

Directed evolution of genetic parts and circuits by compartmentalized partnered replication

Jared W Ellefson^{1,5}, Adam J Meyer^{1,5}, Randall A Hughes^{1,2}, Joe R Cannon³, Jennifer S Brodbelt³ & Andrew D Ellington¹⁻⁴

Most existing directed evolution methods, both *in vivo*¹⁻³ and *in vitro*⁴⁻⁶, suffer from inadvertent selective pressures (i.e., altering organism fitness), resulting in the evolution of products with unintended or suboptimal function. To overcome these barriers, here we present compartmentalized partnered replication (CPR). In this approach, synthetic circuits are linked to the production of *Taq* DNA polymerase so that evolved circuits that most efficiently drive *Taq* DNA polymerase production are enriched by exponential amplification during a subsequent emulsion PCR step. We apply CPR to evolve a T7 RNA polymerase variant that recognizes an orthogonal promoter and to reengineer the tryptophanyl tRNA-synthetase:suppressor tRNA pair from *Saccharomyces cerevisiae*⁷ to efficiently and site-specifically incorporate an unnatural amino acid into proteins. In both cases, the CPR-evolved parts were more orthogonal and/or more active than variants evolved using other methods. CPR should be useful for evolving any genetic part or circuit that can be linked to *Taq* DNA polymerase expression.

With the emergence of sophisticated DNA synthesis and cloning techniques, the creation of *in vivo*-based synthetic circuitry has become commonplace^{8,9}. However, in contrast to the relative ease with which one can design electronic circuits exhibiting predictable and precise behavior, it is not yet possible to reliably design biological pathways of equivalent complexity¹⁰. Most methods for refining synthetic circuits so that they behave in a predictable and reliable manner rely on the bioengineering principles of screening and selection; however, this process has proven challenging, in part because of the inherent linkage of the host organism's fitness and the circuit's function¹¹. Evolution and engineering techniques, including phage-assisted continuous evolution (PACE)³ and multiplex automated genomic engineering (MAGE)¹², were developed to better facilitate engineering of synthetic circuits, but each technique has limitations. PACE relies on a viral replication cycle, thus limiting control over the evolutionary process, whereas MAGE has relied on manual screening of the libraries generated and is therefore time consuming and laborious. Here we present CPR, an engineering platform capable of efficiently improving synthetic gene circuitry by using advantages of both *in vivo* and *in vitro* approaches.

In a previously described, related method called compartmentalized self-replication, DNA polymerases produced in cells facilitate the compartmentalized *in vitro* amplification of their own genes^{13,14}. Although this technique has proven to be effective for the evolution of novel DNA polymerases, the method is of limited utility for the evolution of other genes or genetic circuits. CPR addresses this limitation by allowing the coupling of *Taq* DNA polymerase production to a variety of other gene functions; these gene functions form the synthetic circuit of interest. As such, synthetic circuits that most efficiently drive *Taq* DNA polymerase production in cells in the *in vivo* step will be preferentially amplified during the subsequent compartmentalized *in vitro* PCR step (Fig. 1a). More specifically, *Escherichia coli* is transformed by a library of genetic circuits or parts designed to drive expression of *Taq* DNA polymerase. Host cells are then separated into emulsion compartments that contain primers specific for the genetic circuit or part. Upon thermal cycling, cell lysis occurs and the compartment-to-compartment variations in the amount of *Taq* DNA polymerase result in preferential amplification of genes in compartments containing the most *Taq* DNA polymerase; these compartments correspond to the cells transformed with the most efficient genetic circuits or parts. CPR is a modular platform that should in theory be capable of evolving almost any genetic part or circuit that can influence *Taq* DNA polymerase production or activity. Here, to demonstrate the capacity of CPR to effectively evolve genetic circuitry and to facilitate comparison to other molecular engineering methods, we wired CPR to evolve circuitry for two important and ongoing synthetic biology efforts—the generation of orthogonal transcription machinery (Fig. 1b) and expansion of the genetic code (Fig. 1c).

We first constructed a genetic circuit in which transcription of *Taq* DNA polymerase relies on T7 RNA polymerase (T7 RNAP) binding to and activating its promoter (Fig. 1b). In a proof-of-principle experiment, we initially coupled the production of *Taq* DNA polymerase to a wild-type (WT) T7 RNA polymerase promoter (P_{T7}), and generated a library of T7 RNA polymerase variants in which the six amino acids of the T7 RNAP specificity loop¹⁵ (R746, L747, N748, R756, L757, Q758; RLN...RLQ) were fully randomized. *E. coli* was transformed by the library, and the cells were emulsified in the presence of primers specific for the T7 RNAP gene. Presumably, cells expressing

¹Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA. ²Applied Research Laboratories, University of Texas at Austin, Austin, Texas, USA. ³Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas, USA. ⁴Center for Systems & Synthetic Biology, University of Texas at Austin, Austin, Texas, USA. ⁵These authors contributed equally to this work. Correspondence should be addressed to A.D.E. (andy.ellington@mail.utexas.edu).

Received 25 July; accepted 5 September; published online 3 November 2013; doi:10.1038/nbt.2714

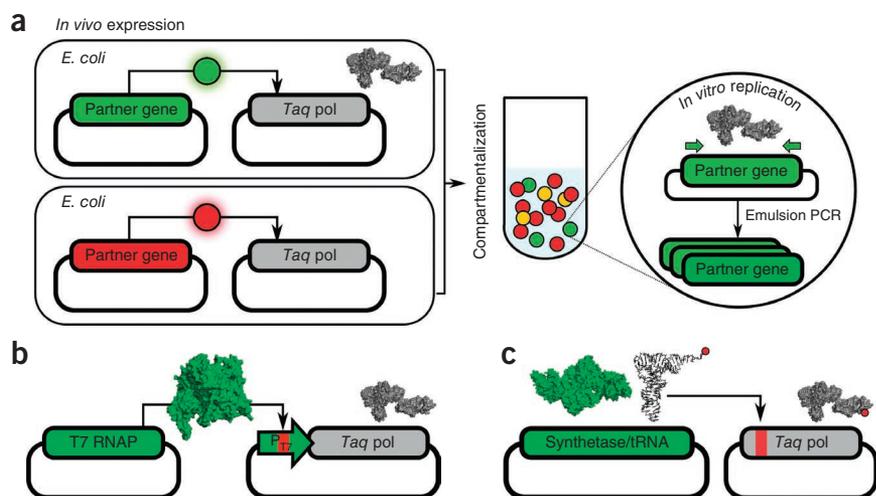


Figure 1 General CPR concept. (a) In the *in vivo* expression step, *E. coli* are transformed with genetic circuits or parts (partner genes) designed to drive production of *Taq* DNA polymerase (*Taq pol*). The partner gene–encoded biomolecules that display an active phenotype (green) produce *Taq pol* whereas inactive biomolecules (red) do not. Whole *E. coli* cells are compartmentalized by a water-in-oil emulsion along with primers, dNTPs and *Taq* DNA polymerase buffer. Emulsions are thermal cycled, leading to *E. coli* cell lysis and preferential *in vitro* PCR amplification of partner genes that drove production of the most *Taq* DNA polymerase during the *in vivo* expression step. (b) *In vivo* CPR design for the evolution of orthogonal T7 RNA polymerase:promoter pairs. A T7 RNA polymerase library drives the expression of *Taq pol* from a mutant promoter sequence. (c) *In vivo* CPR diagram for the evolution of tRNA synthetase:suppressor tRNAs. A tRNA and/or tRNA synthetase library suppresses amber codons in the *Taq pol* gene to generate active polymerase.

those T7 RNAP variants that most efficiently bound to and activated P_{T7} contained higher amounts of *Taq* DNA polymerase. Heat lysis released *Taq* DNA polymerase and T7 RNAP gene variants into individual emulsion bubbles, and the emulsion bubbles were subjected to thermal cycling. After four iterative rounds of this process, the library was enriched for variants driving robust expression from the P_{T7} (as shown by GFP expression from P_{T7} ; **Supplementary Fig. 1**) and converged on a consensus sequence resembling that of wild-type T7 RNAP, although some variation remained at the noncritical positions L747 and L757 (**Supplementary Fig. 2**).

