LETTER

doi:10.1038/nature13314

A semi-synthetic organism with an expanded genetic alphabet

Denis A. Malyshev[1], Kirandeep Dhami[1], Thomas Lavergne[1], Tingjian Chen[1], Nan Dai[2], Jeremy M. Foster[2], Ivan R. Corrêa Jr[2] & Floyd E. Romesberg[1]

Organisms are defined by the information encoded in their genomes, and since the origin of life this information has been encoded using a two-base-pair genetic alphabet (A–T and G–C). *In vitro*, the alphabet has been expanded to include several unnatural base pairs (UBPs)[1–3]. We have developed a class of UBPs formed between nucleotides bearing hydrophobic nucleobases, exemplified by the pair formed between d5SICS and dNaM (d5SICS–dNaM), which is efficiently PCR-amplified[1] and transcribed[4,5] *in vitro*, and whose unique mechanism of replication has been characterized[6,7]. However, expansion of an organism's genetic alphabet presents new and unprecedented challenges: the unnatural nucleoside triphosphates must be available inside the cell; endogenous polymerases must be able to use the unnatural triphosphates to faithfully replicate DNA containing the UBP within the complex cellular milieu; and finally, the UBP must be stable in the presence of pathways that maintain the integrity of DNA. Here we show that an exogenously expressed algal nucleotide triphosphate transporter efficiently imports the triphosphates of both d5SICS and dNaM (d5SICSTP and dNaMTP) into *Escherichia coli*, and that the endogenous replication machinery uses them to accurately replicate a plasmid containing d5SICS–dNaM. Neither the presence of the unnatural triphosphates nor the replication of the UBP introduces a notable growth burden. Lastly, we find that the UBP is not efficiently excised by DNA repair pathways. Thus, the resulting bacterium is the first organism to propagate stably an expanded genetic alphabet.

To make the unnatural triphosphates available inside the cell, we previously suggested using passive diffusion of the free nucleosides into the cytoplasm followed by their conversion to the corresponding triphosphate via the nucleoside salvage pathway[8]. Although we have shown that analogues of d5SICS and dNaM are phosphorylated by the nucleoside kinase from *Drosophila melanogaster*[8], monophosphate kinases are more specific[9], and in *E. coli* we found that overexpression of the endogenous nucleoside diphosphate kinase results in poor growth. As an alternative, we focused on the nucleotide triphosphate transporters (NTTs) of obligate intracellular bacteria and algal plastids[10–14]. We expressed eight different NTTs in *E. coli* C41(DE3)[15–17] and measured the uptake of [α-^{32}P]-dATP as a surrogate for the unnatural triphosphates (Extended Data Fig. 1). We confirmed that [α-^{32}P]-dATP is efficiently transported into cells by the NTTs from *Phaeodactylum tricornutum* (*Pt*NTT2)[18] and *Thalassiosira pseudonana* (*Tp*NTT2)[18]. Although NTTs from *Protochlamydia amoebophila* (*Pam*NTT2 and *Pam*NTT5)[15] also import [α-^{32}P]-dATP, *Pt*NTT2 showed the most activity, and both it and *Tp*NTT2 are known to have broad specificity[18], making them the most promising NTTs for further characterization.

Transport via an NTT requires that the unnatural triphosphates are sufficiently stable in culture media; however, preliminary characterization of d5SICSTP and dNaMTP indicated that decomposition occurs in the presence of actively growing *E. coli* (Extended Data Fig. 2). Similar behaviour was observed with [α-^{32}P]-dATP, and the dephosphorylation products detected by thin-layer chromatography (TLC) for [α-^{32}P]-dATP, or by high-performance liquid chromatography (HPLC) and matrix-assisted laser desorption/ionization (MALDI) for d5SICSTP and dNaMTP, suggest that decomposition is mediated by phosphatases. As no degradation was observed upon incubation in spent media, decomposition seems to occur within the periplasm. No increase in stability was observed in cultures of single-gene-deletion mutants of *E. coli* BW25113 lacking a specific periplasmic phosphatase[19] (as identified by the presence of a Sec-type amino-terminal leader sequence), including *phoA*, *ushA*, *appA*, *aphA*, *yjjX*, *surE*, *yfbR*, *yjjG*, *yfaO*, *mutT*, *nagD*, *yggV*, *yrfG* or *ymfB*, suggesting that decomposition results from the activity of multiple phosphatases. However, the extracellular stability of [α-^{32}P]-dATP was significantly greater when 50 mM potassium phosphate (KPi) was added to the growth medium (Extended Data Fig. 3). Thus, we measured [α-^{32}P]-dATP uptake from media containing 50 mM KPi after induction of the transporter with isopropyl-β-D-thiogalactoside (IPTG) (Extended Data Fig. 4). Although induction with 1 mM IPTG resulted in slower growth, consistent with the previously reported toxicity of NTTs[17], it also resulted in maximal [α-^{32}P]-dATP uptake. Thus, after addition of 1 mM IPTG, we analysed the extracellular and intracellular stability of [α-^{32}P]-dATP as a function of time (Extended Data Fig. 5). Cells expressing *Pt*NTT2 were found to have the highest levels of intracellular [α-^{32}P]-dATP, and although both extra- and intracellular dephosphorylation was still observed, the ratio of triphosphate to dephosphorylation products inside the cell remained roughly constant, indicating that the extracellular concentrations and *Pt*NTT2-mediated influx are sufficient to compensate for intracellular decomposition.

Likewise, we found that the addition of KPi increased the extracellular stability of d5SICSTP and dNaMTP (Extended Data Fig. 2), and

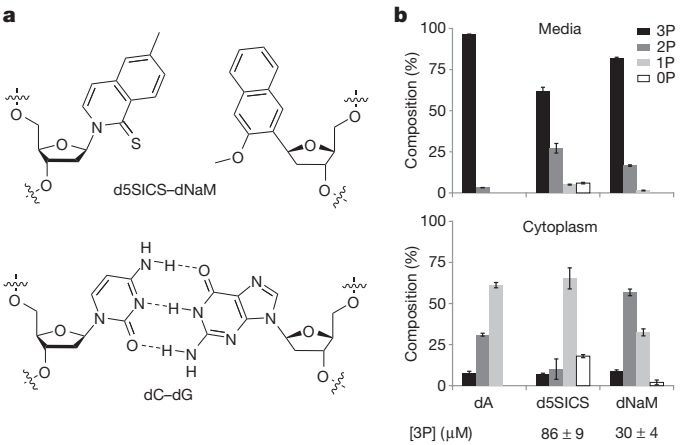


Figure 1 | Nucleoside triphosphate stability and import. a, Chemical structure of the d5SICS–dNaM UBP compared to the natural dG–dC base pair. **b**, Composition analysis of d5SICS and dNaM in the media (top) and cytoplasmic (bottom) fractions of cells expressing *Pt*NTT2 after 30 min incubation; dA shown for comparison. 3P, 2P, 1P and 0P correspond to triphosphate, diphosphate, monophosphate and nucleoside, respectively; [3P] is the intracellular concentration of triphosphate. Error bars represent s.d. of the mean, $n = 3$.

[1]Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [2]New England Biolabs, 240 County Road, Ipswich, Massachusetts 01938, USA.

when a stationary phase culture was diluted 100-fold into fresh media, the half-lives of both unnatural triphosphates (initial concentrations of 0.25 mM) were found to be approximately 9 h, which seemed sufficient for our purposes. Thirty minutes after their addition to the media, neither of the unnatural triphosphates was detected in cells expressing *Tp*NTT2; in contrast, 90 μM of d5SICSTP and 30 μM of dNaMTP were found in the cytoplasm of cells expressing *Pt*NTT2 (Fig. 1b). Although intracellular decomposition was still apparent, the intracellular concentrations of intact triphosphate are significantly above the sub-micromolar K_M values of the unnatural triphosphates for DNA polymerases[20], setting the stage for replication of the UBP in a living bacterial cell.

The replication of DNA containing d5SICS-dNaM has been validated *in vitro* with different polymerases, primarily family A polymerases, such as the Klenow fragment of *E. coli* DNA polymerase I (pol I)[20,21]. As the majority of the *E. coli* genome is replicated by pol III, we engineered a plasmid to focus replication of the UBP to pol I. Plasmid pINF (the information plasmid) was constructed from pUC19 using solid-phase DNA synthesis and circular-extension PCR to replace the dA–dT pair at position 505 with dNaM paired opposite an analogue of d5SICS (dTPT3[22]) (Fig. 2a, b). This positions the UBP 362 bp downstream of the ColE1 origin of replication where leading-strand replication is mediated by pol I[23], and within the TK-1 Okazaki processing site[24], where lagging-strand synthesis is also expected to be mediated by pol I. Synthetic pINF was constructed using the d5SICS analogue because it should be efficiently replaced by d5SICS if replication occurs *in vivo*, making it possible to differentiate *in vivo* replicated pINF from synthetic pINF.

To determine whether *E. coli* can use the imported unnatural triphosphates to stably propagate pINF, C41(DE3) cells were first transformed with a pCDF-1b plasmid encoding *Pt*NTT2 (hereafter referred to as pACS, for accessory plasmid, Fig. 2a) and grown in media containing 0.25 mM of both unnatural triphosphates, 50 mM KPi and 1 mM IPTG to induce transporter production. Cells were then transformed with pINF, and after a 1-h recovery period, cultures were diluted tenfold with the same media supplemented with ampicillin, and growth was monitored via culture turbidity (Extended Data Table 1). As controls, cells were also transformed with pUC19, or grown without either IPTG or without the unnatural triphosphates. Again, growth was significantly slower in the presence of IPTG, but the addition of d5SICSTP and dNaMTP resulted in only a slight further decrease in growth in the absence of pINF, and interestingly, it eliminated a growth lag in the presence of pINF (Fig. 2c), suggesting that the unnatural triphosphates are not toxic and are required for the efficient replication of pINF.

