Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2019 January 17.

Published in final edited form as:

J Am Chem Soc. 2018 January 17; 140(2): 758–765. doi:10.1021/jacs.7b11488.

Reprograming the replisome of a semi-synthetic organism for the expansion of the genetic alphabet

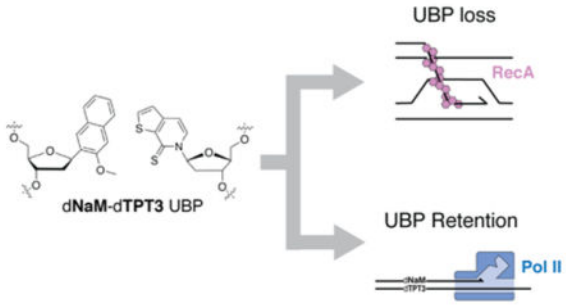
Michael P. Ledbetter, **Rebekah J. Karadeema**, and **Floyd E. Romesberg***

Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037 USA

Abstract

Semi-synthetic organisms (SSOs) created from *Escherichia coli* can replicate a plasmid containing an unnatural base pair (UBP) formed between the synthetic nucleosides d**NaM** and d**TPT3** (d**NaM**-d**TPT3**) when the corresponding unnatural triphosphates are imported via expression of a nucleoside triphosphate transporter. The UBP can also be transcribed and used to translate proteins containing unnatural amino acids. However, UBPs are not well retained in all sequences, limiting the information that can be encoded, and are invariably lost upon extended growth. Here we explore the contributions of the *E. coli* DNA replication and repair machinery to the propagation of DNA containing d**NaM**-d**TPT3** and show that replication by DNA polymerase III, supplemented with the activity of polymerase II and methyl-directed mismatch repair contribute to retention of the UBP and that recombinational repair of stalled forks is responsible for the majority of its loss. This work elucidates fundamental aspects of how bacteria replicate DNA and we use this information to reprogram the replisome of the SSO for increased UBP retention, which then allowed for the first time the construction of SSOs harboring a UBP in their chromosome.

Graphical Abstract



*Corresponding Author: floyd@scripps.edu.

Notes

The authors declare the following competing financial interests: a patent application has been filed based on the use of UBPs in SSOs and F.E.R. has a financial interest (shares) in Synthorx, Inc., a company that has commercial interests in the UBP.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Additional Methods and Supporting Figures (PDF)

Table of plasmids and primers (XLSX)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

INTRODUCTION

An expanded genetic alphabet would increase the information that can be stored in a cell and facilitate the creation of semi-synthetic organisms (SSOs) that use this increased information to create novel forms and functions, the central goal of synthetic biology.[1] Towards this goal, we have developed a family of synthetic nucleotides that form unnatural base pairs (UBPs) via only hydrophobic and packing forces, most notably the UBP d**NaM**-d**TPT3** (Figure 1*A*).[2,3] We have shown that when the corresponding triphosphates (d**NaM**TP and d**TPT3**TP) are imported into *Escherichia coli* via transgenic expression of the nucleoside triphosphate transporter *Pt*NTT2,[4] they are used by cellular polymerases to replicate a plasmid containing the UBP.[5] Recently we have also demonstrated that such SSOs can transcribe DNA containing the UBP into mRNAs and tRNAs with cognate unnatural codonanticodon pairs that efficiently direct the incorporation of non-canonical amino acids into proteins.[6] It is also of particular interest that the UBP forms without the aid of complementary hydrogen-bonds, demonstrating that hydrophobic and packing forces can productively participate in every step of information storage and retrieval.

While the reported SSO demonstrated that a man-made part may be designed to function within what is perhaps the most central of all biological processes, its retention of the UBP is sequence-dependent, which limits the number of unnatural codons available, and the SSO invariably lost the UBP during extended growth.[3] Both of these limitations may be mitigated by applying selection pressure for triphosphate up-take and UBP retention via expression of Cas9 directed to cleave, and thus degrade DNA sequences that have lost the UBP.[3] However, even with this error elimination mechanism, retention remains challenging in some sequence contexts, and moreover, this approach requires optimizing different guide RNAs for every sequence to be retained, which is impossible or challenging with many applications, for example, those involving propagation of random DNA sequences. In addition, encoding information with the UBP in the chromosome as opposed to a plasmid, a long term goal of the project, was expected to be incompatible with applying this selection pressure due to undesired cleavage of UBP-containing sequences and/or because cleavage would result in destruction of the chromosome as opposed to the less consequential elimination of one of many copies of a plasmid. Thus, elucidating the mechanism of how DNA containing the UBP is replicated might not only provide fundamental insights into cellular physiology, but might also identify approaches to optimize the SSO, and possibly even facilitate the creation of SSOs with UBPs in their chromosome.

Under steady-state conditions, DNA containing the d**NaM**-d**TPT3** UBP is replicated *in vitro* with an efficiency approaching that of a fully natural counterpart;[2,7] however, these rates are likely limited by product dissociation. *In vivo* replication is more processive, and correspondingly less likely to be limited by product dissociation. Therefore, replication of DNA containing the UBP in the SSO may be less efficient than that of fully natural DNA, and in turn, may cause replication forks to stall. Additionally, structural studies have indicated that the UBP adopts a Watson-Crick-like structure during triphosphate insertion, but once inserted, the UBP adopts a cross-strand intercalated structure that induces local helix distortions.[8,9] Cells interpret both stalled replication forks and helix distortions as signs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of DNA damage and initiate programs to repair or tolerate the offending nucleotides, which we suspected might contribute to UBP loss.

To determine how cells retain or lose the UBP, we examined the effects of disabling these pathways. We found that neither nucleotide excision repair (NER) nor the SOS response contribute significantly to UBP retention or loss. Conversely, the normal replisome polymerase, DNA polymerase III (Pol III), Pol II, and methyl-directed mismatch repair (MMR), all contribute to UBP retention; while recombinational repair (RER) of replication forks that stall provides the major route to UBP loss. This understanding allowed us to reprogram the replisome of the SSO and impart it with the ability to not only better retain the UBP on a plasmid, but also to stably harbor a UBP in its chromosome.

