UNIVERSITY OF CALIFORNIA,
IRVINE

Genetic tools for probing long evolutionary pathways

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Ziwei Zhong

Dissertation Committee:
Associate Professor Chang C. Liu, Chair
Associate Professor Elliot Hui
Assistant Professor Tim Downing

2022

DEDICATION

To

Evelyn Hoover

for

helping me with life, the universe, and everything

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Chang Liu, for all the advice, encouragement, mentorship, and support that he has given me over the years, not just in science, but also other aspects of life. His unwavering support, through successes and failures, have enabled me to develop into the scientist that I am today with skills that will undoubtedly carry me through the rest of my career. I will forever be grateful for the patience and understanding during which this degree was obtained. I also want to thank my committee members, Dr. Elliot Hui and Dr. Timothy Downing, for their genuine interest that resulted in insightful feedback and discussions.

I also want to thank the Medical Scientist Training Program at University of California, Irvine for their support, funding, and mentorship throughout my entire training career here. With Dr. Alan Goldin at the helm for many years now, the success of the program can be heavily attributed to his efforts and dedication. Additionally, I would like to thank Dr. Edwin Monuki for his generous support and mentorship throughout my training as my MSTP faculty advisor. Finally, all the efforts of the MSTP have been enabled by the wonderful and thorough administrator, Joanne Ham, who has kindly watched over me through my time in the program.

Additionally, all the work contained herein did not happen in a vacuum, but with the support of an entire lab. I cannot thank Drs. Arjun Ravikumar and Alex Javanpour enough for their work in establishing OrthoRep and for setting the tone and work ethic that has enabled my success. Further, the camaraderie and friendships of the many other members of the Liu and Hui labs have made this journey one worth remembering. I will forever cherish the memories that we have made. I am also especially grateful for the post-doctorate scholars who have been a source of guidance, inspiration, and mentorship, especially that of Dr. Alon Wellner, who has guided me through the field of protein evolution.

I am especially grateful to our collaborators at Boston University, Drs. Ahmad Khalil and Brandon Wong. Their support through my use of eVOLVER was instrumental for the development of Automated Continuous Evolution as well as further projects.

It goes without saying that life goes on, even outside of research. In that, I thank my wife, Dr. Evelyn Hoover, who has been a source of support both in research and in life. She has been a steadying influence in all matters big and small, and my life would not be the same without her. Thank you Evelyn.

VITA

Ziwei Zhong

EDUCATION

2014-2022	PhD Biomedical Engineering, *University of California, Irvine*
2012-2022	Medical Scientist Training Program, *University of California, Irvine*
2008-2012	BS Biomedical Engineering, *Purdue University*

RESEARCH EXPERIENCE

May 2014 – January 2022	Graduate Research Assistant; Biomedical Engineering; University of California, Irvine
August 2011 – August 2012	Undergraduate Research Assistant; Speech, Language, and Hearing Sciences; Purdue University
May 2011 – August 2011	SUNFEST Scholar; McKay Orthopaedic Research Laboratory; University of Pennsylvania
May 2009 – May 2011	Undergraduate Research Assistant; School of Biomedical Engineering, Purdue University

PUBLICATIONS

Peer-reviewed Journal Articles

Zhong Z.*, Wong B.G.*, Ravikumar A., Arzumanyan G.A., Khalil A.S., Liu C.C.. Automated Continuous Evolution of Proteins in Vivo. *ACS Synthetic Biology* **2020**; 9:1270-1276.
* Denotes equal contribution

Zhong Z, Liu CC. Probing Pathways of Adaptation with Continuous Evolution. *Current Opinion in Systems Biology* **2019**;14: 18-24.

Zhong Z, Ravikumar A, Liu CC. Tunable Expression Systems for Orthogonal DNA Replication. *ACS Synthetic Biology* **2019**;7: 2930-2934.

Zhong Z, Henry KS, Heinz, MG. Sensorineural hearing loss amplifies neural coding of envelope information in the central auditory system of chinchillas. *Hearing Research* **2014**;309: 55-62.

Zhong Z*, Muckley M*, Agcaoglu S, Grisham ME, Zhao H, Orth M, Lilburn MS, Akkus O, Karcher DM. The morphological, material-level, and ash properties of turkey femurs from 3 different genetic strains during production. *Poultry Science* **2012**;91: 2736-2746.
* Denotes equal contribution

Zhong Z, Akkus O. Effect of Age and Shear Rate on the Rheological Properties of Human Yellow Bone Marrow. *Biorheology* **2011**;48: 89-97.

Conference Proceedings

Zhong Z, Ravikumar A, Liu CC. OrthoRep Reveals Drift Potentiates Novel Adaptation Pathways in Protein Evolution. In: 2018 Synthetic Biology: Engineering, Evolution & Design, Scottsdale, AZ: SEED; **2018**.

Zhong Z, Ravikumar A, Liu CC. Continuous Evolution of a TIM Barrel Protein to Complement Amino Acid Auxotrophies. In: EBRC 2018 Spring Symposium, Seattle, WA: EBRC; **2018**.

Zhong Z, Ravikumar A, Liu CC. Rapid and Continuous Evolution of a Surface Displayed Protein in Yeast. In: Synberc 2014 Spring Symposium, Berkeley, CA: Synberc; **2014**.

Zhong Z, Ravikumar A, Liu CC. Rapid and Continuous Evolution of a Surface Displayed Protein in Yeast. In: Synberc 2013 Fall Symposium, Boston, MA: Synberc; **2013**.

Vesuna S, Jain N, **Zhong Z**, Maguire G. Emotional Content of Real-Time Twitter Status Updates Associated with Stuttering. In: *26th Annual Psychiatric Conference*, Las Vegas, NV: APNA; **2013**.

Heo SJ, Driscoll TP, **Zhong Z**, Mauck RL. Altered Mechanosensitivity of Fibrochondrogenic Mesenchymal Stem Cells with Modulation of Nuclear Mechanics. In: *58th Annual Meeting of the Orthopaedic Research Society*, San Francisco, CA: ORS; **2012**.

Zhong Z, Agcaoglu S, Muckley M, Zhao H, Karcher D, Orth M, Lilburn M, Akkus O. Changes In Turkey Femora Mechanical Properties Resulting From Selective Breeding For Body Weight. In: *ASME Summer Bioengineering Conference*. Farmington, PA: ASME; **2011**.

Muckley M, Agcaoglu S, **Zhong Z**, Zhao H, Grisham M, Karcher D, Orth M, Lilburn M, Akkus O. The Effect Of Selective Breeding For Body Weight On Geometric Properties In Turkey Femurs. In: *American Society of Mechanical Engineers Summer Bioengineering Conference*. Farmington, PA: ASME; **2011**.

Uquillas JA, Kishore V, **Zhong Z**, Breur GJ, Snyder P, Akkus O. Electrochemically Aligned Collagen Scaffolds Induce Tendon Hypertrophy In Vivo In: *56th Annual Meeting of the ORS*. New Orleans, LA; **2010**.

Mentorship and Teaching

Undergraduate research mentor 2016-2018
UCI MCBS Foundations Short Course Teaching Assistant 2018-2019
BME50A Teaching Assistant Winter 2018, 2022

ABSTRACT OF THE DISSERTATION

Genetic tools for probing long evolutionary pathways

by

Ziwei Zhong

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2022

Professor Chang C. Liu, Chair

Directed evolution is a powerful tool that has been used for novel drug discovery, commodity chemical synthesis, biodegradation, and many other medical and industrial advancements. However, one challenge with traditional directed evolution experiments is the divide between *ex vivo* diversification and *in vivo* selection, as the two are kept separate both to take advantage of their respective settings and to avoid unintentional off-target effects. This results in labor and time-intensive evolution experiments, limiting the number of replicates, rounds of evolution, or both. To address this, previous work in our lab sought to develop OrthoRep, a plasmid system in yeast enabling continuous *in vivo* mutagenesis of genes at high rates, allowing for the coupling of diversification and selection. Expanding on this, we developed a panel of constructs allowing genes encoded on OrthoRep to be expressed at levels spanning a wide range comparable to genes encoded on nuclear promoters. This has expanded OrthoRep's capabilities for more ambitious targets, including nanobodies and poorly functioning enzymes. Next, we paired OrthoRep with a continuous culture device to enable Automated Continuous Evolution, a hands-free evolution device that automatically adjusts culture conditions to maintain a programmable

level of selection. We show that this pairing enables faster adaptation compared to manual passaging and prevents overly stringent conditions that can lead to extinction of evolving cultures. Finally, we used OrthoRep to evolve HisA from *Thermotoga maritima* (*Tm*HisA) to catalyze the related Trp1 activity in yeast and demonstrated the ability to survey a broad fitness landscape and probe multiple selection conditions. We further reveal that after reaching a plateau for Trp1 activity, we can escape this plateau through an alternative selection for the original activity of *Tm*HisA while continued strong selection for Trp1 activity is ineffective. These advancements have enabled OrthoRep to be a capable alternative to traditional mutagenesis, as well as offered insights into the design of selection schemes that may facilitate reaching higher catalytic peaks.

Chapter 1. Probing pathways of adaptation with continuous evolution

Ziwei Zhong[1], Arjun Ravikumar[1], Chang C. Liu[1,2,3,*]

[1] Department of Biomedical Engineering, University of California, Irvine, California 92697, United States

[2] Department of Chemistry, University of California, Irvine, California 92697, United States

[3] Department of Molecular Biology & Biochemistry, University of California, Irvine, California 92697, United States

Introduction

Directed evolution is an effective approach to engineering enzymes and proteins for industrial, medical, and biotech applications [1,2] and was recently recognized with a Nobel Prize in Chemistry for its extraordinary practical impact. What is perhaps less well-known is the role directed evolution has played in elucidating and testing evolutionary mechanisms and theories of gene adaptation [3–6]. Understanding how a gene evolves is traditionally done retrospectively, by examining natural sequences and structures and working backwards to reconstruct descent and key evolutionary intermediates [7–11]. However, inferred histories are incomplete, can never be fully validated, and usually represent an N=1 experiment, as the conditions of natural evolution do not systematically repeat. By accelerating the process of evolution in the laboratory, directed evolution offers a way to study the evolution of genes in the forward direction, enabling researchers to observe adaptation in controlled environments, often in many replicates, and armed with the ability to sample and characterize the entire "fossil record" of each experiment. Such studies have yielded critical insights into the mechanisms by which genes, particularly proteins, evolve [12–15], the importance of stability in protein evolvability [16–21], the catalytic promiscuity of enzymes in the evolution of new activities [22–25], the complex fitness landscapes of proteins [26–35], and the role of neutral drift and fluctuating environments in crossing fitness valleys [36–39], to name only a few – and has solidified directed evolution as a powerful tool for understanding adaptation [5,6,40–42].

Despite these significant successes, there are limitations to using classical directed evolution techniques to study evolutionary mechanisms. Commonly, directed evolution mimics natural evolution by subjecting one or more genes of interest (GOIs) to multiple rounds of *ex*

vivo diversification (*e.g.* error-prone PCR), transformation into cells, and selection [2]. Each round of this process represents a step in an adaptive trajectory but requires significant manual intervention that restricts the extent and scale of experiments. This keeps three tantalizing categories of experiments largely outside our reach. First are experiments requiring the traversal of long mutational pathways such as ambitious adaptations or studies aimed at probing gene evolution under varying conditions over extended periods of time. (Indeed, most directed evolution experiments reach outcomes less than 5-10 non-synonymous mutations away from the parent sequence [16,38,43], with some exceptions that study the effects of extensive mutagenesis in one or few rounds [44–47].) Second are experiments requiring high statistical power through replication, such as studying drug resistance pathways, comparing the effects of different conditions on adaptation, detecting rare outcomes and rare adaptive trajectories, or mapping rugged fitness landscapes. (Currently, most directed evolution experiments are limited to only a few replicates [39,43,48]). Third are experiments that wish to capture or test complex population dynamics, since the technical idiosyncrasies of transformation and *ex vivo* diversification can cause population bottlenecks and perturb dynamics in artificial ways that influence evolutionary trajectories.

To address these limitations, synthetic biologists are working to establish a new paradigm in directed evolution through the construction of so-called continuous evolution systems. Continuous evolution achieves diversification of GOIs *in vivo* such that manual rounds of *ex vivo* diversification, transformation, and *in vivo* GOI expression and selection are not needed [49–52]. Instead, rapid diversification of GOIs occurs concurrently with their expression and functional selection, converting labor-intensive stepwise directed evolution processes into ones requiring

only the serial passaging of cells under selection conditions. This allows for evolution experiments that require long mutational pathways, large-scale replication, and the ability to capture complex population dynamics, categories particularly useful for testing evolutionary mechanisms and theory. In this opinion, we will briefly discuss the current state of continuous evolution systems and present their early successes and potential in reinvigorating the use of directed evolution to study basic questions in gene and protein adaptation.

State of continuous evolution systems

We will center our discussion in this section around three key properties that continuous evolution systems should have: targeting, durability, and scalability. There are others, discussed in depth elsewhere [52,53], but targeting, durability, and scalability are uniquely important if the goal is to study mechanisms of how a GOI evolves, as these three properties enable the rapid exploration of long mutational trajectories with statistical power. First, targeting. Certainly, if one wishes to explore how a GOI evolves, one will not want other loci contributing to the evolved function. But beside this practical reason for targeting is a deeper one. Very high rates of diversification are needed to see adaptation at the gene level on laboratory timescales, but such high rates harm or destroy host genomes, since there is a general inverse relationship between the rate at which an information polymer can be mutated under selection for function and its size [54–57]. For example, a GOI of size 1 kb can likely withstand a continuous mutation rate on order ~10^{-3} substitutions per base (s.p.b.) while a host genome of size 10^{7} bp (*e.g.* for *Saccharomyces cerevisiae*) will likely accumulate a lethal mutation every generation at mutation rates around ~f10^{-6} s.p.b. and experience clear fitness defects at mutation rates above ~10^{-8} s.p.b. [53,55]. Indeed, mutation rates of microbes and mammalian cells are evolutionarily optimized to

be in the 10^{-9}-10^{-10} s.p.b. range to prevent deterioration of fitness through high mutational loads over time [57–59]. Therefore, to evolve a GOI rapidly *in vivo*, mutations must be targeted to the GOI with extreme specificity. Most continuous evolution systems achieve incomplete targeting of GOIs relative to host genomes and other DNA [60–62], but two systems have either managed to achieve complete targeting or avoid the problems of genomic mutation physically [49,53]. OrthoRep, consisting of an orthogonal DNA polymerase (DNAP)-plasmid pair in *S. cerevisiae*, can mutate target GOIs at ~10^{-5} s.p.b. without any increase in host genomic rates (~10^{-10} s.p.b.) [53]; and phage assisted continuous evolution (PACE) elevates phage genome mutation rates along with host mutation rates, but ingeniously disregards host mutation effects by removing host *Escherichia coli* cells fast enough to prevent host propagation but slow enough to ensure phage propagation [49,63].