To demonstrate the replication of pINF, we recovered the plasmid from cells after 15 h of growth. The introduction of the UBP resulted in

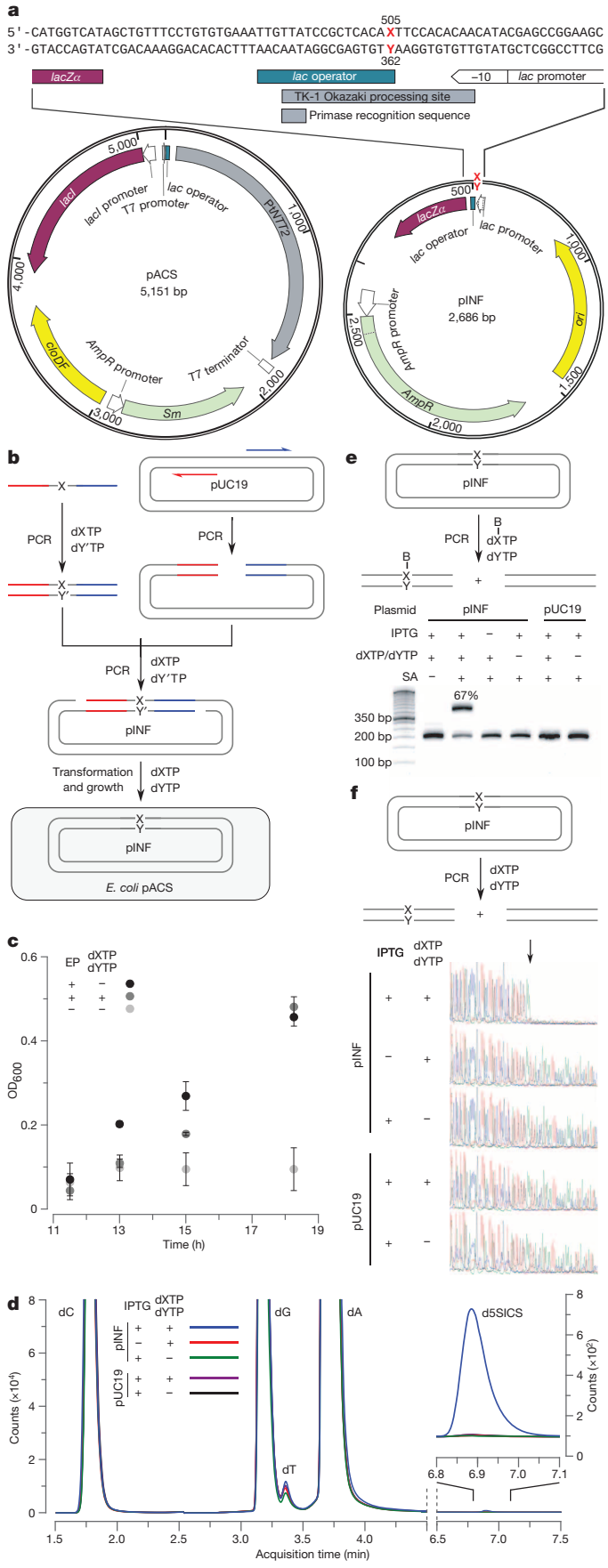


Figure 2 | Intracellular UBP replication. **a**, Structure of pACS and pINF. dX and dY correspond to dNaM and a d5SICS analogue[22] that facilitated plasmid construction (see Methods). *cloDF*, origin of replication; *Sm*, streptomycin resistance gene; *AmpR*, ampicillin resistance gene; *ori*, ColE1 origin of replication; *lacZα*, β-galactosidase fragment gene. **b**, Overview of pINF construction. A DNA fragment containing the unnatural nucleotide was synthesized via solid-phase DNA synthesis and then used to assemble synthetic pINF via circular-extension PCR[29]. X, dNaM; Y′, dTPT3 (an analogue of d5SICS[22]); y, d5SICS (see text). Colour indicates regions of homology. The doubly nicked product was used directly to transform *E. coli* harbouring pACS. **c**, The addition of d5SICSTP and dNaMTP eliminates a growth lag of cells harbouring pINF. EP, electroporation. Error bars represent s.d. of the mean, $n = 3$. **d**, LC-MS/MS total ion chromatogram of global nucleoside content in pINF and pUC19 recorded in dynamic multiple reaction monitoring (DMRM) mode. pINF and pUC19 (control) were propagated in *E. coli* in the presence or absence of unnatural triphosphates, and with or without *Pt*NTT2 induction. The inset shows a 100-fold expansion of the mass-count axis in the d5SICS region. **e**, Biotinylation only occurs in the presence of the UBP, the unnatural triphosphates and transporter induction. After growth, pINF was recovered, and a 194-nucleotide region containing the site of UBP incorporation (nucleotides 437–630) was amplified and biotinylated. B, biotin; SA, streptavidin. The natural pUC19 control plasmid was prepared identically to pINF. A 50-bp DNA ladder is shown to the left. **f**, Sequencing analysis demonstrates retention of the UBP. An abrupt termination in the Sanger sequencing reaction indicates the presence of UBP incorporation (site indicated with arrow).

a small (approximately twofold) reduction in the copy number of pINF, as gauged by its ratio to pACS (Extended Data Table 1); we determined that the plasmid was amplified 2×10^7-fold during growth (approximately 24 doublings) based on the amount of recovered plasmid and the transformation efficiency. To determine the level of UBP retention, the recovered plasmid was digested, dephosphorylated to single nucleosides, and analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS)[25]. Although the detection and quantification of dNaM were precluded by its poor fragmentation efficiency and low product ion counts over background, signal for d5SICS was clearly observable (Fig. 2d). External calibration curves were constructed using the unnatural nucleoside and validated by determining its ratio to dA in synthetic oligonucleotides (Extended Data Table 2). Using the resulting calibration curve, we determined the ratio of dA to d5SICS in recovered pINF was 1,106 to 1, which when compared to the expected ratio of 1,325 to 1, suggests the presence of approximately one UBP per plasmid. No d5SICS was detected in control experiments in which the transporter was not induced, or when the unnatural triphosphates were not added to the media, or when pUC19 was used instead of pINF (Fig. 2d, inset), demonstrating that its presence results from the replication of the UBP and not from misinsertion of the unnatural triphosphates opposite a natural nucleotide. Importantly, as the synthetic pINF contained an analogue of d5SICS, and d5SICS was only provided as a triphosphate added to the media, its presence in pINF confirms *in vivo* replication.

To independently confirm and quantify the retention of the UBP in the recovered plasmid, the relevant region was amplified by PCR in the presence of d5SICSTP and a biotinylated dNaMTP analogue[4] (Fig. 2e). Analysis by streptavidin gel shift showed that 67% of the amplified DNA contained biotin. No shift was observed in control experiments where the transporter was not induced, or when unnatural triphosphates were not added, or when pUC19 was used instead of pINF, demonstrating that the shift results from the presence of the UBP. Based on a calibration curve constructed from the shifts observed with the amplification products of controlled mixtures of DNA containing dNaM or its fully natural counterpart (Methods and Extended Data Fig. 6), the observed gel shift corresponds to a UBP retention of 86%. Similarly, when the amplification product obtained with d5SICSTP and dNaMTP was analysed by Sanger sequencing in the absence of the unnatural triphoshates[1,26,27], the sequencing chromatogram showed complete termination at the position of UBP incorporation, which with an estimated lower limit of read-through detection of 5%, suggests a level of UBP retention in excess of 95% (Fig. 2f). In contrast, amplification products obtained from pINF recovered from cultures grown without *Pt*NTT2 induction, without added unnatural triphosphates, or obtained from pUC19 propagated under identical conditions, showed no termination. Overall, the data unambiguously demonstrate that DNA containing the UBP was replicated *in vivo* and allow us to estimate that replication occurred with fidelity (retention per doubling) of at least 99.4% (24 doublings; 86% retention; $0.994^{24} = 0.86$). This fidelity corresponds to an error rate of approximately 10^{-3}, which is comparable to the intrinsic error rate of some polymerases with natural DNA[28].

The high retention of the UBP over a 15-h period of growth (approximately 24 doublings) strongly suggests that it is not efficiently excised by DNA repair pathways. To test further this hypothesis and to examine retention during prolonged stationary phase growth, we repeated the experiments, but monitored UBP retention, cell growth and unnatural triphosphate decomposition for up to 6 days without providing any additional unnatural triphosphates (Fig. 3 and Extended Data Fig. 7). At 15 and 19 h of growth, the cultures reached an optical density at 600 nm (OD_{600}) of approximately 0.9 and 1.2, respectively, and both d5SICSTP and dNaMTP decomposed to 17–20% and 10–16% of their initial 0.25-mM concentrations (Extended Data Fig. 7a). In agreement with the experiments described above, retention of the UBP after 15 h was 97 ± 5% and >95%, as determined by gel shift and sequencing, respectively, and after 19 h it was 91 ± 3% and >95%. As the cultures entered stationary phase and the triphosphates decomposed completely, plasmid loss began to compete

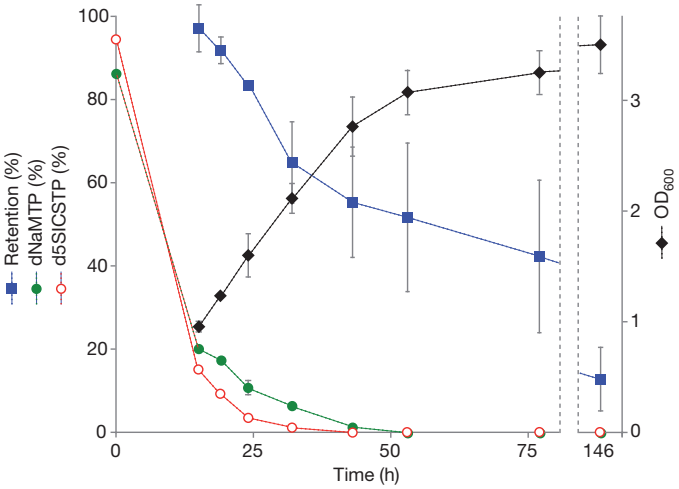


Figure 3 | Intracellular stability of the UBP. *E. coli* C41(DE3) pACS was transformed with pINF and grown after a single dose of d5SICSTP and dNaMTP was provided in the media. UBP retention in recovered pINF, OD_{600}, and relative amount of d5SICSTP and dNaMTP in the media (100% = 0.25 mM), were determined as a function of time. Error bars represent s.d. of the mean, $n = 3$.

with replication (Extended Data Fig. 7b, c, d), but even then, retention of the UBP remained at approximately 45% and 15%, at days 3 and 6 respectively. Moreover, when d5SICS-dNaM was lost, it was replaced by dA–dT, which is consistent with the mutational spectrum of DNA pol I[20]. Finally, the shape of the retention versus time curve mirrors that of the growth versus time curve. Taken together, these data suggest that in the absence of unnatural triphosphates, the UBP is eventually lost by replication-mediated mispairing, and not from the activity of DNA repair pathways.