Based on this result, we attempted to evolve more orthogonal circuitry. Specifically, the randomized specificity loop library was tasked with driving *Taq* DNA polymerase expression from a promoter, P_{CGG} , which differed from P_{T7} at positions 11, 10 and 9. After seven rounds of CPR, a single sequence “RVH...EMQ” dominated the population (**Supplementary Fig. 3**), and transcriptional activity of an evolved variant bearing this motif (CGG-R7-8) was analyzed. Whereas the CGG-R7-8 T7 RNAP activated expression of GFP driven by the P_{CGG} promoter, this activation was equivalent to only 2% of the activity of WT T7 RNAP on the WT P_{T7} promoter (**Supplementary Fig. 4**). Therefore, we reasoned that additional T7 RNAP mutations, not present in the initial library, were needed to compensate for the changes to the promoter sequence.

To further improve the activity of the CGG-R7-8 polymerase on the P_{CGG} promoter, we performed an additional five rounds of selection by CPR, with a larger (~500 bp) region of the CGG-R7-8 polymerase immediately flanking the specificity loop subjected to error-prone PCR before the first and fourth round. We transformed *E. coli* with the round-12 population, so they expressed a P_{CGG} -driven GFP reporter and characterized the eight T7 RNAP variants that drove the highest GFP expression (**Supplementary Fig. 5**). The most active T7 RNA polymerase variant from this population, which we termed CGG-R12-KI,

displayed 20% of the activity of the WT T7 RNAP:promoter pair (**Supplementary Fig. 4**).

Additional mutations beyond those seen in CGG-R12-KI were also frequently observed in variants in the round-12 population (**Supplementary Fig. 5**). We inserted combinations of three of these additional, highly represented mutations into CGG-R12-KI; this further improved its *in vivo* activity to roughly 40–60% of the WT T7 RNAP:promoter pair (**Supplementary Fig. 4**). The five most active variants were mixed in equal ratios and selected by CPR for an additional four rounds (with error-prone PCR before the first, third and fourth round). The resulting round-16 population closely resembled the mutant CGG-R12-KIRV (**Supplementary Fig. 6**), which robustly drove the expression of GFP from the P_{CGG} promoter but demonstrated only minimal activity on the P_{T7} promoter (~1% cross-reactivity) (**Fig. 2a**). In turn, the WT T7 RNAP did not markedly drive expression of the P_{CGG} -GFP reporter *in vivo*. When expressed, purified and assayed *in vitro*, CGG-R12-KIRV and WT T7 RNAP also demonstrated <0.1% promoter cross-reactivity (**Fig. 2b** and **Supplementary Fig. 7**).

Although T7 RNAP variants bearing altered promoter specificity have been evolved and engineered over two decades by various means^{3,16–18}, the CPR-evolved CGG-R12-KIRV is the most orthogonal and among the most active T7 RNAP:promoter pair when compared with previously described T7 RNAP variants that have been engineered and evolved (**Fig. 2c**, **Supplementary Figs. 8–10** and **Supplementary Table 1**).

Next, CPR was applied to evolve an orthogonal aminoacyl tRNA synthetase:tRNA pair facilitating site-directed incorporation of an unnatural amino acid, 5-hydroxy-L-tryptophan (5HTP). Previously, the tryptophanyl tRNA synthetase from *S. cerevisiae* (ScWRS) and its corresponding suppressor tRNA have been adapted to the orthogonal suppression of amber codons in *E. coli*⁷. To redirect CPR for evolving synthetic translation machinery, we generated a *Taq* DNA polymerase variant containing amber stop codons and attempted to evolve variants of both a tRNA synthetase and its cognate suppressor tRNA with altered substrate specificity and improved amber suppression efficiency (**Fig. 1c**).

Based on the previously described crystal structure of the yeast tryptophanyl tRNA synthetase, we randomized three residues adjacent to the presumed position of the 5-hydroxy moiety in the tryptophan analog (T107, P254 and C255)¹⁹. The library was used to transform an *E. coli* strain that expressed an amber codon-containing *Taq* DNA polymerase, and the transformed *E. coli* was grown in media supplemented with 5HTP. After each of three rounds of CPR, the library became more enriched for tRNA synthetase variants capable of amber suppression (**Supplementary Fig. 11**). Mock selections indicated that we could expect up to 100-fold enrichment per round of CPR (**Supplementary Fig. 12**); thus three rounds of selection and amplification should have substantively narrowed the initial pool of tRNA synthetases. Individual tRNA synthetase variants were picked and screened for activity using a modified *E. coli* strain with an amber codon in the β -galactosidase gene. Several tRNA synthetase variants effectively suppressed the amber codon, resulting in a visibly blue