Second, durability. Ideally, a continuous evolution system will mutate target GOIs indefinitely so that long mutational paths (*e.g.* >10 non-synonymous mutations) can be traversed over extended periods of strong selection or more complex sequences of selection that mimic natural evolution [64,65]. So far, both OrthoRep and PACE have proven to be quite durable – we have used OrthoRep in several evolution experiments to evolve GOIs for >300 generations and still observe rapid adaptation, accumulating 10-20 mutations (manuscripts in preparation); and PACE has been used in experiments that adapt over hundreds of phage generations, accumulating 10-20 mutations [64,65]. Durability in other continuous evolution systems [60–62,66–73] remain untested, but one can predict durability based on the architecture of the system. For example, in OrthoRep, the only way a GOI gets replicated is through an error-prone DNAP, encoded on a host plasmid or genome that doesn't experience elevated mutagenesis.

This, combined with the fact that OrthoRep achieves complete mutational targeting to avoid selection against elevated mutation rates through mutational loads on the genome, favors durability. Likewise, in PACE, durability is favored, because the only way a phage genome encoding a GOI is replicated is through error-prone means.

Third, scalability. Especially important for experimental evolution, a continuous evolution system should ideally be scalable in nature. Evolving a GOI with a large number of replicates is crucial for observing low frequency events [53,74,75], inferring beneficial mutations [76,77], and determining the extent to which evolutionary trajectories are reproducible [31,78]. Continuous evolution systems that are fully *in vivo*, such as OrthoRep, offer scalablity, because evolution experiments can be carried out simply through serial passaging, amenable to extensive replication or parallelization [60,61,71,72]. Although PACE usually requires chemostat or turbidostat setups that limit scale, recent experiments demonstrate that PACE may be conducted via bulk passaging without such setups and should be amenable to extensive replication [79].

In short, a number of continuous evolution systems, including OrthoRep and PACE, are at a stage of development where they should be able to routinely drive GOI evolution at the speeds, durations, and scale required to study mechanisms of gene evolution through forward evolution experiments.

Early applications of continuous evolution to studying evolutionary pathways and mechanisms

PACE has been the most successful continuous evolution system for proteins to date. In PACE, a GOI is encoded on a phage's genome and through coupling an improvement in the GOI's function to phage survival and infectivity, GOIs with beneficial mutations rapidly propagate in a pool of *E. coli*. By having a continual influx of *E. coli* at a rate that is between the doubling time of phage and *E. coli*, GOIs can rapidly accumulate mutations while mutated host cells are removed. Although most PACE experiments have focused on protein engineering applications, some have aimed to understand the details of evolutionary mechanisms. In 2013, Leconte and colleagues examined the effect of selection stringency and rate of mutagenesis, key parameters of in evolutionary theory, on the evolution of T7 RNA polymerase (RNAP) towards recognition of the T3 promoter [80]. While the effect of mutation rate on adaptive pathways has been studied in other contexts [81–87], PACE enabled Leconte *et al.* to isolate the effects of mutation rate on a single gene in freely evolving replicate cultures [80]. They demonstrated that path choice is largely dictated by mutation rate, and that high mutation rates result in more deterministic fixation of mutations, confirming predictions made *in silico* [84]. Further, Leconte *et al.* revealed that selection stringency also affects mutation path choice, as the strength of selection resulted in in substantial differences in adaptive trajectories. Significantly, only with the benefit of replicate cultures, Leconte *et al*. were able to show that while mutational patterns indeed appeared across replicates, both the specific adaptive mutations, and even successful adaptation itself, can be stochastic in nature.

In a separate study, Dickinson *et al.* used PACE to explore contingency in evolution [88]. Previous research has shown that while the stepwise evolution of a single gene or a small set of genes could be practically deterministic [26,31,89], convergent evolution from dissimilar proteins and histories can lead to vastly different sequences, structures, and activities [32,90,91]. In their 2013 study, Dickinson *et al.* asked just how much dissimilarity in the history of a protein's evolution was needed to result in significant changes in evolutionary outcomes. With PACE, they were able to conduct replicate evolution experiments where T7 RNAP was first diverged to recognize either the T3 or SP6 promoter, and subsequently pressured to recognize the same final promoter, a hybrid T3/SP6 promoter. Surprisingly, Dickenson *et al*. showed that a single divergent evolutionary step was sufficient to drastically alter the mutational trajectory as well as the maximum catalytic efficiency of the evolved enzyme. Specifically, populations of T7 RNAP that were first evolved for T3 promoter recognition evolved lower activity for the final promoter compared to populations first evolved for the SP6 promoter, with differences persisting even after extensive continued selection (~40 generations at high mutational load) for recognition of the final promoter. This study elegantly shows the importance of contingency in evolution and how historical effects cannot be overcome through strong selection alone. Further, through characterization of mutations present in different cultures, Dickenson *et al*. identified key epistatic interactions between mutations that result in the two distinct evolutionary outcomes. With these results, PACE gives us a sense of the types of questions in evolutionary theory that continuous evolution can address, questions that would be difficult to study with traditional directed evolution techniques.

Another continuous evolution technology, OrthoRep, has recently enabled a detailed mapping of adaptive trajectories on a fitness landscape, including low probability events, and demonstrated the effects of epistasis and clonal interference on the reproducibility of adaptation [92]. OrthoRep uses an orthogonal error-prone DNA polymerase-plasmid pair in *S. cerevisiae* to achieve targeted mutagenesis of GOIs [53,92]. By encoding *Plasmodium falciparum* dihydrofolate reductase (*Pf*DHFR) on OrthoRep, Ravikumar *et al.* rapidly evolved resistant pyrimethamine-resistant *Pf*DHFR variants simply by passaging 0.5 mL yeast cultures in media containing increasing concentrations of pyrimethamine. Owing to the scalability of OrthoRep, this experiment was easily repeated 90 times to abundantly sample adaptive trajectories. From this, Ravikumar *et al.* uncovered a more complex fitness landscape than previously realized, including new mutants as resistant as those widely studied. One mutant occurred frequently due to a highly adaptive first-step mutation (S108N) that exhibited sign epistasis with a highly-adaptive later mutation (D54N), requiring one or two additional mutations (C59R and Y57H) to occur between S108N and D54N to resolve the negative sign epistasis. This led to convergence of adaptive trajectories across replicates. However, in a few replicates, rare mutations steered populations towards other equally-fit outcomes, including ones lacking S108N, and suboptimal local fitness peaks. Since these alternative variants are expected to respond differently to secondary drugs, population structures and strategies that favor rare mutational pathways may be important for drug schedule design, which we are currently exploring. In short, by exploiting rapid and scalable continuous evolution, one can explore adaptation on rugged fitness landscapes to tease out both the stochastic and deterministic nature of evolution.

Future potential

Continuous evolution systems hold great promise in studying the mechanisms and pathways of gene adaptation. The early studies described above give a glimpse into how continuous evolution can be used to carry out controlled forward evolution experiments that discover and map interesting regions of fitness landscapes, test the reproducibility of adaptation, and compare how different parameters of evolution and selection schedules result in different mutational trajectories and outcomes. As more researchers use continuous evolution to carry out forward evolution experiments with previously inaccessible speed, depth, and scale, significant insights should be made. These should not only include exquisite details of how specific genes adapt through the interplay among mutations, but also general insights into the most fundamental questions in molecular evolution – the reproducibility of adaptation [93,94], how fitness valleys are crossed [95,96], the importance of fluctuating environments or population structure in adaptation [97], the prevalence and role of epistasis in protein evolution [98–100], the existence of tradeoffs among different gene functions [19,41], the determinants of evolvability [40,101], the high prevalence of certain folds or structures in enzymes [102–104], the evolutionary basis of protein-protein interactions [105,106], and the role of both intracellular and environmental conditions in dictating how a gene adapts [107,108]. With the number of powerful systems available and ongoing development in each, such as the inclusion of gene-specific sexual recombination into OrthoRep (unpublished data), continuous evolution should become a staple technology for probing the fundamentals of adaptation.

Advancement to directed evolution presented in this work

My first contribution was to engineer a series of constructs to allow expression of proteins on OrthoRep to span a range comparable to genomic expression. Based on preliminary evolution experiments conducted in our lab, I screened a panel of mutations in the 5' UCR of genes encoded on OrthoRep to examine if any mutations increased expression. I found a combination of mutations, termed 10B2, which modestly increased expression of genes encoded on OrthoRep three fold. I combined this with a genetically encoded poly-A tail to further increase expression by 15-fold. Combined, these developments allow genes encoded on OrthoRep to be expressed at levels comparable to this strongest nuclear promoters when paired with the wild-type DNA polymerase, and at levels comparable to a medium nuclear promoter when using a mutagenic DNA polymerase. This has expanded the possible evolutionary targets to include a plethora of proteins, including nanobodies and weakly functioning enzymes.

Additionally, I paired OrthoRep with a continuous culture device, eVOLVER, to develop Automated Continuous Evolution (ACE). ACE uses the eVOLVER framework and a custom algorithm that continuously measures culture conditions, including growth rate, and adjusts culture conditions accordingly to maintain a given selection pressure, therefore removing the need for manual passaging and adjustment of selection stringency. I demonstrate the potential of ACE through evolving *Plasmodium falciparum* DHFR to resist the competitive inhibitor pyrimethamine and to adapt the thermophilic HisA from *Thermotoga maritima* to function at mesophile temperatures by complementing a His6 deletion in *Saccharomyces cerevisiae*. Through the course of evolution, I demonstrate that using ACE with both of these selections is superior to manual passaging experiments by the time to reach full fitness and by the amount of manual input needed. Consequently, ACE is a powerful tool when selection stringencies are

unknown or highly sensitive, as well as enabling the control of evolutionary pressures in experimental evolution.

Finally, I use OrthoRep to evolve HisA from *Thermotoga maritima* (*Tm*HisA), which naturally is involved in the biosynthesis of histidine, to catalyze a similar isomerization reaction naturally catalyzed by Trp1 in the biosynthesis of tryptophan. Through this experiment, I demonstrate the ability of OrthoRep to explore a complicated fitness landscape and show that strong selection of *Tm*HisA for Trp1 activity consistently reaches a plateau in activity that cannot be bypassed by continued strong selection. Rather, only through an alternative selection for His6 activity, the native function of *Tm*HisA, followed by reselection for Trp1 activity can *Tm*HisA achieve mutations that allow for full complementation of Trp1 in *S. cerevisiae*. Which the mechanism for this need for alternative selection has not been elucidated, this offers a potential selection scheme that may allow enzymes obtained through directed evolution to achieve higher catalytic efficiencies.

References

(1) Turner, N. J. Directed Evolution Drives the next Generation of Biocatalysts. *Nature Chemical Biology*. 2009, pp 567–573. https://doi.org/10.1038/nchembio.203.

(2) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet.* **2015**, *16* (7), 379–394. https://doi.org/10.1038/nrg3927.

(3) Peisajovich, S. G.; Tawfik, D. S. Protein Engineers Turned Evolutionists. *Nature Methods*. 2007, pp 991–994. https://doi.org/10.1038/nmeth1207-991.

(4) Bloom, J. D.; Arnold, F. H. In the Light of Directed Evolution: Pathways of Adaptive Protein Evolution. *Proc. Natl. Acad. Sci.* **2009**, *106* (Supplement_1), 9995–10000. https://doi.org/10.1073/pnas.0901522106.

(5) Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (12), 866–876. https://doi.org/10.1038/nrm2805.

(6) Arnold, F. H.; Wintrode, P. L.; Miyazaki, K.; Gershenson, A. How Enzymes Adapt: Lessons from Directed Evolution. *Trends Biochem. Sci.* **2001**, *26* (2), 100–106. https://doi.org/10.1016/S0968-0004(00)01755-2.

(7) Merkl, R.; Sterner, R. Ancestral Protein Reconstruction: Techniques and Applications. *Biological Chemistry*. 2016, pp 1–21. https://doi.org/10.1515/hsz-2015-0158.

(8) Koshi, J. M.; Goldstein, R. A. Probabilistic Reconstruction of Ancestral Protein Sequences. *J. Mol. Evol.* **1996**, *42* (2), 313–320. https://doi.org/10.1007/BF02198858.

(9) Zuckerkandl, E.; Pauling, L. Molecules as Documents of Evolutionary History. *J. Theor. Biol.* **1965**, *8*, 357–366. https://doi.org/10.1016/0022-5193(65)90083-4.

(10) *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*;

Salemi, M., Vandamme, A.-M., Eds.; Cambridge University Press, Cambridge, UK, 2003.

(11) Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*; Oxford University Press, 2000.

(12) Tokuriki, N.; Jackson, C. J.; Afriat-Jurnou, L.; Wyganowski, K. T.; Tang, R.; Tawfik, D. S. Diminishing Returns and Tradeoffs Constrain the Laboratory Optimization of an Enzyme. *Nat. Commun.* **2012**, *3*, 1257–1259. https://doi.org/10.1038/ncomms2246.

(13) Yadid, I.; Kirshenbaum, N.; Sharon, M.; Dym, O.; Tawfik, D. S. Metamorphic Proteins Mediate Evolutionary Transitions of Structure. *Proc. Natl. Acad. Sci.* **2010**, *107* (16), 7287–7292. https://doi.org/10.1073/pnas.0912616107.

(14) Peisajovich, S. G.; Rockah, L.; Tawfik, D. S. Evolution of New Protein Topologies through Multistep Gene Rearrangements. *Nat. Genet.* **2006**, *38* (2), 168–174. https://doi.org/10.1038/ng1717.

(15) Khersonsky, O.; Malitsky, S.; Rogachev, I.; Tawfik, D. S. Role of Chemistry versus Substrate Binding in Recruiting Promiscuous Enzyme Functions. *Biochemistry* **2011**, *50* (13), 2683–2690. https://doi.org/10.1021/bi101763c.

(16) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci.* **2006**, *103* (15), 5869–5874. https://doi.org/10.1073/pnas.0510098103.

(17) Bloom, J. D.; Raval, A.; Wilke, C. O. Thermodynamics of Neutral Protein Evolution. *Genetics* **2007**, *175* (1), 255–266. https://doi.org/10.1534/genetics.106.061754.

(18) Tokuriki, N.; Tawfik, D. S. Stability Effects of Mutations and Protein Evolvability. *Curr. Opin. Struct. Biol.* **2009**, *19* (5), 596–604. https://doi.org/10.1016/j.sbi.2009.08.003.

(19) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How Protein Stability and New

Functions Trade Off. *PLoS Comput. Biol.* **2008**, *4* (2), 35–37. https://doi.org/10.1371/journal.pcbi.1000002.

(20) Wang, X.; Minasov, G.; Shoichet, B. K. Evolution of an Antibiotic Resistance Enzyme Constrained by Stability and Activity Trade-Offs. **2002**, *2836* (02), 85–95. https://doi.org/10.1016/S0022-2836(02)00400-X.

(21) Beadle, B. M.; Shoichet, B. K. Structural Bases of Stability-Function Tradeoffs in Enzymes. *J. Mol. Biol.* **2002**, *321* (2), 285–296. https://doi.org/10.1016/S0022-2836(02)00599-5.