We have demonstrated that *Pt*NTT2 efficiently imports d5SICSTP and dNaMTP into *E. coli* and that an endogenous polymerase, possibly pol I, efficiently uses the unnatural triphosphates to replicate DNA containing the UBP within the cellular environment with reasonable efficiency and fidelity. Moreover, the UBP appears stable during both exponential and stationary phase growth despite the presence of all DNA repair mechanisms. Remarkably, although expression of *Pt*NTT2 results in a somewhat reduced growth rate, neither the unnatural triphosphates nor replication of the UBP results in significant further reduction in growth. The resulting bacterium is the first organism that stably harbours DNA containing three base pairs. In the future, this organism, or a variant with the UBP incorporated at other episomal or chromosomal loci, should provide a synthetic biology platform to orthogonally re-engineer cells, with applications ranging from site-specific labelling of nucleic acids in living cells to the construction of orthogonal transcription networks and eventually the production and evolution of proteins with multiple, different unnatural amino acids.

METHODS SUMMARY

To prepare electrocompetent C41(DE3) pACS cells, freshly transformed *E. coli* C41(DE3) pACS was grown overnight in 2 × YT medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl) supplemented with streptomycin and KPi. After 100-fold dilution into the same medium and outgrowth at 37 °C to $OD_{600} = 0.20$, IPTG was added to induce expression of *Pt*NTT2. After 40 min, cultures were rapidly cooled, washed with sterile water and resuspended in 10% glycerol. An aliquot of electrocompetent cells was mixed with pINF and electroporated. Pre-warmed 2 × YT medium containing streptomycin, IPTG and KPi was added, and an aliquot was diluted 3.3-fold in the same media supplemented with 0.25 mM each of dNaMTP and d5SICSTP. The resulting mixture was allowed to recover at 37 °C with shaking. After recovery, cultures were centrifuged. Spent media was analysed for nucleotide composition by HPLC (Extended Data Fig. 7a); cells were resuspended in fresh medium containing streptomycin, ampicillin, IPTG, KPi and 0.25 mM each of dNaMTP and d5SICSTP, and grown with shaking. At defined time points, OD_{600} was determined and aliquots were removed and centrifuged. Spent media were analysed for nucleotide

composition, and pINF was recovered by spin column purification. UBP retention was characterized by LC-MS/MS, PCR amplification and gel electrophoresis, or sequencing, as described in the Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 November 2013; accepted 8 April 2014.
Published online 7 May 2014.

1. Malyshev, D. A. *et al.* 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,** 12005–12010 (2012).
2. Yang, Z., Chen, F., Alvarado, J. B. & Benner, S. A. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **133,** 15105–15112 (2011).
3. Yamashige, R. *et al.* Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res.* **40,** 2793–2806 (2012).
4. Seo, Y. J., Malyshev, D. A., Lavergne, T., Ordoukhanian, P. & Romesberg, F. E. Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs. *J. Am. Chem. Soc.* **133,** 19878–19888 (2011).
5. Seo, Y. J., Matsuda, S. & Romesberg, F. E. Transcription of an expanded genetic alphabet. *J. Am. Chem. Soc.* **131,** 5046–5047 (2009).
6. Betz, K. *et al.* Structural insights into DNA replication without hydrogen bonds. *J. Am. Chem. Soc.* **135,** 18637–18643 (2013).
7. Betz, K. *et al.* KlenTaq polymerase replicates unnatural base pairs by inducing a Watson–Crick geometry. *Nature Chem. Biol.* **8,** 612–614 (2012).
8. Wu, Y., Fa, M., Tae, E. L., Schultz, P. G. & Romesberg, F. E. Enzymatic phosphorylation of unnatural nucleosides. *J. Am. Chem. Soc.* **124,** 14626–14630 (2002).
9. Yan, H. & Tsai, M. D. Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. *Adv. Enzymol.* **73,** 103–134 (1999).
10. Winkler, H. H. & Neuhaus, H. E. Non-mitochondrial ATP transport. *Trends Biochem. Sci.* **24,** 64–68 (1999).
11. Amiri, H., Karlberg, O. & Andersson, S. G. Deep origin of plastid/parasite ATP/ADP translocases. *J. Mol. Evol.* **56,** 137–150 (2003).
12. Hatch, T. P., Al-Hossainy, E. & Silverman, J. A. Adenine nucleotide and lysine transport in *Chlamydia psittaci*. *J. Bacteriol.* **150,** 662–670 (1982).
13. Winkler, H. H. Rickettsial permeability: an ADP-ATP transport system. *J. Biol. Chem.* **251,** 389–396 (1976).
14. Horn, M. & Wagner, M. Bacterial endosymbionts of free-living amoebae. *J. Eukaryot. Microbiol.* **51,** 509–514 (2004).
15. Haferkamp, I. *et al.* Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of *Protochlamydia amoebophila*. *Mol. Microbiol.* **60,** 1534–1545 (2006).
16. Miroux, B. & Walker, J. E. 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,** 289–298 (1996).
17. Haferkamp, I. & Linka, N. Functional expression and characterisation of membrane transport proteins. *Plant Biol.* **14,** 675–690 (2012).
18. Ast, M. *et al.* Diatom plastids depend on nucleotide import from the cytosol. *Proc. Natl Acad. Sci. USA* **106,** 3621–3626 (2009).
19. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).
20. Lavergne, T., Malyshev, D. A. & Romesberg, F. E. Major groove substituents and polymerase recognition of a class of predominantly hydrophobic unnatural base pairs. *Chemistry* **18,** 1231–1239 (2012).
21. Seo, Y. J., Hwang, G. T., Ordoukhanian, P. & Romesberg, F. E. Optimization of an unnatural base pair toward natural-like replication. *J. Am. Chem. Soc.* **131,** 3246–3252 (2009).
22. Li, L. *et al.* Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. *J. Am. Chem. Soc.* **136,** 826–829 (2014).
23. Tomizawa, J. & Selzer, G. Initiation of DNA synthesis in *Escherichia coli*. *Annu. Rev. Biochem.* **48,** 999–1034 (1979).
24. Allen, J. M. *et al.* Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nucleic Acids Res.* **39,** 7020–7033 (2011).
25. Hashimoto, H. *et al.* Structure of a *Naegleria* Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506,** 391–395 (2013).
26. Malyshev, D. A., Seo, Y. J., Ordoukhanian, P. & Romesberg, F. E. PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.* **131,** 14620–14621 (2009).
27. Hirao, I. *et al.* An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nature Methods* **3,** 729–735 (2006).
28. Goodman, M. F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **71,** 17–50 (2002).
29. Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nature Protocols* **6,** 242–251 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank I. Haferkamp and J. Audia for kindly providing the NTT plasmids and helpful discussions, and P. Ordoukhanian for providing access to the Center for Protein and Nucleic Acid Research at TSRI. This work was supported by the US National Institutes of Health (NIH) (GM 060005).

Author Contributions D.A.M., K.D., T.C. and F.E.R. designed the experiments. D.A.M., K.D. and T.L. performed the experiments. N.D., J.M.F. and I.R.C.J. performed LC-MS/MS analysis. D.A.M., K.D. and F.E.R. analysed data and D.A.M. and F.E.R. wrote the manuscript with assistance from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.E.R. (floyd@scripps.edu).