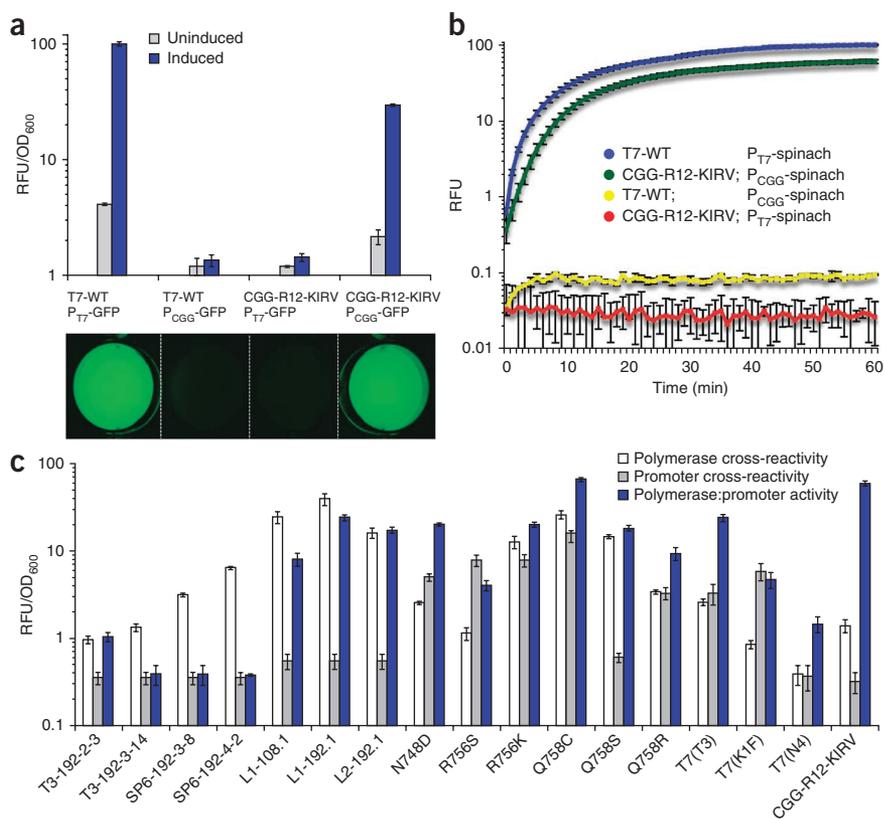
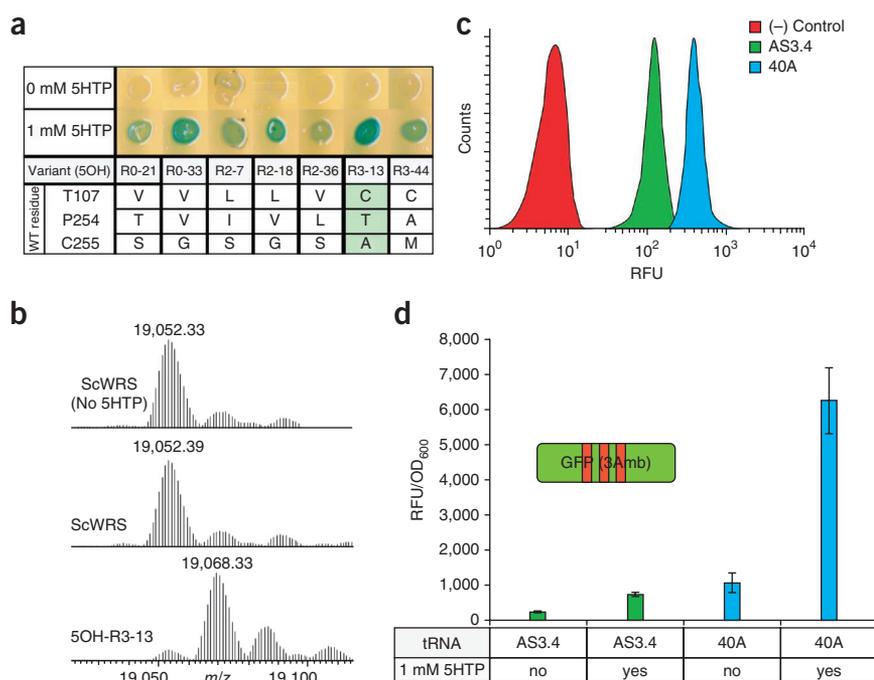


Figure 2 CPR selection of an orthogonal T7 RNA polymerase. **(a)** Activity of T7-WT RNAP and the CPR-evolved variant CGG-R12-KIRV in *E. coli* cells expressing P_{T7} and P_{CGG} -driven GFP reporters (top). Induced cultures were imaged with UV transillumination and a digital camera (bottom). **(b)** Activity of T7-WT RNAP and the CPR-evolved variant CGG-R12-KIRV in an *in vitro* transcription assay using P_{T7} and P_{CGG} -driven expression of the spinach aptamer²⁷ as a readout. Spinach fluorescence was read every minute and plotted as a function of time. **(c)** Activity of several evolved or engineered T7 RNAP variants (**Supplementary Table 1**) in *E. coli* cells expressing P_{T7} and P_{cog} -driven GFP reporters. P_{cog} refers to the cognate promoter for the mutant being assayed (**Supplementary Table 1**). Polymerase cross-reactivity refers to activity of the mutant T7 RNAP on the WT promoter. Promoter cross-reactivity refers to activity of the T7-WT RNAP on the mutant promoter. Polymerase:promoter activity refers to activity of the mutant polymerase on the mutant promoter. Fluorescence was quantified on a Tecan Safire monochromator. The WT polymerase:promoter pair's value was defined as 100 in each experiment. *In vivo* fluorescence was normalized to OD_{600} ; fluorescence/ OD_{600} ratio reported is the average of three independently grown cultures. *In vitro* fluorescence reported is the average of three independently assembled transcription reactions. Error bars represent one s.d.

colony, only when 5HTP was present (**Fig. 3a**). The most active and specific utilizer of 5HTP, 5OH-R3-13, had the amino acid substitutions T107C, P254T and C255A. Mutational analysis of single and double mutants of 5OH-R3-13 demonstrated that although it does not directly contact the 5-hydroxy moiety, the T107C mutation was the crucial step toward specificity (**Supplementary Fig. 13**). In the wild-

type tRNA synthetase binding pocket, T107 appears to be hydrogen-bonding with the carbonyl oxygen between residues P254 and C255. Mutation of T107 may eliminate this bond and instead allow the 5-hydroxyl moiety to hydrogen-bond with the carbonyl oxygen. Further specificity is gained after mutating P254T, potentially stabilizing the pocket with an additional hydrogen bond to T127 or perhaps

Figure 3 CPR evolved 5-hydroxy-L-tryptophan-using tRNA synthetase and optimized tRNA. **(a)** tRNA synthetase variants from rounds of CPR selection were assayed using β -galactosidase containing a single amber codon with and without supplemented 5HTP in the growth media. **(b)** Deconvoluted, intact, whole-protein mass spectra of DHFR (containing an amber codon at position 10) with WT *S. cerevisiae* synthetase (ScWRS), which incorporates tryptophan (expected mass: ~19,052 Da), or with the evolved 5OH-R3-13 variant, which demonstrates the ~16-Da mass shift expected from incorporation of 5HTP. **(c)** tRNA amber suppression efficiency was quantified by the ability to suppress three amber codons in GFP in conjunction with the WT tRNA synthetase. The parental AS3.4 and CPR-evolved 40A tRNA are compared by flow cytometry, demonstrating that the optimized tRNA increases GFP production. RFU, relative fluorescent units. **(d)** Amber suppression efficiency of the 5OH-R3-13 synthetase with the parental AS3.4 or CPR-evolved 40A tRNA in the presence and absence of supplemented 5HTP. Expression of GFP containing three amber codons was measured. Fluorescence/ OD_{600} ratio reported is the average of three independently grown cultures; error bars represent one s.d.



allowing greater flexibility to the beta-sheet adjacent to the 5-hydroxyl moiety, whereas C255A may prevent a possible steric clash.

To determine if the variant 5OH-R3-13 incorporated 5HTP site-specifically, the enzyme was co-expressed with a modified version of dihydrofolate reductase (DHFR) that contained an amber codon at position 10 (V10amber)⁷ in the presence of rich 2×YT media supplemented with 5HTP. Purified DHFR proteins were characterized by top-down ultraviolet photodissociation mass spectrometry (Supplementary Methods)²⁰. The presence of a single mass shift of ~16 Da, indicative of the single incorporation of 5HTP in place of tryptophan, was detected in the 5OH-R3-13 samples but not in the WT tRNA synthetase samples (Fig. 3b). Fragmentation analysis verified that 5HTP was in fact incorporated at position 10, corresponding to the location of the amber codon (Supplementary Fig. 14). Incorporation fidelity was estimated to be ~85%, despite growth in rich media in the presence of abundant tryptophan. When compared with the WT ScWRS enzyme, the evolved 5OH-R3-13 synthetase showed 1,500-fold improvement in incorporating 5HTP (ratio of peak density in Fig. 3b).