(22) Bloom, J. D.; Romero, P. A.; Lu, Z.; Arnold, F. H. Neutral Genetic Drift Can Alter Promiscuous Protein Functions, Potentially Aiding Functional Evolution. *Biol. Direct* **2007**, *2*, 7–10. https://doi.org/10.1186/1745-6150-2-17.

(23) Kaltenbach, M.; Emond, S.; Hollfelder, F.; Tokuriki, N. Functional Trade-Offs in Promiscuous Enzymes Cannot Be Explained by Intrinsic Mutational Robustness of the Native Activity. *PLoS Genet.* **2016**, *12* (10), 1–18. https://doi.org/10.1371/journal.pgen.1006305.

(24) Khersonsky, O.; Tawfik, D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annual Review of Biochemistry*. 2010, pp 471–505. https://doi.org/10.1146/annurev-biochem-030409-143718.

(25) Aharoni, A.; Gaidukov, L.; Khersonsky, O.; Gould, S. M.; Roodveldt, C.; Tawfik, D. S. The "evolvability" of Promiscuous Protein Functions. *Nat. Genet.* **2005**, *37* (1), 73–76. https://doi.org/10.1038/ng1482.

(26) Weinreich, D. M.; Delaney, N. F.; Depristo, M. a; Hartl, D. L. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science (80-.).* **2006**, *312*,

111–114. https://doi.org/10.1126/science.1123539.

(27) Tan, L.; Serene, S.; Chao, H. X.; Gore, J. Hidden Randomness between Fitness Landscapes Limits Reverse Evolution. *Phys. Rev. Lett.* **2011**, *106* (19), 1–4. https://doi.org/10.1103/PhysRevLett.106.198102.

(28) Hietpas, R. T.; Jensen, J. D.; Bolon, D. N. A. Experimental Illumination of a Fitness Landscape. *Proc. Natl. Acad. Sci.* **2011**, *108* (19), 7896–7901. https://doi.org/10.1073/pnas.1016024108.

(29) Bershtein, S.; Segal, M.; Bekerman, R.; Tokuriki, N.; Tawfik, D. S. Robustness-Epistasis Link Shapes the Fitness Landscape of a Randomly Drifting Protein. *Nature* **2006**, *444* (7121), 929–932. https://doi.org/10.1038/nature05385.

(30) Haddox, H. K.; Dingens, A. S.; Hilton, S. K.; Overbaugh, J.; Bloom, J. D. Mapping Mutational Effects along the Evolutionary Landscape of HIV Envelope. *Elife* **2018**, *7*, 1–29. https://doi.org/10.7554/eLife.34420.

(31) De Visser, J. A. G. M.; Krug, J. Empirical Fitness Landscapes and the Predictability of Evolution. *Nat. Rev. Genet.* **2014**, *15* (7), 480–490. https://doi.org/10.1038/nrg3744.

(32) Salverda, M. L. M.; Dellus, E.; Gorter, F. A.; Debets, A. J. M.; van der Oost, J.; Hoekstra, R. F.; Tawfik, D. S.; de Visser, J. A. G. M. Initial Mutations Direct Alternative Pathways of Protein Evolution. *PLoS Genet.* **2011**, *7* (3). https://doi.org/10.1371/journal.pgen.1001321.

(33) Dellus-Gur, E.; Elias, M.; Caselli, E.; Prati, F.; Salverda, M. L. M.; De Visser, J. A. G. M.; Fraser, J. S.; Tawfik, D. S. Negative Epistasis and Evolvability in TEM-1 β-Lactamase - The Thin Line between an Enzyme's Conformational Freedom and Disorder. *J. Mol. Biol.* **2015**, *427* (14), 2396–2409. https://doi.org/10.1016/j.jmb.2015.05.011.

(34) Bank, C.; Matuszewski, S.; Hietpas, R. T.; Jensen, J. D. On the (Un)Predictability of a Large Intragenic Fitness Landscape. *Proc. Natl. Acad. Sci.* **2016**, *113* (49), 14085–14090. https://doi.org/10.1073/pnas.1612676113.

(35) Kaltenbach, M.; Jackson, C. J.; Campbell, E. C.; Hollfelder, F.; Tokuriki, N. Reverse Evolution Leads to Genotypic Incompatibility despite Functional and Active Site Convergence. *Elife* **2015**, *4* (AUGUST2015), 1–20. https://doi.org/10.7554/eLife.06492.

(36) Amitai, G.; Gupta, R. D.; Tawfik, D. S. Latent Evolutionary Potentials under the Neutral Mutational Drift of an Enzyme. *HFSP J.* **2007**, *1* (1), 67–78. https://doi.org/10.2976/1.2739115/10.2976/1.

(37) Smith, W. S.; Hale, J. R.; Neylon, C. Applying Neutral Drift to the Directed Molecular Evolution of a B-Glucuronidase into a b-Galactosidase: Two Different Evolutionary Pathways Lead to the Same Variant. *BMC Res. Notes* **2011**, *4* (1), 138. https://doi.org/10.1186/1756-0500-4-138.

(38) Bershtein, S.; Goldin, K.; Tawfik, D. S. Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J. Mol. Biol.* **2008**, *379* (5), 1029–1044. https://doi.org/10.1016/j.jmb.2008.04.024.

(39) Steinberg, B.; Ostermeier, M. Environmental Changes Bridge Evolutionary Valleys. *Sci. Adv.* **2016**, *2* (1). https://doi.org/10.1126/sciadv.1500921.

(40) Soskine, M.; Tawfik, D. S. Mutational Effects and the Evolution of New Protein Functions. *Nat. Rev. Genet.* **2010**, *11* (8), 572–582. https://doi.org/10.1038/nrg2808.

(41) Tawfik, D. S. Accuracy-Rate Tradeoffs: How Do Enzymes Meet Demands of Selectivity and Catalytic Efficiency? *Curr. Opin. Chem. Biol.* **2014**, *21*, 73–80. https://doi.org/10.1016/j.cbpa.2014.05.008.

(42) Tokuriki, N.; Tawfik, D. S. Protein Dynamism and Evolvability. *Science (80-.).* **2012**, *324* (April), 203–207. https://doi.org/10.1126/science.1221339.

(43) Goldsmith, M.; Tawfik, D. S. Enzyme Engineering: Reaching the Maximal Catalytic Efficiency Peak. *Curr. Opin. Struct. Biol.* **2017**, *47*, 140–150. https://doi.org/10.1016/j.sbi.2017.09.002.

(44) Drummond, D. A.; Iverson, B. L.; Georgiou, G.; Arnold, F. H. Why High-Error-Rate Random Mutagenesis Libraries Are Enriched in Functional and Improved Proteins. *J. Mol. Biol.* **2005**, *350* (4), 806–816. https://doi.org/10.1016/j.jmb.2005.05.023.

(45) Zaccolo, M.; Gherardi, E. The Effect of High-Frequency Random Mutagenesis on in Vitro Protein Evolution: A Study on TEM-1 β-Lactamase. *J. Mol. Biol.* **1999**, *285* (2), 775–783. https://doi.org/10.1006/jmbi.1998.2262.

(46) Kunichika, K.; Hashimoto, Y.; Imoto, T. Robustness of Hen Lysozyme Monitored by Random Mutations. *Protein Eng.* **2002**, *15* (10), 805–809. https://doi.org/10.1093/protein/15.10.805.

(47) Yang, J.; Ruff, A. J.; Arlt, M.; Schwaneberg, U. Casting EpPCR (CepPCR): A Simple Random Mutagenesis Method to Generate High Quality Mutant Libraries. *Biotechnol. Bioeng.* **2017**, *114* (9), 1921–1927. https://doi.org/10.1002/bit.26327.

(48) Tokuriki, N.; Tawfik, D. S. Chaperonin Overexpression Promotes Genetic Variation and Enzyme Evolution. *Nature* **2009**, *459* (7247), 668–673. https://doi.org/10.1038/nature08009.

(49) Badran, A. H.; Liu, D. R. In Vivo Continuous Directed Evolution. *Curr. Opin. Chem. Biol.* **2015**, *24*, 1–10. https://doi.org/10.1016/j.cbpa.2014.09.040.

(50) d'Oelsnitz, S.; Ellington, A. Continuous Directed Evolution for Strain and Protein

Engineering. *Curr. Opin. Biotechnol.* **2018**, *53*, 158–163. https://doi.org/10.1016/j.copbio.2017.12.020.

(51) Zheng, X.; Xing, X.; Zhang, C. Targeted Mutagenesis : A Sniper-like Diversity Generator in Microbial Engineering. *Synth. Syst. Biotechnol.* **2017**, *2* (2), 75–86. https://doi.org/10.1016/j.synbio.2017.07.001.

(52) Wellner, A.; Ravikumar, A.; Liu, C. C. Continuous Evolution of Proteins in Vivo; 2018; p in press.

(53) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C. Scalable Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **2018**, *175*, 1–12. https://doi.org/10.1101/313338.

(54) Bull, J. J.; Sanjuan, R.; Wilke, C. O. Theory of Lethal Mutagenesis for Viruses. *J. Virol.* **2007**, *81* (6), 2930–2939. https://doi.org/10.1128/jvi.01624-06.

(55) Herr, A. J.; Ogawa, M.; Lawrence, N. A.; Williams, L. N.; Eggington, J. M.; Singh, M.; Smith, R. A.; Preston, B. D. Mutator Suppression and Escape from Replication Error-Induced Extinction in Yeast. *PLoS Genet.* **2011**, *7* (10). https://doi.org/10.1371/journal.pgen.1002282.

(56) Wilke, C. O.; Wang, J. L.; Ofria, C.; Lenski, R. E.; Adami, C. Evolution of Digital Organisms at High Mutation Rates Leads to Survival of the Flattest. *Nature* **2001**, *412*, 331–333.

(57) Nowak, M.; Schuster, P. Error Thresholds of Replication in Finite Populations Mutation Frequencies and the Onset of Muller's Ratchet. *J. Theor. Biol.* **1989**, *137*, 375–395.

(58) Felsenstein, J. The Evolutionary Advantage of Recombination. *Genetics* **1976**, *83* (4), 845–859. https://doi.org/S1090-0233(10)00296-0 [pii]\r10.1016/j.tvjl.2010.09.005 [doi].

(59) Muller, H. J. The Relation of Recombination to Mutational Advance. *Mutat. Res.* **1964**, *1*, 2–9. https://doi.org/10.1117/12.722789.

(60) Halperin, S. O.; Tou, C. J.; Wong, E. B.; Modavi, C.; Schaffer, D. V.; Dueber, J. E. CRISPR-Guided DNA Polymerases Enable Diversification of All Nucleotides in a Tunable Window. *Nature* **2018**, *560* (7717), 248–252. https://doi.org/10.1038/s41586-018-0384-8.

(61) Moore, C. L.; Papa, L. J.; Shoulders, M. D. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions in Vivo. *J. Am. Chem. Soc.* **2018**, *140* (37), 11560–11564. https://doi.org/10.1021/jacs.8b04001.

(62) Camps, M.; Naukkarinen, J.; Johnson, B. P.; Loeb, L. A. Targeted Gene Evolution in Escherichia Coli Using a Highly Error-Prone DNA Polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (17), 9727–9732. https://doi.org/10.1073.

(63) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **2011**, *472* (7344), 499–503. https://doi.org/10.1038/nature09929.

(64) Packer, M. S.; Rees, H. A.; Liu, D. R. Phage-Assisted Continuous Evolution of Proteases with Altered Substrate Specificity. *Nat. Commun.* **2017**, *8* (1). https://doi.org/10.1038/s41467-017-01055-9.

(65) Badran, A. H.; Guzov, V. M.; Huai, Q.; Kemp, M. M.; Vishwanath, P.; Kain, W.; Nance, A. M.; Evdokimov, A.; Moshiri, F.; Turner, K. H.; Wang, P.; Malvar, T.; Liu, D. R. Continuous Evolution of Bacillus Thuringiensis Toxins Overcomes Insect Resistance. *Nature* **2016**, *533* (7601), 58–63. https://doi.org/10.1038/nature17938.

(66) Fabret, C.; Poncet, S.; Danielsen, S.; Borchert, T. V.; Ehrlich, S. D.; Janniere, L. Efficient Gene Targeted Random Mutagenesis in Genetically Stable Escherichia Coli Strains.

Nucleic Acids Res **2000**, *28* (21), 1–5. https://doi.org/10.1080/00202967.2000.11871320.

(67) Wang, C. L.; Harper, R. A.; Wabl, M. Genome-Wide Somatic Hypermutation. *Proc. Natl. Acad. Sci.* **2004**, *101* (19), 7352–7356. https://doi.org/10.1073/pnas.0402009101.

(68) Romanini, D. W.; Peralta-yahya, P.; Mondol, V.; Cornish, V. W. A Heritable Recombination System for Synthetic Darwinian Evolution in Yeast. *ACS Synth. Biol.* **2012**, *1*, 602–609.

(69) Finney-Manchester, S. P.; Maheshri, N. Harnessing Mutagenic Homologous Recombination for Targeted Mutagenesis in Vivo by TaGTEAM. *Nucleic Acids Res.* **2013**, *41* (9), 1–10. https://doi.org/10.1093/nar/gkt150.

(70) Crook, N.; Abatemarco, J.; Sun, J.; Wagner, J. M.; Schmitz, A.; Alper, H. S. In Vivo Continuous Evolution of Genes and Pathways in Yeast. *Nat. Commun.* **2016**, *7*. https://doi.org/10.1038/ncomms13051.

(71) Hess, G. T.; Frésard, L.; Han, K.; Lee, C. H.; Li, A.; Cimprich, K. A.; Montgomery, S. B.; Bassik, M. C. Directed Evolution Using DCas9-Targeted Somatic Hypermutation in Mammalian Cells. *Nat. Methods* **2016**, *13* (12), 1036–1042. https://doi.org/10.1038/nmeth.4038.

(72) Ma, Y.; Zhang, J.; Yin, W.; Zhang, Z.; Song, Y.; Chang, X. Targeted AID-Mediated Mutagenesis (TAM) Enables Efficient Genomic Diversification in Mammalian Cells. *Nat. Methods* **2016**, *13* (12), 1029–1035. https://doi.org/10.1038/nmeth.4027.

(73) Smith, S. N.; Wang, Y.; Baylon, J. L.; Singh, N. K.; Baker, B. M.; Tajkhorshid, E.; Kranz, D. M. Changing the Peptide Specificity of a Human T-Cell Receptor by Directed Evolution. *Nat. Commun.* **2014**, *5*, 1–13. https://doi.org/10.1038/ncomms6223.

(74) Burke, M. K.; Dunham, J. P.; Shahrestani, P.; Thornton, K. R.; Rose, M. R.; Long, A. D.

Genome-Wide Analysis of a Long-Term Evolution Experiment with Drosophila. *Nature* **2010**, *467* (7315), 587–590. https://doi.org/10.1038/nature09352.

(75) Imhof, M.; Schlo, C. Fitness Effects of Advantageous Mutations in Evolving Escherichia Coli Populations. **2001**, *98* (3), 1113–1117.

