not due to contamination by RNA or noncovalent mRNA hybrids; (3) covalent linkages exist between deoxyribonucleotides and ribonucleotides in the isolated genome; and (4) multiple genomes are likely not present in the HR-1 strain. Based on the levels of rNMPs in the genome, the nature of the chimeric di- and trinucleotides present, the isolated deoxyribonucleotide-only fragment lengths, and the current understanding of Okazaki fragments involved in DNA replication, it is unlikely that primers are responsible for the high levels of ribonucleotides found in the genomes. Previous studies suggest that an increased dNTP level results in an overall decrease in the mean size of Okazaki fragments produced by both the primosome and the DNAB replication forks. However, to the best of our knowledge, a correlation between a drop in the dNTP pools and its effects on initiation, length, and repair of Okazaki fragments, and consequently ribonucleotide genome content, has yet to be investigated. Further, a significant digestion of intact chimeric genomes by RNase I suggests a relatively uniform distribution of rNMPs in the chimeric genomes. It is possible the observed rNMP levels could be the maximum possible levels that an E. coli cell could tolerate in terms of genome chemical structure and conformation, function, and viability of the corresponding cells.

Intriguingly, the strains that show chimeric genomes do not show the presence of ShmC in the genome, which could be due to loss of expression of T4 gp 42 from plasmid pAM38 or due to defects in other pathways (dNTP biosynthesis, repair enzymes, etc.) and needs further investigation. Unfortunately, our attempts at whole genome sequencing were unsuccessful, likely due to the lability of the chimeric genome. However, sequencing of candidate genes involved in DNA synthesis and repair did not show any mutations. Similarly, these strains could not be transformed with plasmids, likely due to their impaired growth rates.

Metabolomic analysis of the HR-1 strain revealed a significant drop in the deoxyribonucleoside levels. On the basis of this observation, we hypothesize that triggering of ribonucleotide incorporation is caused by a significant drop in the dNTP/rNTP ratio coupled with defects in DNA repair. Consistent with this notion, we showed that E. coli DNA polymerase I (Klenow fragment) incorporates ribonucleotides in vitro as the ratio of rNTPs:dNTPs increases. We also show that mutants of archaeal polymerases can efficiently copy the entire M13 genome in the presence of rNTPs. We are currently attempting to generate mutants of archaeal and E. coli polymerases that incorporate rNTPs more efficiently and are exploring the effects of ShmC on ribonucleotide reductase activity.

We also analyzed E. coli ΔrnhA and ΔrnhB mutants and observed higher levels of rNMPs in the chromosome as compared to the E. coli DH10B genome. However, these levels were significantly below those of the HR-1 strain, and attempts to increase the ribonucleotide content of the RNase H mutant strains were unsuccessful. This is likely due to the inability of such strains to survive with a highly compromised RNA repair mechanism when combined with expressed genes affecting dNTP pools. The hmC-ntg-31 strain, which was generated by a distinct protocol, differs from HR-1 in the presence of U in the genome, which may occur if there are additional defects in the UDG gene in the hmC-ntg-31 strain. Finally, proteomic studies raised the possibility of a correlation between RNA incorporation in the genome and a high proteome mutation rate, which will be explored in future studies involving transcriptomics analysis to determine transcriptional fidelity of chimeric DNA–RNA genomes.

■ CONCLUSION

The observations reported herein raise several important questions related to the ability of chimeric DNA–RNA sequences to function in replication, recombination, restriction, and repair, as well as in transcription. Previous in vitro studies and this report indicate that the DNA polymerases can incorporate ribonucleotides as the levels of deoxyribonucleotides drop. In addition earlier studies have demonstrated that chimeric DNA–RNA templates are competent for transcription. A more recent report suggests that homogeneous DNA genomes may have evolved from heterogeneous chimeric
RNA–DNA templates. Further investigations are required to investigate the origin and function(s) of these chimeric DNA–RNA templates, their relevance to replication, transcription, and repair, and the possibility that they may provide insights into the transition from the RNA to DNA world.