Optimization of orthogonal tRNAs (as opposed to tRNA synthetases) can also have a substantial impact on the efficiency of unnatural amino acid incorporation²¹. Although the *S. cerevisiae* tRNA is already highly efficient, it is unlikely that its interaction with other parts of the *E. coli* translation machinery, such as elongation factors or the ribosome, will be optimal. Three tRNA libraries were constructed randomizing either the anticodon stem, acceptor stem or loop sequences (Supplementary Fig. 15). Each library contained roughly 10⁶ different tRNA variants, and we subjected each library to ten rounds of CPR. To increase selection pressure, we progressively increased the number of amber codons in the open reading frame of the *Taq* DNA polymerase (up to six), as well as reduced the expression of the orthogonal tRNA synthetase (Supplementary Table 2). Moreover, to show that CPR could potentially be used to co-evolve entire genetic circuits (as opposed to a single part in a circuit), the WT tRNA synthetase was also allowed to mutate during the selection of the suppressor tRNA libraries. After ten rounds, several circuits containing co-evolved tRNA synthetase and tRNA pairs were assayed by flow cytometry on the basis of their ability to suppress GFP bearing three amber codons (Supplementary Fig. 16). Although some neutral and silent mutations were detected in the tRNA synthetases, as might be expected, most of the mutations accrued within the tRNA libraries. Seven of ten tRNA variants displayed more efficient amber suppression than the parental suppressor tRNA (which itself had previously been improved over the WT yeast suppressor tRNA by rational mutagenesis⁷). The two best tRNA variants (40A and 49A) were roughly three- and fourfold more active, respectively, than the parental tRNA (AS3.4) and in consequence likely 12-fold better than the WT tRNA (Fig. 3c). Both suppressor tRNA variants were assayed for cross-utilization by *E. coli* tRNA synthetases by determining whether they could suppress a single amber codon in a β-galactosidase or GFP, in the absence of functional ScWRS. The 49A tRNA variant displayed some background charging or loading of amino acids by endogenous synthetases, but the 40A variant (loop mutations U16G, G43U, U58G) appeared to be completely orthogonal (Supplementary Figs. 17 and 18).

We hypothesized that the efficiency of the CPR-evolved 5HTP incorporating 5OH-R3-13 tRNA synthetase would be improved by pairing it with the CPR-evolved 40A suppressor tRNA. When combined, these optimized parts drove 8.5-fold higher expression of GFP containing three amber codons in the presence of 5HTP relative to the parental AS3.4 tRNA (Fig. 3d). Similarly, when DHFR (containing

one amber codon) was expressed in cells expressing both the 5OH-R3-13 synthetase and 40A suppressor tRNA, mass spectrometry experiments confirmed no loss of fidelity of 5HTP incorporation despite the improved suppression efficiency (Supplementary Fig. 19). The efficiency with which 5HTP is incorporated into these proteins containing multiple amber codons is particularly impressive given that other evolved synthetase:tRNA pairs have been shown to be capable of suppressing only one amber codon with an unnatural amino acid²².

One advantage of CPR is its ability to select for a gene part or circuit function *in vivo* without being confounded by the need to simultaneously select for organismal fitness. This is because the short duration of the *in vivo* functional selection limits fitness effects that may disfavor highly active circuits. Another advantage is that the subsequent *in vitro* replication phase enables the exponential amplification of the most active circuits. These attributes likely lead to better functioning circuits being selected over a range of functionality (Supplementary Fig. 20), meaning that even genetic parts or circuits resulting in initially weak phenotypes (in other words, inducing only a small difference in *Taq* DNA polymerase expression) can be established in the population and that the most active variants can ultimately come to dominate. Finally, the discontinuous nature of the selection process offers a much-needed element of control that limits parasite accumulation^{23,24}, avoids extraneous mutations affecting the selection^{25,26} and enables a part to be evolved largely in the context in which it will be used.

The evolution of transcription and translation machinery suggests that CPR can evolve enzymes and RNAs that are highly active and specific. By simply rewiring the *in vivo* circuit-based architecture to favor the production of *Taq* DNA polymerase in other ways, CPR can potentially be adapted to the evolution of regulatory parts such as transcription factors, repressors and riboswitches, as well as larger genetic ensembles, such as operons and biosynthetic pathways.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

This work was supported by the National Security Science and Engineering Faculty (FA9550-10-1-0169) the Welch Foundation (F-1654 to A.D.E. and F-1155 to J.S.B.), the National Science Foundation (CHE1012622 to J.S.B.), and the Defense Advanced Research Projects Agency (HR-0011-12-C-0066). J.S.B. thanks Thermo Fisher Scientific with helping on the modifications to the Orbitrap Elite mass spectrometer to allow UVPD.

AUTHOR CONTRIBUTIONS

J.W.E. conceived of the selection scheme. A.J.M. carried out experiments with T7 RNA polymerase. J.W.E. carried out experiments on the ScWRS and tRNA with input from R.A.H. J.R.C. and J.S.B. performed mass spectrometry and analysis. J.W.E., A.J.M. and A.D.E. wrote the manuscript with contributions from R.A.H., J.R.C. and J.S.B.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Collins, C.H., Leadbetter, J.R. & Arnold, F.H. Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* **24**, 708–712 (2006).
- Sinha, J., Reyes, S.J. & Gallivan, J.P. Reprogramming bacteria to seek and destroy an herbicide. *Nat. Chem. Biol.* **6**, 464–470 (2010).

3. Esvelt, K.M., Carlson, J.C. & Liu, D.R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
4. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
5. Tawfik, D.S. & Griffiths, A.D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **16**, 652–656 (1998).
6. Levy, M., Griswold, K.E. & Ellington, A.D. Direct selection of trans-acting ligase ribozymes by *in vitro* compartmentalization. *RNA* **11**, 1555–1562 (2005).
7. Hughes, R.A. & Ellington, A.D. Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. *Nucleic Acids Res.* **38**, 6813–6830 (2010).
8. Hughes, R.A., Miklos, A.E. & Ellington, A.D. Gene synthesis: methods and applications. *Methods Enzymol.* **498**, 277–309 (2011).
9. Gibson, D.G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **498**, 349–361 (2011).
10. Temme, K., Zhao, D. & Voigt, C.A. Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc. Natl. Acad. Sci. USA* **109**, 7085–7090 (2012).
11. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* **5**, 842–848 (2009).
12. Wang, H.H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
13. Ghadessy, F.J., Ong, J.L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. USA* **98**, 4552–4557 (2001).
14. Ghadessy, F.J. & Holliger, P. Compartmentalized self-replication: a novel method for the directed evolution of polymerases and other enzymes. *Methods Mol. Biol.* **352**, 237–248 (2007).
15. Cheetham, G.M., Jeruzalmi, D. & Steitz, T.A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80–83 (1999).
16. Raskin, C.A., Diaz, G.A. & McAllister, W.T. T7 RNA polymerase mutants with altered promoter specificities. *Proc. Natl. Acad. Sci. USA* **90**, 3147–3151 (1993).
17. Chelliserrykattil, J., Cai, G. & Ellington, A.D. A combined *in vitro/in vivo* selection for polymerases with novel promoter specificities. *BMC Biotechnol.* **1**, 13 (2001).
18. Temme, K., Hill, R., Segall-Shapiro, T.H., Moser, F. & Voigt, C.A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* **40**, 8773–8781 (2012).
19. Zhou, M., Dong, X., Shen, N., Zhong, C. & Ding, J. Crystal structures of *Saccharomyces cerevisiae* tryptophanyl-tRNA synthetase: new insights into the mechanism of tryptophan activation and implications for anti-fungal drug design. *Nucleic Acids Res.* **38**, 3399–3413 (2010).
20. Shaw, J.B. *et al.* Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J. Am. Chem. Soc.* **135**, 12646–12651 (2013).
21. Young, T.S., Ahmad, I., Yin, J.A. & Schultz, P.G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **395**, 361–374 (2010).
22. Johnson, D.B.F. *et al.* RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **7**, 779–786 (2011).
23. Breaker, R.R. & Joyce, G.F. Emergence of a replicating species from an *in vitro* RNA evolution reaction. *Proc. Natl. Acad. Sci. USA* **91**, 6093–6097 (1994).
24. Bull, J.J. & Pease, C.M. Why is the polymerase chain reaction resistant to *in vitro* evolution? *J. Mol. Evol.* **41**, 1160–1164 (1995).
25. Goldsmith, M. & Tawfik, D.S. Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proc. Natl. Acad. Sci. USA* **106**, 6197–6202 (2009).
26. Dickinson, B.C., Leconte, A.M., Allen, B., Esvelt, K.M. & Liu, D.R. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proc. Natl. Acad. Sci. USA* **110**, 9007–9012 (2013).
27. Paige, J.S., Wu, K.Y. & Jaffrey, S.R. RNA mimics of green fluorescent protein. *Science* **333**, 642–646 (2011).