(76) Lang, G. I.; Rice, D. P.; Hickman, M. J.; Sodergren, E.; Weinstock, G. M.; Botstein, D.; Desai, M. M. Pervasive Genetic Hitchhiking and Clonal Interference in Forty Evolving Yeast Populations. *Nature* **2013**, *500* (7464), 571–574. https://doi.org/10.1038/nature12344.

(77) Barrick, J. E.; Yu, D. S.; Yoon, S. H.; Jeong, H.; Oh, T. K.; Schneider, D.; Lenski, R. E.; Kim, J. F. Genome Evolution and Adaptation in a Long-Term Experiment with Escherichia Coli. *Nature* **2009**, *461* (7268), 1243–1247. https://doi.org/10.1038/nature08480.

(78) Wichman, H. A.; Badgett, M. R.; Scott, L. A.; Boulianne, C. M.; Bull, J. J. Different Trajectories of Parallel Evolution during Viral Adaptation. *Science (80-.).* **1999**, *285* (5426), 422–424. https://doi.org/10.1126/science.285.5426.422.

(79) Bryson, D. I.; Fan, C.; Guo, L. T.; Miller, C.; Söll, D.; Liu, D. R. Continuous Directed Evolution of Aminoacyl-TRNA Synthetases. *Nat. Chem. Biol.* **2017**, *13* (12), 1253–1260. https://doi.org/10.1038/nchembio.2474.

(80) Leconte, A. M.; Dickinson, B. C.; Yang, D. D.; Chen, I. a; Allen, B.; Liu, D. R. A Population-Based Experimental Model for Protein Evolution: Effects of Mutation Rate and Selection Stringency on Evolutionary Outcomes. *Biochemistry* **2013**, *52*, 1490–1499. https://doi.org/10.1021/bi3016185.

(81) Elena, S. F.; Wilke, C. O.; Ofria, C.; Lenski, R. E. Effects of Population Size and

Mutation Rate on the Evolution of Mutational Robustness. *Evolution (N. Y).* **2007**, *61* (3), 666–674. https://doi.org/10.1111/j.1558-5646.2007.00064.x.

(82) Rouzine, I. M.; Rodrigo, A.; Coffin, J. M. Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. *Microbiol. Mol. Biol. Rev.* **2001**, *65* (1), 151–185. https://doi.org/10.1128/mmbr.65.1.151-185.2001.

(83) Vahdati, A. R.; Sprouffske, K.; Wagner, A. Effect of Population Size and Mutation Rate on the Evolution of RNA Sequences on an Adaptive Landscape Determined by RNA Folding. **2017**, *13*. https://doi.org/10.7150/ijbs.19436.

(84) Szendro, I. G.; Franke, J.; de Visser, J. A. G. M.; Krug, J. Predictability of Evolution Depends Nonmonotonically on Population Size. *Proc. Natl. Acad. Sci.* **2013**, *110* (2), 571–576. https://doi.org/10.1073/pnas.1213613110.

(85) Jiang, X.; Mu, B.; Huang, Z.; Zhang, M.; Wang, X.; Tao, S. Impacts of Mutation Effects and Population Size on Mutation Rate in Asexual Populations : A Simulation Study. **2010**.

(86) Daugherty, P. S.; Chen, G.; Iverson, B. L.; Georgiou, G. Quantitative Analysis of the Effect of the Mutation Frequency on the Affinity Maturation of Single Chain Fv Antibodies. *Proc. Natl. Acad. Sci.* **2000**, *97* (5), 2029–2034. https://doi.org/10.1073/pnas.030527597.

(87) Dijk, T. Van; Hwang, S.; Krug, J.; Visser, A. G. M. De; Zwart, M. P. Mutant Library Size and the Repeatability of Selection for Antibiotic Resistance. *Phys. Biol.* **2017**, *14*, 1–29.

(88) 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.* **2013**, *110* (22), 9007–9012.

https://doi.org/10.1073/pnas.1220670110.

(89) Woods, R.; Schneider, D.; Winkworth, C. L.; Riley, M. A.; Lenski, R. E. Tests of Parallel Molecular Evolution in a Long-Term Experiment with Escherichia Coli. *Proc. Natl. Acad. Sci.* **2006**, *103* (24), 9107–9112. https://doi.org/10.1073/pnas.0602917103.

(90) Kuriyan, J.; Krishna, T. S. R.; Wong, L.; Guenther, B.; Pahler, A.; Williams, C. H.; Model, P. Convergent Evolution of Similar Function in Two Structurally Divergent Enzymes. *Nature*. 1991, pp 172–174. https://doi.org/10.1038/352172a0.

(91) Graumann, P.; Marahiel, M. A. A Case of Convergent Evolution of Nucleic Acid Binding Modules. *BioEssays* **1996**, *18* (4), 309–315.

(92) Ravikumar, A.; Arrieta, A.; Liu, C. C. An Orthogonal DNA Replication System in Yeast. *Nat. Chem. Biol.* **2014**, *10* (3), 175–177. https://doi.org/10.1038/nchembio.1439.

(93) Achaz, G.; Rodriguez-Verdugo, A.; Gaut, B. S.; Tenaillon, O. The Reproducibility of Adaptation in the Light of Experimental Evolution with Whole Genome Sequencing. In *Ecological Genomics. Advances in Experimental Medicine and Biology, vol 781*; Landry, C. R., Aubin-Horth, N., Eds.; Springer, Dordrecht, 2014; pp 211–231.

(94) Lässig, M.; Mustonen, V.; Walczak, A. M. Predicting Evolution. *Nat. Ecol. Evol.* **2017**, *1* (3), 1–9. https://doi.org/10.1038/s41559-017-0077.

(95) *The Adaptive Landscape in Evolutionary Biology*; Svensson, E., Calsbeek, R., Eds.; Oxford University Press, 2012.

(96) Obolski, U.; Ram, Y.; Hadany, L. Key Issues Review: Evolution on Rugged Adaptive Landscapes. *Reports Prog. Phys.* **2018**, *81* (1), 1–31. https://doi.org/10.1088/1361-6633/aa94d4.

(97) Wright, S. The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution.

Proc. 6th Int. Congr. Genet. **1932**, *1*, 356–366. https://doi.org/10.1016/j.tig.2013.09.001.

(98) Starr, T. N.; Thornton, J. W. Epistasis in Protein Evolution. *Protein Sci.* **2016**, *25*, 1204–1218. https://doi.org/10.1002/pro.2897.

(99) Poelwijk, F. J.; Kiviet, D. J.; Weinreich, D. M.; Tans, S. J. Empirical Fitness Landscapes Reveal Accessible Evolutionary Paths. *Nature* **2007**, *445* (7126), 383–386. https://doi.org/10.1038/nature05451.

(100) Storz, J. F. Compensatory Mutations and Epistasis for Protein Function. *Curr. Opin. Struct. Biol.* **2018**, *50*, 18–25. https://doi.org/10.1016/j.sbi.2017.10.009.

(101) Pal, C.; Papp, B.; Lercher, M. J. An Integrated View on Protein Evolution. *Nat. Rev. Genet.* **2006**, *7*, 337–348.

(102) Tóth-Petróczy, Á.; Tawfik, D. S. The Robustness and Innovability of Protein Folds. *Curr. Opin. Struct. Biol.* **2014**, *26* (1), 131–138. https://doi.org/10.1016/j.sbi.2014.06.007.

(103) Glasner, M. E.; Gerlt, J. A.; Babbitt, P. C. Evolution of Enzyme Superfamilies. *Curr. Opin. Chem. Biol.* **2006**, *10* (5), 492–497. https://doi.org/10.1016/j.cbpa.2006.08.012.

(104) Storz, J. F. Causes of Molecular Convergence and Parallelism in Protein Evolution. *Nat. Rev. Genet.* **2016**, *17* (4), 239–250. https://doi.org/10.1038/nrg.2016.11.

(105) De Juan, D.; Pazos, F.; Valencia, A. Emerging Methods in Protein Co-Evolution. *Nat. Rev. Genet.* **2013**, *14* (4), 249–261. https://doi.org/10.1038/nrg3414.

(106) Szurmant, H.; Weigt, M. Inter-Residue, Inter-Protein and Inter-Family Coevolution: Bridging the Scales. *Curr. Opin. Struct. Biol.* **2018**, *50*, 26–32. https://doi.org/10.1016/j.sbi.2017.10.014.

(107) Melbinger, A.; Vergassola, M. The Impact of Environmental Fluctuations on Evolutionary Fitness Functions. *Nat. Publ. Gr.* **2015**, 1–11. https://doi.org/10.1038/srep15211.

(108) Poelwijk, F. J.; De Vos, M. G. J.; Tans, S. J. Tradeoffs and Optimality in the Evolution of Gene Regulation. *Cell* **2011**, *146* (3), 462–470. https://doi.org/10.1016/j.cell.2011.06.035.

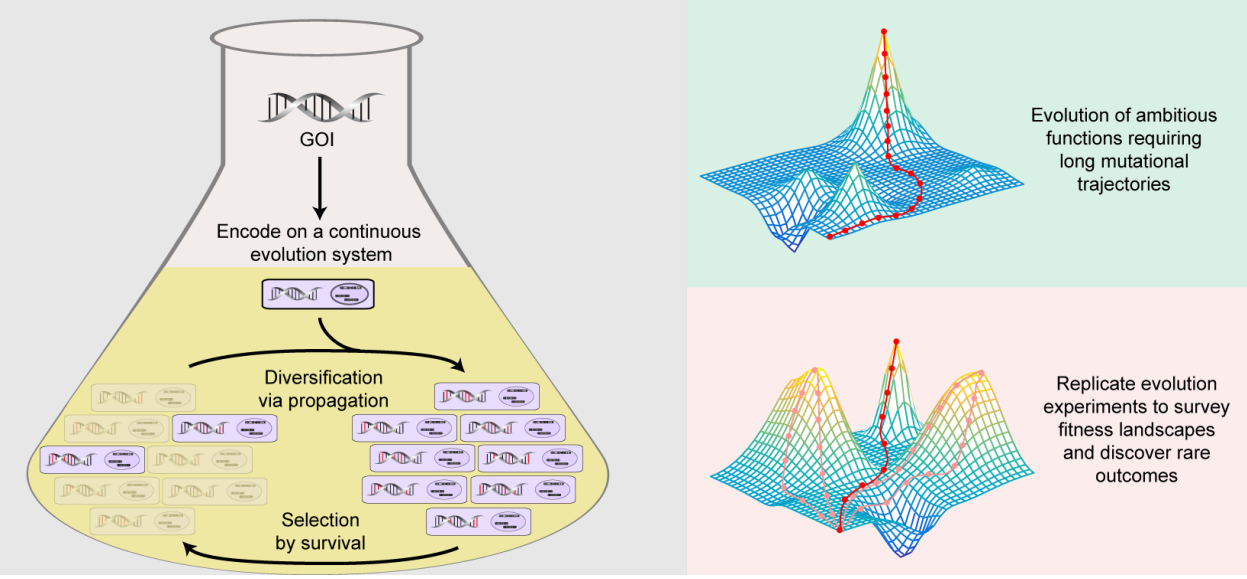


Figure 1.1. Continuous *in vivo* evolution systems enable the rapid continuous diversification of genes of interest in multiple replicate cultures. Through coupling continuous diversification with selection, simply passaging cultures can drive protein evolution on laboratory timescales. This allows proteins to achieve ambitious functions that may require high numbers of mutation (>10-20). Further, the ability to run replicate evolution experiments allows for the detailed mapping of fitness landscapes, discovery of rare outcomes, exploration of multiple environmental conditions and population structures, and statistical power in testing evolutionary reproducibility and basic evolutionary theories.

Chapter 2. Tunable expression systems for orthogonal DNA replication

Ziwei Zhong[1], Arjun Ravikumar[1], Chang C. Liu[1,2,3,*]

[1] Department of Biomedical Engineering, University of California, Irvine, California 92697, United States

[2] Department of Chemistry, University of California, Irvine, California 92697, United States

[3] Department of Molecular Biology & Biochemistry, University of California, Irvine, California 92697, United States

Introduction

We have previously described an orthogonal replication system in *Saccharomyces cerevisiae*, OrthoRep, which allows for the *in vivo* directed evolution of genes in a continuous and rapid manner.[1,2] OrthoRep is an engineered derivative of the linear killer cytoplasmic plasmids from *Kluyveromyces lactis*, pGKL1 (p1) and pGKL2 (p2).[3] Since each of the plasmids encodes its own dedicated DNA polymerase (DNAP) and replicates independently of the host genome (Figure 2.1a),[2–4] the resulting property of replicative orthogonality has allowed us to engineer highly-error prone versions of p1's DNAP (TP-DNAP1). As a result, genes encoded on p1 continuously mutate ~100,000 times more rapidly than genes encoded on the genome while chromosomal genes remain at their natural mutation rate of 10^{-10} substitutions per base (s.p.b.). Consequently, genes encoded on OrthoRep rapidly evolve under selection, resulting in a scalable system for continuous *in vivo* evolution. Indeed, using OrthoRep, our group has mapped drug-target-based resistance in scores of replicates to elucidate new adaptive trajectories and their interplay[1] and are currently carrying out several other rapid evolution experiments.

Genes encoded on OrthoRep need to be expressed for selection to act on their function, and this occurs through transcription by a dedicated set of enzymes encoded on p2. These enzymes, which include an RNA polymerase (RNAP) and a capping enzyme, appear distinct and isolated from nuclear transcription. Nuclear promoters are nonfunctional on p1,[5] and p1/p2 promoter sequences fail to yield proteins when encoded on nuclear plasmids.[6] Additionally, OrthoRep transcripts lack canonical 3' mRNA polyadenylation, as northern and RACE analysis of p2 transcripts show they are similar in size to their encoded gene,[7,8] suggesting that host mRNA

processing does not act on OrthoRep transcripts. Therefore, special cis-elements need to be used when encoding genes for OrthoRep.

To drive transcription, our previous experiments with OrthoRep have relied on the addition of upstream control regions (UCRs) to any heterologous gene of interest (GOI) that we wish to continuously evolve. These UCRs are simply defined as the 100 bp sequence upstream of the start codon of endogenous open reading frames (ORFs) encoded on unmodified p1 and p2 plasmids.[1,2,9] We found that the strengths of the promoters in these natural UCRs are relatively low. Indeed, the strongest promoter identified, derived from p**G**L**2** **O**RF**10** (K2O10), only drives expression, summed across the many copies of p1, at a level equivalent to a low-medium strength nuclear promoter.[2,9] While these promoters may allow for evolution of proteins whose initial selectable function does not require high expression to achieve (*e.g.* transcription factors, biosensors, and certain enzymes), we sought to expand the range of expression levels available to OrthoRep to make it more versatile in the evolution of proteins or pathways. By identifying promoter mutants from our previous OrthoRep-based continuous evolution experiments and combining these promoters with a genetically encoded 3' poly(A) tail, we now report a panel of expression cassettes that span the range of expression strengths achievable on standard yeast nuclear plasmids.