**MATERIALS AND METHODS**

**Growth Conditions and Isolation of Genomic Nucleic Acid.** For each experiment cells were plated under antibiotic selection (5 μg/mL tetacycline, 50 μg/mL spectinomycin). Individual colonies were picked and inoculated in LB media in the presence of antibiotics. Cultures were incubated at 37 °C for 18 to 22 h. Cells were harvested by centrifugation. Unless otherwise noted, genomic DNA was isolated by a Purelink genomic DNA isolation mini kit using the manufacturer’s protocol. For the HR-1 strain, the proteinase K lysis was performed for 45 min. Minimal media consisted of M9 minimal salts along with glucose or glycerol as the carbon source, MgSO4, KCl, and deoxyribonuclease (25 μg/mL). Agarose plugs were then washed twice with TE buffer (10 mM Tris-HCl, pH 8, 0.2% sodium deoxycholate, 1% sodium lauryl sarcosine, 1 mg/mL proteinase K), 500 μL reaction volume), and analyzed by LCMS.

**Digestion of the Genomic Nucleic Acid to Individual Nucleosides.** Genomic nucleic acids (25 ng/μL) were incubated with phosphodiesterase I (12 μM, MP Biomedical), calf intestinal phosphatase (30 U, NEB), and Benzonase (25 U, Millipore) in the reaction buffer (10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl2, pH 8.0, 100 μL total reaction volume) at 37 °C for 4 h for complete breakdown to individual nucleosides. The small molecule pool was isolated using 3 kDa cutoff filters (VWR) and analyzed by HPLC. HPLC conditions: Luna 5 μm C18 100 Å 250 × 2 mm; solvent A: water; solvent B: 5 mM potassium phosphate buffer, pH 6.7; solvent C: methanol; flow rate: 0.3 mL/min; 0 min, 100% B; 7 min, 100% B; 12 min, 25% A, 60% B, 15% C; 18 min, 25% A, 10% B, 65% C; 21 min, 100% A; 30 min, 100% A. Agilent 1260 HPLC. For hmC-ntg-31 digestion, 100 μL total reaction volume) at 37 °C for 3 h. The small molecule pool was isolated using 3 kDa cutoff filters (VWR). The oligonucleotide fraction (fraction 2, Figure S3A) was HPLC purified. HPLC conditions: Agilent Zorbax SB-C18 (5 μm, 4.6 × 150 mm); solvent A: water; solvent B: 5 mM ammonium acetate buffer, pH 6.7; solvent C: 0.1% formic acid in acetonitrile; flow rate: 0.5 mL/min; 0 min, 100% B; 7 min, 100% B; 12 min, 70% A, 30% B; 17 min, 30% A, 70% B; 19 min, 100% A; 28 min, 100% A.

**Sample Preparation for LCMS Detection of Di- and Trinucleotides.** Genomic nucleic acids (25 ng/μL) were incubated with calf intestinal phosphatase (30 U, NEB) and Benzonase (25 U, Millipore) in reaction buffer (10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl2, pH 8.0, 100 μL total reaction volume) at 37 °C for 3 h. The small molecule pool was isolated using 3 kDa cutoff filters (VWR). The oligonucleotide fraction (fraction 2, Figure S3A) was HPLC purified. HPLC conditions: Agilent Zorbax SB-C18 (5 μm, 4.6 × 150 mm); solvent A: water; solvent B: 5 mM ammonium acetate buffer, pH 6.7; solvent C: methanol; flow rate: 0.5 mL/min; 0 min, 100% B; 7 min, 100% B; 12 min, 25% A, 60% B, 15% C; 18 min, 25% A, 10% B, 65% C; 21 min, 100% A; 30 min, 100% A. Instrument: Agilent 1260 HPLC. The collected fractions were lyophilized and dissolved in water. These samples were then analyzed by LCMS. The sequences of IDT-synthesized chimeric standards are listed in Table S2.

**Agarose Plugs for PFGE.** Cells were resuspended in PBS and incubated at 55 °C for 10 min. An equal volume of 2% agarose was then mixed with cells. A 100 μL amount of this mixture was then added to well plug molds (Bio-Rad # 1703713), and the plugs were allowed to solidify. Plugs were then transferred to plug buffer (100 mM EDTA, pH 8, 0.2% sodium deoxycholate, 1% sodium lauryl sarcosine, 1 mg/mL proteinase K, 500 μL/plug), and the plugs were incubated overnight at 55 °C. Plugs were then washed three times with TE buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), incubated in TE buffer containing 1 mM phenylmethylsulfonyl fluoride (PMSF) for 30 min at room temperature, and stored at 4 °C prior to PFGE analysis. PFGE settings: 0.5X TBE buffer (Corning # 46-011-CM), initial pulse 5 s, final pulse 5 s, voltage 6 V/cm, time 24 h, 14 °C. Ethidium bromide (1 mg/mL) solution was used for staining.