METHODS

Materials. 2 × YT, 2 × YT agar, IPTG, ampicillin and streptomycin were obtained from Fisher Scientific. Ampicillin and streptomycin were used at 100 μg ml^{-1} and 50 μg ml^{-1}, respectively. All pET-16b constructs containing the nucleotide transporters were kindly provided by I. Haferkamp (Technische Universität Kaiserslautern, Germany) with the exception of pET16b-*Rp*NTT2, which along with the C41(DE3) *E. coli* strain, was provided by J. P. Audia (University of South Alabama, USA). Plasmids pUC19 and pCDF-1b were obtained from Thermo Scientific and EMD Millipore, respectively. Plasmids were purified using the PureLink Quick Plasmid DNA Miniprep Kit (Life Technologies). OneTaq, DeepVent, Q5 Hot Start High-Fidelity DNA Polymerases, and all restriction endonucleases were obtained from New England Biolabs. In general, PCR reactions were divided into multiple aliquots with one followed in real time using 0.5 × Sybr Green I (Life Technologies); following PCR, the aliquots were recombined, purified by spin column (DNA Clean and Concentrator-5; Zymo Research, Irvine, California, USA) with elution in 20 μl of water, then separated by agarose gel electrophoresis, followed by band excision and recovery (Zymoclean Gel DNA Recovery Kit), eluting with 20 μl of water unless stated otherwise. Polyacrylamide gels were stained with 1 × Sybr Gold (Life Technologies) for 30 min, agarose gels were cast with 1 × Sybr Gold. All gels were visualized using a Molecular Imager Gel Doc XR+ equipped with 520DF30 filter (Bio-Rad) and quantified with Quantity One software (Bio-Rad). The sequences of all DNA oligonucleotides used in this study are provided in Supplementary Information. Natural oligonucleotides were purchased from IDT (San Diego, California, USA). The concentration of dsDNA was measured by fluorescent dye binding (Quant-iT dsDNA HS Assay kit, Life Technologies) unless stated otherwise. The concentration of ssDNA was determined by UV absorption at 260 nm using a NanoDrop 1000 (Thermo Scientific). [α-^{32}P]-dATP (25 μCi) was purchased from PerkinElmer (Shelton, Connecticut, USA). Polyethyleneimine cellulose pre-coated Bakerflex TLC plates (0.5 mm) were purchased from VWR. dNaM phosphoramidite, dNaM and d5SICS nucleosides were obtained from Berry & Associates Inc. (Dexter, Michigan, USA). Free nucleosides of dNaM and d5SICS (Berry & Associates) were converted to the corresponding triphosphates under Ludwig conditions[30]. After purification by anion exchange chromatography (DEAE Sephadex A-25) followed by reverse phase (C18) HPLC and elution through a Dowex 50WX2-sodium column, both triphosphates were lyophilized and kept at −20 °C until use. The d5SICSTP analogue dTPT3TP[22] and the biotinylated dNaMTP analogue dmmo2^{SSBIO}TP[4] were made as reported previously. MALDI-TOF mass spectrometry (Applied Biosystems Voyager DE-PRO System 6008) was performed at the TSRI Center for Protein and Nucleic Acid Research.

Construction of NTT expression plasmids. The Pt*NTT2* gene was amplified from plasmid pET-16b-*Pt*NTT2 using primers PtNTT2-forward and PtNTT2-reverse; the TpNTT2 gene was amplified from plasmid pET-16b-*Tp*NTT2 using primers TpNTT2-forward and TpNTT2-reverse. A linear fragment of pCDF-1b was generated using primers pCDF-1b-forward and pCDF-1b-reverse. All fragments were purified as described in Materials. The pCDF-1b fragment (100 ng, 4.4 × 10^{-14} mol) and either the Pt*NTT2* (78 ng, 4.4 × 10^{-14} mol) or Tp*NTT2* (85 ng, 4.4 × 10^{-14} mol) fragment were then assembled together using restriction-free circular polymerase extension cloning[29] in 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, and 0.02 U μl^{-1} of OneTaq DNA under the following thermal cycling conditions: initial denaturation (96 °C, 1 min); 10 cycles of denaturation (96 °C, 30 s), touchdown annealing (54 °C to 49.5 °C for 30 s (−0.5 °C per cycle)), extension of 68 °C for 5 min, and final extension (68 °C, 5 min). Upon completion, the samples were purified and used for heat-shock transformation of *E. coli* XL10. Individual colonies were selected on lysogeny broth (LB)-agar containing streptomycin, and assayed by colony PCR with primers PtNTT2-forward/reverse or TpNTT2-forward/reverse. The presence of the *NTT* genes was confirmed by sequencing and double digestion with ApaI/EcoO109I restriction endonucleases with the following expected pattern: pCDF-1b-*Pt*NTT2 (2,546/2,605 bp), pCDF-1b-*Tp*NTT2 (2,717/2,605 bp), pCDF-1b (1,016/2,605 bp). The complete nucleotide sequence of the pCDF-1b-*Pt*NTT2 plasmid (pACS) is provided in Supplementary Information.

Growth conditions to quantify nucleoside triphosphate uptake. *E. coli* C41(DE3)[16] freshly transformed with pCDF-1b-*Pt*NTT2 was grown in 2 × YT with streptomycin overnight, then diluted (1:100) into fresh 2 × YT medium (1 ml of culture per uptake with [α-^{32}P]-dATP; 2 ml of culture per uptake with d5SICSTP or dNaMTP) supplemented with 50 mM potassium phosphate (KPi) and streptomycin. A negative control with the inactive transporter pET-16b-*Rp*NTT2, was treated identically except ampicillin was used instead of streptomycin. Cells were grown to an OD_{600} of approximately 0.6 and the NTT expression was induced by the addition of IPTG (1 mM). The culture was allowed to grow for another hour (final OD_{600} approximately 1.2) and then assayed directly for uptake as described below using a method adapted from a previous paper[15].

Preparation of media fraction for unnatural nucleoside triphosphate analysis. The experiment was initiated by the addition of either dNaMTP or d5SICSTP (10 mM each) directly to the media to a final concentration of 0.25 mM. Cells were incubated with the substrate with shaking at 37 °C for 30 min and then pelleted (8,000 r.c.f. (relative centrifugal force) for 5 min, 4 °C). An aliquot of the media fraction (40 μl) was mixed with acetonitrile (80 μl) to precipitate proteins[31], and then incubated at 22 °C for 30 min. Samples were either analysed immediately by HPLC or stored at −80 °C until analysis. Analysis began with centrifugation (12,000 r.c.f. for 10 min at 22 °C), then the pellet was discarded, and the supernatant was reduced to approximately 20 μl by SpeedVac, resuspended in buffer A (see below) to a final volume of 50 μl, and analysed by HPLC (see below).

Preparation of cytoplasmic fraction for nucleoside triphosphate analysis. To analyse the intracellular desphosphorylation of the unnatural nucleoside triphosphate, cell pellets were subjected to 3 × 100 μl washes of ice-cold KPi (50 mM). Pellets were then resuspended in 250 μl of ice cold KPi (50 mM) and lysed with 250 μl of lysis buffer L7 of the PureLink Quick Plasmid DNA Miniprep Kit (200 mM NaOH, 1% w/v SDS), after which the resulting solution was incubated at 22 °C for 5 min. Precipitation buffer N4 (350 μl, 3.1 M potassium acetate, pH 5.5) was added, and the sample was mixed to homogeneity. Following centrifugation (>12,000 r.c.f. for 10 min, at 22 °C) the supernatant containing the unnatural nucleotides was applied to a Hypersep C18 solid phase extraction column (Thermo Scientific) prewashed with acetonitrile (1 ml) and buffer A (1 ml, see HPLC protocol for buffer composition). The column was then washed with buffer A and nucleotides were eluted with 1 ml of 50% acetonitrile:50% triethylammonium bicarbonate (TEAB) 0.1 M (pH 7.5). The eluent was reduced to approximately 50 μl in a SpeedVac and its volume was adjusted to 100 μl with buffer A before HPLC analysis.

HPLC protocol and nucleoside triphosphate quantification. Samples were applied to a Phenomenex Jupiter LC column (3 μm C18 300 Å, 250 × 4.6 mm) and subjected to a linear gradient of 0–40% B over 40 min at a flow rate of 1 ml min^{-1}. Buffer A: 95% 0.1 M TEAB, pH 7.5; 5% acetonitrile. Buffer B: 20% 0.1 M TEAB, pH 7.5; 80% acetonitrile. Absorption was monitored at 230, 273, 288, 326 and 365 nm.

Each injection series included two extra control samples containing 5 nmol of dNaMTP or d5SICSTP. The areas under the peaks that corresponded to triphosphate, diphosphate, monophosphate and free nucleoside (confirmed by MALDI-TOF) were integrated for both the control and the unknown samples (described above). After peak integration, the ratio of the unknown peak to the control peak adjusted for the loss from the extraction step (62% and 70% loss for dNaM and d5SICS, respectively, Extended Data Table 3), provided a measure of the amount of each of the moieties in the sample. To determine the relative concentrations of unnatural nucleotide inside the cell, the amount of imported unnatural nucleotide (d**X**TP, μmol) was then divided by the volume of cells, which was calculated as the product of the volume of a single *E. coli* cell (1 μm^3 based on a reported average value[32]; that is, 1 × 10^{-9} μl per cell) and the number of cells in each culture (OD_{600} of 1.0 equal to 1 × 10^9 cells per ml (ref. 32)). The *Rp*NTT2 sample was used as a negative control and its signal was subtracted to account for incomplete washing of nucleotide species from the media.

dATP uptake. To analyse the intracellular desphosphorylation of dATP, after induction of the transporter, the uptake reaction was initiated by the addition of dATP (spiked with [α-^{32}P]-dATP) to a final concentration of 0.25 mM, followed by incubation at 37 °C with shaking for 30 min. The culture was then centrifuged (8,000 r.c.f. for 5 min at 22 °C). Supernatant was analysed by TLC. Cell pellets were washed three times with ice-cold KPi (50 mM, 100 μl) to remove excess radioactive substrate, lysed with NaOH (0.2 M, 100 μl) and centrifuged (10,000 r.c.f. for 5 min at 22 °C) to remove cell debris; supernatant was analysed by TLC.

TLC analysis. Samples (1 μl) were applied on a 0.5 mm polyethyleneimine cellulose TLC plate and developed with sodium formate pH 3.0 (0.5 M, 30 s; 2.5 M, 2.5 min; 4.0 M, 40 min). Plates were dried using a heat gun and quantified by phosphorimaging (Storm Imager, Molecular Dynamics) and Quantity One software.

Optimization of nucleotide extraction from cells for HPLC injection. To minimize the effect of the lysis and triphosphate extraction protocols on the decomposition of nucleoside triphosphate within the cell, the extraction procedure was optimized for the highest recovery with the lowest extent of decomposition (Extended Data Table 3). To test different extraction methods, cells were grown as described above, washed, and then 5 nmol of either dNaMTP or d5SICSTP was added to the pellets, which were then subjected to different extraction protocols including boiling water, hot ethanol, cold methanol, freeze and thaw, lysozyme, glass beads, NaOH, trichloroacetic acid (TCA) with Freon, and perchloric acid (PCA) with KOH[33]. The recovery and composition of the control was quantified by HPLC as described above to determine the most effective procedure. Method 3—that is, cell lysis with NaOH (Extended Data Table 3)—was found to be most effective and reproducible, thus we further optimized it by resuspension of the pellets in ice-cold KPi (50 mM, 250 μl) before addition of NaOH to decrease dephosphorylation after cell lysis (Method 4). Cell pellets were then processed as described above. See above for the final extraction protocol.