RESULTS AND DISCUSSION

Nucleotide excision repair does not contribute to UBP retention or loss

Generally, *E. coli* responds to DNA damage via direct damage reversal, base excision repair, NER, MMR, RER, and the SOS response. Neither direct damage reversal nor base excision repair is likely to contribute to UBP retention or loss, because these pathways rely on enzymes that recognize specific forms of DNA damage which are not likely to be mimicked by the UBP. In contrast, NER, MMR, RER, and the SOS response, are induced by less structure-specific signals. Thus, to begin to explore how cells manage to retain the UBP in their DNA we examined NER, which is mediated in a replication-independent manner by a complex of proteins that scan DNA for distortions resulting from bulky lesions that may be mimicked by the UBP.[10] Contributions of NER to UBP retention or loss were explored by deleting *uvrC*, which encodes an essential component of NER, from the parental SSO (*E. coli* BL21(DE3)+pACS2 (Figure S1)). Replication of DNA containing the d**NaM**-d**TPT3** UBP positioned in two different sequence contexts in plasmids pINF1 and pINF2 was unaffected by deletion of *uvrC*, indicating that NER makes no contribution to UBP retention or loss (Figure 1*B*).

Methyl-directed mismatch repair increases UBP retention

We next examined MMR, which provides the critical first check of newly synthesized DNA as it emerges from a DNA polymerase during replication and is mediated by a complex of proteins that recognizes helix distortions caused by mismatched natural nucleotides.[11] Upon detection of a mismatch, the MMR complex nicks the newly synthesized, unmethylated strand, which in turn leads to gap formation and subsequent resynthesis of the DNA. In contrast to NER, deactivation of MMR via deletion of *mutH* resulted in a reduction in UBP retention with both pINF1 and pINF2 (Figure 1*B*). These results indicate that the helix distortions associated with the UBP are not sufficiently severe to activate MMR or that the unnatural nucleotides cannot be excised, but that the distortions caused by the pairing of an unnatural and a natural nucleotide are recognized and processed by MMR. Thus, MMR appears to effectively recognize the UBP as natural-like and selectively removes mispaired natural nucleotides, thereby supporting the stable expansion of the genetic alphabet.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Recombinational repair provides the major route to UBP loss

RER is mediated by RecA, which forms filaments on single-stranded DNA ahead of stalled replication forks, in turn, facilitating the formation of recombination intermediates and switching to a homologous template for continued DNA replication.[12,13] The SOS response is induced when the same RecA filaments promote cleavage of the SOS repressor LexA, which leads to the derepression of a variety of genes involved in the tolerance and/or repair of the damaged DNA that caused the fork to stall.[14] We explored the combined contribution of RER and the SOS response through the deletion of *recA* and observed a small but significant increase in UBP retention with pINF1 (Figure 1*B*). To further explore the contribution of RecA, retention of the UBP in the more challenging sequences provided by pINF3, pINF4, and pINF5,[3] was measured in the Δ*recA* SSO (Figure 1*C*). In these sequence contexts, the absence of *recA* resulted in a more dramatic increase in UBP retention.

To discern if *recA* deletion facilitates UBP retention by ablating RER or by preventing the induction of the SOS response we examined an SSO that is unable to induce the SOS response, but which is competent for RER (SSO *lexA*(S119A)) (Figure 1*C*). While selective suppression of the SOS response resulted in moderately increased UBP retention with pINF3, the increase was less than that observed with the Δ*recA* SSO. With pINF4 and pINF5, selective SOS suppression resulted in only modest increases in UBP retention that were well below those observed with the *recA* SSO. These results demonstrate that the majority of UBP loss mediated by RecA occurs via RER and not via induction of the SOS response.

Pol II contributes to the replication of DNA containing the UBP

While the data clearly suggests that the majority of UBP loss is mediated via RER, the marginal and sequence-specific increase in UBP retention with the *lexA*(S119A) SSO suggests that one or more SOS regulated proteins may also contribute. While *recA* is itself partially repressed by LexA[14]) we were particularly interested in characterizing the contribution of the three SOS-regulated DNA polymerases, Pol II, Pol IV, and Pol V. Indeed, Pol IV and Pol V are “translesion” polymerases that are well known for their ability to mediate DNA synthesis across “non-instructional” damaged nucleotides.[15] However, deletion of both *dinB* and *umuCD* (which encode Pol IV and the precursor of Pol V, respectively) did not impact UBP retention with either pINF1 or pINF2 (Figure 1*D*). In contrast to the Δ*dinB*Δ*umuDC* SSO, the deletion of *polB* (which encodes Pol II) resulted in a dramatic increase in UBP loss with both pINF1 and pINF2 (Figure 1*D*). Overall, these data demonstrate that RER constitutes the major route to UBP loss and that Pol II provides an important route to UBP retention. While the production of Pol II is increased by the induction of SOS[16], the data suggests that its beneficial role is overwhelmed by the deleterious effects of the concomitantly induced RER.

DNA polymerase III also contributes to the replication of DNA containing the UBP

The reduced but still detectable retention of the UBP in the Δ*polB* SSO, along with the negligible effects of deleting the genes encoding Pol IV and Pol V, strongly suggest that one or both of the remaining DNA polymerases, Pol I and Pol III, must also contribute to retention of the UBP. To specifically examine whether Pol I or Pol III contribute to the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

replication of DNA containing the UBP, we constructed and characterized strains in which their 3′—5′ exonuclease "proofreading" activity was eliminated or impaired via mutation (Pol I^{exo-}, *polA*(D424A, K890R)[17] and Pol III^{exo-}, *dnaQ*(D12N),[18] respectively (see Supporting Information and Figures S1 and S3) (Figure 1*E*, Figure S2). While the deletion of Pol I exonuclease activity had no effect on UBP retention, the Pol III exonuclease deficient mutant showed a dramatic reduction in UBP retention. This data clearly indicates that in wild type cells, Pol III, but not Pol I, contributes to the replication of DNA containing the UBP.

To determine if any effects of the Pol I or Pol III mutations were masked by the activities of Pol II and/or RER, we examined UBP retention in the Δ*polB*, or Δ*polB*Δ*recA* SSO. Indeed, we found that the UBP was well retained with the Δ*polB*Δ*recA* SSO, clearly demonstrating that polymerases other than Pol II are capable of mediating high-level UBP retention in the absence of competition with RER-mediated loss (Figure 1*D*). The Pol III exonuclease mutant again showed decreased UBP retention in both Δ*polB* and Δ*polB*Δ*recA* SSOs. However, in contrast to wild type cells, the deletion of Pol I exonuclease activity had significant and opposite effects with the Δ*polB* and Δ*polB*Δ*recA* SSO, in which retention increased and decreased, respectively. These data demonstrate that in addition to Pol II, Pol III contributes to the retention of the UBP, and in the absence of RER, Pol I does as well.