**Metabolome Analysis.** Cell cultures of 5 mL were harvested by centrifugation and resuspended in 1 mL of buffer (10 mM Tris-HCl, pH 8.0 buffer containing 4 mg of lysozyme). The mixture was incubated at 4 °C for 1 h and then sonicated to lyse cells. The mixture was spun down at 20000g for 15 min, and the supernatant was added to 3 kDa cut-off filters (VWR). The flow-through was collected, treated with alkaline phosphatase (30 U) in reaction buffer (10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl2, pH 8.0, 100 μL total reaction volume), and analyzed by LCMS.

**Primer Extension Assays on the M13 ssDNA Genome with the Thermococcus sp. 9N' exon- (9N) Polymerase.** A 10 pmol primer–template mixture was annealed in a 50 μL reaction volume of Thermopol buffer (20 mM Tris-HCl, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1% Triton X-100, pH 8.8) by heating for 5 min at 90 °C and cooling for 10 min at 4 °C. For primer extension assays the 9N polymerase (5 μL volume, 0.7 mg/mL) was added directly to the reaction mixture. Nucleotide triphosphates (rNTPs, 400 μM) or deoxynucleotide triphosphates (dNTPs, 400 μM) were added to the reaction mixture to initiate the reactions at 55 °C (total reaction volume 57 μL). The extension times for RNA and DNA synthesis were 8 and 4 h, respectively. The following primers were annealed to M13 ssDNA:


For the rNTP:dNTP competition experiments in the primer extension assays, the following concentrations were used: 1:1 rNTP:dNTP = 1 mM rNTP and 1 mM dNTP; 10:1 rNTP:dNTP = 10 mM rNTP and 1 mM dNTP; 100:1 rNTP:dNTP = 10 mM rNTP and 0.1 mM dNTP.

**RNase Treatment of Agarose Plugs.** E. coli DH10B, HR-1, hmc-nrng-31, E. coli ΔnraA, and E. coli ΔnraB strains were grown in LB medium (with antibiotics for hrCas-31, HR-1, E. coli ΔnraA, and E. coli ΔnraB strains) Genomic nucleic acid was isolated using an agarose-plug-based isolation protocol described above. Agarose plugs were washed twice with TE buffer (10 mM Tris-HCl pH 7.4, 1 mM EDTA) and once with water. Plugs were then treated with RNase I (150 U, New England Biolabs), RNase H (20 U, New England Biolabs), NEB), or Benzonase (0.5 U) with NEB recommended buffers (500 μL total reaction volume). Reactions were performed overnight at 37 °C for RNase H treatment and 6 h for RNase I treatment. The genomic nucleic acids were then analyzed by PFGE gels followed by ethidium bromide staining.

**ASSOCIATED CONTENT**

**Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b07046.

Additional experimental procedures, data figures, and proteomics data (PDF)

Peptide analysis data (XLSX)

**AUTHOR INFORMATION**

**Corresponding Author**

*schultz@scripps.edu*


This work was supported by the California Institute of Biomedical Research. We would like to acknowledge Prof. Eros Lasserini Denchi and Tatevik Simavonian for their assistance with PFGE and Joseph H. Davis for his assistance with microscopy. We would like to thank Prof. Thomas Cech, Prof. David Wemmer, Prof. Gerald Joyce, Prof. Harry Noller, Prof. Jack Szostak, Dr. Richard Glynne, Prof. Alan Grossman, Prof. David R. Liu, Prof. Christopher T. Walsh, Prof. JoAnne Stubbe, Prof. Utaam RajBhandary, and Prof. Jamie Williamson for their critical comments and suggestions. We would like to acknowledge Kristen Williams for her assistance in manuscript preparation and Faith Hark for her assistance with graphics.

This work was supported by the California Institute of Biomedical Research. We would like to acknowledge Prof. Eros Lasserini Denchi and Tatevik Simavonian for their assistance with PFGE and Joseph H. Davis for his assistance with microscopy. We would like to thank Prof. Thomas Cech, Prof. David Wemmer, Prof. Gerald Joyce, Prof. Harry Noller, Prof. Jack Szostak, Dr. Richard Glynne, Prof. Alan Grossman, Prof. David R. Liu, Prof. Christopher T. Walsh, Prof. JoAnne Stubbe, Prof. Utaam RajBhandary, and Prof. Jamie Williamson for their critical comments and suggestions. We would like to acknowledge Kristen Williams for her assistance in manuscript preparation and Faith Hark for her assistance with graphics.

ACKNOWLEDGMENTS

The authors declare no competing financial interest.

REFERENCES