Preparation of the unnatural insert for pINF construction. The TK-1-dNaM oligonucleotide containing dNaM was prepared using solid-phase DNA synthesis

with ultra-mild DNA synthesis phosphoramidites on CPG ultramild supports (1 μmol, Glen Research, Sterling, Virginia, USA) and an ABI Expedite 8905 synthesizer. After the synthesis, the DMT-ON oligonucleotide was cleaved from the solid support, deprotected and purified by Glen-Pak cartridge according to the manufacturer's recommendation (Glen Research), and then subjected to 8 M urea 8% PAGE. The gel was visualized by ultraviolet shadowing, the band corresponding to the 75-mer was excised, and the DNA was recovered by crush and soak extraction, filtration (0.45 μm), and final desalting over Sephadex G-25 (NAP-25 Columns, GE Healthcare). The concentration of the single stranded oligonucleotide was determined by ultraviolet absorption at 260 nm assuming that the extinction coefficient of dNaM at 260 nm is equal to that of dA. TK-1-dNaM (4 ng) was next amplified by PCR under the following conditions: 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, 0.1 mM of dNaMTP, 0.1 mM of the d5SICSTP analogue dTPT3TP, 1 μM of each of the primers pUC19-fusion-forward and pUC19-fusion-reverse, and 0.02 U μl^{-1} of OneTaq DNA Polymerase (in a total of 4 × 50 μl reactions) under the following thermal cycling conditions: initial denaturation (96 °C, 1 min) followed by 12 cycles of denaturation (96 °C, 10 s), annealing (60 °C, 15 s), and extension (68 °C, 2 min). An identical PCR without the unnatural triphosphates was run to obtain fully natural insert under identical conditions for the construction of the natural control plasmid. Reactions were subjected to spin column purification and then the desired PCR product (122 bp) was purified by a 4% agarose gel.

pUC19 linearization for pINF construction. pUC19 (20 ng) was amplified by PCR under the following conditions: 1 × Q5 reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, 1 μM of each primers pUC19-lin-forward and pUC19-lin-reverse, and 0.02 U μl^{-1} of Q5 Hot Start High-Fidelity DNA Polymerase (in a total of 4 × 50 μl reactions with one reaction containing 0.5 × Sybr Green I) under the following thermal cycling conditions: initial denaturation (98 °C, 30 s); 20 cycles of denaturation (98 °C, 10 s), annealing (60 °C, 15 s), and extension (72 °C, 2 min); and final extension (72 °C, 5 min). The desired PCR product (2,611 bp) was purified by a 2% agarose gel.

PCR assembly of pINF and the natural control plasmid. A linear fragment was amplified from pUC19 using primers pUC19-lin-forward and pUC19-lin-reverse. The resulting product (800 ng, 4.6×10^{-13} mol) was combined with either the natural or unnatural insert (see above) (56 ng, 7.0×10^{-13} mol) and assembled by circular overlap extension PCR under the following conditions: 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, 0.1 mM of dNaMTP, 0.1 mM of the d5SICSTP analogue dTPT3TP, and 0.02 U μl^{-1} of OneTaq DNA Polymerase (in a total of 4 × 50 μl reactions with one reaction containing 0.5 × Sybr Green I) using the following thermal cycling conditions: initial denaturation (96 °C, 1 min); 12 cycles of denaturation (96 °C, 30 s), annealing (62 °C, 1 min), and extension (68 °C, 5 min); final extension (68 °C, 5 min); and slow cooling (68 °C to 10 °C at a rate of -0.1 °C s^{-1}). The PCR product was analysed by restriction digestion on 1% agarose and used directly for *E. coli* transformation. The d5SICS analogue dTPT3[22] pairs with dNaM, and dTPT3TP was used in place of d5SICSTP as DNA containing dTPT3–dNaM is better PCR amplified than DNA containing d5SICS–dNaM, and this allowed for differentiation of synthetic and *in vivo* replicated pINF, as well as facilitated the construction of high-quality pINF (UBP content >99%).

Preparation of electrocompetent cells for pINF replication in *E. coli*. C41(DE3) cells were transformed by heat shock[34] with 200 ng of pACS plasmid, and the transformants were selected overnight on 2 × YT-agar supplemented with streptomycin. A single clone of freshly transformed C41(DE3) pACS was grown overnight in 2 × YT medium (3 ml) supplemented with streptomycin and KPi (50 mM). After 100-fold dilution into the same fresh 2 × YT media (300 ml), the cells were grown at 37 °C until they reached an OD_{600} of 0.20 at which time IPTG was added to a final concentration of 1 mM to induce the expression of *Pt*NTT2. Cells were grown for another 40 min and then growth was stopped by rapid cooling in ice water with intensive shaking. After centrifugation in a prechilled centrifuge (2,400 r.c.f. for 10 min, 4 °C), the spent media was removed, and the cells were prepared for electroporation by washing with ice-cold sterile water (3 × 150 ml). After washing, the cells were resuspended in ice-cold 10% glycerol (1.5 ml) and split into 50-μl aliquots. Although we found that dry ice yielded better results than liquid nitrogen for freezing cells to store for later use, freshly prepared cells were used for all reported experiments as they provided higher transformation efficiency of pINF and higher replication fidelity of the UBP.

Electroporation and recovery for pINF replication in *E. coli*. The aliquot of cells was mixed with 2 μl of plasmid (400 ng), transferred to 0.2 cm gap electroporation cuvette and electroporated using a Bio-Rad Gene Pulser according to the manufacturer's recommendations (voltage 25 kV, capacitor 2.5 μF, resistor 200 Ω, time constant 4.8 ms). Pre-warmed 2 × YT media (0.95 ml, streptomycin, 1 mM IPTG, 50 mM KPi) was added, and after mixing, 45 μl was removed and combined with 105 μl of the same media (3.33-fold dilution) supplemented with 0.25 mM of dNaMTP and d5SICSTP. The resulting mixture was allowed to recover for 1 h at 37 °C with shaking (210 revolutions per min (r.p.m.)). The original transformation media (10 μl) was spread onto 2 × YT-agar containing streptomycin with 10- and 50-fold dilutions for the determination of viable colony forming units after overnight growth at 37 °C to calculate the number of the transformed pINF molecules (see the section on calculation of the plasmid amplification). Transformation, recovery and growth were carried out identically for the natural control plasmid. In addition, a negative control was run and treated identically to pINF transformation except that it was not subjected to electroporation (Extended Data Fig. 7b). No growth in the untransformed negative control samples was observed even after 6 days. No PCR amplification of the negative control was detected, which confirms that unamplified pINF plasmid is not carried through cell growth and later detected erroneously as the propagated plasmid.

Analysis of pINF replication in *E. coli*. After recovery, the cells were centrifuged (4,000 r.c.f. for 5 min, 4 °C), and spent media (0.15 ml) was removed and analysed for nucleotide composition by HPLC (Extended Data Fig. 7a). The cells were resuspended in fresh 2 × YT media (1.5 ml, streptomycin, ampicillin, 1 mM IPTG, 50 mM KPi, 0.25 mM dNaMTP, 0.25 mM d5SICSTP) and grown overnight at 37 °C while shaking (250 r.p.m.), resulting in tenfold dilution compared to recovery media or 33.3-fold dilution compared to the originally transformed cells. Aliquots (100 μl) were taken after 15, 19, 24, 32, 43, 53, 77 and 146 h, OD_{600} was determined, and the cells were centrifuged (8,000 r.c.f. for 5 min, 4 °C). Spent media were analysed for nucleotide composition by HPLC (Extended Data Fig. 7a), and the pINF and pACS plasmid mixtures were recovered and linearized with *NdeI* restriction endonuclease; pINF plasmid was purified by 1% agarose gel electrophoresis (Extended Data Fig. 7b) and analysed by LC-MS/MS. The retention of the UBP on the pINF plasmid was quantified by biotin gel shift mobility assay and sequencing as described below.

Mass spectrometry of pINF. Linearized pINF was digested to nucleosides by treatment with a mixture of nuclease P1 (Sigma-Aldrich), shrimp alkaline phosphatase (NEB), and DNase I (NEB), overnight at 37 °C, following a previously reported protocol[25]. LC-MS/MS analysis was performed in duplicate by injecting 15 ng of digested DNA on an Agilent 1290 UHPLC equipped with a G4212A diode array detector and a 6490A Triple Quadrupole Mass Detector operating in the positive electrospray ionization mode (+ESI). UHPLC was carried out using a Waters XSelect HSS T3 XP column (2.1 × 100 mm, 2.5 μm) with the gradient mobile phase consisting of methanol and 10 mM aqueous ammonium formate (pH 4.4). MS data acquisition was performed in Dynamic Multiple Reaction Monitoring (DMRM) mode. Each nucleoside was identified in the extracted chromatogram associated with its specific MS/MS transition: dA at *m/z* 252⟶136, d5SICS at *m/z* 292⟶176, and dNaM at *m/z* 275⟶171. External calibration curves with known amounts of the natural and unnatural nucleosides were used to calculate the ratios of individual nucleosides within the samples analysed. LC-MS/MS quantification was validated using synthetic oligonucleotides[1] containing unnatural d5SICS and dNaM (Extended Data Table 2).