A model for the replication of DNA containing the UBP

Based on the data presented above, we propose the following model for replication of DNA containing the d**NaM**-d**TPT3** UBP in the *E. coli* SSO (Figure 2). When the replisome with Pol III encounters an unnatural nucleotide during processive leading or lagging strand replication, Pol III incorporates either a natural or an unnatural nucleotide. If a natural nucleotide is incorporated, the rate of proofreading is competitive with, and perhaps more efficient than continued extension, and thus the natural nucleotide is commonly excised via the proofreading activity of Pol III. However, if a correct UBP is synthesized, more efficient extension prevents excision and the replisome continues synthesizing DNA. As it exits the polymerase, the nascent duplex is scanned by the MMR complex, which further increases UBP retention by preferentially eliminating any mispaired natural nucleotides that escaped proofreading.

Because extension of even a correct UBP is likely to be less efficient than natural synthesis, Pol III may also dissociate. The stalled fork, likely with the extending strand terminated immediately before the unnatural nucleotide in the template, is a now a substrate for RER, which reinitiates synthesis using a homologous natural sequence and thus provides the dominant mechanism for UBP loss. However, in competition with RecA-mediated RER, Pol II can rescue the stalled fork and reinitiate synthesis with high UBP retention, after which it presumably yields to Pol III and the reestablishment of a normal replication fork. The contribution of Pol I is more complex. In wild type cells, Pol I does not appear to contribute to the replication of DNA containing the UBP. In contrast, in the absence of Pol II and RecA, Pol I does contribute and correspondingly, the deletion of Pol I exonuclease activity results in decreased UBP retention. However, if the exonuclease activity is eliminated, Pol I

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

can contribute if Pol II is eliminated, and in this case it increases retention by competing with RER.

Perhaps the most interesting aspect of this mechanism is the clear role played by Pol II, whose natural function has remained enigmatic despite the nearly fifty years that have passed since its discovery.[19] Two putative roles for Pol II have been broadly accepted. One is in replication restart where Pol II rescues stalled forks after Pol III synthesizes a mispair that it cannot efficiently extend.[20,21] The second broadly accepted role for Pol II is to compete with RER to fill in gaps created by NER as part of the cellular response to interstrand cross-linked DNA.[22] Interestingly, the evoked role of Pol II in rescuing replication forks stalled at the UBP in competition with RER is strikingly similar to aspects of both of the putative natural roles. However, to our knowledge this effect on the replication of DNA containing the UBP is the most significant phenotype ever observed with its elimination.

Optimization of the SSO

An intriguing consequence of the model proposed above is that it suggests that UBP retention might be optimized through the manipulation of RecA and Pol II. To explore this possibility, we constructed SSOs lacking *recA* and with or without Pol II constitutively expressed at SOS-derepressed levels (Δ*recA* and Pol II^{+}Δ*recA*, respectively (see Supporting Information and Figure S3). These strains also expressed an optimized *Pt*NTT2 transporter from a chromosomal locus (Δ*lacZYA::*P_{lacUV5}-*PtNTT2*(66–575))[3] (see Supporting Information and Figure S1). For comparison, we constructed the wild type strain with the same chromosomally integrated transporter (WT-Opt). SSOs were transformed with pINF1, pINF5, or pINF6 (Figure 3*A*), with pINF6 embedding the UBP in a sequence where its retention is particularly challenging, and plasmids were recovered from individual colonies to characterize UBP retention. In this case, selection on solid growth media was introduced to allow for analysis of UBP retention in individual clones, as opposed to the average UBP retentions determined in the previous experiments. A distribution of UBP retentions was observed with each plasmid in all SSOs, however, the distributions were shifted toward higher retention with the Δ*recA*-Opt and especially the Pol II^{+}Δ*recA* SSOs, compared to the WT-Opt SSO. Additionally, only the Pol II^{+}Δ*recA* SSO produced clones with undetectable UBP loss in each sequence context examined. Notably, this was even true with pINF6, for which retention in the wild type SSO was undetectable, and only moderate (<60%) when enforced with Cas9 selection.[3]

With the genetically optimized Δ*recA*-Opt and Pol II^{+}Δ*recA* SSOs in hand, we next evaluated whether it could facilitate integration of the UBP into the chromosome. We constructed an integration cassette that targets the sequence GTA**X**TGA (**X** = **NaM**) to the *arsB* locus, and used lambda red recombineering to integrate the cassette into the chromosomes of the WT-Opt, Δ*recA*-Opt, and Pol II^{+}Δ*recA* SSOs. Screening of integrants for UBP retention identified clones with 100% retention from the Δ*recA*-Opt and Pol II^{+}Δ*recA* SSOs, but despite significant effort we were unable to isolate WT-Opt clones with greater than 91% UBP retention (see Figure S4), suggesting that significant UBP loss occurred during the required growth step. To characterize the effect of the chromosomally integrated UBP, aliquots of mid-log phase cells were inoculated into growth media with or

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

without d**NaM**TP and d**TPT3**TP (Figure 3*B*, Figure S5). The Δ*recA*-Opt and Pol II^{+}Δ*recA* integrants grew poorly when the unnatural triphosphates were not provided, consistent with the model that RER is required to efficiently bypass an unnatural nucleotide in the template. However, this growth defect was almost entirely eliminated in both SSOs when d**NaM**TP and d**TPT3**TP were provided. Thus, the deletion of *recA* and the overexpression of Pol II facilitate high-level retention of the UBP in the chromosome with only minimal consequence to fitness.

Finally, we explored whether the genetically optimized strains facilitate the long-term stability of the chromosomally integrated UBP. Previous studies have demonstrated that without Cas9-mediated selection for retention, a plasmid-borne UBP is lost during extended growth.[3] The WT-Opt, Δ*recA*-Opt, and Pol II^{+}Δ*recA* integrants were serially passaged over many generations of growth and UBP retention characterized (Figure 3*C*). With WT-Opt, the UBP was slowly lost until approximately the 40th generation, and then lost more rapidly with complete loss observed by the 90th generation. The apparently biphasic kinetics of loss suggest that at least one additional process contributes in addition to RER. Indeed, sequencing revealed a gross chromosomal rearrangement that eliminated the *Pt*NTT2 gene at the time of the precipitous drop in UBP retention (Figure S7). In contrast to WT-Opt, both the Δ*recA*-Opt and Pol II^{+}Δ*recA* SSOs, the *Pt*NTT2 remained intact and retention of the genomic UBP remained high, especially with the Pol II^{+}Δ*recA* SSO, where it remained >55% after 137 generations. These results demonstrate that not only does *recA* deletion facilitate UBP retention during replication, it significantly increases transporter stability during extended growth. The observed retention corresponds to a fidelity per doubling in excess of 99.6%, which in turn corresponds to loss of the chromosomal UBP in only a small fraction of the cells (<0.4%) per doubling. Thus, along with the Cas9-error elimination system, which was not employed in the current work, this error prevention system should allow for the retention of the UBP in a wide range of sequence contexts, which in turn should enable the storage of the entirety of the new information made possible by the UBP.