ONLINE METHODS

T7 RNAP library design and selection. Site-saturation mutagenesis was used to randomize the residues R746, L747, N748, R756, L757 and Q758 of the T7 RNAP promoter specificity loop. The degenerate oligonucleotide (RAH.1) was synthesized in-house on an Expedite 8900 synthesizer using reagents and phosphoramidites purchased from Glen Research (Sterling, VA) at a 40-nmol synthesis scale (**Supplementary Table 3**). Degeneracy was introduced into the oligonucleotides by the use of trimer phosphoramidites (Glen Research) containing a mixture of 20-trimer (codon) phosphoramidites encoding all 20 amino acids. The oligonucleotides were synthesized with the 5' DMT (4,4'-dimethoxytrityl) group retained. The oligonucleotides were cleaved from the solid synthesis supports using ammonium hydroxide, and deprotected at 55 °C overnight in the same solution. The cleaved, deprotected oligonucleotides were run over a Glen-Pak (Glen Research) purification column to remove aborted oligonucleotide sequences and eluted following the manufacturer's protocol. The oligonucleotides were then ethanol precipitated and reconstituted in water for further use. RAH.1 was PCR amplified using the primers AJM.1 and AJM.2 for selections on the WT promoter and the primers AJM.3 and AJM.4 for selections using the CGG-promoter. These primer sets each introduced different silent mutations (watermarks) up- and downstream of the randomized regions. These allowed for specific amplification as well as identification upon sequencing.

The nonrandomized portions of T7 RNAP were amplified from pQE-RSS (**Supplementary Table 4** and **Supplementary Fig. 21**). "T7 RSS" is a previously synthesized version of the T7 RNAP that was optimized for reduced secondary structure of the mRNA²⁸. All selections and initial characterizations (**Fig. 2a,c** and **Supplementary Figs. 1–6**) were performed with this codon set to avoid contamination during the selection process. All protein purification and comparison to other mutants (**Fig. 2b** and **Supplementary Figs. 7–10**) were performed with WT codon set to provide a fairer representation of mutants. In all cases, every polymerase used the same codon set for a given assay.

The N-terminal portion of the T7 RNAP coding sequence was amplified from pQE-RSS using the primers AJM.5 and AJM.6; the C-terminal portion of the T7 RNAP coding sequence was amplified from pQE-RSS using the primers AJM.7 and AJM.8. The three portions were then assembled by overlap PCR (N-terminal:library:C-terminal at a 1:2:1 molar ratio) using AJM.5 and AJM.8. The assembly PCR and an empty pQE vector were each digested with BamHI and HindIII (New England BioLabs). Insert and vector were mixed at a 2.3:1 molar ratio (~2.5 µg total), incubated with 2,000 U T4 DNA ligase (New England BioLabs) and at 14 °C for at least 15 h. All PCR steps (other than those in emulsion PCR) were performed using Accuprime Pfx DNA Polymerase (Life Technologies) per manufacturer's instructions. All PCR and digestion products were gel purified using QIAquick Gel Extraction Kit (Qiagen) and ligations were purified with SV Wizard PCR clean-up (Promega) before electroporation.

The WT *Taq* DNA polymerase gene was cloned into a modified pACYC-duet (Novagen) backbone with a single T7 promoter (**Supplementary Table 4** and **Supplementary Fig. 21**). BL21 gold cells (Agilent) were transformed with pACYC-Taq (or its derivative with altered promoter) and grown in bulk overnight. 250 µl of this culture was subcultured in 20 ml 2×YT medium and grown at 37 °C for 2 h (OD₆₀₀ ~ 0.5). The culture was then spun and washed with ice cold 10% glycerol four times, with the fourth resuspension in 100 µl 10% glycerol. This cell slurry (~200 µl total) was combined with 2–10 µl purified ligation and using electroporated 0.2 cm cuvettes at 2.5 kV in an *E. coli* pulser (Bio-Rad). This routinely resulted in 2 × 10⁷ CFUs (multiple replicates were pooled for early rounds in order to attain full coverage).

100 µl overnight transformation cultures were subcultured in 2 ml 2×YT medium, grown for 1 h (OD₆₀₀ ~ 0.6) and induced with 0.3 mM IPTG at 37 °C for 4 h. 200 µl of the induction culture was centrifuged (10 min: 5,000g) to pellet the cells. The supernatant was removed and cells were gently resuspended in 20 µl 10× PCR buffer (500 mM KCl, 100 mM Tris-HCl pH 8.3, 15 mM MgCl₂) 10 µl dNTP mix (4 mM each), 4 µl AJM.9 (20 µM), 4 µl AJM.10 (20 µM) and 162 µl water.

Emulsification was performed by slowly adding resuspended cells to 600 µl of spinning oil mix (438 µl Tegoseft DEC (Evonik), 42 µl AbilWE09 (Evonik) and 120 µl mineral oil (Sigma)). The oil mixture was constantly spun in a tube (Sarstedt 13 ml 95 mm × 16.8 mm) on ice using a stirbar (Spinplus 9.5 mm ×

9.5 mm Teflon, Bel-Art) on a magnetic plate (Corning) at the maximum setting (1150 r.p.m.). The cell mixture was slowly added over a 1-min interval and spun for an additional 4 min. The emulsified cells were thermal cycled (95 °C:3min, 20 cycles (95 °C:30 s, 55 °C:30 s, 72 °C:2 min/kb), 72 °C:5 min) such that cells containing the most active enzymes will also contain the most *Taq* DNA polymerase and will preferentially PCR amplify. The emulsion was broken in two steps. First, it was spun down by centrifugation (5 min: 10,000g) and the oil (upper) phase was removed. Second, 300 µl of H₂O and 500 µl chloroform was added and the mixture was vortexed vigorously. The mixture was transferred to a heavy-gel phase-lock tube (5 Prime) and upon centrifugation (2 min: 16,000g) the aqueous (upper) phase was collected along with any nucleic acids present. To purify PCR-amplified DNA from plasmid DNA we used a 5' biotinylated primer such that products amplified from *Taq* DNA polymerase can be purified away from plasmid DNA using streptavidin-coated beads (MyOne Streptavidin C1 Dynabeads, Invitrogen).