DNA biotinylation by PCR to measure fidelity by gel shift mobility assay. Purified mixtures of pINF and pACS plasmids (1 ng) from growth experiments were amplified by PCR under the following conditions: 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.3 mM of dNTP, 0.1 mM of the biotinylated dNaMTP analogue dMMO2^{SSBIO}TP, 0.1 mM of d5SICSTP, 1 μM of each of the primers pUC19-seq-forward and pUC19-seq-reverse, 0.02 U μl^{-1} of OneTaq DNA Polymerase, and 0.0025 U μl^{-1} of DeepVent DNA Polymerase in a total volume of 25 μl in an CFX Connect Real-Time PCR Detection System (Bio-Rad) under the following thermal cycling conditions: initial denaturation (96 °C, 1 min); 10 cycles of denaturation (96 °C, 30 s), annealing (64 °C, 30 s), and extension (68 °C, 4 min). PCR products were purified, and the resulting biotinylated DNA duplexes (5 μl, 25–50 ng) were mixed with streptavidin (1 μl, 1 μg μl^{-1}, Promega) in phosphate buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1 mM EDTA), incubated for 30 min at 37 °C, mixed with 5 × non-denaturing loading buffer (Qiagen), and loaded onto 6% non-denaturing PAGE. After running at 110 V for 30 min, the gel was visualized and quantified. The resulting fragment (194 bp) with primer regions underlined and the unnatural nucleotide in bold (X represents dNaM or its biotinylated analogue dMMO2^{SSBIO}) is 5′-GCAGGCATGCAAGCTTGGCGTAATCATGG TCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA**X**TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAG.

Streptavidin shift calibration for gel shift mobility assay. We have already reported a calibration between streptavidin shift and the fraction of sequences with UBP in the population (see Supplementary Fig. 8 of ref. 1). However, we found that spiking the PCR reaction with DeepVent improves the fidelity with which DNA containing d5SICS-dMMO2^{SSBIO} is amplified, and thus we repeated the calibration with added DeepVent. To quantify the net retention of the UBP, nine defined mixtures of the TK-1-dNaM template and its fully natural counterpart were prepared (Extended Data Fig. 6a), subjected to biotinylation by PCR and analysed by mobility-shift assay on 6% non-denaturing PAGE as described above. For calibration, the mixtures

TK-1-dNaM template and its fully natural counterpart with a known ratio of unnatural and natural templates (0.04 ng) were amplified under the same conditions over nine cycles of PCR with pUC19-fusion primers and analysed identically to samples from the growth experiment (see the section on DNA biotinylation by PCR). Each experiment was run in triplicate (a representative gel assay is shown in Extended Data Fig. 6b), and the streptavidin shift (SAS, %) was plotted as function of the UBP content (UBP, %). The data was then fit to a linear equation, SAS = $0.77 \times \mathrm{UBP} + 2.0$ ($R^2 = 0.999$), where UBP corresponds to the retention of the UBP (%) in the analysed samples after cellular replication and was calculated from the SAS shift using the equation above.

Calculation of plasmid amplification. The cells were plated on 2 × YT-agar containing ampicillin and streptavidin directly after transformation with pINF, and the colonies were counted after overnight growth at 37 °C. Assuming each cell is only transformed with one molecule of plasmid, colony counts correspond to the original amount of plasmid that was taken up by the cells. After overnight growth, the plasmids were purified from a specific volume of the cell culture and quantified. As purified plasmid DNA represents a mixture of the pINF and pACS plasmids, digestion restriction analysis with NdeI exonuclease was performed to linearize both plasmids, followed by 1% agarose gel electrophoresis (Extended Data Fig. 7b). An example of calculations for the 19-h time point with one of three triplicates is provided in Supplementary Information.

Fragment generation for Sanger sequencing to measure fidelity. Purified mixtures of pINF and pACS plasmids (1 ng) after the overnight growth were amplified by PCR under the following conditions: 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, 0.1 mM of dNaMTP, 0.1 mM of the d5SICSTP analogue dTPT3TP, 1 μM of each of the primers pUC19-seq2-forward and pUC19-seq-reverse (see below), and 0.02 U μl^{-1} of OneTaq DNA Polymerase in a total volume of 25 μl under the following thermal cycling conditions: initial denaturation (96 °C, 1 min); and 10 cycles of denaturation (96 °C, 30 s), annealing (64 °C, 30 s), and extension (68 °C, 2 min). Products were purified by spin column, quantified to measure DNA concentration and then sequenced as described below. The sequenced fragment (304 bp) with primer regions underlined and the unnatural nucleotide in bold (X, dNaM) is 5′-GCTGCAAGGCGATTAAGTTGGGTAACGCC AGGGT TTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCG GTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCG TAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA**X** TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTG CCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCG CTTTCCAGTCGGGAAACCTGTCGTGCCAG.

Sanger sequencing. The cycle sequencing reactions (10 μl) were performed on a 9800 Fast Thermal Cycler (Applied Biosystems) with the Cycle Sequencing Mix (0.5 μl) of the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) containing 1 ng template and 6 pmol of sequencing primer pUC19-seq-reverse under the following thermal cycling conditions: initial denaturation (98 °C, 1 min); and 25 cycles of denaturation (96 °C, 10 s), annealing (60 °C, 15 s), and extension (68 °C, 2.5 min). Upon completion, the residual dye terminators were removed from the reaction with Agencourt CleanSEQ (Beckman-Coulter, Danvers, Massachusetts, USA). Products were eluted off the beads with deionized water and sequenced directly on a 3730 DNA Analyzer (Applied Biosystems). Sequencing traces were collected using Applied Biosystems Data Collection software v3.0 and analysed with the Applied Biosystems Sequencing Analysis v5.2 software.

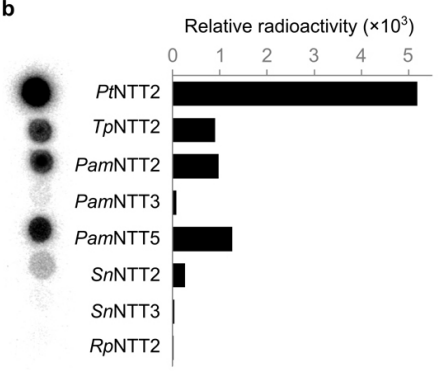
Analysis of Sanger sequencing traces. Sanger sequencing traces were analysed as described previously[1,26] to determine the retention of the unnatural base pair. In brief, the presence of an unnatural nucleotide leads to a sharp termination of the sequencing profile, whereas mutation to a natural nucleotide results in 'read-through'. The extent of this read-through after normalization is inversely correlated with the retention of the unnatural base pair. Raw sequencing traces were analysed by first adjusting the start and stop points for the Sequencing Analysis software (Applied Biosystems) and then determining the average signal intensity individually for each channel (A, C, G and T) for peaks within the defined points. This was done separately for the parts of the sequencing trace before (section *L*) and after (section *R*) the unnatural nucleotide. The *R*/*L* ratio after normalization $(R/L)_{\mathrm{norm}}$ for sequencing decay and read-through in the control unamplified sample ($R/L = 0.55(R/L)_{\mathrm{norm}} + 7.2$, see ref. 26 for details) corresponds to the percentage of the natural sequences in the pool. Therefore, an overall retention (F) of the incorporation of the unnatural base pair during PCR is equal to $1 - (R/L)_{\mathrm{norm}}$. As significant read-through (over 20%) was observed in the direction of the pUC19-seq2-forward primer even with the control plasmid (synthetic pINF); sequencing of only the opposite direction (pUC19-seq-reverse) was used to gauge fidelity. Raw sequencing traces are shown in Fig. 2f and provided as Supplementary Data.

30. Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5′-0-(1-thiotriphosphates), 5′-triphosphates and 2′,3′-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **54,** 631–635 (1989).
31. Alpert, A. & Shukla, A. Precipitation of Large, High-Abundance Proteins from Serum with Organic Solvents in *ABRF 2003: Translating Biology using Proteomics and Functional Genomics* Poster no. P111-W http://www.abrf.org/Other/ABRFMeetings/ABRF2003/Alpert.pdf (2003).
32. Kubitschek, H. E. & Friske, J. A. Determination of bacterial cell volume with the Coulter Counter. *J. Bacteriol.* **168,** 1466–1467 (1986).
33. Yanes, O., Tautenhahn, R., Patti, G. J. & Siuzdak, G. Expanding coverage of the metabolome for global metabolite profiling. *Anal. Chem.* **83,** 2152–2161 (2011).
34. Seidman, C. E., Struhl, K., Sheen, J. & Jessen, T. Introduction of plasmid DNA into cells. *Curr. Prot. Mol. Biol.* Chapter 1, Unit 1.8 (2001).
35. Knab, S., Mushak, T. M., Schmitz-Esser, S., Horn, M. & Haferkamp, I. Nucleotide parasitism by *Simkania negevensis* (Chlamydiae). *J. Bacteriol.* **193,** 225–235 (2011).
36. Audia, J. P. & Winkler, H. H. Study of the five *Rickettsia prowazekii* proteins annotated as ATP/ADP translocases (Tlc): Only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. *J. Bacteriol.* **188,** 6261–6268 (2006).
37. Hofer, A., Ekanem, J. T. & Thelander, L. Allosteric regulation of *Trypanosoma brucei* ribonucleotide reductase studied *in vitro* and *in vivo*. *J. Biol. Chem.* **273,** 34098–34104 (1998).
38. Reijenga, J. C., Wes, J. H. & van Dongen, C. A. M. Comparison of methanol and perchloric acid extraction procedures for analysis of nucleotides by isotachophoresis. *J. Chromatogr. B Biomed. Sci. Appl.* **374,** 162–169 (1986).