CONCLUSIONS

Since the last common ancestor of all life on earth, biological information has been stored in a four-letter alphabet. The reprogramed replisome of the Pol II^{+}Δ*recA* SSO represents significant progress toward the unrestricted expansion of this alphabet, and the first progress mediated through the optimization of the cell itself. While the primary goals of the research were to understand how the UBP is replicated and to use that information to optimize the SSO, the results also provide a novel route to the study of how challenging replication is normally managed. For example, while the data suggests that a significant fraction of the DNA containing the UBP is replicated by Pol III, it also clearly reveals that a significant amount is not, and in these cases, the data reveal an interesting competition between Pol II-mediated replication restart and RecA-mediated RER. Such competitions may be common during challenging replication, which may have contributed to the challenges in identifying the normal roles of Pol II. Moreover, the inability of MMR to recognize the UBP suggests that helix distortions alone are insufficient and that the process requires specific interactions with the nucleobases that are not available with the unnatural nucleotides. Finally, the increased genetic stability afforded by deletion of *recA* may also have significant

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

implications for methods directed at expansion of the genetic code via amber suppression, as these methods also suffer from genetic instability with extended growth.[23] Regardless of these interesting issues, the reprogrammed SSO now allows for the more stable retention of increased biological information, including within its chromosome, and with the previous demonstration that this information can be retrieved in the form of proteins with non-canonical amino acids, should provide a platform to achieve the central goal of synthetic biology – the creation of life with new forms and functions.

METHODS

pINF/UBP containing DNA construction

pINFs (Figure S8) were constructed through Golden Gate assembly of pUCX2 and insert dsDNA containing a d**NaM**-d**TPT3** pair as described previously[3] with the following modifications. UBP containing dsDNA was produced with a 50-μL PCR with chemically synthesized UBP containing oligonucleotides (0.025 ng/μL), primers introducing BsaI sites and vector homology (1 μM, Table S1), d**TPT3**TP (100 μM), d**NaM**TP (100 μM), dNTPs (200 μM), $MgSO_4$ (1.2 mM), One*Taq* DNA Polymerase (0.025 U/μL), and One*Taq* Standard Reaction Buffer (1×, New England Biolabs). The reaction was cycled through the following temperature regime on an MJ Research PTC-200 system (time in mm:ss): [94 °C 00:30 | 25 × (94 °C 00:30 | 47 °C 00:30 | 68 °C 04:00)]. The resulting UBP containing dsDNA was purified using a DNA Clean & Concentrator-5 (Zymo Research) according to manufacturer recommendations. For pINF assembly, pUCX2 (1 μg) and insert DNA were combined at a 1:4 molar ratio in a 80 μL reaction with ATP (1 mM), T4 DNA ligase (6.65 U/μL, New England Biolabs), BsaI-HF (0.66 U/μL, New England Biolabs), and CutSmart Buffer (1×, New England Biolabs) and subjected to the following temperature regime: [37 °C 20 min | 40 × (37 °C 5 min | 16 °C 10 min | 22 °C 5 min) | 37 °C 20 min | 50 °C 15 min | 70 °C 30 min]. BsaI-HF (0.33 U/μL) and T5 exonuclease (0.16 U/μL, New England Biolabs) were then added, and the reaction was incubated at 37 °C for 1 h to remove any pUCX2 without an insert. This reaction was purified using a DNA Clean & Concentrator-5 according to manufacturer recommendations except that reactions were mixed with 3 volumes of 1:1 DNA Wash:DNA Binding Buffer before binding to the silica column.

The UBP knock-in cassette for the *arsB* locus (Figure S4) was produced through overlapping PCR of a 150-bp dsDNA containing a UBP and the kanamycin resistance gene of pKD13. The 150 bp DNA was produced with a 50-μL PCR using the same reaction solution conditions as above and the following temperature regime (time in mm:ss): [98 °C 02:00 | 5 × (98 °C 00:10 | 50 °C 00:10 | 68 °C 04:00) | 15 × (98 °C 00:10 | 58 °C 00:10 | 68 °C 04:00)]. The kanamycin resistance gene amplicon was produced through PCR amplification off pKD13 using Q5 DNA polymerase as per manufacturer recommendations. The amplification of long DNAs (approximately 200 bp or longer) is inhibited by the presence of d**TPT3**TP. Therefore, the overlap assembly PCR of the UBP containing amplicon and kanamycin resistance gene amplicon was performed on large-scale (2 mL of reaction mixture split into 40 individual 50-μL reactions) with the following solution conditions: UBP-containing amplicon (0.02 ng/μL), kanamycin resistance gene amplicon (0.02 ng/μL), primers (1 μM, Table S1), d**TPT3**TP (5 μM), d**NaM**TP (100 μM), dNTPs (200

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

μM), $MgSO_4$ (1.2 mM), One*Taq* DNA Polymerase (0.025 U/μL), and One*Taq* Standard Reaction Buffer (1×). The reactions were subjected to the following temperature regime (time in mm:ss): [98 °C 02:00 | 5 × (98 °C 00:10 | 50 °C 00:10 | 68 °C 04:00) | 15 × (98 °C 00:10 | 58 °C 00:10 | 68 °C 04:00)]. These reactions were pooled and concentrated using a DNA Clean & Concentrator-5 according to manufacturer recommendations.