Results

UCR mutations increase OrthoRep expression

In many OrthoRep evolution experiments, the repeated occurrence of mutations in the UCR regions attached to the GOI undergoing evolution led us to hypothesize that these mutations increased expression of target genes. For example, in a replicate evolution experiment for DHFR-dependent pyrimethamine resistance, we found several mutations in DHFR's K2O10 UCR,[1] and it is known that DHFR overexpression is a mechanism for increasing resistance.[10] Using a fluorescent protein reporter as a proxy for expression strength, we individually verified six different substitution mutations in K2O10 UCR, and observed that they indeed significantly increase fluorescence relative to wild type (wt) K2O10 UCR (Figure S2.2a,b, Table S2.1). These individual mutations were strategically combined to form a non-exhaustive panel of double, triple, and quadruple K2O10 UCR mutants for further analysis (Figure 2.1b, S2.2c,d, Table S2.1). One set of mutations, referred to as **10B2**, resulted in an up to three-fold increase in expression over the wt K2O10 UCR sequence (Figure 2.1b, S2.2c,d, Table S2.1).

Genetically encoded poly(A) tails boost expression of genes encoded in OrthoRep

Due to the lack of 3' polyadenylation on native OrthoRep transcripts, we hypothesized that the synthetic addition of 3' poly(A) tails would improve translation efficiency. Poly(A) tails are critical determinants of mRNA stability and are important for ribosome cycling, both of which contribute to protein expression.[11] In canonical eukaryotic transcription, poly(A) polymerases coordinate *in cis* with RNA polymerase II to tail nascent mRNA transcripts.[12] Therefore, nuclear poly(A) polymerases cannot be simply repurposed for OrthoRep. Instead, Dower and colleagues reported an alternative, synthetic route for polyadenylation wherein a genetically encoded

poly(A) tract and self-cleaving hammerhead ribozyme (RZ) was positioned directly downstream of a GOI and transcribed by an RNA polymerase to mimic a canonical poly(A) tail.[13] We sought to implement this synthetic polyadenylation strategy for OrthoRep.

A genetically encoded poly(A) tail containing at least 48 adenosines significantly increased expression of GOIs encoded on p1 by up to an order of magnitude (Figure 2.1d). Consistent with previous reports that natural yeast mRNAs have a poly(A) length of ~75 nucleotides,[14,15] maximal protein expression was seen with the synthetic 75 adenosine tail (A75-RZ). As expected, the identity of nucleotides play a role in this expression level increase, as a similarly long tract of thymidines fails to achieve a significant increase. As a result, A75-RZ was used for all subsequent experiments. In particular, we combined A75-RZ with the various promoter mutants identified above, and found a consistent 10- to 20-fold increase in protein expression owing to A75-RZ. This was observed across different UCRs and independent experiments (Figure 2.1e), suggesting that promoter strength and transcript poly-adenylation independently affect protein expression, resulting in two means by which expression of GOIs can be tuned.

Promoters and poly(A) tails can be combined to yield a panel of genetic constructs for customized and high expression levels from OrthoRep

By combining different promoter mutants with A75-RZ, we created a panel of OrthoRep expression constructs that span the ~280-fold range achieved with standard constitutive genomic promoters on CEN/ARS plasmids (Figure 2.2a). We identified three promoters, K2O1, K2O10, and the evolved promoter **10B2**, which allow for custom-tailored expression of genes encoded on OrthoRep. In fact, **10B2** used in combination with A75-RZ drives gene expression at a minimum

of 39% (95% CI [27.6, 53.6]) of the level of a strong pTDH3 nuclear promoter when summed across the copy numbers of p1 (~100)[1,2] and the CEN/ARS plasmid (~3)[16] on which pTDH3 was tested (Figure 2.2a, S3a). To ensure that this panel of promoters would be generalizable and stable for experimental evolution, we validated several promoters with a second fluorescent protein (Figure S2.3) and confirmed stability over at least 60 generations (Figure 2.2b).

10B2 and the poly(A) tail also increase expression levels over previous expression systems in the presence of TP-DNAP1-4-2, a highly error-prone DNAP.[1] Error-prone DNAPs are supplied *in trans* during evolution to introduce mutations onto genes encoded on OrthoRep (Figure S2.1b). However, the activity of these engineered error-prone DNAPs is lower than the wt TP-DNAP1, so they sustain lower copy numbers of p1, and consequently, lower protein expression from p1. For example, the copy number of p1 when replicated by TP-DNAP1-4-2 is ~6.[1] We tested our panel of promoters with error-prone TP-DNAP1-4-2 supplied *in trans* to characterize expression levels as well as to validate the panel of promoters for use during rapid evolution. Although the absolute levels of fluorescence were reduced in accordance with the previously seen copy number decrease, expression levels were still able to span over an 18-fold range and achieve at least 68% (95% CI [50.0, 92.6])) the levels of a medium strength pRPL18B nuclear promoter encoded on a standard CEN/ARS plasmid (Figure 2.2a, S2.3b). Since p1 copy number under replication by TP-DNAP1-4-2 is similar to the copy number of a CEN/ARS plasmid, this suggests that the per-copy expression strength for GOIs on OrthoRep can reach pRPL18B levels, which is much higher relative to our previous expression systems for OrthoRep and should be sufficient to achieve selectable function from most GOIs one might wish to evolve.

Discussion and Conclusion

We have engineered a panel of OrthoRep expression cassettes that allows expression levels of OrthoRep-encoded GOIs to span a range comparable to traditional yeast promoters on nuclear plasmids. Interestingly, the mutations that were able to combine to form **10B2** are 24 or 27 bases upstream of the start codon and are adjacent to the conserved ATNTGA sequence (Table S1, highlighted in yellow).[9] This suggests that these mutations may directly influence RNAP recruitment or initiation and that further optimization of RNAP recruitment may be possible as native UCRs are not optimized for high expression. Already, these improvements in expression should extend OrthoRep's application space to include some classic categories of directed evolution experiments that have been conducted using high expression promoters, including yeast surface display[17] and pathway engineering.[18] One consideration when choosing an expression cassette for OrthoRep is the effect of polymerase choice on cumulative expression. Our previous characterization of OrthoRep polymerases has shown that highly error-prone DNAPs reduce p1 copy number ~20 fold, with aggregate expression of GOIs decreasing comparably. However, since p2 remains unaffected by the presence of error-prone TP-DNAP1s,[4] the per-copy expression of GOIs encoded on OrthoRep remain unchanged, thus allowing beneficial mutations on GOIs both to achieve homoplasmy more rapidly as well as to have a proportionally higher impact on fitness. Further, we note that since OrthoRep is a continuous evolution system, the utility of these strong expression systems is mostly at the beginning of an evolution experiment, where initial diversity in the GOI generated through neutral drift with OrthoRep must yield a variant with enough function for selection to act upon. After that, continuous evolution will take hold, and the function and expression of the GOI will further improve as necessary.

These expression cassettes also establish a foundation for using OrthoRep as an orthogonal transcription system. This property of transcriptional orthogonality has important implications for the use of yeast strains in protein production or the construction of synthetic biological circuits. In particular, genomic promoters designated as constitutive are still often affected by cellular regulation, leading to less predictable behavior of synthetic genetic circuits using such parts. However, OrthoRep's orthogonal replication and transcription properties should be insulated from such cellular regulation to a greater degree. Therefore, outside of OrthoRep's use as a continuous evolution system, it may be useful in the establishment of "virtual machine" platforms for encoding synthetic genes and circuits in complex hosts.[19] Finally, OrthoRep's replicative and transcriptional orthogonality should enable future engineering efforts aimed at creating polymerization systems that use only non-natural substrates in all nucleic acid information transfer steps in a GOI's expression,[20] and eventually association with non-natural ribosomes[21] to ultimately realize a full orthogonal central dogma.

References

(1) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C. Scalable Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **2018**, *175*, 1–12. https://doi.org/10.1101/313338.

(2) Ravikumar, A.; Arrieta, A.; Liu, C. C. An Orthogonal DNA Replication System in Yeast. *Nat. Chem. Biol.* **2014**, *10* (3), 175–177. https://doi.org/10.1038/nchembio.1439.

(3) Stark, M. J. R.; Boyd, A.; Milehams, A. J.; Romanosg, A. N. D. M. A. The Plasmid-Encoded Killer System of Kluyveromyces Lactis : A Review. **1990**.

(4) Arzumanyan, G. A.; Gabriel, K. N.; Ravikumar, A.; Javanpour, A. A.; Liu, C. C. Mutually Orthogonal DNA Replication Systems in Vivo. *ACS Synth. Biol.* **2018**, *7* (7), 1722–1729. https://doi.org/10.1021/acssynbio.8b00195.

(5) Kämper, J.; Meinhardt, F.; Gunge, N.; Esser, K. In Vivo Construction of Linear Vectors Based on Killer Plasmids from Kluyveromyces Lactis: Selection of a Nuclear Gene Results in Attachment of Telomeres. *Mol. Cell. Biol.* **1989**, *9* (9), 3931–3937. https://doi.org/10.1128/mcb.9.9.3931-3937.1989.

(6) Stam, J. C.; Kwakman, J.; Meijer, M.; Stuitje, A. R. Efficient Isolation of the Linear DNA Killer Plasmid of Kluyveromyces Lactis: Evidence for Location and Expression in the Cytoplasm and Characterization of Their Terminally Bound Proteins. *Nucleic Acids Res.* **1986**, *14* (17), 6871–6884. https://doi.org/10.1093/nar/14.17.6871.

(7) Schaffrath, R.; Soond, S. M.; Meacock, P. A. The DNA and RNA Polymerase Genes of Yeast Plasmid PGKL2 Are Essential Loci for Plasmid Integrity and Maintenance.

Microbiology **1995**, *141* (10), 2591–2599. https://doi.org/10.1099/13500872-141-10-2591.

(8) Vopalensky, V.; Sykora, M.; Masek, T.; Pospisek, M. Messenger RNAs Transcribed from Yeast Linear Cytoplasmic Plasmids Possess Unconventional 5' and 3' UTRs and Suggest a Novel Mechanism of Translation. *BioRxiv* **2018**.

(9) Schickel, J.; Helmig, C.; Meinhardt, F. Kluyveromyces Lactis Killer System: Analysis of Cytoplasmic Promoters of the Linear Plasmids. *Nucleic Acids Res.* **1996**, *24* (10), 1879–1886. https://doi.org/10.1093/nar/24.10.1879.

(10) Thaithong, S.; Chan, S.; Songsomboon, S.; Wilairat, P.; Seesod, N.; Sueblinwong, T.; Goman, M.; Ridley, R. Pyrimethamine Resistant Mutations in Plasmodium Falciparum. **1992**, *52* (108), 149–157.

(11) Sachs, A. The Role of Poly(A) in the Translation and Stability of MRNA. *Curr. Opin. Cell Biol.* **1990**, *2* (6), 1092–1098. https://doi.org/10.1016/0955-0674(90)90161-7.

(12) Runner, V. M.; Podolny, V.; Buratowski, S. The Rpb4 Subunit of RNA Polymerase II Contributes to Cotranscriptional Recruitment of 3 J Processing Factors □ †. **2008**, *28* (6), 1883–1891. https://doi.org/10.1128/MCB.01714-07.

(13) Dower, K.; Kuperwasser, N.; Merrikh, H.; Rosbash, M. A Synthetic A Tail Rescues Yeast Nuclear Accumulation of a Ribozyme-Terminated Transcript. *RNA* **2004**, *10* (12), 1888–1899. https://doi.org/10.1261/rna.7166704.

(14) Brown, C. E.; Sachs, A. B. Poly(A) Tail Length Control in Saccharomyces Cerevisiae Occurs by Message-Specific Deadenylation . *Mol. Cell. Biol.* **1998**, *18* (11), 6548–6559. https://doi.org/10.1128/mcb.18.11.6548.

(15) Eckmann, C. R.; Rammelt, C.; Wahle, E. Control of Poly(A) Tail Length. *Wiley Interdiscip. Rev. RNA* **2011**, *2* (3), 348–361. https://doi.org/10.1002/wrna.56.

(16) Karim, A. S.; Curran, K. A.; Alper, H. S. Characterization of Plasmid Burden and Copy Number in Saccharomyces Cerevisiae for Optimization of Metabolic Engineering Applications. *FEMS Yeast Res.* **2013**, *13* (1), 107–116. https://doi.org/10.1111/1567-1364.12016.

(17) Boder, E. T.; Wittrup, K. D. Yeast Surface Display for Screening Combinatorial Polypeptide Libraries. *Nat. Biotechnol.* **1997**, *15*, 553–557.

(18) Lee, S. M.; Jellison, T.; Alper, H. S. Directed Evolution of Xylose Isomerase for Improved Xylose Catabolism and Fermentation in the Yeast Saccharomyces Cerevisiae. *Appl. Environ. Microbiol.* **2012**, *78* (16), 5708–5716. https://doi.org/10.1128/AEM.01419-12.

(19) Liu, C. C.; Jewett, M. C.; Chin, J. W.; Voigt, C. A. Toward an Orthogonal Central Dogma. *Nature Chemical Biology*. Nature Publishing Group 2018, pp 103–106. https://doi.org/10.1038/nchembio.2554.

(20) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science (80-.).* **2012**, *336* (6079), 341–344. https://doi.org/10.5061/dryad.5t110.Supplementary.

(21) Orelle, C.; Carlson, E. D.; Szal, T.; Florin, T.; Jewett, M. C.; Mankin, A. S. Protein Synthesis by Ribosomes with Tethered Subunits. *Nature* **2015**, *524* (7563), 119–124.

https://doi.org/10.1038/nature14862.