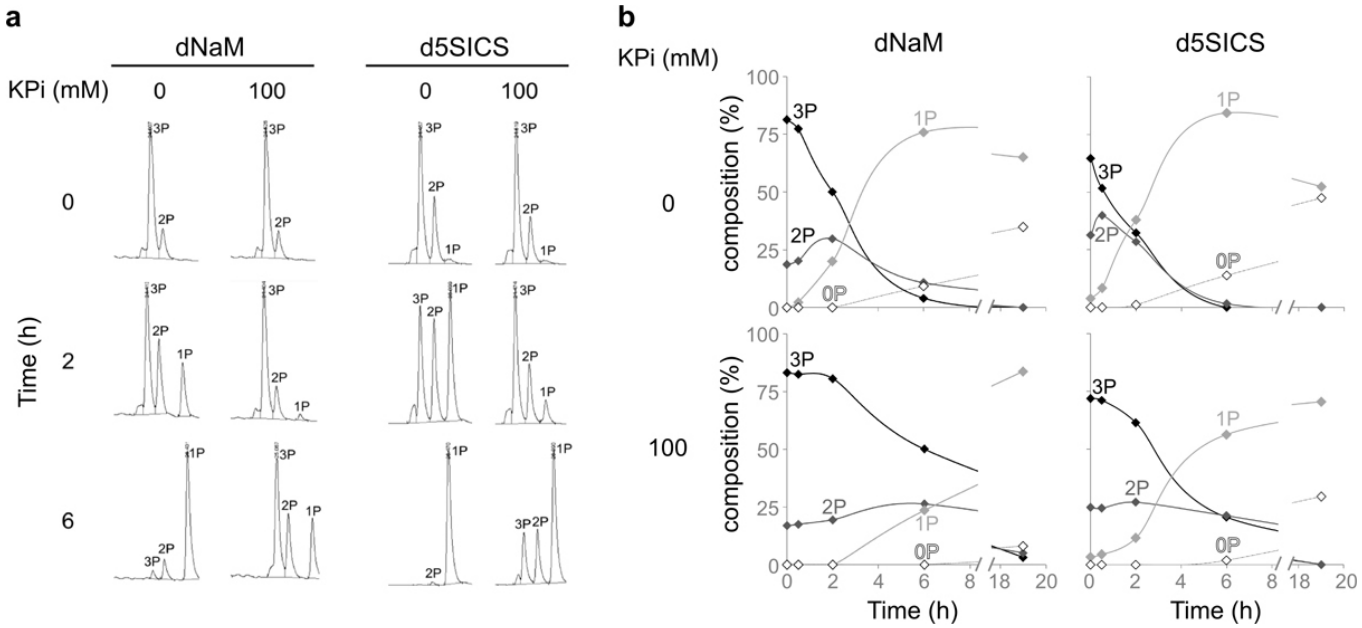
a

Transporter	Deoxyribotriphosphates				Ribotriphosphates				Reference
	dG	dA	dC	T	G	A	C	U	
*Pt*NTT2	82	271	31	428	78	197	49	86	18
*Tp*NTT2	74	665	60	251	59	49	33	90	18
*Pam*NTT2					156	437	570	676	15
*Pam*NTT3								1320	15
*Pam*NTT5	121				22	360			15
*Sn*NTT2					179	654			35
*Sn*NTT3			42		407	375	9	34	35
*Rp*NTT2	substrate is unknown								36

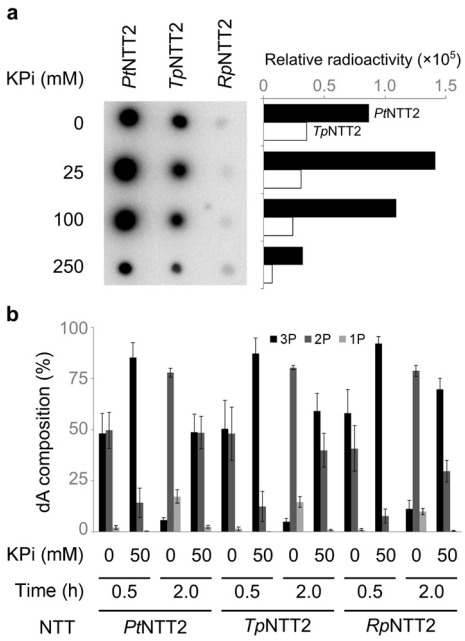
b



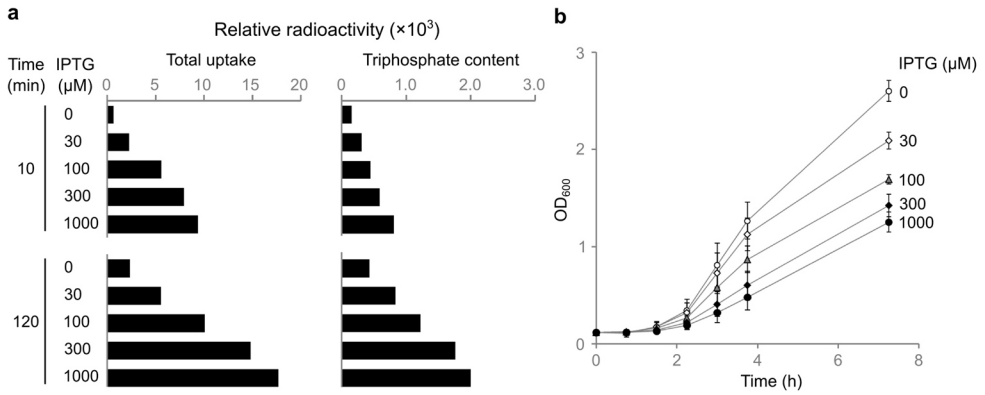
Extended Data Figure 1 | Natural triphosphate uptake by NTTs. a, Survey of reported substrate specificity (K_M, μM) of the NTTs assayed in this study. **b**, *Pt*NTT2 is significantly more active in the uptake of [α-^{32}P]-dATP compared to other nucleotide transporters. Raw (left) and processed (right) data are shown. Relative radioactivity corresponds to the total number of counts produced by each sample. Interestingly, both *Pam*NTT2 and *Pam*NTT5 exhibit a measurable uptake of dATP although this activity was not reported before. This can possibly be explained by the fact that substrate specificity was only characterized using competition experiments, and assay sensitivity might not have been adequate to detect this activity[15]. References 35, 36 are cited in this figure.



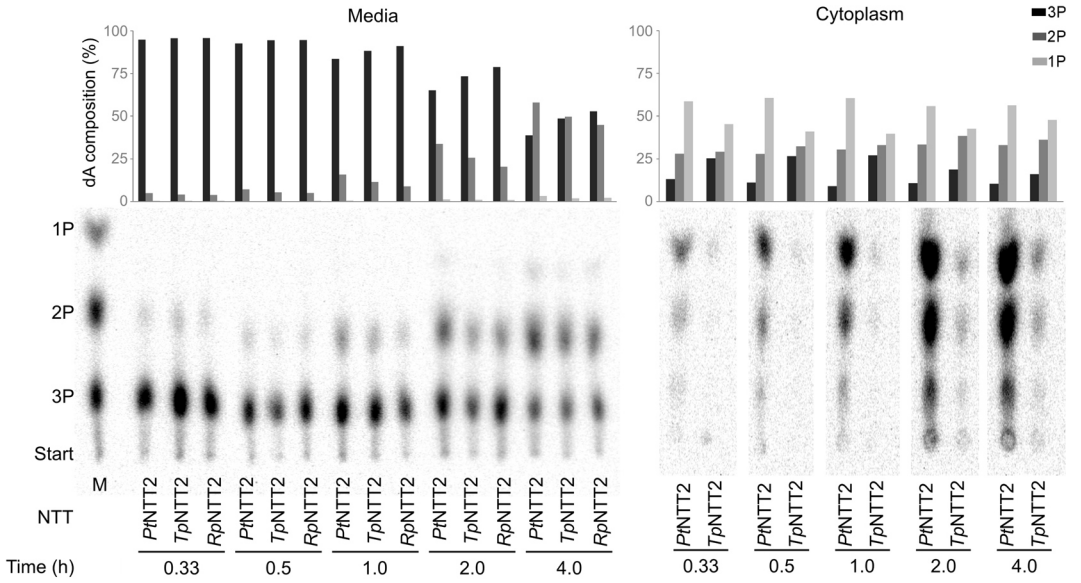
Extended Data Figure 2 | Degradation of unnatural triphosphates in growth media. Unnatural triphosphates (3P) of dNaM and d5SICS are degraded to diphosphates (2P), monophosphates (1P) and nucleosides (0P) in the growing bacterial culture. Potassium phosphate (KPi) significantly slows down the dephosphorylation of both unnatural triphosphates. **a**, Representative HPLC traces (for the region between ~20 and 24 min). dNaM and d5SICS nucleosides are eluted at approximately 40 min and not shown. **b**, Composition profiles.



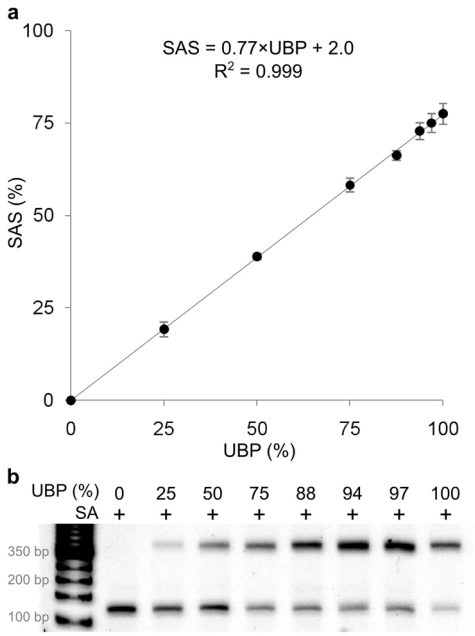
Extended Data Figure 3 | Effect of potassium phosphate on dATP uptake and stability in growth media. **a**, KPi inhibits the uptake of [α-^{32}P]-dATP at concentrations above 100 mM. Raw (left) and processed (right) data are shown. The NTT from *Rickettsia prowazekii* (*Rp*NTT2) does not mediate the uptake of any of the dNTPs and was used as a negative control: its background signal was subtracted from those of *Pt*NTT2 (black bars) and *Tp*NTT2 (white bars). Relative radioactivity corresponds to the total number of counts produced by each sample. **b**, KPi (50 mM) significantly stabilizes [α-^{32}P]-dATP in the media. Triphosphate stability in the media is not significantly affected by the nature of the NTT expressed. 3P, 2P and 1P correspond to triphosphate, diphosphate and monophosphate states, respectively. Error bars represent s.d. of the mean, $n = 3$.