In vivo UBP replication in genetic knockouts

All genetic knockouts (Figure 1 and Figure S2) were assayed for their ability to replicate pINF-borne UBPs according to the following protocol. Electrocompetent cells were prepared from a 45-mL culture of mid-log phase cells (OD_{600} 0.35–0.7) by pelleting cells and washing twice with 50 mL of 4 °C sterile diH_2O. Washed cells were resuspended in 4 °C sterile diH_2O at a final OD_{600} of 40–60. 50 μL of cells were mixed with 2 ng of a Golden Gate assembled pINF and transferred to an electroporation cuvette (2 mm gap, Cat. #FB102, Fisher Scientific). Electroporation was performed using a Gene Pulser II (BioRad) according to manufacturer recommendations (voltage 25 kV, capacitor 2.5 μF, resistor 200 Ω). Transformed cells were diluted in 950 μL of 2×YT containing chloramphenicol (33 μg/mL) and potassium phosphate (50 mM, pH 7). 40 μL of diluted cells were further diluted into a final volume of 200 μL of 2×YT containing chloramphenicol (33 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7), transferred to a 1.5 mL tube and allowed to recover for 1 h at 37 °C and 230 RPM. 10 μL of recovered cells were diluted into a final volume of 100 μL of 2×YT containing chloramphenicol (33 μg/mL) and ampicillin (100 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM) and potassium phosphate (50 mM, pH 7) in the well of a 96-well plate (Ref. # 655161, Greiner Bio-One). Additionally, recovered cells were plated on 2×YT Agar (2%) containing ampicillin (100 μg/mL) and potassium phosphate (50 mM, pH 7) to estimate transformation efficiency. The 96-well and transformation efficiency plates were kept at 4 °C and 37 °C overnight (approximately 12 h), respectively. The transformation efficiency plate was inspected to ensure that all samples in the 96-well plate received at least 50 colony forming units before refrigeration. The 96-well plate was then transferred to 37 °C and 230 RPM. Cells were pelleted, decanted, and frozen after reaching 0.6–0.92 OD_{600}. *In vivo* replicated pINFs were isolated using a ZR Plasmid Miniprep-Classic kit (Zymo Research) and a 5-μg silica column (Cat. #D4003, Zymo Research) according to manufacturer recommendations and advanced to biotin-shift PCR analysis (see Supporting Information). This procedure was preformed in at least triplicate for each knockout strain starting from preparation of electrocompetent cells.

It should be noted that under these conditions replicates and strains undergo a similar but not identical number of cell doublings during the pINF replication experiment. However due to the pINF's unregulated origin of replication, matching cell doublings between replicates and strains does correspond to matching the number of pINF replication events. Therefore, the data in Figures 1 and 3A are reported as % Retention values (see Supplementary Information for further discussion) as opposed to estimated fidelities and should be interpreted as such.

Examination of clonal pINFs

The ability of the optimized strains to clone pINFs was assessed (Figure 3*A*) as described above with the following modifications. After recovery, dilutions of the recovered culture

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

were plated on 2×YT containing agar (2%), carbenicillin (100 μg/mL), chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7). Plates were incubated at 37 °C for approximately 12 h. Individual colonies were picked and transferred to 100 μL of 2×YT containing carbenicillin (100 μg/mL), chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7) in the well of a 96-well plate. The 96-well plate was kept at 4 °C for approximately 12 h and then transferred to 37 °C and 230 RPM. Cells were pelleted, decanted, and frozen after reaching an OD_{600} of 0.6–0.9. *In vivo* replicated pINFs were isolated using a ZR Plasmid Miniprep-Classic kit according to manufacturer recommendations and advanced to biotin-shift PCR analysis (see Supporting Information).

It should be noted that the Pol $II^{+}\Delta recA$ strain used in these experiments (Figure 3*A*) had a *neo* cassette at the former *recA* locus (P_polB(−)lexA-polB+FRT+ΔrecA+KanR+lacZYA::P_lacUV5-ΔΔ(CoOp) col 2.1, Table S1).

UBP integration at *arsB*

The UBP integration cassette for the *arsB* locus was constructed as described above and depicted in Figure S4. Integration of this cassette was performed using standard lambda red recombineering [24] with the following modifications. Overnight cultures of strains (WT-Opt, Δ*recA*-Opt, and Pol $II^{+}\Delta recA$-Opt in 2×YT containing chloramphenicol (5 μg/mL), and KPi (50 mM, pH 7)) possessing pKD46 were diluted to 0.03 OD_{600} in 2×YT containing ampicillin (100 μg/mL), chloramphenicol (5 μg/mL), and KPi (50 mM, pH 7). Cultures were grown to approximately 0.1 OD_{600} then induced with 0.4% L-(+)-arabinose and allowed to continue to grow to approximately 0.4 OD_{600}. Electrocompetent cells were prepared from these cultures as described above. 50 μL of electrocompetent cells were mixed with 960 ng (5 μL at 192 ng/μL) of the integration cassette described above and electroporated as described above. Transformed cells were diluted to a final volume of 1 mL of 2×YT containing chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7), transferred to a 1.5 mL tube, and allowed to recover for 2 h at 37 °C and 230 RPM. Cells were pelleted and resuspended in 115 μL of 2×YT containing chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), KPi (50 mM, pH 7). 15 μL samples of this cell suspension were plated on 2×YT containing agar (2%), kanamycin (50 μg/mL), chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7). Plates were incubated for 14–24 h at 37 °C. Colonies were picked and transferred to 500 μL of 2×YT containing kanamycin (50 μg/mL), chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), KPi (50 mM, pH 7) in a 48-well plate (Ref. # 677180, Greiner Bio-One). Plates were either refrigerated at 4 °C for ~ 12 h followed by incubation at 37 °C at 230 RPM or advanced directly to incubation. After reaching 0.6–1 OD_{600} cultures were sampled as follows: 100 μL was combined with 100 μL glycerol (50%) and frozen at −80 °C; 350 μL was pelleted and frozen for later isolation of genomic DNA; 50 μL was pelleted, washed once with 200 μdiH_2O, pelleted, and resuspended in 200 μL.

The cell suspensions were analyzed by colony biotin-shift PCR (see Supporting Information). Genomic DNA was isolated from saved frozen cell pellets for samples that displayed high colony biotin-shift PCR percent shift values (≥80%) with a PureLink

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Genomic DNA Mini Kit (Thermo Fisher Scientific) according to manufacturer recommendations. Genomic DNA was analyzed by biotin-shift PCR (see Supporting Information). This analysis revealed high retention clones ($Retention_B \geq 90\%$) for all genetic backgrounds. While these results confirmed successful chromosomal integration of the UBP and remarkably high retention of the UBP in chromosomal DNA, it was suspected that the cells depleted their media of d**TPT3**TP and d**NaM**TP during the integration protocol given the protocol's requirement to incubate cells at high cell density. Actively growing cultures of *E. coli* are known to degrade extracellular d**TPT3**TP and d**NaM**TP to their corresponding di- and mono- phosphate and nucleoside species [5]. To address this possibility the glycerol stocks of the highest retention samples were used to inoculate 100 μL of 2×YT containing kanamycin (50 μg/mL), chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7) in a 96-well plate. Cultures were grown to approximately 0.6 OD_{600} at 37 °C at 230 RPM. Cells from this culture were plated, picked, grown, and sampled as described above. This "replating" procedure quickly revealed clones for the Δ*recA*-Opt and Pol II$^+$Δ*recA*-Opt SSOs with undetectable chromosomal UBP loss ($Retention_B = 100\%$). However despite screening 12 clones for the WT-Opt SSO, no clones with $Retention_B > 91\%$ were discovered. Therefore, we chose to use a WT-Opt integrant ($Retention_B = 91\%$) that did not undergo the replating procedure for the doubling time and passaging experiments. For Δ*recA*-Opt and Pol II$^+$Δ*recA*-Opt we selected one clone each with $Retention_B=100\%$ for the doubling time and passaging experiments.

