# A semisynthetic organism engineered for the stable expansion of the genetic alphabet

Yorke Zhang[a,1], Brian M. Lamb[a,1], Aaron W. Feldman[a], Anne Xiaozhou Zhou[a], Thomas Lavergne[b], Lingjun Li[c], and Floyd E. Romesberg[a,2]

[a]Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037; [b]Département de Chimie Moléculaire, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5250, Université Grenoble Alpes, F-38000 Grenoble, France; and [c]School of Chemistry and Chemical Engineering, Henan Normal University, Henan 453007, People's Republic of China

Edited by Clyde A. Hutchison III, The J. Craig Venter Institute, San Diego, CA, and approved December 21, 2016 (received for review October 6, 2016)

**All natural organisms store genetic information in a four-letter, two-base-pair genetic alphabet. The expansion of the genetic alphabet with two synthetic unnatural nucleotides that selectively pair to form an unnatural base pair (UBP) would increase the information storage potential of DNA, and semisynthetic organisms (SSOs) that stably harbor this expanded alphabet would thereby have the potential to store and retrieve increased information. Toward this goal, we previously reported that *Escherichia coli* grown in the presence of the unnatural nucleoside triphosphates dNaMTP and d5SICSTP, and provided with the means to import them via expression of a plasmid-borne nucleoside triphosphate transporter, replicates DNA containing a single dNaM-d5SICS UBP. Although this represented an important proof-of-concept, the nascent SSO grew poorly and, more problematically, required growth under controlled conditions and even then was unable to indefinitely store the unnatural information, which is clearly a prerequisite for true semisynthetic life. Here, to fortify and vivify the nascent SSO, we engineered the transporter, used a more chemically optimized UBP, and harnessed the power of the bacterial immune response by using Cas9 to eliminate DNA that had lost the UBP. The optimized SSO grows robustly, constitutively imports the unnatural triphosphates, and is able to indefinitely retain multiple UBPs in virtually any sequence context. This SSO is thus a form of life that can stably store genetic information using a six-letter, three-base-pair alphabet.**

unnatural base pair | CRISPR | Cas9 | DNA replication | nucleotide transporter

The natural genetic alphabet is composed of four letters whose selective pairing to form two base pairs underlies the storage and retrieval of virtually all biological information. This alphabet is essentially conserved throughout nature, and has been since the last common ancestor of all life on Earth. Significant effort has been directed toward the development of an unnatural base pair (UBP), formed between two synthetic nucleotides, that functions alongside its natural counterparts (1–3), which would represent a remarkable integration of a man-made, synthetic component into one of life's most central processes. Moreover, semisynthetic organisms (SSOs) that stably harbor such a UBP in their DNA could store and potentially retrieve the increased information, and thereby lay the foundation for achieving the central goal of synthetic biology: the creation of new life forms and functions (4).

For over 15 years, we have sought to develop such a UBP (1), and these efforts eventually yielded a family of predominantly hydrophobic UBPs, with that formed between d**NaM** and d**5SICS** (d**NaM**-d**5SICS**; Fig. 1*A*) being a particularly promising example (5–7). Despite lacking complementary hydrogen bonding, we demonstrated that the d**NaM**-d**5SICS** UBP is well replicated by a variety of DNA polymerases in vitro (7–10), and that this efficient replication is mediated by a unique mechanism that draws upon interbase hydrophobic and packing interactions (11, 12). These efforts then culminated in the first progress toward the creation of an SSO in 2014, when we reported that *Escherichia coli* grown in the presence of the corresponding unnatural nucleoside triphosphates (d**NaM**TP and d**5SICS**TP), and provided with a plasmid-encoded nucleoside triphosphate transporter (NTT2) from *Phaeodactylum tricornutum* (which we denote as *Pt*NTT2) (13), is able to import the unnatural triphosphates and replicate a single d**NaM**-d**5SICS** UBP on a second plasmid (14).

Although this first SSO represented an important proof-of-concept, the generality of the expanded genetic alphabet remained unclear, as retention of the UBP was explored at only a single locus and in only a single sequence context. True expansion of the genetic alphabet requires the unrestricted retention of multiple UBPs at any loci and in any sequence context. Moreover, several limitations were already apparent with the nascent SSO (14). First, although expression of the nucleoside triphosphate transporter enabled *E. coli* to import d**NaM**TP and d**5SICS**TP, its expression caused the SSO to grow poorly, with doubling times twice that of the parental strain. Second, the UBP was not well retained during high-density liquid growth or during growth on solid media, presumably due to the secretion of phosphatases that degrade the unnatural triphosphates. Finally, even under optimal conditions, the nascent SSO was unable to retain the UBP with extended growth. Clearly, the ability to robustly grow under the standard repertoire of culture conditions and indefinitely retain the UBP is a prerequisite for true semisynthetic life. Here, we used genetic and chemical approaches to optimize different components of the SSO, ultimately resulting in a simplified and optimized SSO that grows robustly and is capable of the virtually unrestricted storage of increased information.

## Significance

**The genetic alphabet encodes all biological information, but it is limited to four letters that form two base pairs. To expand the alphabet, we developed synthetic nucleotides that pair to form an unnatural base pair (UBP), and used it as the basis of a semisynthetic organism (SSO) that stores increased information. However, the SSO grew poorly and lost the UBP under a variety of standard growth conditions. Here, using chemical and genetic approaches, we report the optimization of the SSO so that it is healthy, more autonomous, and able to store the increased information indefinitely. This SSO constitutes a stable form of semisynthetic life and lays the foundation for efforts to impart life with new forms and functions.**

Author contributions: Y.Z., B.M.L., and F.E.R. designed research; Y.Z., B.M.L., A.W.F., A.X.Z., T.L., and L.L. performed research; Y.Z., B.M.L., and F.E.R. analyzed data; and Y.Z. and F.E.R. wrote the paper.

Conflict of interest statement: Y.Z., B.M.L., and F.E.R. have filed a patent application based on the use of Cas9 for enforced retention of the UBP. Y.Z. and F.E.R. have filed a patent application for the truncated transporter. F.E.R. has a financial interest (shares) in Synthorx Inc., a company that has commercial interests in the UBP. The other authors declare no other competing financial interests.

This article is a PNAS Direct Submission.

[1]Y.Z. and B.M.L. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: floyd@scripps.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616443114/-/DCSupplemental.