Purified DNA was used as a template for re-amplification using primers specific to the watermark introduced above (AJM.11 and AJM.12 for the PT7 selection and AJM.13 and AJM.14 for the P_{CGG} selection). This PCR product was used in an assembly PCR, followed by digestion and ligation as above.

In the later rounds of the P_{CGG} selection (after isolation of CGG-R7-8), a larger region of the polymerase coding sequence was recovered (and thus allowed to evolve). The biotinylated primers AJM.9 and AJM.10 were still used for the emulsion PCR, but the post-recovery PCR was performed using AJM.15 and AJM.16. The assembly PCR N- and C-terminal pieces were made as before with primers AJM.5 and AJM.17 (N-terminal) and AJM.18 and AJM.8 (C-terminal).

CGG-R7-8 was subject to error-prone PCR and used as the input for CGG-R8. CGG-R12-KIV, CGG-R12-KIR, CGG-R12-KIRV, CGG-R12-KIGR and CGG-R12-KIGRV were combined in equal amounts, subject to error-prone PCR and used as the input for CGG-R13. The recovered product of CGG-R10, CGG-14 and CGG-R15 were subject to error-prone PCR as described²⁹. Briefly, the reaction mixture was composed of 50 mM KCl, 10 mM Tris-HCl pH 8.3, 2.5 mM MgCl₂, 5 µg/ml BSA, 0.35 mM dATP, 0.4 mM dCTP, 0.2 mM dGTP, 1.35 mM dTTP, 0.5 mM MnCl₂, 0.5 µM each primer (AJM.15 and AJM.16), 2 ng/µl template and 0.8 U/µl *Taq* DNAP (New England BioLabs) and was thermal cycled (95 °C:4 min, 25 cycles (95 °C:30 s, 55 °C:30 s, 72 °C:2 min), 72 °C:5 min). This achieved the expected 1 mutation per 500 bp.

Individual variants from PT7-R4, CGG-R7, CGG-R12 and CGG-R16 were sequenced and analyzed using Geneious software (Biomatters Ltd.).

***In vivo* RNAP activity assays.** A FACS-optimized variant of GFP, GFP mut2 (ref. 30), was cloned in place of the *Taq* DNAP open reading frame in pACYC-*Taq*. For measures of *in vivo* activity of T7 RNAP, variants (plasmid) or pools (ligation) were electroporated into BL21 gold cells containing pACYC-GFPmut2 (or its derivative with altered promoter). Transformations were grown at 37 °C overnight. 100 µl of the culture was grown in 2 ml 2×YT medium at 37 °C for 1 h (OD₆₀₀ ~ 0.6) and induced at 0.05 mM IPTG for 4 h. This concentration of IPTG was chosen to limit metabolic overload on the host and prevent saturation of signal. After induction, cells were measured for OD₆₀₀ on a Synergy-HT plate reader (Bio-Tek) and GFP fluorescence (excitation/emission 481/507) on a Safire monochromator (Tecan).

Images of T7 RNA polymerase-driven GFP expression shown in **Figure 2a** and **Supplementary Figure 1** were generated by spinning down 10 ml of induced culture, decanting the supernatant and resuspending cells in 500 µl PBS. The resuspended cells were excited with a UV transilluminator and imaged by Canon DSLR 500D digital camera.

Images of T7 RNA polymerase-driven GFP expression shown in **Supplementary Figure 8** were generated by spinning down 2 ml of induced culture, decanting the supernatant and resuspending cells in 200 µl PBS. The resuspended cells were excited with a UV transilluminator and imaged by Canon DSLR 500D digital camera.

***In vivo* RNAP cross-reactivity assays.** A promiscuous mutant of the phenylalanine aminoacyl-tRNA synthetase, PheS A294G^{31,32}, was cloned in place of the *Taq* DNAP open reading frame in pACYC-*Taq*. T7 RNAP mutant plasmids were electroporated into BL21-gold cells containing pACYC-PheS with PheS A294G driven by the WT T7 promoter. Transformations were grown at 37 °C overnight. 100 µl of the culture was grown in 2 ml 2×YT medium at 37 °C

for 1 h ($OD_{600} \sim 0.6$) and induced at 0.05 mM IPTG for 4 h. Cells were diluted with media (containing the same antibiotics and IPTG as the growth media) to OD_{600} of 0.1, 0.01 and 0.001. 5 μ l of each dilution was plated on 0 mM, 5 mM, 10 mM, 15 mM or 20 mM 4-chloro-DL-phenylalanine (Cl-Phe; Sigma). The plating media also contained 0.4% glycerol, 0.5% yeast extract, 1% NaCl, 1.5% agar, 50 μ g/ml kanamycin, 24 μ g/ml chloramphenicol and 0.05 mM IPTG. Plates were grown at 37 °C for 20 h and imaged with ambient white light on a FluorChem Q (Protein Simple). Mutant cross-reactivity may be judged by the dose-dependent cytotoxicity of Cl-Phe, which is only lethal (at the concentrations used) when PheS A294G is expressed.

In vitro transcription assays. For *in vitro* transcription assays, T7 RNAP variants were purified by standard Ni-NTA 6 \times His (N-terminal) methods. The plasmid pQE-T7RSS (or a derivative thereof for T7 RNAP mutant) was transformed in BL21-gold (Agilent). Cells were grown in 2 \times YT media at 37 °C until reaching $OD_{600} \sim 0.7$ –0.8, at which point 1 mM IPTG was added. Cells were grown 4 h at 37 °C. Following induction, cells were harvested by centrifugation and resuspended in binding buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole). Resuspended cells were lysed via sonication on ice using 40% probe amplitude for 2 min (1 s ON, 1 s OFF). Cell debris was pelleted by centrifugation (30 min:20,000g). His-tagged T7 RNAP was purified by immobilized metal affinity chromatography (IMAC). The lysate was run over 1 ml (bead volume) Ni-NTA gravity column pre-equilibrated with binding buffer. The column was washed with 10 \times column volumes of wash buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 20 mM imidazole). T7 RNAP was eluted off the column by the addition of 4 \times column volumes of elution buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 250 mM imidazole). Dialysis was performed in final storage buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DDT, 1 mM EDTA). Dialysates were adjusted to 1 mg/ml and added to an equal volume of glycerol (final concentration 0.5 mg/ml).

Transcription reactions contained 40 mM Tris-HCl pH 7.0, 30 mM MgCl₂, 6 mM spermidine, 6 mM each NTP, 10 mM DTT, 0.5 μ M T7 RNAP, 0.5 μ M DNA template and 0.17 mg/ml DFHBI²⁷ in DMSO. Reactions were incubated for up to 2 h at 37 °C with spinach fluorescence (excitation/emission 469/501) reading taken every minute in a Safire monochromator (Tecan). Spinach templates were made by thermal cycling 2 μ M AJM.19 with 2 μ M AJM.20 (for PT7-spinach) or 2 μ M AJM.21 (for P_{CGG}-spinach) with Accuprime Pfx in its standard buffer (94 °C:2 min, 12 cycles (94 °C:15 s, 50 °C:30 s, 68 °C:30 s), 68 °C:1 min). Templates were purified by QIAquick Gel Extraction Kit (Qiagen).

For all T7 RNA polymerase studies: investigators were not blinded to the identity of the samples during analysis; samples were not randomized; no samples were excluded from analysis.