It should be noted that the Pol II$^+$Δ*recA* strain used in these experiments (Figure 3B, C) did not have a *neo cassette* at the former *recA* locus (P_polB(−)lexA-polB+ΔrecA+FRT +lacZYA::P_lacUV5-ΔΔ(CoOp) col 1.1, Table S1).

Determination of strain doubling time

Mid-log phase cells WT-Opt, Δ*recA*-Opt, and Pol II$^+$Δ*recA*-Opt SSOs and their corresponding chromosomal UBP integrants (described above) were prepared using the following procedure. Saturated over-night cultures were prepared by inoculation of 2×YT containing chloramphenicol (5 μ/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7) from glycerol stock stabs and overnight growth (approximately 14 h) at 37 °C at 230 RPM. These cells were diluted to 0.03 OD_{600} in 500 μL 2×YT containing chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7) and grown at 37 °C at 230 RPM. Growth was monitored by OD_{600}. Once cells reached mid-log phase (0.3–0.5 OD_{600}), they were diluted to 0.013 OD_{600} in 500 μL 2×YT containing chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7) or 2×YT containing chloramphenicol (5 μg/mL) and KPi (50 mM, pH 7) in a 48-well plate and grown at 37 °C at 230 RPM. OD_{600} was measured every 30 min. This procedure was performed in triplicate for each strain starting from inoculation of overnight cultures.

OD_{600} data from each experiment was analyzed to obtain a theoretical cell doubling time (Figure 3B and Figure S5). OD_{600} measurements corresponding to the exponential growth phase (0.01–0.9) were fit to the following exponential growth model using R version 3.2.4:[25]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$OD_i = OD_0 * 2^{C_{Growth} * t}$$

Where OD_i is the OD_{600} at time (t), OD_0 is minimum OD_{600} value for a given data set, and C_{Growth} is the growth constant. C_{Growth} was fit using the "nls()" command. Doubling times (DT) were calculated using the following equation:

$$DT = \frac{1}{C_{Growth}}$$

Passaging of strains bearing a genomic UBP

Glycerol stock stabs of chromosomal UBP integrants from the WT-Opt, Δ*recA*-Opt, and Pol II$^+$Δ*recA*-Opt SSOs (described above) were used to inoculate 500 μL of 2×YT containing kanamycin (50 μg/mL), chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7). Cells were grown to mid log phase (0.5–0.8 OD_{600}) at 37 °C at 230 RPM and then diluted to 0.03 OD_{600} in 500 μL of 2×YT containing kanamycin (50 μg/mL), chloramphenicol (5 μg/mL), d**TPT3**TP (37.5 μM), d**NaM**TP (150 μM), and KPi (50 mM, pH 7) in a 48-well plate and grown at 37 °C at 230 RPM. The cultures inoculated at 0.03 OD_{600} were considered the starting point (Doublings = 0) for passaging. The cultures were grown to 1–1.5 OD_{600} corresponding to approximately 5 cell doublings. This growth from 0.03 to 1–1.5 OD_{600} was considered one "passage" with one passage corresponding to approximately 5 cell doublings. After these samples reached 1–1.5 OD_{600}, another passage was started by diluting cells to 0.03 OD_{600} in fresh media of the same composition. After dilution, the 1–1.5 OD_{600} culture was sampled as follows: 100 μL was combined with 100 μL glycerol (50%) and frozen at −80 °C; 350 μL was pelleted and frozen for later isolation of genomic DNA; and 50 μL was pelleted, washed once with 200 μL diH_2O, pelleted, and resuspended in 200 μL. The passaging process was repeated for a total of 15 passages, corresponding to approximately 80 cell doublings for all three strains.

Throughout passaging, colony biotin-shift PCR analysis (see Supporting Information) was performed on the cell suspension samples. This revealed that retention had declined to <10% in WT-Opt after 15 passages. Therefore, this strain was no longer passaged. In contrast, retention remained at 60–80% in Δ*recA*-Opt and Pol II$^+$Δ*recA*-Opt. Therefore, an additional passage was performed as above for these strains. Retention remained unchanged now a total 16 passages. Therefore, these strains were subjected to 4 additional passages at a higher dilution factor that corresponded to approximately 13 cell doublings per passage (growth from approximately 0.0001 to 1–1.5 OD_{600}). At this point Δ*recA*-Opt and Pol II$^+$Δ*recA*-Opt integrants had experience approximately 130 cell doublings and UBP retention remained >40% according to colony biotin-shift PCR analysis. Further passaging was deemed unnecessary and the experiment was stopped for more rigorous analysis of the genomic DNA samples gathered during passaging. This experiment was performed in triplicate starting from inoculation of media with the genomic integrant glycerol stock stabs.

After completing the passaging experiment, genomic DNA was isolated and analyzed by biotin-shift PCR (Figure 3C) (see Supporting Information). The slow, then rapid loss of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

UBP in WT-Opt suggested that multiple processes contributed to UBP loss. It was suspected that the P_{lacUV5}-*PtNTT2*(66–575) may have been mutated during the experiment, as expression of *Pt*NTT2 causes a slight growth defect.[3] Thus, cells that inactivate the transporter through mutation gain a fitness advantage and can rapidly dominate the experimental population. This hypothesis was explored through isolation of individual clones from the end of WT-Opt passaging and PCR analysis of purified genomic DNA (see Supporting Information and Figure S7). Primer walking for several clones revealed that all genes between *cat* and *insB*-4 including *PtNTT2*(66–575) had been deleted in these cells. The *insB*-4 gene encodes one of two proteins required for the transposition of the IS1 transposon.[26] Sequencing of one clone confirmed that IS1 inserted at *PtNTT2*(66–575) (T1495) corresponding to a 15890 base pair deletion.

After confirmation of the *PtNTT2*(66–575) mutation event, the emergence of deletion mutants was assessed by PCR analysis of genomic DNA samples from WT-Opt integrant passaging (see Supporting Information and Figure S7B). This analysis revealed that several amplicons of sizes corresponding to IS1-mediated *PtNTT2*(66–575) deletion events appear in passaging samples during the rapid phase of UBP loss.