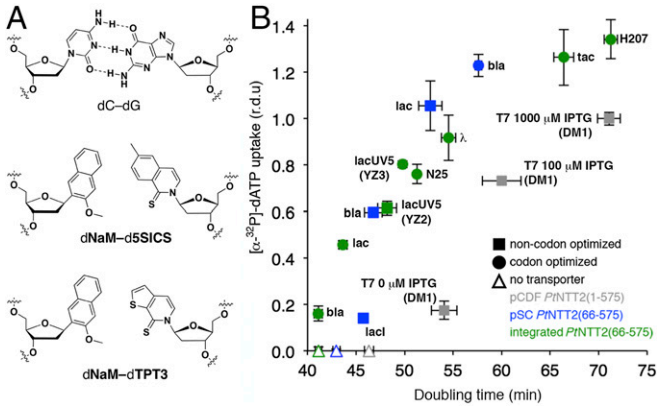


**Fig. 1.** UBPs and transporter optimization. (*A*) Chemical structure of the d**NaM**-d**5SICS** and d**NaM**-d**TPT3** UBPs with a natural dC–dG base pair included for comparison. (*B*) Comparison of fitness and [α-$^{32}$P]-dATP uptake in DM1 and the various constructed strains: pCDF and inducible *Pt*NTT2(1–575) (gray); pSC and constitutive *Pt*NTT2(66–575) (blue); and chromosomally integrated and constitutive *Pt*NTT2(66–575) (green). The promoters from which *Pt*NTT2 is expressed are indicated by the labels next to their corresponding markers. Open triangles denote corresponding control strains without *Pt*NTT2. The pCDF plasmids are in *E. coli* C41(DE3); pSC plasmids and integrants are in *E. coli* BL21(DE3). All *Pt*NTT2 strains are non-codon-optimized for plasmid-based expression and codon-optimized for chromosomal expression unless otherwise indicated; r.d.u. is relative decay units, which corresponds to the total number of radioactive counts per minute normalized to the average $OD_{600}$ across a 1-h window of uptake, with the uptake in DM1 induced with 1,000 μM IPTG set to 1 (see *Materials and Methods* for additional details). Error bars represent SD of the mean, $n = 3$ cultures grown and assayed in parallel; the error bars on some data points are smaller than their marker.

## Results and Discussion

In our previously reported SSO (hereafter referred to as DM1), the transporter was expressed from a T7 promoter on a multicopy plasmid (pCDF-1b) in *E. coli* C41(DE3), and its toxicity mandated carefully controlled induction (14). In its native algal cell, *Pt*NTT2's N-terminal signal sequences direct its subcellular localization and are ultimately removed by proteolysis. However, in *E. coli,* they are likely retained, and could potentially contribute to the observed toxicity. Using the cellular uptake of [α-$^{32}$P]-dATP as a measure of functional transporter expression and as a proxy for the uptake of the unnatural triphosphates, we found that removal of amino acids 1 to 65 and expression of the resulting N-terminally truncated variant *Pt*NTT2(66-575) in *E. coli* C41(DE3) resulted in significantly lower toxicity, but also significantly reduced uptake (Fig. S1 *A* and *B*), possibly due to reduced expression (15). Expression of *Pt*NTT2(66-575) in *E. coli* BL21(DE3) resulted in significant levels of [α-$^{32}$P]-dATP uptake with little increase in toxicity relative to an empty vector control, but the higher level of T7 RNAP in this strain (16) was itself toxic (Fig. S1 *A* and *C*).

We next explored constitutive expression of *Pt*NTT2(66-575) from a low-copy plasmid or a chromosomal locus, which we anticipated would not only eliminate the need to use T7 RNAP but would also impart the SSO with greater autonomy (eliminating the need to induce transporter production), and, importantly, would result in more homogeneous transporter expression and triphosphate uptake across a population of cells, which we reasoned might improve UBP retention. We explored expression of *Pt*NTT2 (66-575) in *E. coli* BL21(DE3) with the promoters $P_{lacI}$, $P_{bla}$, and $P_{lac}$ from a pSC plasmid, and with $P_{bla}$, $P_{lac}$, $P_{lacUV5}$, $P_{H207}$, $P_{\lambda}$, $P_{tac}$, and $P_{N25}$ from the chromosomal *lacZYA* locus (Dataset S1). We also explored the use of a codon-optimized variant of the truncated transporter (see Dataset S1). Although increasing expression of *Pt*NTT2(66-575) (as measured by uptake of [α-$^{32}$P]-dATP) was correlated with increasing doubling time, indicating that expression of *Pt*NTT2(66-575) still exhibited some toxicity (uptake of [α-$^{32}$P]-dATP is itself not toxic), each strain exhibited an improved ratio of uptake to fitness compared with DM1 (Fig. 1*B*). Strain YZ3, which expresses the codon-optimized, chromosomally integrated *Pt*NTT2(66-575) from the $P_{lacUV5}$ promoter, exhibited an optimal compromise of robust growth (<20% increased doubling time relative to the isogenic strain without the transporter), and [α-$^{32}$P]-dATP uptake, and was thus selected for further characterization.

To determine whether the optimized transporter system of YZ3 facilitates high UBP retention, we constructed three plasmids that position the UBP within the 75-nt TK1 sequence (14) [with a local sequence context of d(A-**NaM**-T)]. These include two high-copy pUC19-derived plasmids, pUCX1 [referred to in previous work as pINF (14)] and pUCX2, as well as one low-copy pBR322-derived plasmid, pBRX2 (Fig. S2). In addition to allowing us to examine the effect of copy number on UBP retention, these plasmids position the UBP at proximal (pUCX1) and distal (pUCX2 and pBRX2) positions relative to the origin of replication, which we previously speculated might be important (14). Strains YZ3 and DM1 were transformed with pUCX1, pUCX2, or pBRX2 and directly cultured in liquid growth media supplemented with d**NaM**TP and d**5SICS**TP [and isopropyl β-D-1-thiogalactopyranoside (IPTG) for DM1 to induce the transporter], and growth and UBP retention were characterized at an $OD_{600}$ of ∼1 (Fig. 2*A* and Fig. S3*A*; see also *Materials and Methods*). Although DM1 showed variable levels of retention and reduced growth, especially with the high-copy plasmids, YZ3 showed uniformly high levels of UBP retention and robust growth with all three UBP-containing plasmids (Fig. 2*B* and Fig. S3*A*).