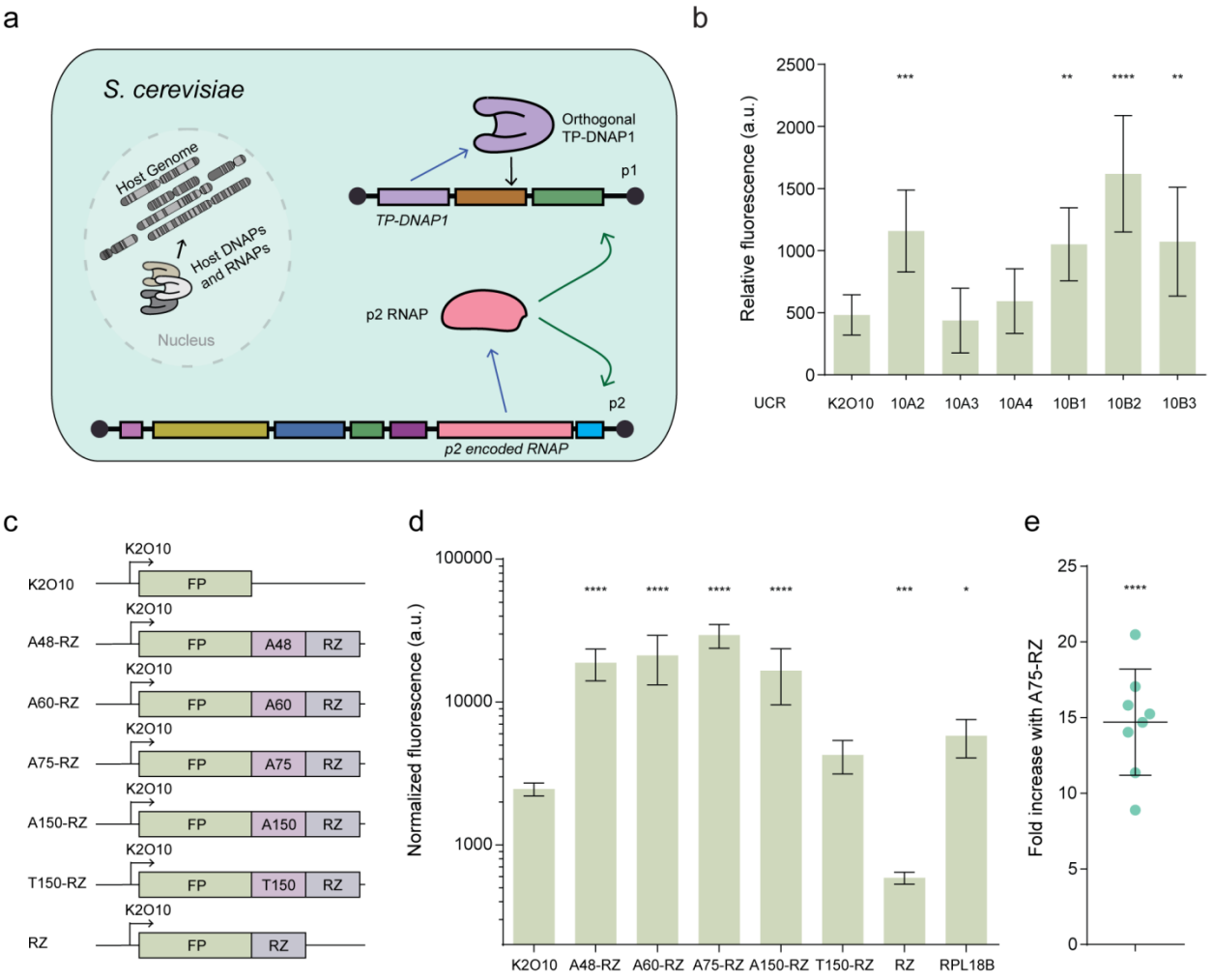


Figure 2.1. A genetically encoded poly(A) tail increases expression of genes encoded on OrthoRep. (**a**) A schematic of OrthoRep, an orthogonal transcription/translation system in *Saccharyomyces cerevisiae*. The natural p1 plasmid encodes TP-DNAP1 for its own replication and p2 encodes the RNA polymerase for transcribing genes encoded on both p1 and p2. **(b)** Relative fluorescence of single mutants (10A prefix) or combination mutations (10B prefix) in K2O10 UCR. Statistical comparisons are vs. wt K2O10 by one-way ANOVA. (**c**) A graphical depiction of the constructs used to assay a genetically encoded 3' poly(A) tail. K2O10, the UCR of p2ORF10; FP, a fluorescent protein used as a reporter; A48, A60, etc., a tract of 48

adenosines, 60 adenosines, etc.; RZ, a 5' self-cleaving hammerhead ribozyme. (**d**) OD_{600}-normalized fluorescence of constructs shown in **c**. RPL18B denotes the use of pRPL18B on a CEN/ARS plasmid as a control for nuclear expression. N = 2-4. Statistical comparisons are vs. wt K2O10 by one-way ANOVA (**e**) Ratios of fluorescence with and without a poly(A) tail across 8 independent expression cassettes and experiments. Statistical comparison is vs. no difference by one-way t-test. Bars and errors denote mean and standard deviation or range. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

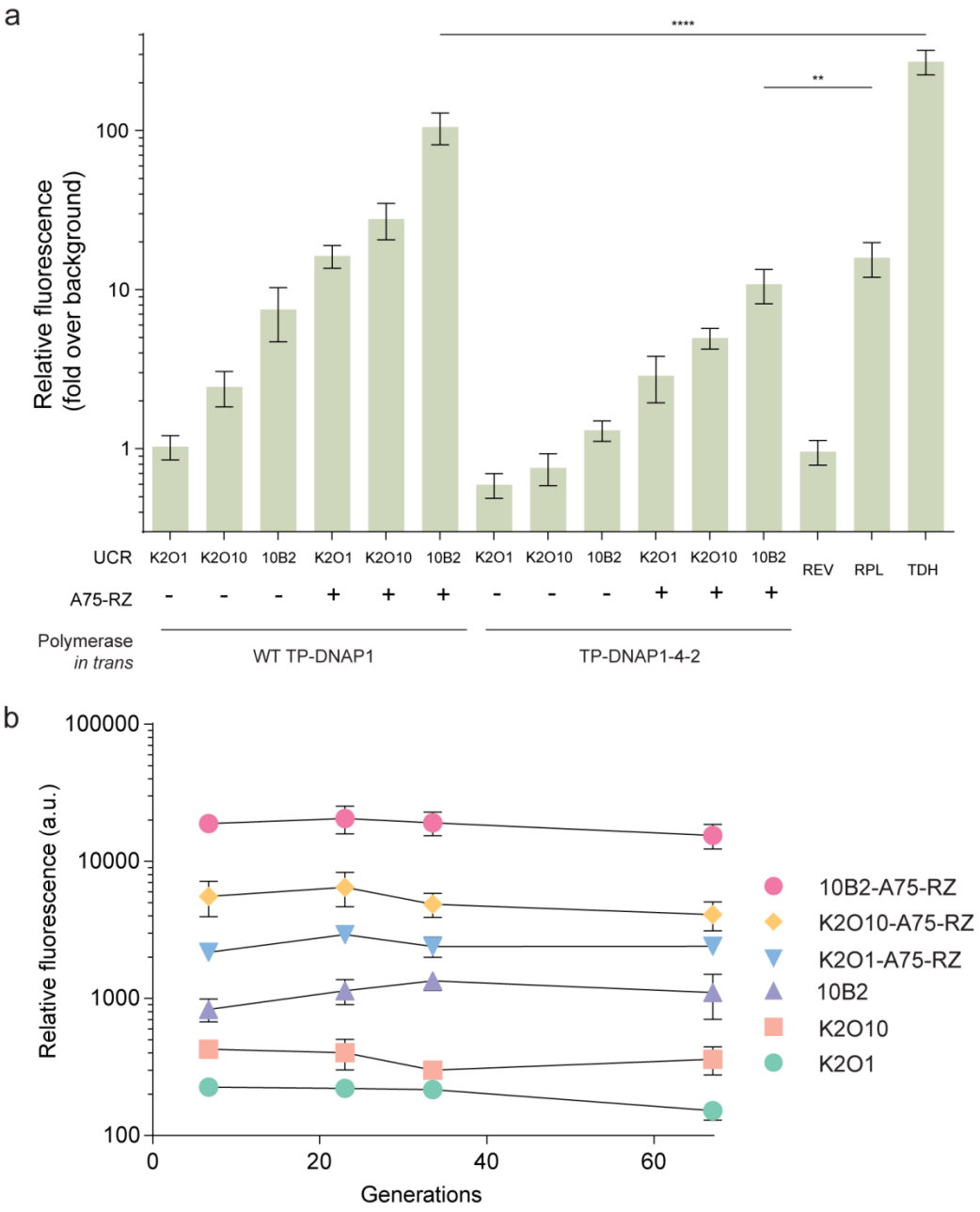


Figure 2.2. A collection of OrthoRep expression cassettes spans a range of expression comparable to nuclear promoters. (**a**) Fluorescence fold change over background autofluorescence of three OrthoRep UCRs (K2O1, K2O10, and **10B2**) with or without a genetically encoded poly(A) tract and ribozyme (A75-RZ) in the context of wild-type (WT TP-

DNAP1) or an error-prone polymerase (TP-DNAP1-4-2). REV, RPL, and TDH denote the use of pREV1, pRPL18B, or pTDH3 on nuclear CEN/ARS plasmids to drive the fluorescent reporter to serve as expression controls. Select comparisons are shown. ** $p < 0.01$, **** $p < 0.0001$ by one-way ANOVA. (**b**) Relative fluorescence measurements of OrthoRep constructs over 67 generations. N = 6-8. Bars and errors denote mean and standard deviation.

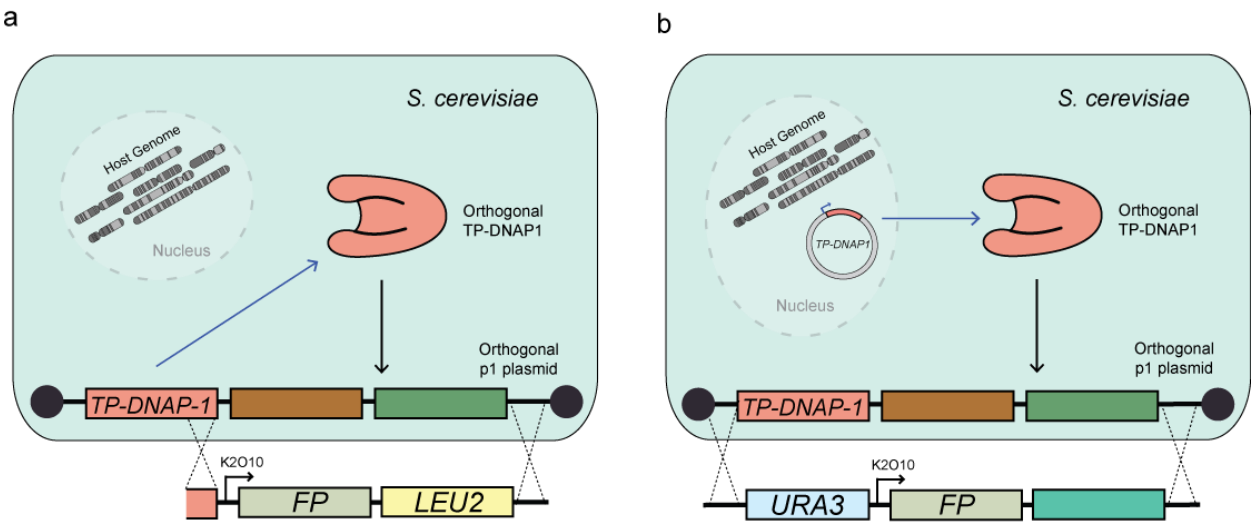


Figure S2.1. Integration cassettes to rapidly assay protein expression levels. (**a**) An integration cassette that does not disrupt the ORF encoding wt *TP-DNAP1* on p1. A selection marker, *LEU2*, was used to select for the correct integration of the fluorescent protein (FP). No DNAP is supplied *in trans* as the recombinant p1 still encodes its own DNA polymerase. (**b**) An integration cassette that disrupts *TP-DNAP1* requires *in trans* expression of TP-DNAP1. A selection marker, *URA3*, was used to select for the corrct integration of the FP. A CEN/ARS expression vector with pREV1 driving either wt TP-DNAP1 or TP-DNAP1-4-2 is used to complement a *TP-DNAP1* disruption on p1.

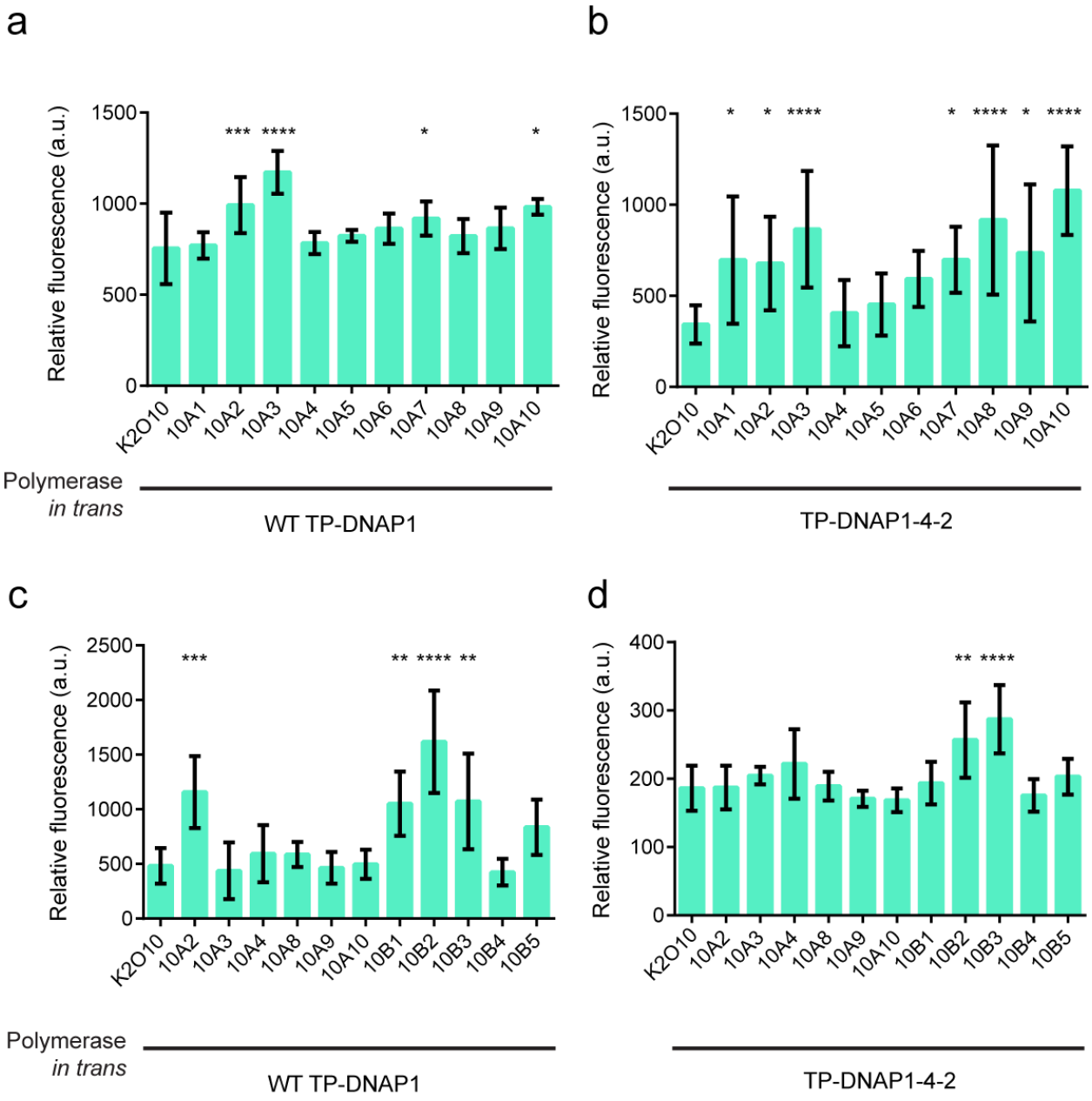


Figure S2.2. Individual testing of K2O10 UCR mutants resulted in the discovery of promoter **10B2**. (**a**) Single mutations in K2O10 UCR tested in the context of wt TP-DNAP1. (**b**) Single mutations in K2O10 UCR tested in the context of TP-DNAP1-4-2, an error-prone TP-DNAP1 used in previous evolution studies. (**c**) Single and combination mutations in K2O10 UCR tested in the context of wt TP-DNAP1. (**d**) Single and combination mutations in K2O10 UCR tested in

the context of TP-DNAP1-4-2. N = 6-12. Bars and errors denote mean and standard deviation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ vs. wt K2O10 by one-way ANOVA.