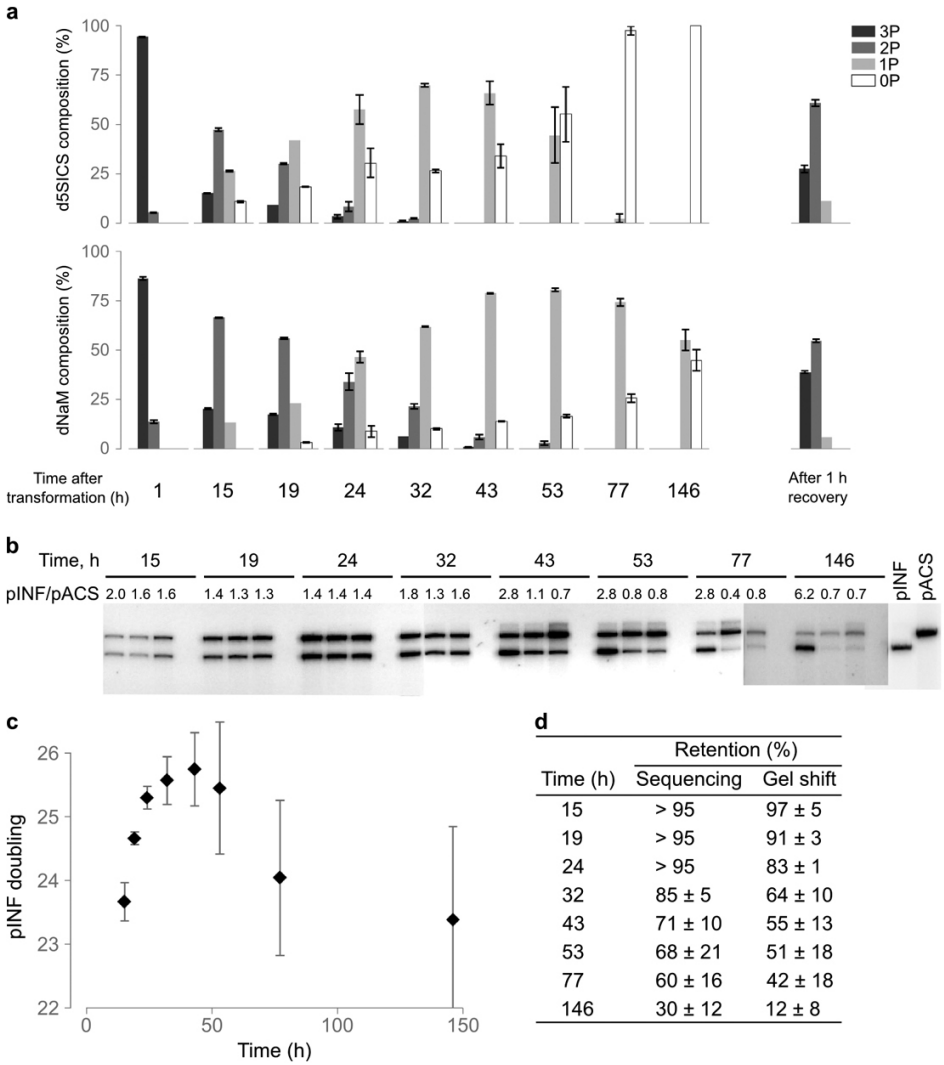
Extended Data Figure 4 | dATP uptake and growth of cells expressing *Pt*NTT2 as a function of inducer (IPTG) concentration. Growth curves and [α-^{32}P]-dATP uptake by bacterial cells transformed with pCDF-1b-*Pt*NTT2 (pACS) plasmid as a function of IPTG concentration. **a**, Total uptake of radioactive substrate (left) and total intracellular triphosphate content (right) are shown at two different time points. Relative radioactivity corresponds to the total number of counts produced by each sample. **b**, A stationary phase culture of C41(DE3) pACS cells was diluted 100-fold into fresh 2 × YT media containing 50 mM KPi, streptomycin, and IPTG at the indicated concentrations and were grown at 37°C. Error bars represent s.d. of the mean, $n = 3$.



Extended Data Figure 5 | Stability and uptake of dATP in the presence of 50 mM KPi and 1 mM IPTG. Composition of [α-^{32}P]-dATP in the media (left) and cytoplasmic fraction (right) as a function of time. TLC images and their quantifications are shown at the bottom and the top of each of the panels, respectively. 3P, 2P and 1P correspond to nucleoside triphosphate, diphosphate and monophosphate, respectively. M refers to a mixture of all three compounds that was used as a TLC standard. The position labelled 'Start' corresponds to the position of sample spotting on the TLC plate.



Extended Data Figure 6 | Calibration of the streptavidin shift (SAS). **a**, The SAS is plotted as a function of the fraction of template containing the UBP. Error bars represent s.d. of the mean, $n = 3$. **b,** Representative data. SA, streptavidin.



d

Time (h)	Retention (%)	
	Sequencing	Gel shift
15	> 95	97 ± 5
19	> 95	91 ± 3
24	> 95	83 ± 1
32	85 ± 5	64 ± 10
43	71 ± 10	55 ± 13
53	68 ± 21	51 ± 18
77	60 ± 16	42 ± 18
146	30 ± 12	12 ± 8

Extended Data Figure 7 | Decomposition of unnatural triphosphates, pINF quantification, and retention of the UBP with extended cell growth. **a**, Dephosphorylation of the unnatural nucleoside triphosphate. 3P, 2P, 1P and 0P correspond to triphosphate, diphosphate, monophosphate and nucleoside states, respectively. The composition at the end of the 1 h recovery is shown at the right. **b**, Restriction analysis of pINF and pACS plasmids purified from *E. coli*, linearized with NdeI restriction endonuclease and separated on a 1% agarose gel (assembled from independent gel images). Molar ratios of pINF/pACS plasmids are shown at the top of each lane. For each time point, triplicate data are shown in three lanes with the untransformed control shown in the fourth, rightmost lane (see Methods). **c**, Number of pINF doublings as a function of time. The decrease starting at approximately 50 h is due to the loss of the pINF plasmid that also results in increased error. See the section on pINF replication in *E. coli* in the Methods for details. **d**, UBP retention (%) as a function of growth as determined by gel shift (data shown in Fig. 3) and Sanger sequencing (sequencing traces are available as Supplementary Data). In **a**, **c** and **d**, error shown is the s.d. of mean, $n = 3$.

Extended Data Table 1 | OD_{600} of *E. coli* cultures and relative copy number of plasmid (pINF or control pUC19) as determined by its molar ratio to pACS after 19 h of growth

Plasmid	IPTG	dXTP/dYTP	Relative Copy Number	OD_{600} (15 h)	OD_{600} (19 h)
pINF	+	+	1.8	0.34	0.75
	+	–	4.6	0.15	0.75
	–	+	8.9	3.13	3.98
pUC19	+	+	2.8	0.54	1.25
	+	–	2.6	0.73	1.39

X, NaM; Y, 5SICS.

Extended Data Table 2 | Relative quantification by LC-MS/MS using synthetic oligonucleotides containing d5SICS and dNaM

Oligonucleotide (ssDNA)	Size (nt)	Sequence	dA/d5SICS Exp. (calcd.)	dA/dNaM Exp. (calcd.)
D6-NaM	82	CAC ACA GGA AAC AGC TAT GAC CCG GGT TAT TAC ATG CGC TAG CAC TTG GAA TTC ACC AG ACG NNN NaM NNN CGG GAC CCA TAG T		**22.5** (23.5)*
D6-5SICS	87	GAA ATT AAT ACG ACT CAC TAT AGG GTT AAG CTT AAC TTT AAG AAG GAG ATT TAC TAT GGG TCC CG NNN 5SICS NNN CGT CTG GTG AAT TCC	**23.4** (25.5)*	
D13-NaM×2	130	CAC ACA GGA AAC AGC TAT GAC CCG GGT TAT TAC ATG CGC TAG CAC TTG GAA TTC ACT ATC AC NaM AGT CAC NaM AGT AAT CCA TAG TAA ATC TCC TTG TTA AGC TTA ACC CTA TAG TGA GTC GTA TTA ATT TCT		**16.1** (19.5)

* dA/d5SICS and dA/dNaM ratios were calculated assuming that randomized nucleotides (N) around the unnatural base are distributed equally.

Extended Data Table 3 | Summary of the most successful extraction methods

Method		Protocol summary	Total recovery (%)*		Triphosphate stability (%)†		Ref.
			dNaM	d5SICS	dNaM	d5SICS	
1.	TCA with Freon	• Lyse with cold TCA • Extract aqueous phase using Freon with trioctylamine solution	38	23	92	99	Adapted from Ref. 37
2.	PCA w/KOH	• Lyse with cold PCA • Precipitate proteins with KOH and KPi	36	21	98	77	Adapted from Ref. 38
3.	NaOH w/ KOAc	• Lyse with NaOH and SDS • Precipitate proteins with potassium acetate	21	26	86	100	see footnotes
4.	NaOH w/ KOAc supplemented w/ KPi (50 mM)	• Suspend cells in KPi • Lyse with NaOH and SDS • Precipitate proteins with potassium acetate	38	30	99	100	see footnotes

* Recovery of all nucleotides (3P, 2P, 1P and nucleoside).
† Calculated as a ratio of 3P composition (%) before and after the extraction.
References 37, 38 are cited in this figure. Details of methods 3 and 4 can be found online (http://2013.igem.org/wiki/images/e/ed/BGU_purelink_quick_plasmid_qrc.pdf).