Given that no plasmid locus or copy number biases on UBP retention were observed in YZ3, we chose pUCX2 as a representative UBP-containing plasmid to explore the effect of local sequence context on UBP retention, and we constructed 16 pUCX2 variants in which the UBP was flanked by each possible combination of natural base pairs within a fragment of the GFP gene (see Dataset S1). Under the same growth conditions as above, we observed a wide range of UBP retentions, with some sequence contexts showing complete loss of the UBP (Fig. 2*C*). However, since the development of DM1 with the d**NaM**-d**5SICS** UBP, we have determined that ring contraction and sulfur derivatization of d**5SICS**, yielding the d**NaM**-d**TPT3** UBP (Fig. 1*A*), results in more efficient replication in vitro (17). To explore the in vivo use of d**NaM**-d**TPT3**, we repeated the experiments with YZ3 and each of the 16 pUCX2 plasmids, but with growth in media supplemented with d**NaM**TP and d**TPT3**TP. UBP retentions were clearly higher with d**NaM**-d**TPT3** than with d**NaM**-d**5SICS** (Fig. 2*C*).

Although d**NaM**-d**TPT3** is clearly a more optimal UBP for the SSO than d**NaM**-d**5SICS**, its retention is still moderate to poor in some sequence contexts (Fig. 2*C*). Moreover, several sequences that show good retention in YZ3 cultured in liquid media show poor retention when growth includes culturing on solid media (Fig. S3*B*). To further increase UBP retention with even these challenging sequences and/or growth conditions, we sought to selectively eliminate plasmids that lose the UBP. In prokaryotes, the clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated (Cas) system provides adaptive immunity against viruses and foreign plasmids (18–20). In type II CRISPR-Cas systems, such as that from *Streptococcus pyogenes* (21, 22), the endonuclease Cas9 uses encoded RNAs [or their artificial mimics known as single-guide RNAs (sgRNAs) (23)] to introduce double-strand breaks in cDNA upstream of a 5′-NGG-3′ protospacer adjacent motif (PAM) (24), which then results in DNA degradation by exonucleases (25) (Fig. 3*A*). In vitro, we found that the presence of a UBP in the target DNA generally reduces Cas9-mediated cleavage relative to sequences that are fully complementary to the provided sgRNA (Fig. S4). We thus

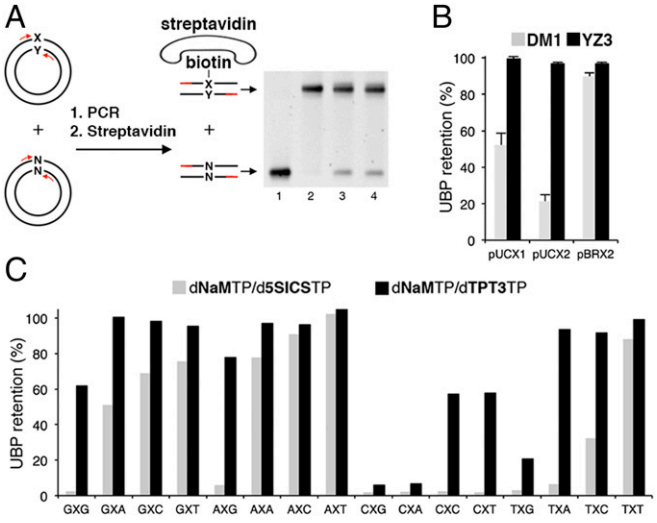


**Fig. 2.** UBP retention assay and the effects of transporter and UBP optimization. (*A*) Schematic representation of the biotin shift assay used to determine UBP retention. The plasmid DNA to be analyzed is first amplified in a PCR supplemented with the unnatural triphosphates, and the resulting products are then incubated with streptavidin and subjected to PAGE analysis. X = d**NaM**, or, in the PCR, its biotinylated analog d**MMO2**$^{biotin}$. Y = d**5SICS** in the PCR, whereas Y = d**TPT3** or d**5SICS** in the plasmid DNA, depending on the experimental conditions. Lane 1 is the product from the oligonucleotide analogous to that used to introduce the UBP during plasmid assembly, but with the UBP replaced by a natural base pair (negative control). This band serves as a marker for DNA that has lost the UBP. Lane 2 is the product from the synthetic oligonucleotide containing the UBP that was used for plasmid assembly. The shift of this band serves as a marker for the shift of DNA containing the UBP. Lane 3 is the product from the in vitro-assembled plasmid before SSO transformation (positive control). The unshifted band results from DNA that has lost the UBP during in vitro plasmid assembly. Lane 4 is the product from an in vivo replication experiment. (*B*) UBP retentions of plasmids pUCX1, pUCX2, and pBRX2 in strains DM1 and YZ3. Error bars represent SD of the mean, $n = 4$ transformations for pUCX1 and pUCX2, $n = 3$ for DM1 pBRX2, and $n = 5$ for YZ3 pBRX2. (*C*) UBP retentions of pUCX2 variants, wherein the UBP is flanked by all possible combinations of natural nucleotides (NXN, where $N$ = d(G, C, A, or T) and X = d**NaM**), in strain YZ3 grown in media supplemented with either d**NaM**TP and d**5SICS**TP (gray bars) or d**NaM**TP and d**TPT3**TP (black bars).

hypothesized that, within a cell, Cas9 programmed with sgRNA(s) complementary to natural sequences that arise from UBP loss would enforce retention in a population of plasmids (by eliminating those that lose the UBP), which we refer to as immunity to UBP loss. To test this, we used a p15A plasmid to construct pCas9, which expresses Cas9 via an IPTG-inducible *lacO* promoter, as well as an 18-nt sgRNA that is complementary to the TK1 sequence containing the most common d**NaM**-d**TPT3** mutation (dT-dA) via the constitutive *proK* promoter (Figs. S2 and S5*A*). Initial exploratory experiments were carried out with strain YZ2 (a forerunner of YZ3 with slightly less optimal transporter performance, but which was available before strain YZ3; Fig. 1*B* and Fig. S1 *D* and *E*). Strain YZ2 was transformed with corresponding pairs of pCas9 and pUCX2 plasmids (i.e., the pUCX2 variant with the UBP embedded within the TK1 sequence such that loss of the UBP produces a sequence targeted by the sgRNA encoded on pCas9), grown to an $OD_{600}$ of ∼4, diluted 250-fold, and regrown to the same $OD_{600}$. UBP retention in control experiments with a nontarget sgRNA dropped to 17% after the second outgrowth; in contrast, UBP retention in the presence of the correct sgRNA was 70% after the second outgrowth (Fig. 3*B*). Sequencing revealed that the majority of plasmids lacking a UBP when the correct sgRNA was provided contained a single nucleotide deletion in its place, which was not observed with the nontarget sgRNA (Fig. S5 *B* and *C*). With a pCas9 plasmid that expresses two sgRNAs, one targeting the most common substitution mutation and one targeting the single nucleotide deletion mutation (Fig. S5*A*), and the same growth and regrowth assay, loss of the UBP was undetectable (Fig. 3*B*).