Aminoacyl tRNA synthetase CPR for site-specific 5-hydroxy-L-tryptophan incorporation. The binding pocket of the *S. cerevisiae* tryptophanyl tRNA-synthetase (ScWRS) was mutagenized by site-saturation mutagenesis at residues T107, P254 and C255 using NNS randomized oligonucleotides (Supplementary Table 3) (JE.29 and JE.30). These residues were chosen due to their proximity to the 5' position of tryptophan¹⁹ in the binding pocket of the synthetase. Libraries were prepared by overlap PCR using PFU Ultra II fusion HS (Agilent) and cloned into pRST.2TAA, a modified version of pRST.11B.AS3.4 (ref. 7) containing two ochre (TAA) stop codons, using HindIII and XhoI restriction endonucleases (Supplementary Table 4 and Supplementary Fig. 21). *E. coli* BL21(DE3) (Invitrogen) harboring the plasmid pACYC.Taq.1Amb (W167Amber), which contains an amber codon in the open reading frame of *Taq* DNA polymerase, was transformed by the resulting library. The efficiency of transformation was >10⁶ for each round of selection indicating several fold coverage of the library. Transformed cells were grown in 2 \times YT media (Sigma) overnight at 37 °C in carbenicillin (Cellgro) and chloramphenicol (Sigma). The following morning 10 μ l of cells were seeded into 1 ml of fresh 2 \times YT media with appropriate antibiotics and 1 mM 5-hydroxy-L-tryptophan (5HTP) (Sigma) and grown at 37 °C for 2 h. The expression of the library of mutant synthetases and the *Taq* DNA polymerase was initiated by the addition of 1 mM IPTG. Cells were induced at 30 °C for 7 h. Cells were harvested (200 μ l) by centrifugation (8 min: 3,000g) and removal of the supernatant. Cells were resuspended in 168 μ l CPR buffer (45 mM KCl,

9 mM Tris-HCl (pH 8.3), 1.4 mM MgCl₂, 0.5 nM each CPR primer (JE.33 and JE.35) and 250 μ M each dNTP).

Emulsification was performed by slowly adding resuspended cells to 600 μ l of spinning oil mix (438 μ l Tegosoft DEC (Evonik), 42 μ l AbilWE09 (Evonik), and 120 μ l mineral oil (Sigma)). The oil mixture was constantly spun in a tube (Sarstedt 13 ml 95 mm \times 16.8 mm) on ice using a stirbar (Spinplus 9.5 mm \times 9.5 mm Teflon, Bel-Art) on a magnetic plate (Corning) at the maximum setting (1,150 r.p.m.). The cell mixture was slowly added over a 1-min interval and spun for an additional 4 min. The emulsified cells were thermal cycled (95 °C:3 min, 20 cycles (95 °C:30 s, 55 °C:30 s, 72 °C:2 min/kb), 72 °C:5 min) to selectively amplify functional variants. The emulsion was broken in two steps. First, it was spun down by centrifugation (5 min: 10,000g) and the oil (upper) phase was removed. Second, 300 μ l of H₂O and 500 μ l chloroform was added and the mixture was vortexed vigorously. The mixture was transferred to a heavy-gel phase-lock tube (5 Prime) and upon centrifugation (2 min: 16,000g) the aqueous (upper) phase was collected. To purify CPR amplified DNA, we used a 5' biotinylated primers, thus, products amplified by *Taq* DNA polymerase can be purified away from plasmid DNA using streptavidin-coated beads (MyOne Streptavidin C1 Dynabeads, Invitrogen). Purified DNA was used as a template for re-amplification using nested primers (JE.34 and JE.37). Reamplification products were cloned into the pRST.2TAA; completing one round of CPR selection. Three rounds of selection were carried out using the same reaction conditions. The pool activity was assayed using a β -galactosidase colony spot assay (Fig. 3a). Blue colonies were counted for each of the rounds and divided by the total, indicating that each round of selection enriched for active synthetases. Synthetase variants were assayed for 5HTP specificity by patch-plating colonies onto plates with or without the unnatural amino acid 5HTP (1 mM). Colonies that turned blue only in the presence of 5HTP were putative candidates for selective incorporation. As expected, some active variants were pulled from the initial round 0 pool due to the small library size. Variant 5OH-R3-13 was chosen for further analysis as it displayed the most pronounced β -galactosidase activity.

tRNA CPR for improved amber suppression. CPR was used to select for enhanced amber suppression by randomizing key regions of the tRNA. Libraries were generated using the orthogonal tRNAAS3.4 as a starting point and residues in the anticodon stem (AS), acceptor stem (AX) and the loop sequences were randomized. CPR selections were carried out essentially as described for the 5HTP selection, except the stringency of selection was modulated to increase selective advantage for the best suppressor tRNAs (Supplementary Table 2). To increase selective pressure on the tRNA, we modulated the time of induction, temperature of induction, IPTG concentration and, most importantly, the number of amber codons in the *Taq* DNA polymerase open reading frame. The recovery strategy also changed from streptavidin-based capture to using CPR primers with unique sequences on the 5' end. During reamplification, primers will anneal to the unique sequence preventing amplification of the contaminating plasmid DNA. Additional diversity was introduced by error-prone PCR after rounds 4, 5 and 9. tRNA libraries were pooled after round 3. Amber codons were introduced into the *Taq* DNA polymerase gene at positions W167, W169, W179, W211, W243 and W318. In pACYC-*Taq*, constructs with more than one amber codon were introduced additively from N to C terminus. After ten rounds of selection, tRNAs were screened by flow cytometry for the ability to suppress three amber codons in GFPmut2 (Fig. 3c and Supplementary Fig. 16). The most active tRNA variants, 40A and 49A, were further examined to determine if they worked promiscuously with other synthetases; this was done using the β -galactosidase assay. tRNA variant 40A was completely inactive in the absence of a functional synthetase, whereas variant 49A displayed partial activity, suggesting it might be interacting with other synthetase machinery expressed by the host (data not shown).

β -galactosidase screening. tRNA synthetase variants cloned into pRST.11B backbones were transformed into CA274 *E. coli* cells, which contain an amber codon at position 125 of the *LacZ* gene. Individual colonies were patch plated onto 2 \times YT media plates containing 100 μ g/ml carbenicillin, 40 μ g/ml X-Gal (Sigma) and 0.1 mM IPTG, with or without 1 mM 5HTP. Cells were grown at 37 °C for ~6 h, which resulted in visibly blue colonies for active synthetases.

GFP-FACS screening. Variants from either synthetase or tRNA libraries were cloned into pRST.11B backbones and analyzed by means of GFP. A FACS-optimized variant of GFP, GFP mut2³⁰, was cloned in place of the *Taq* DNAP open reading frame in pACYC-*Taq* and amber codons were introduced at positions Y39, Y151, Y182 (ref. 33), resulting in plasmid pACYC.GFPmut2.3Amb. *E. coli* BL21(DE3) were co-transformed with individual synthetase or tRNA library variants and pACYC.GFPmut2.3Amb and grown overnight at 37 °C in 2×YT media. The following day, 10 μl of cells were diluted into 1 ml fresh 2×YT media (containing appropriate antibiotics and 5HTP when necessary) and grown for 2 h at 37 °C. Expression of both the synthetase and GFP were induced with the addition of 1 mM IPTG and grown for 4 h at 37 °C. Cells were harvested by centrifugation (4 °C) and resuspended in phosphate buffered saline. Fluorescence analysis was performed on either a FACSCalibur (BD Biosciences) flow cytometer or on a Safire monochromator (Tecan) using GFP fluorescence (excitation/emission 481/507).