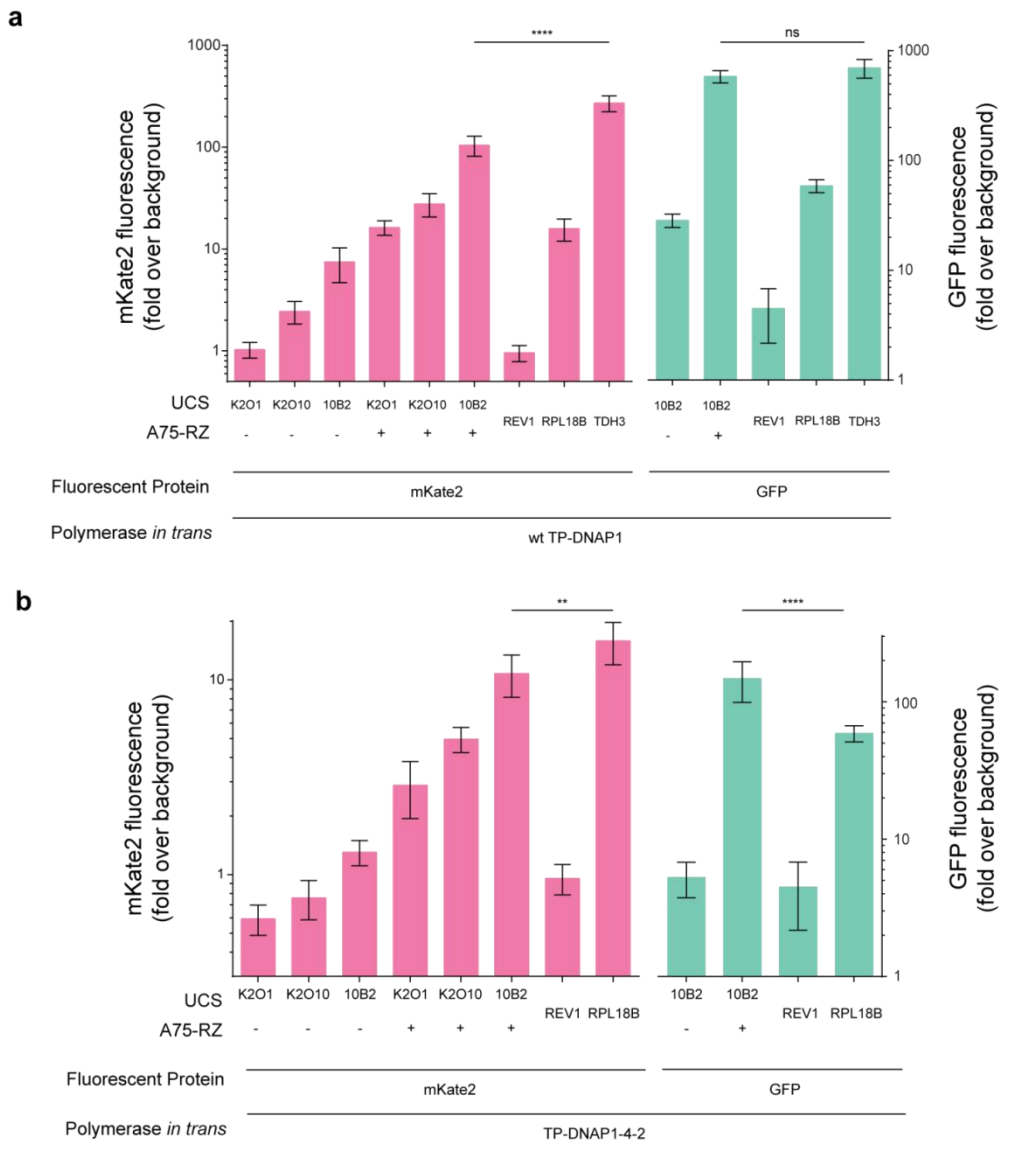


Figure S2.3. Expression of genes encoded on OrthoRep is consistent across two fluorescent reporters. Our highest expression mutant, **10B2**, was confirmed with a second fluorescent protein and both with and without a poly(A) tail (A75-RZ) in the context of **(a)** wt TP-DNAP1 or **(b)** a highly error-prone DNAP, TP-DNAP1-4-2. A fluorescent protein driven by pREV1, pRPL18B, or pTDH3 on nuclear CEN/ARS plasmids served as controls for expression levels. N = 6-8. Bars

and errors denote mean and standard deviation. Select comparisons are shown. ns: no significance, ** $p < 0.01$, **** $p < 0.0001$

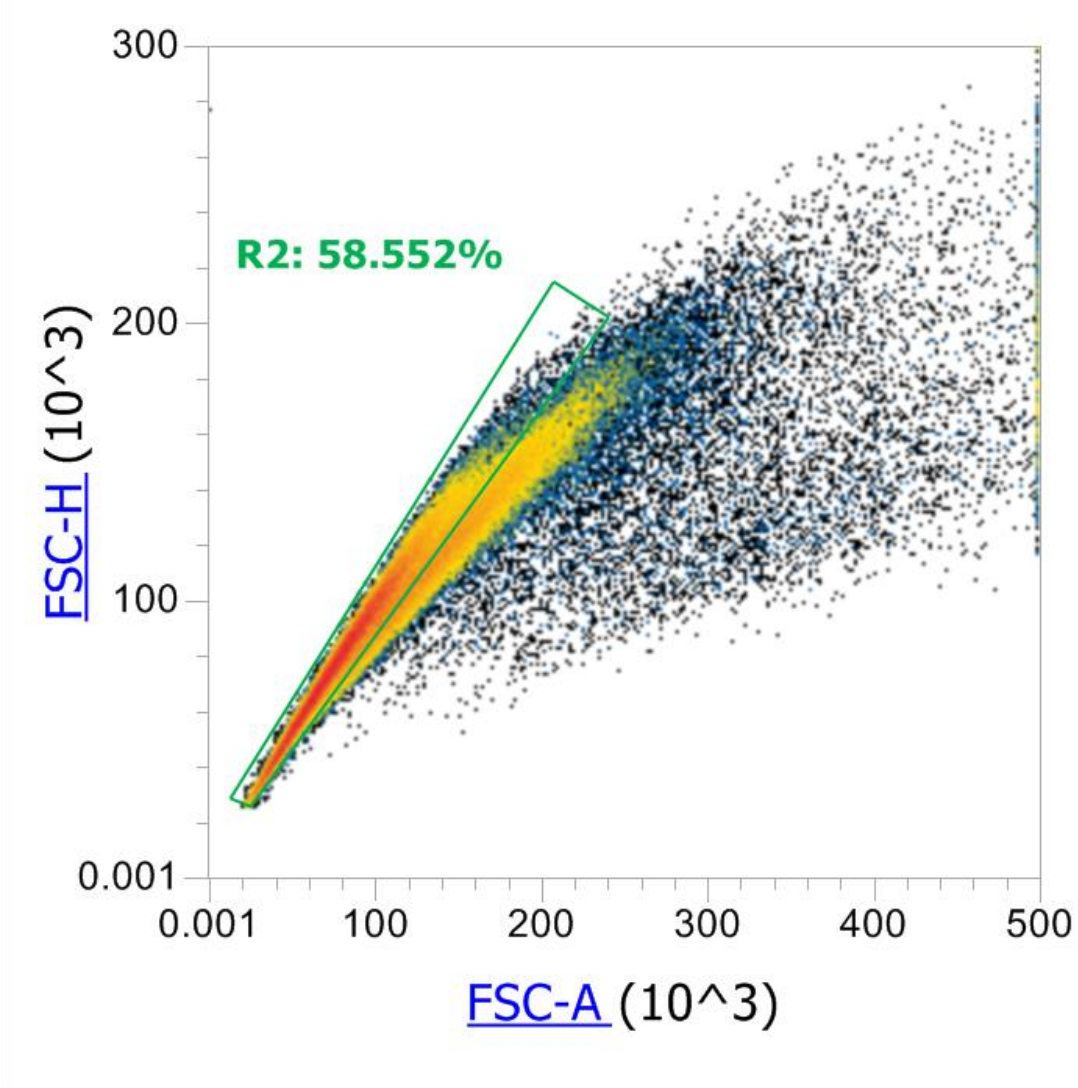


Figure S2.4. Illustration of gating scheme used for all fluorescence measurements on the flow cytometer. A polygon gate (R2) was constructed in the forward scatter height (FSC-H) versus forward scatter area (FSC-A) to remove abnormally large cells and clumps of cells. For each sample, the reported fluorescence value represents the mean fluorescence of gated cells.

UCR	Sequence
K2O1	tttttcgccggagtcaattaggtcatacttttctatataatccaaatcccaaaaatcaattgaatgattcttaatatgatttaatagtttatgattataa
K2O10	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgtagaaatatatgataagctcatagacatgtaaa
10A1	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgtagaaatatatgataagctcacagacatgtaaa
10A2	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgtagaaatatgtgataagctcatagacatgtaaa
10A3	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgtagaaacatatgataagctcatagacatgtaaa
10A4	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgtaggaatatatgataagctcatagacatgtaaa
10A5	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgtggaaatatatgataagctcatagacatgtaaa
10A6	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgcagaaatatatgataagctcatagacatgtaaa
10A7	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagagaaatgtagaaatatatgataagctcatagacatgtaaa
10A8	gatgacctatacataggaagatctatagaaacaaaaagattaataacttccaaatatcagaaaaatgtagaaatatatgataagctcatagacatgtaaa
10A9	gatgacctatacataggaagatctatagaaacaaaaagattaataaccttcaaatatcagaaaaatgtagaaatatatgataagctcatagacatgtaaa
10A10	gatgacctatacataggaagatctacagaaacaaaaagattaataactttcaaatatcagaaaaatgtagaaatatatgataagctcatagacatgtaaa
10B1	gatgacctatacataggaagatctacagaaacaaaaagattaataacttccaaatatcagaaaaatgtagaaacatatgataagctcatagacatgtaaa
10B2	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgtagaaacatgtgataagctcatagacatgtaaa
10B3	gatgacctatacataggaagatctatagaaacaaaaagattaataactttcaaatatcagaaaaatgtaggaacatgtgataagctcatagacatgtaaa
10B4	gatgacctatacataggaagatctacagaaacaaaaagattaataacttccaaatatcagaaaaatgtagaaatatatgataagctcatagacatgtaaa
10B5	gatgacctatacataggaagatctacagaaacaaaaagattaataacctccaaatatcagaaaaatgtagaaacatatgataagctcatagacatgtaaa

Table S2.1. List of UCRs used in this report. Highlight indicates the conserved ATNTGA sequence present in all p1/p2 UCRs.

Mutations in K2O10 variants that differ from wt K2O10 are underlined in red.

Chapter 3. Automated continuous evolution of proteins *in vivo*

Ziwei Zhong[1,*], Brandon G. Wong[2,3,*], Arjun Ravikumar[1,2,3], Garri A. Arzumanyan[1], Ahmad S. Khalil[2,3,4], and Chang C. Liu[1,5,6]

[1]Department of Biomedical Engineering, University of California, Irvine, California, USA.

[2]Biological Design Center, Boston University, Boston, Massachusetts, USA.

[3]Department of Biomedical Engineering, Boston University, Boston, Massachusetts, USA.

[4]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA.

[5]Department of Chemistry, University of California, Irvine, California, USA.

[6]Department of Molecular Biology & Biochemistry, University of California, Irvine, California, USA.

*These authors contributed equally to this work.

Introduction

Continuous evolution has emerged as a powerful paradigm for the evolution of proteins and enzymes[1–4] towards challenging functions.[5,6] In contrast to classical directed evolution approaches that rely on stepwise rounds of *ex vivo* mutagenesis, transformation into cells, and selection,[7] continuous evolution systems achieve rapid diversification and functional selection autonomously, often through *in vivo* targeted mutagenesis systems (Figure 3.1a).[2,7-14] The result is a mode of directed evolution that requires only the basic culturing of cells, in theory, enabling extensive speed, scale, and depth in evolutionary search.[3] In practice, however, developing a continuous evolution method that realizes all three properties has been challenging. Recently, our groups made two advances, OrthoRep and eVOLVER, that can pair to achieve continuous evolution at significant speed, scale, and depth.

OrthoRep is an engineered genetic system for continuous *in vivo* targeted mutagenesis of genes of interest (GOIs).[2,14] OrthoRep uses a highly error-prone, orthogonal DNA polymerase-plasmid pair in yeast that replicates GOIs at a mutation rate of 10^{-5} substitutions per base (spb) without increasing the genomic mutation rate of 10^{-10} spb (Figure 3.1a). This ~100,000-fold increase in the mutation rate of GOIs drives their accelerated evolution (speed). Because the OrthoRep system functions entirely *in vivo* and culturing yeast is straightforward, independent GOI evolution experiments can be carried out in high-throughput (scale). In addition, long multi-mutation pathways can be traversed using OrthoRep, owing to the durability of mutagenesis over many generations (depth). However, to practically realize depth in evolutionary search, *in vivo* mutagenesis with OrthoRep must be coupled with a functional selection that can be tuned over the course of a continuous evolution experiment. This tuning is necessary to precisely and efficiently guide populations to the desired evolutionary search depth. For example, evolution of

novel functions requiring long mutational trajectories may demand frequent modification of selection conditions in order to maintain strong selection,[5,6,15] guide evolution through strategic intermediate functions,[1,6] or impose periods of neutral drift or alternating selection to promote crossing of fitness valleys (Figure 3.1c).[16,17] Yet, selection schedules cannot be determined *a priori* as the generation of beneficial mutations is a fundamentally stochastic process. Therefore, selection schedules should be adjusted dynamically based on how populations adapt, rendering manual implementation of continuous evolution experiments onerous. Further, each functional selection demands its own selection schedule, necessitating empirical probing of conditions that are appropriately stringent to generate selection pressures, yet sufficiently lenient to allow for mutational accumulation. Previous continuous evolution campaigns approached the challenge of optimizing selection schedules by either limiting the number of parallel evolution experiments being conducted so that selection can be manually tuned on the fly,[1,5] or by setting a fixed but conservative selection schedule to buffer against variations in adaptation rate across a large number of replicate experiments.[2] However, even with conservative selection schedules, a proportion of replicates in high-throughput evolution studies went extinct when the rate of selection stringency increase outpaced the rate of adaption.[2] Indeed, streamlining selection schedules for experimental evolution remains an open challenge.[18–20]

To address this challenge, we turned to eVOLVER. eVOLVER is a versatile continuous culture platform that enables multiparameter control of growth and selection conditions across independent microbial cultures (Figure 3.1b).[21] eVOLVER's flexible hardware and software permit development of "algorithmic selection routines" that apply selective pressures based on real-time monitoring and feedback from culture growth characteristics. Additionally,

eVOLVER's robust framework ensures experimental durability over long timeframes, and its unique scalable design allows independent control over tens to hundreds of cultures. Combining OrthoRep and eVOLVER should therefore enable continuous evolution with speed, depth, and scale.