To more broadly explore Cas9-mediated immunity to UBP loss, we examined retention using 16 pUCX2 variants with sequences that flank the UBP with each possible combination of natural base pairs but also vary its position relative to the PAM, and vary which unnatural nucleotide is present in the strand recognized by the sgRNAs (Table S1). We also constructed a corresponding set of 16 pCas9 plasmids that express two sgRNAs, one targeting a substitution mutation and one targeting the single nucleotide deletion mutation, for each pUCX2 variant. Strain YZ2 carrying a pCas9 plasmid was transformed with its corresponding pUCX2 variant and grown in the presence of the unnatural triphosphates and IPTG (to induce Cas9), and UBP retention was assessed after cells reached an $OD_{600}$ of ∼1. As a control, the 16 pUCX2 plasmids were also propagated in YZ2 carrying a pCas9 plasmid with a nontarget sgRNA. For 4 of the 16 sequences explored, UBP loss was already minimal without immunity (nontarget sgRNA), but was undetectable with expression of the correct sgRNA (Fig. 4*A*). The remaining sequences showed moderate to no retention without immunity, and significantly higher retention with it, including at positions up to 15 nts from the PAM.

To further simplify and streamline the SSO, we next constructed strain YZ4 by integrating an IPTG-inducible Cas9 gene at the *arsB* locus of the YZ3 chromosome, which allows for the use of a single plasmid that both carries a UBP and expresses the sgRNAs that enforce its retention. Sixteen such "all-in-one" plasmids (pAIO) were constructed by replacing the Cas9 gene in each of the pCas9 variants with a UBP sequence from the corresponding pUCX2 variant (Fig. S2 and Table S1). YZ4 and YZ3 (included as a no-Cas9 control due to leaky expression of Cas9 in YZ4) were transformed with a single pAIO plasmid and cultured on solid growth media supplemented with the unnatural triphosphates and with or without IPTG to induce Cas9. Single colonies were used to inoculate liquid media of the same composition, and UBP retention was assessed after cells reached an

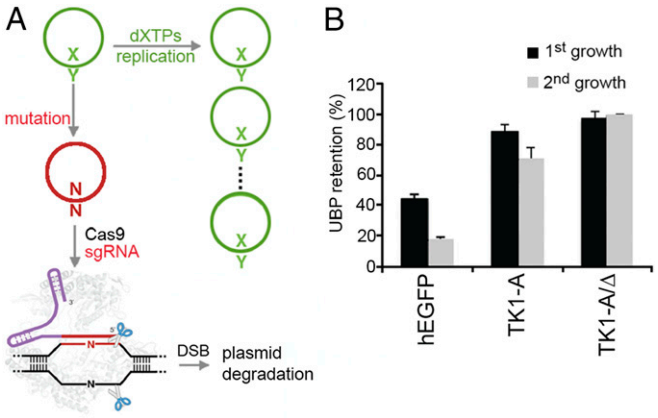


**Fig. 3.** Cas9-based editing system. (*A*) Model for Cas9-mediated immunity to UBP loss. (*B*) UBP retention for pUCX2 TK1 is enhanced by targeting Cas9 cleavage to plasmids that have lost the UBP. The bars labeled hEGFP correspond to UBP retention with growth (black) and regrowth (gray) with an sgRNA that has a sequence taken from the hEGFP gene and thus does not target DNA containing the TK1 sequence (negative control). The bars labeled TK1-A or TK1-A/Δ correspond, respectively, to growth and regrowth with an sgRNA that targets the d**NaM** to dT mutation or two sgRNAs that individually target the d**NaM** to dT mutation and a single nucleotide deletion of the UBP.

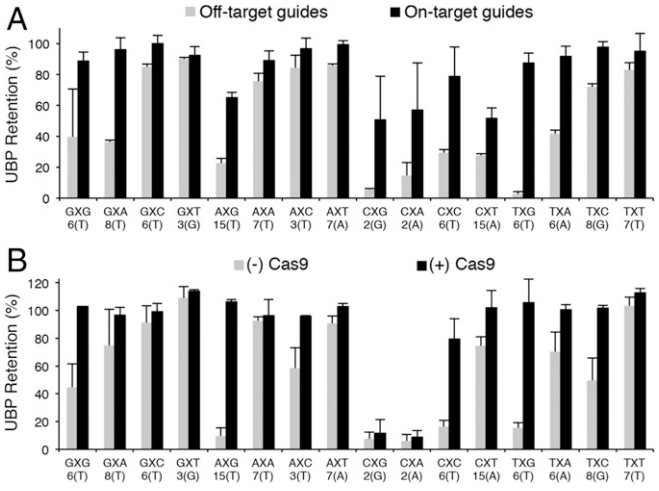


**Fig. 4.** Variation in Cas9-mediated immunity with sequence context. (*A*) UBP retentions of pUCX2 variants in strain YZ2 with a pCas9 plasmid that expresses a nontarget sgRNA (gray) or an on-target sgRNA (black). Error bars represent SD of the mean, and $n = 3$ transformations for all sequences except on-target CXA and CXG, where $n = 5$. (*B*) UBP retentions of pAIO plasmids in strain YZ3 (gray), which does not express Cas9, or in strain YZ4 (black) with expression of Cas9. In *A* and *B*, the nucleotides immediately flanking X = d**NaM** are indicated, as is distance to the PAM. "(*N*)" denotes the nucleotide *N* in the sgRNA that targets a substitution mutation of the UBP; all pCas9 and pAIO plasmids also express an sgRNA that targets a single nucleotide deletion of the UBP. Error bars represent SD of the mean, and $n \geq 3$ colonies; see Table S1 for exact values of *n*, sequences, and IPTG concentrations used to induced Cas9 in YZ4. See *Materials and Methods* for additional experimental details.

$OD_{600}$ of ∼1 to 2 (Fig. 4*B*). Despite variable levels of retention in the absence of Cas9 (YZ3), with induction of Cas9 expression in YZ4, loss was minimal to undetectable in 13 of the 16 sequences. Although retention with the three problematic sequences—d(C-**NaM**-C), d(C-**NaM**-A), and d(C-**NaM**-G)—might be optimized, for example, through alterations in Cas9 or sgRNA expression, the undetectable loss of the UBP with the majority of the sequences after a regimen that included growth both on solid and in liquid media, which was not possible with our previous SSO DM1, attests to the vitality of YZ4.