DHFR purification and mass spectrometry. *E. coli* DHFR was used to assess the incorporation of the 5HTP amino acid. Plasmid pACYC.DHFR_V10Amb⁷ contains an amber codon at V10 of DHFR. Plasmids pRST.11B, pRST.5HTP and pRST.5HTP.40A were co-expressed with pACYC.DHFR_v10Amb in BL21(DE3). The strains were grown in 2×YT media at 37 °C until reaching OD₆₀₀ ~0.7–0.8, at which point 1 mM 5HTP and 1 mM IPTG were added. Cells were grown overnight at 30 °C. Following induction cells were harvested by centrifugation and resuspended in binding buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole). Resuspended cells were lysed by means of sonication on ice using 40% probe amplitude for 2 min (1 s ON, 1 s OFF). Cell debris was pelleted by centrifugation (30 min: 20,000g). The His-tagged DHFR was purified by immobilized metal affinity chromatography (IMAC). The lysate was run over 1 ml (bead volume) Ni-NTA gravity column pre-equilibrated with binding buffer. The column was washed with 10× column volumes of binding buffer, 3× column volumes of wash 1 buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 20 mM imidazole) and an additional 3× column volumes of wash 2 buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 30 mM imidazole). DHFR was eluted off the column by the addition of 4× column volumes of elution buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 250 mM imidazole). DHFR samples were centrifuged (10 min:16,000g) to remove insoluble protein and then loaded onto a fast protein liquid chromatography column and fractionated by size-exclusion chromatography into 10 mM Tris-pH 8.0. The purest DHFR fractions were used for mass spectrometry analysis.

Top-down ultraviolet photodissociation MS. DHFR (V10Amber) was expressed and purified as described above. Following purification, the proteins were buffer exchanged into LC-MS grade water using 3 kDa molecular weight cutoff filters. The proteins were diluted to 10 μM in a solution of 50:49:1 MeOH/water/formic acid. Proteins were infused at a flow rate of 5 μl/min and ionized by electrospray ionization on a Thermo Scientific Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) modified for ultraviolet photodissociation (UVPD) in the higher energy collisionally activated dissociation cell as described previously^{20,34}. A 193-nm excimer laser was used for UVPD. Intact molecular weight measurements at maximum resolution were undertaken to confirm the presence or absence of the expected single 5-hydroxy-L-tryptophan incorporation. The site of modification was confirmed using UVPD by means of a single 5-ns, 193-nm laser pulse. The UVPD product ion spectra were also acquired at maximum resolution. Precursor and product ion spectra were interpreted manually and using the Xtract (Thermo Fisher) deconvolution algorithm in conjunction with a beta version of ProSightPC 3.0. The Sequence Gazer tool was used to assign possible locations of the ~16 Da observed mass shift. Incorporation efficiency was calculated by subtracting the area of artifactual oxidation (ScWRS - 5HTP control) from the combined areas of the 5HTp and singly oxidized 5HTP peaks, and dividing by the summed areas of all peaks (Trp, 5HTP and 5HTP+oxidation). Peak area integration was performed using the five most abundant peaks for each isotope cluster. Fold

enhancement was calculated by comparing the incorporation rates of 5HTP for ScWRS and R3-13. Incorporation rates for each were arrived at by dividing the area of the 5HTP containing peak by the combined areas of the 5HTP and naturally occurring reduced Trp containing peaks, and then applying a correction to account for artifactual oxidation during sample handling. The correction factor was attained by performing the same calculations as above for the singly oxidized peak using ScWRS in the absence of 5HTP in the media.

For all tRNA-synthetase studies: investigators were not blinded to the identity of the samples during analysis; samples were not randomized; no samples were excluded from analysis.

Comparison of tRNAs generated by existing methods and CPR. Existing approaches for generating orthogonal tRNAs have relied upon an *in vivo* life-death selection system. A recent publication³⁵ used this traditional approach to generate optimized versions of the *S. cerevisiae* orthogonal suppressor tRNA. To demonstrate the effectiveness of CPR, tRNAs were compared to each other for cross-reactivity to the native *E. coli* translation machinery and for amber suppression activity. The optimized tRNAs were cloned into the starting vector (sequences of tRNAs: AS3.4, JWE4.8; 40A, JWE.49; H13, JWE.50; H14, JWE.51) in place of the AS3.4 tRNA as well as a vector containing a nonfunctional version of the synthetase (to test for tRNA orthogonality). The tRNAs (compared to AS3.4 and 40A) are in the AS3.51 background with acceptor stem mutations H13 and H14. To test for tRNA cross-reactivity to *E. coli* tRNA synthetases, tRNAs (without active *S. cerevisiae* aminoacyl tRNA synthetase) were expressed with GFPmut2 (Y39TAG; 1Amb), as described above and measured on a monochromator. Resulting fluorescence reflects tRNA charging with endogenous *E. coli* machinery and therefore lack of orthogonality. To further test orthogonality, tRNAs were transformed into *E. coli* strain CA274 (LacZ 1Amb). The relative efficiency of amber suppression was tested as described above (**Supplementary Fig. 18**).

Dynamic range of CPR. CPR mixes were processed as described (above) but purified *Taq* DNA polymerase (NEB) was added exogenously. A mixture consisting of an abundance of *Taq* DNA polymerase (0.8 μg; 5.1×10^{12} molecules) and 20 ng (3.8×10^9 molecules) of a DNA template “A” was emulsified and added in equal volume to a second emulsion consisting of a variable concentration of *Taq* DNA polymerase and 20 ng template “B” (which is identical to template A but contains an internal HindIII restriction endonuclease site). Emulsion PCRs were performed with various ratios of polymerase concentrations (A (0.8 μg):B (N μg)). Emulsions were broken and recovery PCR performed, followed by a HindIII digest. The DNA was run on a gel resulting in two distinct bands, one corresponding to template A and the other template B. Intensities of template A and B were measured (ImageJ), normalized to the equal ratio condition (1:1) and plotted (**Supplementary Fig. 20**).

28. Davidson, E.A., Meyer, A.J., Ellefson, J.W., Levy, M. & Ellington, A.D. An *in vitro* Autogene. *ACS Synth. Biol.* **1**, 190–196 (2012).
29. Fromant, M., Blanquet, S. & Plateau, P. Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Anal. Biochem.* **224**, 347–353 (1995).
30. Cormack, B.P., Valdivia, R.H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
31. Kast, P. & Hennecke, H. Amino acid substrate specificity of *Escherichia coli* phenylalanyl-tRNA synthetase altered by distinct mutations. *J. Mol. Biol.* **222**, 99–124 (1991).
32. Thyer, R., Filipovska, A. & Rackham, O. Engineered rRNA enhances the efficiency of selenocysteine incorporation during translation. *J. Am. Chem. Soc.* **135**, 2–5 (2013).
33. Wang, L., Brock, A., Herberich, B. & Schultz, P.G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500 (2001).
34. Han, S.-W. *et al.* Tyrosine sulfation in a Gram-negative bacterium. *Nat. Commun.* **3**, 1153 (2012).
35. Chatterjee, A., Xiao, H., Yang, P.-Y., Soundararajan, G. & Schultz, P.G.A. Tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in *E. coli*. *Angew Chem. Int. Ed. Engl.* **52**, 5106–5109 (2013).