It was also observed that one replicate of the Pol $II^{+}\Delta recA$-Opt integrant rapidly lost the UBP at the same time as the WT-Opt integrants, strongly suggesting that this replicate may have been contaminated with WT-Opt cells during the passaging. This possibility was confirmed using colony PCR analysis, which revealed that this replicate became contaminated with WT-Opt cells at passages corresponding rapid loss of the UBP (see Supporting Information and Figure S6). Therefore, data from this replicate was only used from samples without WT-Opt cell contamination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (GM118178 to F.E.R). M.P.L. was supported by a National Science Foundation Graduate Research Fellowship (Grant No. NSF/DGE-1346837) We acknowledge Synthorx Inc. for providing unnatural triphosphates.

References

1. Leduc, S. The Mechanisms of Life. Rebman Company; New York: 1911.
2. Li L, Degardin M, Lavergne T, Malyshev DA, Dhami K, Ordoukhanian P, Romesberg FE. J Am Chem Soc. 2014; 136:826–829. [PubMed: 24152106]
3. Zhang Y, Lamb BM, Feldman AW, Zhou AX, Lavergne T, Li L, Romesberg FE. Proc Natl Acad Sci USA. 2017; 114:1317–1322. [PubMed: 28115716]
4. Ast M, Gruber A, Schmitz-Esser S, Neuhaus HE, Kroth PG, Horn M, Haferkamp I. Proc Natl Acad Sci USA. 2009; 106:3621–3626. [PubMed: 19221027]
5. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Correa IR Jr, Romesberg FE. Nature. 2014; 509:385–388. [PubMed: 24805238]
6. Zhang Y, Ptacin JL, Fischer EC, Aerni HR, Caffaro CE, San Jose K, Feldman AW, Turner CR, Romesberg FE. Nature. 2017; 551:644–647. [PubMed: 29189780]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

7. Morris SE, Feldman AW, Romesberg FE. ACS Synth Biol. 2017; 6:1834–1840. [PubMed: 28654252]
8. Betz K, Malyshev D, Lavergne T, Welte W, Diederichs K, Romesberg FE, Marx A. J Am Chem Soc. 2013; 135:18637–18643. [PubMed: 24283923]
9. Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Dwyer TJ, Ordoukhanian P, Romesberg FE, Marx A. Nat Chem Biol. 2012; 8:612–614. [PubMed: 22660438]
10. Kisker C, Kuper J, Van Houten B. Cold Spring Harb Perspect Biol. 2013; 5:a012591. [PubMed: 23457260]
11. Jiricny J. Cold Spring Harb Perspect Biol. 2013; 5:a012633. [PubMed: 23545421]
12. Kuzminov A. Microbiol Mol Biol Rev. 1999; 63:751–813. [PubMed: 10585965]
13. Bell JC, Kowalczykowski SC. Trends Biochem Sci. 2016; 41:491–507. [PubMed: 27156117]
14. Michel B. PLoS Biol. 2005; 3:e255. [PubMed: 16000023]
15. Tang M, Pham P, Shen X, Taylor JS, O'Donnell M, Woodgate R, Goodman MF. Nature. 2000; 404:1014–1018. [PubMed: 10801133]
16. Bonner CA, Hays S, McEntee K, Goodman MF. Proc Natl Acad Sci USA. 1990; 87:7663–7667. [PubMed: 2217198]
17. Derbyshire V, Freemont PS, Sanderson MR, Beese L, Friedman JM, Joyce CM, Steitz TA. Science. 1988; 240:199–201. [PubMed: 2832946]
18. Taft-Benz SA, Schaaper RM. Nucleic Acids Res. 1998; 26:4005–4011. [PubMed: 9705512]
19. Kornberg T, Gefter ML. Proc Natl Acad Sci USA. 1971; 68:761–764. [PubMed: 4927672]
20. Rangarajan S, Woodgate R, Goodman MF. Proc Natl Acad Sci USA. 1999; 96:9224–9229. [PubMed: 10430924]
21. Banach-Orlowska M, Fijalkowska IJ, Schaaper RM, Jonczyk P. Mol Microbiol. 2005; 58:61–70. [PubMed: 16164549]
22. Berardini M, Foster PL, Loechler EL. J Bacteriol. 1999; 181:2878–2882. [PubMed: 10217781]
23. Wang Q, Sun T, Xu J, Shen Z, Briggs SP, Zhou D, Wang L. Chembiochem. 2014; 15:1744–1749. [PubMed: 25044429]
24. Datsenko KA, Wanner BL. Proc Natl Acad Sci USA. 2000; 97:6640–6645. [PubMed: 10829079]
25. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: 2016.
26. Escoubas JM, Lane D, Chandler M. J Bacteriol. 1994; 176:5864–5867. [PubMed: 8083181]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

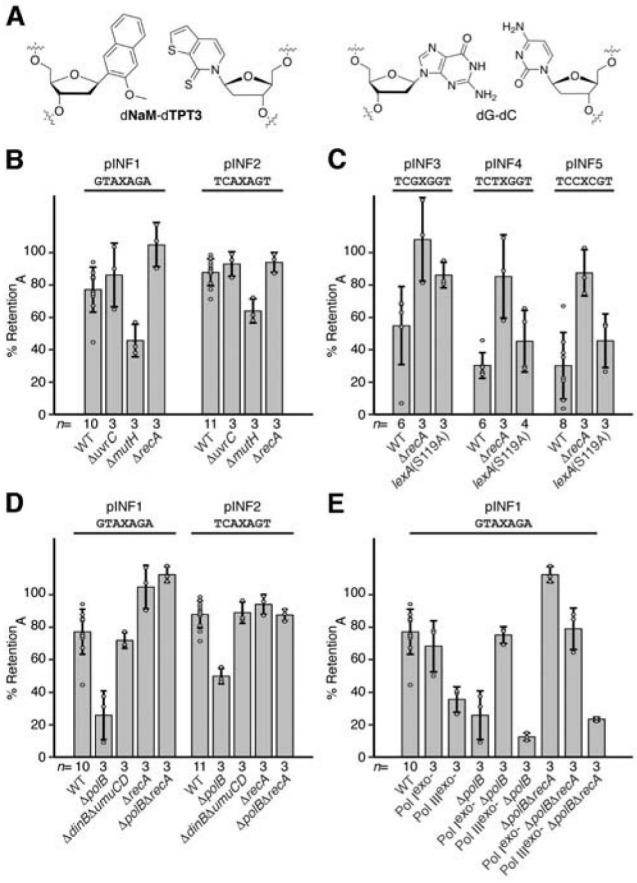


Figure 1.