Finally, we constructed a pAIO plasmid, pAIO2X, containing two UBPs: d**NaM** paired opposite d**TPT3** at position 453 of the sense strand of the GFP gene and d**TPT3** paired opposite d**NaM** at position 36 of the sense strand of the SerT tRNA gene, as well as encoding the sgRNAs targeting the most common substitution mutation expected in each sequence (Fig. S2). YZ4 and YZ3 (again used as a control) were transformed with pAIO2X and subjected to the challenging growth regime depicted in Fig. 5, which included extensive high-density growth on solid and in liquid growth media. Plasmids were recovered and analyzed for UBP retention (Fig. S6) when the $OD_{600}$ reached ∼1 to 2 during each liquid outgrowth. In YZ3, which does not express Cas9, or in the absence of Cas9 induction (no IPTG) in YZ4, UBP retention steadily declined with extended growth (Fig. 5). With induction of immunity (20 or 40 μM IPTG), we observed only a marginal reduction in growth rate (less than 17% increase in doubling time; Fig. S7), and, remarkably, virtually 100% UBP retention (no detectable loss) in both genes.

## Conclusion

Since the last common ancestor of all life on Earth, biological information has been stored with the same four-letter, two-base-pair genetic alphabet. By combining chemical optimization with genetic and immunological engineering, we have created an SSO that is more autonomous (it is naturally competent to import the unnatural triphosphates) and which stores increased information with a fidelity approaching that of natural information. However, unlike any natural organism, the SSO includes an inanimate, man-made component: a UBP that allows it to store information with a six-letter, three-base-pair genetic alphabet. With the virtually unrestricted ability to maintain increased information, the optimized SSO now provides a suitable platform for efforts to retrieve the increased information and create organisms with wholly unnatural attributes and traits not found elsewhere in nature.

## Materials and Methods

**Strains, Plasmids, and Oligonucleotides.** A complete list of strains and plasmids and the sequences of oligonucleotides used in this work can be found in Dataset S2; for information regarding strain construction and plasmid cloning, as well as additional experimental details, see *SI Materials and Methods*. Unless otherwise stated, liquid bacterial cultures were grown in 2×YT (casein peptone 16 g/L, yeast extract 10 g/L, NaCl 5 g/L) supplemented with potassium phosphate (50 mM, pH 7), referred to hereafter as "media," and incubated at 37 °C in a 48-well flat-bottomed plate (CELLSTAR; Greiner Bio-One) with shaking at 200 rpm. Solid growth media was prepared with

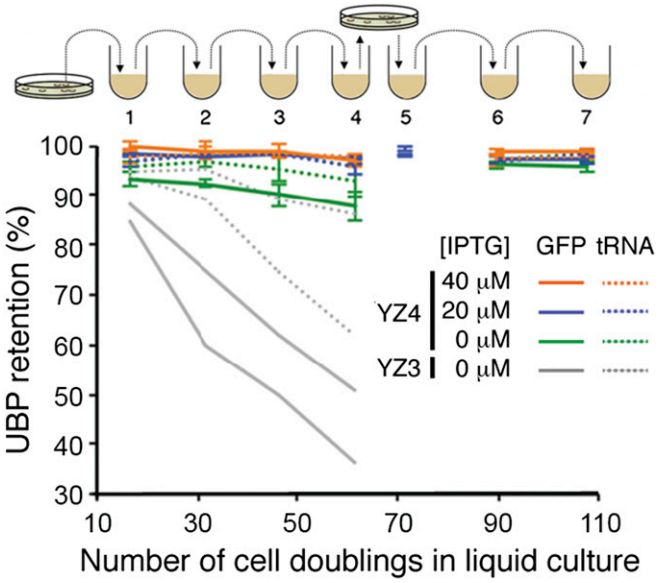


**Fig. 5.** Simultaneous retention of two UBPs during extended growth. Strains YZ3 and YZ4 were transformed with pAIO2X and plated on solid media containing d**NaM**TP and d**TPT3**TP, with or without IPTG to induce Cas9. Single colonies were inoculated into liquid media of the same composition, and cultures were grown to an $OD_{600}$ of ∼2 (point 1). Cultures were subsequently diluted 30,000-fold and regrown to an $OD_{600}$ of ∼2 (point 2), and this dilution–regrowth process was then repeated two more times (points 3 and 4). As a no-immunity control, strain YZ3 was grown in the absence of IPTG, and two representative cultures are indicated in gray. Strain YZ4 was grown in the presence of varying amounts of IPTG, and averages of cultures are indicated in green (0 μM, $n = 5$), blue (20 μM, $n = 5$), and orange (40 μM, $n = 4$). Retentions of the UBP in *gfp* and *serT* are indicated with solid or dotted lines, respectively. After the fourth outgrowth, two of the YZ4 cultures grown with 20 μM IPTG were subcultured on solid media of the same composition. Three randomly selected colonies from each plate ($n = 6$ total) were inoculated into liquid media of the same composition, and each of the six cultures was grown to an $OD_{600}$ of ∼1 (point 5), diluted 300,000-fold into media containing 0, 20, and 40 μM IPTG, and regrown to an $OD_{600}$ of ∼1 (point 6). This dilution–regrowth process was subsequently repeated (point 7). The pAIO2X plasmids were isolated at each of the numbered points and analyzed for UBP retention (Fig. S6). Cell doublings are estimated from $OD_{600}$ (*Materials and Methods*) and did not account for growth on solid media, thus making them an underestimate of actual cell doublings. Error bars represent SD of the mean.

2% (wt/vol) agar. Antibiotics were used, as appropriate, at the following concentrations: carbenicillin, 100 μg/mL; streptomycin, 50 μg/mL; kanamycin, 50 μg/mL; zeocin, 50 μg/mL; and chloramphenicol, 33 μg/mL for plasmids, 5 μg/mL for chromosomal integrants. All selective agents were purchased commercially. Cell growth, indicated as $OD_{600}$, was measured using a PerkinElmer Envision 2103 Multilabel Reader with a 590/20-nm filter.

Natural oligonucleotides were purchased from IDT with standard purification and desalting. Gene synthesis of the codon-optimized *Pt*NTT2 and GFP gene sequences was performed by GeneArt Gene Synthesis (Thermo Fisher) and GenScript, respectively, and kindly provided by Synthorx. Sequencing was performed by Eton Biosciences or Genewiz. Plasmids were isolated using commercial miniprep kits (QIAprep; Qiagen, or ZR Plasmid Miniprep Classic; Zymo Research).

The [α-$^{32}$P]-dATP (3,000 Ci/mmol, 10 mCi/mL) was purchased from PerkinElmer. Triphosphates of d**NaM**, d**5SICS**, d**TPT3**, and d**MMO2**$^{\mathbf{bio}}$ were synthesized as described previously (5, 7, 10, 17) or kindly provided by Synthorx. The d**NaM**-containing TK1 oligonucleotide was described previously (14). All other unnatural oligonucleotides containing d**NaM** were synthesized by Biosearch Technologies with purification by reverse phase cartridge and kindly provided by Synthorx.

**dATP Uptake Assay.** Radioactive uptake assays were conducted as described (26), with the following modifications: C41(DE3) and BL21(DE3) strains carrying plasmid-based transporters and their appropriate empty plasmid controls, as well as BL21(DE3) chromosomal transporter integrants and their appropriate isogenic transporterless control, were grown overnight with appropriate antibiotics (streptomycin for pCDF plasmids and chloramphenicol for pSC plasmids and integrants) in 500 μL of media. Cultures were diluted to an $OD_{600}$ of 0.02 in 500 μL of fresh media, grown for 2.5 h, induced with IPTG (0 mM to 1 mM, pCDF strains only) or grown (all other strains) for 1 h, and incubated with dATP spiked with [α-$^{32}$P]-dATP [final concentration = 250 μM (0.5 μCi/mL)] for ∼1 h. This experimental scheme is analogous to the protocol used to prepare cells for transformation with UBP-containing plasmids, with the 1 h of dATP incubation simulating the 1 h of recovery in the presence of unnatural triphosphates following electroporation. A duplicate 48-well plate without [α-$^{32}$P]-dATP was grown in parallel to monitor growth.

Following incubation with dATP, 200 μL of each culture was collected through a 96-well 0.65-μm glass fiber filter plate (MultiScreen; EMD Millipore) under vacuum and washed with cold potassium phosphate (3 × 200 μL, 50 mM, pH 7) and cold $ddH_2O$ (1 × 200 μL). Filters were removed from the plate and exposed overnight to a storage phosphor screen (BAS-IP MS; GE Healthcare Life Sciences), which was subsequently imaged using a flatbed laser scanner (Typhoon 9410; GE Healthcare Life Sciences). The resulting image was quantified by densitometric analysis using Image Studio Lite (LI-COR). Raw image intensities of each sample were normalized to the length of time and average $OD_{600}$ during dATP incubation (i.e., normalized to an estimate of the area under the growth curve corresponding to the window of uptake), followed by subtracting the normalized signals of the appropriate negative, with no transporter controls.

Doubling times for strains grown in the dATP uptake assay were calculated by doubling time as $(t_2\text{-}t_1)/\log_2(OD_{600,2}/OD_{600,1})$, averaging across three ∼30-min time intervals roughly corresponding to 30 min before dATP uptake and 60 min during dATP uptake.

**Golden Gate Assembly of UBP-Containing Plasmids.** Plasmids containing UBP(s) were generated by Golden Gate Assembly. Inserts containing the UBP were generated by PCR of chemically synthesized oligonucleotides containing d**NaM**, using d**TPT3**TP and d**NaM**TP, and primers that introduce terminal BsaI recognition sites that, when digested, produce overhangs compatible with an appropriate destination plasmid; see Dataset S2 for a full list of primers, templates, and their corresponding Golden Gate destination plasmids. Template oligonucleotides (0.025 ng per 50 μL reaction) were PCR-amplified using reagent concentrations and equipment as described previously (14), under the following thermocycling conditions (times denoted as mm:ss): [96 °C 1:00 | 20 × (96 °C 0:15 | 60 °C 0:15 | 68 °C 4:00)].

To assemble the UBP-containing plasmids, destination plasmid (200 ng to 400 ng), PCR insert(s) (3:1 insert:plasmid molar ratio), T4 DNA ligase (200 U), BsaI-HF (20 U), and ATP (1 mM) were combined in 1× New England Biolabs (NEB) CutSmart buffer (final volume 30 μL) and thermocycled under the following conditions: [37 °C 20:00 | 40 × (37 °C 5:00 | 16 °C 5:00 | 22 °C 2:30) 37 °C 20:00 | 55 °C 15:00 | 80 °C 30:00]. Following the Golden Gate reaction, T5 exonuclease (10 U) and additional BsaI-HF (20 U) were added, and the reaction was incubated (37 °C, 1 h) to digest unincorporated plasmid and insert fragments.

**In Vivo Plasmid Replication Experiments.** Electrocompetent YZ3 cells were prepared by overnight growth in ∼5 mL of media supplemented with chloramphenicol, dilution to $OD_{600}$ of 0.02 in the same media (variable volumes, ∼10 mL of media per transformation), and growth to $OD_{600}$ of ∼0.3 to 0.4. Cells were then rapidly chilled in an ice water bath with shaking, pelleted (2,500 × *g*, 10 min), and washed twice with one culture volume of ice-cold $ddH_2O$. Electrocompetent cells were then resuspended in ice-cold $ddH_2O$ (50 μL per transformation), mixed with a Golden Gate assembled plasmid (∼1 μL, ∼1 ng) containing the UBP, and transferred to a prechilled 0.2-cm-gap electroporation cuvette. Cells were electroporated (Gene Pulser II; Bio-Rad) according to the manufacturer's recommendations (voltage 25 kV, capacitor 2.5 μF, resistor 200 Ω), then immediately diluted with 950 μL of prewarmed media supplemented with chloramphenicol. An aliquot (10 μL to 40 μL) of this dilution was then immediately diluted fivefold with the same prewarmed media, but additionally supplemented with d**NaM**TP (250 μM) and d**5SICS**TP (250 μM). The samples were incubated (37 °C, 1 h), and then ∼15% (vol/vol) of the sample was used to inoculate media (final volume 250 μL to 300 μL) supplemented with chloramphenicol, carbenicillin, d**NaM**TP (250 μM), and d**5SICS**TP (250 μM). Cells were then monitored for growth, collected at the density ($OD_{600}$) indicated in *Results and Discussion*, and subjected to plasmid isolation. Dilutions of the recovery mixture were also spread onto solid media with chloramphenicol and carbenicillin to ascertain transformation efficiencies. Experiments with d**NaM**TP (150 μM) and d**TPT3**TP (37.5 μM) were performed analogously.