UBP and the contribution of DNA damage and tolerance pathways to its retention. (**A**) The d**NaM**-d**TPT3** UBP and a natural dG-dC base pair. (**B**) Strains deficient for NER (Δ*uvrC*), MMR (Δ*mutH*), or RER (Δ*recA*). (**C**) Strains deficient for RER and SOS (Δ*recA*) and strains deficient only for SOS (*lexA*(S119A)). (**D**) Strains deficient for the SOS regulated polymerases Pol II (Δ*polB*) or Pols IV and V (Δ*dinB*Δ*umuCD*) or RER and SOS (Δ*recA*). (**e**) Strains with Pol I^{exo-} (*polA(D424A, K890R)*) or Pol III^{exo-} (*dnaQ*(D12N)) in wild-type, Δ*polB*, or Δ*polB*Δ*recA* backgrounds. In each case the indicated strains were challenged with replicating a plasmid with the UBP embedded within the sequence indicated (**X**=d**NaM**). $n \geq 3$ for all data shown; points represent individual replicates; bars represent sample means; error bars represent S.D.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

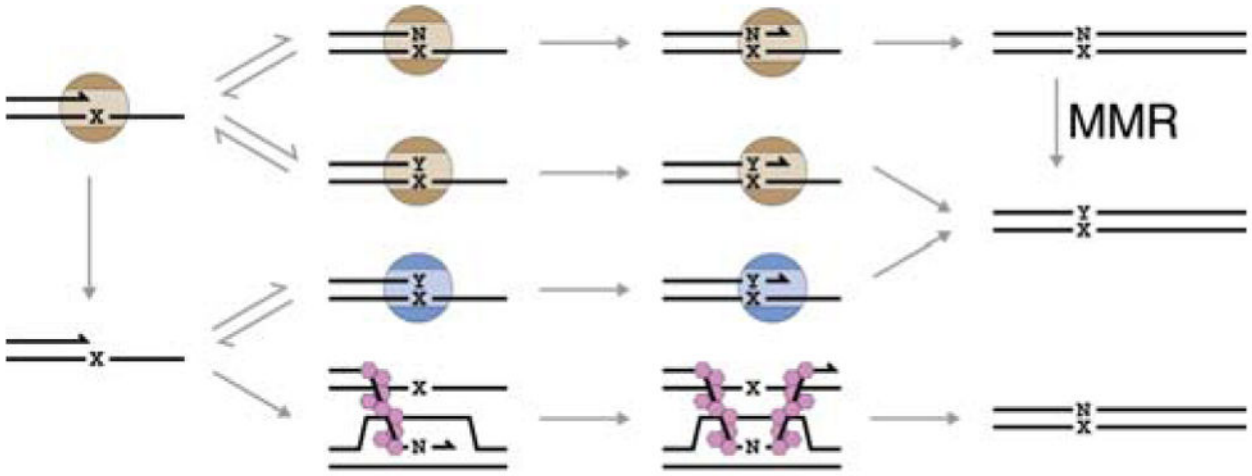


Figure 2.

Proposed mechanism of UBP replication. UBP retention is mediated by the activities of Pol III (tan), Pol II (blue), and MMR. UBP loss is mediated by polymerase replication errors or RecA (pink)-mediated RER. The minor contributions of Pol I and replication errors by Pol II are not shown for clarity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

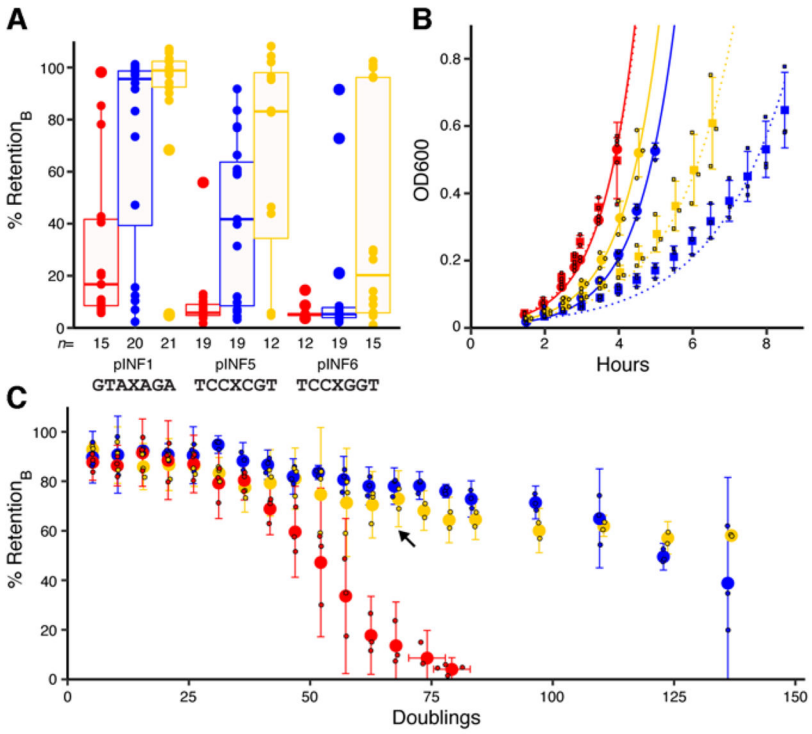


Figure 3.

Replisome reprogramming results in optimized UBP retention. (**A**) Retention of UBP in individual clones of WT-Opt (red), Δ*recA*-Opt (blue), and Pol II$^+$Δ*recA*-Opt (gold) after selection on solid growth media. Each strain was challenged with replicating pINF-borne UBPs in sequence contexts of varying difficulty (GTAXAGA<TCCXCGT<TCCXGGT). Each point represents and individual clone, and $n \geq 12$ for each distribution. (**B**) Growth curves of chromosomal UBP integrants of WT-Opt (red), Δ*recA*-Opt (blue), and Pol II $^+$Δ*recA*-Opt (gold) cells during exponential phase growth in media with (circles/solid lines) and without (squares/dotted lines) d**NaM**TP and d**TPT3**TP. Data is fit with theoretical exponential growth curves. $n = 3$; small points represent individual replicates; large points represent sample means; error bars for time and OD600 represent S.D. (**C**) Retention of the chromosomal d**NaM**-d**TPT3** UBP in WT-Opt (red), Δ*recA*-Opt (blue), and Pol II$^+$Δ*recA*-Opt (gold) cells was measured over long-term growth. $n = 3$; small points represent individual replicates; large points represent sample means; error bars represent two S.D. for both cell doublings and retention except for Pol II$^+$Δ*recA*-Opt data. After approximately seventy doublings, one replicate of Pol II$^+$Δ*recA*-Opt strain was contaminated with WT-Opt cells (see Fig. S6). Therefore, data at and after the black arrow represent the mean of only two independent experiments for Pol II$^+$Δ*recA*-Opt