Experiments with DM1 were performed analogously using media supplemented with streptomycin, with the additional step of inducing transporter expression with IPTG (1 mM, 1 h) before pelleting the cells. All media following electrocompetent cell preparation was also supplemented with streptomycin and IPTG (1 mM) to maintain expression of the transporter.

**In Vivo Plasmid Replication Experiments with Cas9 (Liquid Culture Only).** Electrocompetent YZ2 cells were transformed with various pCas9 guide plasmids, and single clones were used to inoculate overnight cultures. Cells were then grown, prepared, and electroporated as described above for YZ3, with the following modifications: All media was additionally supplemented with zeocin (to select for pCas9) and 0.2% glucose, electrocompetent cells were stored in 10% (vol/vol $ddH_2O$) DMSO at −80 °C until use, and recovery and growth media were supplemented with d**NaM**TP (250 μM) and d**TPT3**TP (75 μM). Varying concentrations of IPTG (0 μM to 100 μM) were added to the growth media (but not the recovery media) to induce Cas9 expression. The sgRNAs corresponding to the d(AXT) sequence are the nontarget guides for all sequences except for the d(AXT)-containing sequence itself, the nontarget guides for which correspond to the d(GXT) sequence, and all experiments with nontarget sgRNAs were conducted with the addition of IPTG (10 μM) to the growth media. For growth and regrowth experiments, cells were grown to an $OD_{600}$ of 3.5 to 4.0, then diluted 1:250 and regrown to an $OD_{600}$ of 3.5 to 4.0, after which plasmids were isolated.

**In Vivo Plasmid Replication Experiments with Cas9 (Plating and Liquid Culture).** Electrocompetent YZ4 cells were grown, prepared, and electroporated as described in *In Vivo Plasmid Replication Experiments with Cas9 (Liquid Culture Only)* for YZ2, with the following modifications: Media for growing cells before electroporation only contained chloramphenicol (i.e., no zeocin), zeocin was used to select for pAIO (i.e., no carbenicillin), and recovery and growth media were supplemented with d**NaM**TP (150 μM) and d**TPT3**TP (37.5 μM). Following transformation with pAIO, dilutions of the recovery mixture were spread onto solid media containing chloramphenicol, zeocin, d**NaM**TP (150 μM), d**TPT3**TP (37.5 μM), 0.2% glucose, and various concentrations of IPTG (0–50 μM). Following overnight growth (37 °C, ∼14 h), individual colonies were used to inoculate liquid media of the same composition as the solid media. Experiments performed with pAIO2X were conducted as described above for YZ4 without using frozen electrocompetent cells or glucose. The second plating depicted in Fig. 5 was performed by streaking cells from liquid culture onto solid media of the same composition as the liquid media, and growth at 37 °C (∼14 h). Six random colonies were selected to continue propagation in liquid culture.

**Cell Doubling Calculation.** Cell doublings for liquid culture growth–dilution–regrowth experiments were calculated by $\log_2$ of the dilution factor (30,000 or 300,000) between growths, except for growths inoculated from a plated colony, the cell doublings for which were calculated by averaging, for each individual clone, the time from inoculation to target $OD_{600}$ (9.4 ± 1.1 h (1 SD) for the first plating inoculation, 10.2 ± 3 h for the second plating inoculation) and dividing these averages by an estimated doubling time of 40 min. Growth times vary for each clone because colonies were isolated when they were barely visible to the naked eye, and thus we did not attempt to control for variability in the number of cells inoculated into the liquid cultures. Note that the reported number of cell doublings is only an estimate of doublings in liquid culture, which underreports the total number of cell doublings, as we did not attempt to estimate the number of cell doublings that occur during each of the growths on solid media.

**Biotin Shift Assay.** The retention of the UBP(s) in isolated plasmids was determined as previously described and validated (14), with the following modifications: Plasmid minipreps or Golden Gate assembled plasmids (0.5 μL, 0.5 ng/μL to 5 ng/μL), or d**NaM**-containing oligonucleotides (0.5 fmol), were PCR-amplified with dNTPs (400 μM), 1× SYBR Green, $MgSO_4$ (2.2 mM), primers (500 nM each), d**5SICS**TP (65 μM), d**MMO2**$^{\mathbf{Bio}}$TP (65 μM), OneTaq DNA polymerase (0.018 U/μL), and DeepVent DNA polymerase (0.007 U/μL) in 1× OneTaq standard reaction buffer (final volume 15 μL), under the following thermocycling conditions: [20 × (95 °C 0:15 | *x*°C 0:15 | 68 °C 4:00)]; see Dataset S2 for a list of primers and their corresponding annealing temperatures (*x* °C) used in this assay. After amplification, 1 μL of each reaction was mixed with streptavidin (2.5 μL, 2 μg/μL; Promega) and briefly incubated at 37 °C. After incubation, samples were mixed with loading buffer and run on a 6% (wt/vol) polyacrylamide (29:1 acrylamide:bis-acrylamide) Tris/borate/EDTA (TBE) gel, at 120 V for ∼30 min. Gels were then stained with 1× SYBR Gold dye (Thermo Fisher) and imaged using a Molecular Imager Gel Doc XR+ (Bio-Rad) equipped with a 520DF30 filter (Bio-Rad).

**Calculation of UBP Retention.** UBP retention was assessed by densitometric analysis of the gels (ImageJ or Image Studio Lite; LICOR) from the biotin shift assay and calculation of a percent raw shift, which equals the intensity of the streptavidin-shifted band divided by the sum of the intensities of the shifted and unshifted bands. See Fig. 2*A* for representative gels. Reported UBP retentions are normalized values.

Unless otherwise indicated, for experiments not involving plating on solid media, UBP retention was normalized by dividing the percent raw shift of each propagated plasmid sample by the percent raw shift of the Golden Gate assembled input plasmid. We assume that the starting UBP content of the cellular plasmid population is equivalent to the UBP content of the input plasmid, which is a valid assumption given direct inoculation of the transformation into liquid culture. Thus, in these experiments, normalized UBP retention is a relative value that relates the UBP content of the propagated plasmid population to the UBP content of the starting population, which is not 100%, due to loss during the PCR used to generate the insert for input plasmid assembly (Fig. 2*A*).

For experiments involving plating on solid media, UBP retention was normalized by dividing the percent raw shift of each propagated plasmid sample by the percent raw shift of the d**NaM**-containing oligonucleotide template used in the assembly of the input plasmid. Plating enables clonal isolation of UBP-containing plasmids from fully natural plasmids that arose during plasmid construction [some of which may contain sequences that are not recognized by the sgRNA(s) used]. Because there is no PCR-mediated loss of the UBP in the oligonucleotide template, normalization to the oligonucleotide template is a better indicator of absolute UBP retention than normalization to the input plasmid. Under the conditions used in the biotin shift assay, most oligonucleotide templates and sequence contexts give >90% raw shift, with <2% shift for a cognate fully natural template (i.e., UBP misincorporation during the biotin shift assay is negligible).

Plating allows for the differentiation between UBP loss that occurs in vivo from loss that occurs in vitro, with the exception of clonally derived samples that give <2% shift, for which we are unable to differentiate between whether the UBP was completely lost in vivo or if the sample comes from a transformant that originally received a fully natural plasmid. Such samples are excluded from reported average values when other samples from the same transformation give higher shifts.

**ACKNOWLEDGMENTS.** This material is based upon work supported by the National Institutes of Health Grant GM060005 (to F.E.R.), the National Science Foundation Graduate Research Fellowship Grant DGE-1346837 (to A.W.F.), the National Natural Science Foundation of China Grant 21472036 (to L.L.), and the Labex ARCANE Grant ANR-11-LABX-0003-01 (to T.L.). The NanoBio-ICMG platforms (FR 2607) are acknowledged for their support. B.M.L. was supported by a postdoctoral fellowship from the American Cancer Society, Illinois Division.

1. Malyshev DA, Romesberg FE (2015) The expanded genetic alphabet. *Angew Chem Int Ed Engl* 54(41):11930–11944.
2. Zhang L, et al. (2015) Evolution of functional six-nucleotide DNA. *J Am Chem Soc* 137(21):6734–6737.
3. Kimoto M, Hirao I (2014) Creation of unnatural base pair systems toward new DNA/RNA biotechnologies. Chemical Biology of Nucleic Acids: Fundamentals and Clinical Applications, eds Erdmann AV, Markiewicz TW, Barciszewski J (Springer, Berlin), pp 131–148.
4. Leduc S (1911) *The Mechanisms of Life* (Rebman, New York).
5. Leconte AM, et al. (2008) Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J Am Chem Soc* 130(7):2336–2343.
6. Dhami K, et al. (2014) Systematic exploration of a class of hydrophobic unnatural base pairs yields multiple new candidates for the expansion of the genetic alphabet. *Nucleic Acids Res* 42(16):10235–10244.
7. Lavergne T, Malyshev DA, Romesberg FE (2012) Major groove substituents and polymerase recognition of a class of predominantly hydrophobic unnatural base pairs. *Chemistry* 18(4):1231–1239.
8. Malyshev DA, et al. (2012) Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc Natl Acad Sci USA* 109(30):12005–12010.
9. Malyshev DA, Seo YJ, Ordoukhanian P, Romesberg FE (2009) PCR with an expanded genetic alphabet. *J Am Chem Soc* 131(41):14620–14621.
10. Seo YJ, Hwang GT, Ordoukhanian P, Romesberg FE (2009) Optimization of an unnatural base pair toward natural-like replication. *J Am Chem Soc* 131(9):3246–3252.
11. Betz K, et al. (2012) KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nat Chem Biol* 8(7):612–614.
12. Betz K, et al. (2013) Structural insights into DNA replication without hydrogen bonds. *J Am Chem Soc* 135(49):18637–18643.
13. Ast M, et al. (2009) Diatom plastids depend on nucleotide import from the cytosol. *Proc Natl Acad Sci USA* 106(9):3621–3626.
14. Malyshev DA, et al. (2014) A semi-synthetic organism with an expanded genetic alphabet. *Nature* 509(7500):385–388.
15. Goodman DB, Church GM, Kosuri S (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science* 342(6157):475–479.
16. Schlegel S, Genevaux P, de Gier JW (2015) De-convoluting the genetic adaptations of *E. coli* C41(DE3) in real time reveals how alleviating protein production stress improves yields. *Cell Reports* 10(10):1758–1766.
17. Li L, et al. (2014) Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. *J Am Chem Soc* 136(3):826–829.
18. Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 11(3):181–190.
19. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327(5962):167–170.
20. Rath D, Amlinger L, Rath A, Lundgren M (2015) The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie* 117:119–128.
21. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157(6):1262–1278.
22. Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346(6213):1258096.
23. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32(3):279–284.
24. Jinek M, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816–821.
25. Simmon VF, Lederberg S (1972) Degradation of bacteriophage lambda deoxyribonucleic acid after restriction by *Escherichia coli* K-12. *J Bacteriol* 112(1):161–169.
26. Haferkamp I, et al. (2006) Tapping the nucleotide pool of the host: Novel nucleotide carrier proteins of *Protochlamydia amoebophila*. *Mol Microbiol* 60(6):1534–1545.
27. Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* 260(3):289–298.
28. Takeshita S, Sato M, Toba M, Masahashi W, Hashimoto-Gotoh T (1987) High-copy-number and low-copy-number plasmid vectors for lacZ alpha-complementation and chloramphenicol- or kanamycin-resistance selection. *Gene* 61(1):63–74.
29. Deuschle U, Kammerer W, Gentz R, Bujard H (1986) Promoters of *Escherichia coli*: A hierarchy of in vivo strength indicates alternate structures. *EMBO J* 5(11):2987–2994.
30. Sabri S, Steen JA, Bongers M, Nielsen LK, Vickers CE (2013) Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the *Escherichia coli* chromosome at well-characterised loci. *Microb Cell Fact* 12:60.
31. Kuhlman TE, Cox EC (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res* 38(6):e92.
32. Khlebnikov A, Keasling JD (2002) Effect of lacY expression on homogeneity of induction from the P(tac) and P(trc) promoters by natural and synthetic inducers. *Biotechnol Prog* 18(3):672–674.
33. Sikorski RS, Hieter P (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122(1):19–27.
34. Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res* 20(6):1425.
35. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97(12):6640–6645.
36. Quan J, Tian J (2011) Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat Protoc* 6(2):242–251.
37. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One* 3(11):e3647.
38. Guo J, Gaj T, Barbas CF, 3rd (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol* 400(1):96–107.
39. Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343–345.
40. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
41. Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33(1):103–119.