Here we describe this pairing of OrthoRep with eVOLVER to achieve Automated Continuous Evolution (ACE) (Figure 3.1c). By implementing a closed-loop feedback routine that dynamically adjusts the strength of selection for a desired function in response to growth rate changes of yeast populations diversifying a GOI on OrthoRep, we demonstrate completely automated continuous evolution over extended periods of time without manual intervention. To illustrate the performance and utility of ACE, we describe its application in two model protein evolution experiments, one yielding drug-resistant *Plasmodium falciparum* dihydrofolate reductases (*Pf*DHFRs) and the other yielding variants of the thermophilic HisA enzyme from *Thermotoga maritima* (*Tm*HisA) that operate well in mesophilic yeast hosts.

Results and Discussion

Establishment of ACE. To establish ACE, we first reconfigured eVOLVER Smart Sleeves[21] so that each culture vial receives two media inputs: (1) 'no selection' base media (*e.g.* media without drug or with the maximum concentration of nutrient in our cases) and (2) 'full selection' media (*e.g.* media with the maximum concentration of drug or without nutrient in our cases). Using eVOLVER software calculations, selection strength can be dynamically tuned by altering the ratios of the two media inputs as cultures are diluted over time (Figure 3.1b,c). We then

implemented a closed-loop control system that seeks to achieve and maintain a target culture growth rate by dynamically adjusting selection strength. Briefly, culture growth rate is continuously measured based on real-time recordings of optical density (OD), and a proportional-integral-derivative (PID) control algorithm[22] is used to determine the percentage of full selection media to add to the culture in order to minimize error between the actual growth rate and a target growth rate (or setpoint) (see **Methods**). Although simpler feedback algorithms[18,19] have been previously used in microbial evolution experiments, these resulted in growth rate oscillations or excessive overshooting in our experiments, frequently driving cultures to extinction (Figure S3.1).

Evolution of *Pf*DHFR resistance using ACE. To validate ACE, we first repeated a continuous evolution experiment that we previously conducted using manual serial passaging. Specifically, we evolved *Plasmodium falciparum* dihydrofolate reductase (*Pf*DHFR) to develop drug resistance to the antimalarial drug, pyrimethamine, by encoding *Pf*DHFR on OrthoRep in a yeast strain that relies on *Pf*DHFR activity for survival (Figure 3.2).[2] We determined appropriate PID constants to tune the concentration of pyrimethamine (Figure 3.2b, S3.2) and keep the measured growth rate of cells at a target growth rate (Figure 3.2a, setpoint = dashed black line). This program forced cells to continuously experience a strong selection pressure imposed by pyrimethamine, which resulted in the rapid evolution of *Pf*DHFR resistance (Figure 3.2c). We observed that after ~550 hours (~100 generations) of continuous hands-free operation of ACE, five out of six replicates adapted to 3 mM pyrimethamine, the highest concentration of pyrimethamine soluble in liquid media (Figure 3.2b, S3.3). ACE maintained cultures near the target growth rate over the entire course of the experiment (Figure 3.2a,b), demonstrating the

effectiveness of the control loop. In contrast to the use of a fixed selection schedule[2] or simpler control algorithms for selection (Figure S3.1) that resulted in occasional extinction caused by too-rapid increases in pyrimethamine concentration, all six ACE experiments reliably adapted to yield multi-mutation pyrimethamine-resistant *Pf*DHFR variants. Validating our method, we found that populations converged on strong resistance mutations in *Pf*DHFR – C50R, D54N, Y57H, C59R, C59Y, and S108N – as observed and characterized previously[2] (Figure 3.2c). Additionally, the monotonically increasing pyrimethamine concentrations we observed for most replicates (Figure 3.2b) are consistent with step-wise fixation of beneficial mutations expected for the evolution of *Pf*DHFR resistance under strong selection.[2,21] Upon examination of one of the evolution replicates (V3), we noted a drop in pyrimethamine at ~200 hours, likely due to a mechanical error. Nevertheless, the selection self-adjusted, resulting in recovery in growth and demonstration of ACE's control algorithm to robustly maintain selection. Finally, ACE demonstrated a substantial increase in speed over our previous evolution campaign performed by manual passaging; with ACE, culture growth rates in 5/6 vials stabilized at the maximum pyrimethamine concentration after ~550 hours, which is over 200 hours faster than for the manual evolution campaign done with serial passaging.[2] Collectively, these results validate the ACE system and highlight its ability to enable reliable and rapid continuous evolution of proteins.

Evolution of *Tm*HisA activity using ACE. We next applied ACE to evolve the thermophilic *Thermotoga maritima* HisA enzyme (*Tm*HisA) to function in *Saccharomyces cerevisiae* at mesophilic temperatures. *Tm*HisA, an ortholog of *S. cerevisiae* HIS6, catalyzes the isomerization of ProFAR to PRFAR in the biosynthesis of histidine. However, *Tm*HisA does not effectively

complement a *his6* deletion in yeast when expressed from a medium-strength yeast promoter (Figure 3.3), likely due to the different temperature niches of *S. cerevisiae* and *T. maritima* (30˚C and 80˚C, respectively). We reasoned that ACE could readily drive the evolution of *Tm*HisA to function in yeast Δ*his6* strains by selecting for growth in media lacking histidine. This evolution serves as a valuable test of the capabilities of ACE for two reasons. First, adapting enzymes from non-model thermophiles to function in model mesophiles is useful for industrial biotechnology whose infrastructure is designed around model organisms like yeast and bacteria. Second, in contrast to drug resistance in *Pf*DHFR, which is driven by a small number of large effect mutations,[2] we reasoned that temperature and host adaptation of enzyme activity would involve a large number of small effect mutations, leading to a more complex fitness landscape. This would act as a more demanding test of ACE's ability to achieve precise feedback-control during selection.

We encoded *Tm*HisA on OrthoRep in a Δ*his6* strain and carried out ACE selection in four independent replicates for a total of 600 hours (~100 generations) (Figure 3.3, S3.4). At the beginning of the experiment, there was no detectable growth in the absence of histidine. At the end of the experiment, all four replicates successfully adapted to media lacking histidine. To confirm that *Tm*HisA evolution was responsible for the observed adaptation, *Tm*HisA variants were isolated from OrthoRep and characterized for their ability to complement a *his6* deletion in fresh yeast strains. Indeed, the evolved *Tm*HisA variants we sampled were able to support growth in media lacking histidine in contrast to wild-type *Tm*HisA (Figure 3.3c). Consistent with a model of a more complex fitness landscape, growth rate traces for the four replicate cultures were noisier (Figure 3.3a) than those of *Pf*DHFR (Figure 3.2a), full adaptation occurred only

after a long period of neutral drift (hours ~100-500 in Figure 3.3b), and the sequences of independently evolved *Tm*HisAs were diverse (Figure 3.3d, Table S3.1). Nevertheless, ACE was able to autonomously adapt *Tm*HisA in all four replicates within 120 fewer hours than manual passaging experiments (unpublished results). Sequencing of sampled clones revealed *Tm*HisA variants harboring between 6 and 15 mutations (Table S3.1), again demonstrating the durability of ACE in carrying out long evolutionary searches to discover high-activity multi-mutation enzyme variants.

Conclusion

In summary, we have developed a fully automated, *in vivo* continuous evolution setup termed ACE that couples OrthoRep-driven continuous mutagenesis and eVOLVER-enabled programmable selection. We demonstrated the evolution of drug resistance in *Pf*DHFR and mesophilic operation of *Tm*HisA, showcasing the ability of ACE to individually control selection schedules in multi-replicate GOI evolution experiments based on real-time measures of adaptation. We further validate the value and generalizability of a PID controlled selection scheme that successfully drives two mechanistically different selections. The result is a system that offers unprecedented speed, depth, and scalability for conducting evolutionary campaigns to achieve ambitious protein functions.

ACE paves the way for an array of complex evolution experiments that can advance both basic and applied protein and enzyme evolution. For example, eVOLVER can be used to program multidimensional selection gradients across OrthoRep experiments, test the effects of

selection strength or different population sizes (and beneficial mutation supply) on the outcomes of adaptation, or explore the relationship between timescales of drift and adaptation. Real-time feedback on growth metrics to adjust selection stringency can ensure that every evolving population is being constantly challenged appropriately or allowed to drift, which is especially relevant when evolving biomolecules with rugged fitness landscapes where predefined selection strategies are prone to driving populations to extinction or local fitness maxima. In the future, many other algorithmic selection routines may be implemented with ACE to more efficiently and intelligently navigate fitness landscapes. For example, machine learning algorithms can take the outcomes of replicate evolution experiments carried out under different selection schedules to train ACE selection programs themselves. Finally, the automated, open-source nature of ACE is well-suited for integration with other open-source hardware and wetware tools to create larger automation pipelines. Overall, we foresee ACE as an enabling platform for rapid, deep, and scalable continuous GOI evolution for applied protein engineering and studying the fundamentals of protein evolution.

Methods

Cloning. All plasmids used in this study are listed in Table S2. Plasmids were cloned using either restriction enzymes if compatible sites were available or using Gibson cloning[23] with 20-40 bps of overlap. Primers and gBlocks were ordered from IDT Technologies. Enzymes for PCR and cloning were purchased from NEB. Plasmids were cloned into either Top10 *E. coli* cells from Thermo Fisher or SS320 *E. coli* from Lucigen.

Yeast transformation and DNA extraction. All yeast strains used in this study are listed in Table S3. Yeast transformations were done with roughly 100 ng – 1 □g of plasmid or donor DNA via the Gietz high-efficiency transformation method.[24] For integration of genes onto the orthogonal plasmid (pGKL1), cassettes were linearized with ScaI and subsequently transformed as described previously.[2,14] Standard preparations of YPD and drop-out synthetic media were obtained from US Biological. When necessary, the following were supplemented at their respective concentrations: 5-FOA at 1 mg/mL, G418 at 400 µg/mL, and Nourseothricin at 200 µg/mL. Yeast DNA extraction of orthogonal plasmids were done as previously reported.[2,14]

eVOLVER feedback control configuration. ACE experiments were performed using the previously described eVOLVER continuous culture system,[21] modified to enable an additional media input into each culture. Specifically, each vessel consists of three connected pumps (two input, one efflux) and are actuated programmatically to implement a so-called "morbidostat" algorithm where the selection stringency is adjusted to maintain a particular rate of cell growth. The custom script of eVOLVER (custom_script.py) was extensively modified to change the behavior of eVOLVER from the default turbidostat to a morbidostat. Briefly, in the new morbidostat mode, eVOLVER dilutes the growing cultures after a defined time, which we set to an hour. At the time of dilution, the growth rate since the last dilution is calculated by fitting the OD measurements to an exponential equation $y = A \cdot e^{Bx}$ where B is the growth rate. Using the current and historical growth rate, a dilution parameter, $r(t)$ was calculated as described below to dilute the morbidostat. The morbidostat algorithm and eVOLVER experimental code are written in Python and included in the supplemental files.

The efflux pump for each vessel is actuated whenever either of the influx pumps are triggered and stay ON for an additional 5 seconds. Therefore, the volume of the culture vessel is determined by the length of the efflux straw and estimated to be at 30 mL. The flow rate of each media input was individually calibrated for accurate metering of drug or nutrient into the culture.

Before each experiment, 40 mL borosilicate glass vessels (Chemglass), stir bars (Fisher), and fluidic straws were assembled and autoclaved. Fluidic lines were sterilized by flushing with 10% bleach and 70% ethanol before use. Culture vessel assemblies were connected to fluidic lines after sterilization and slotted into an eVOLVER Smart Sleeve for monitoring of OD and control of temperature and stir rate.

PID algorithm development and tuning. To control the rate of dilution, we used the following equation to determine the percentage of selection media to add:

$$r(t) = K_P e(t) + K_I \int_{\tau}^{t} e(t)dt + K_D \frac{de(t)}{dt} + K_O$$

where K_P, K_I, K_D, and K_O, are empirically determined constant multipliers of proportional, integral, derivative, and offset terms, and $e(t)$ is the difference between the actual growth rate and the target growth rate. To estimate K_P, K_I, K_D, and K_O, we used the the Ziegler-Nichols method[25] for initially tuning the parameters with the pre-evolution strain, ZZ-Y323. K_I and K_D were first set to zero and K_P was increased until regular oscillations in growth rate were observed (Figure S3.2). This resulted in a $K_P = 4$.

Using the parameters obtained during the oscillation and the Ziegler-Nichols estimation:

$$K_P = X_{osc} * 0.6 = 0.2 * 0.6 = 0.12$$

$$K_I = \frac{1}{T_{osc} * 0.5} = \frac{1}{5 * 0.5} = 0.4$$

$$K_D = T_{osc} * 0.125 = 5 * 0.125 = 0.625$$

These initial values were empirically tuned to achieve the final values of $K_P = 0.07$, $K_I = 0.05$, $K_D = 0.2$ and $K_O = 0$.

These constants were then used to calculate $r(t)$ at any given point during evolution. $r(t)$ would then be used to determine the ratios of media to add during each dilution step by controlling the pump runtime. For example, if an $r(t) = 0.25$ was determined with a pump runtime of 5 seconds, the pump for the base media would run for $[1 - r(t)] * 5 \text{ seconds} = 3.75 \text{ seconds}$ while the pump for the full selection media would run for $r(t) * 5 \text{ seconds} = 1.25 \text{ seconds}$.

The integral error ($\int_{\tau}^{t} e(t)dt$) was reset at every instance the proportional error ($e(t)$) became negative, and the offset (K_O) was updated to equal $r(t)$ at that time. This was done to allow the PID controller to be more sensitive to the integral error and to avoid the bias that would result from the initial conditions having minimal selection pressure.

***Pf*DHFR evolution.** eVOLVER was set to morbidostat mode with the PID settings described above, a target doubling time of 8 hours, and one dilution step per hour. A culture of ZZ-Y435 was grown to saturation in SC-HW and then inoculated 1:50 in eVOLVER vials. SC-HW served as the base media, while SC-HW + 3 mM pyrimethamine served as the full selection media. (3mM was previously determined as the maximum soluble concentration of pyrimethamine in media.[2]) After inoculation, the eVOLVER PID script was initiated and evolution commenced.

During evolution, the only user intervention was media exchange and periodic sampling of cultures. After 725 hrs, all cultures achieved growth rates near wild-type levels in the full selection condition (Figure S3.3), so the experiment was stopped and cultures were frozen in glycerol stocks.

***Tm*HisA evolution.** eVOLVER was set to morbidostat mode with the PID settings described above, a target doubling time of 8 hours, and one dilution step per hour. A culture of ZZ-Y323 was grown to saturation in SC-UL and then inoculated 1:50 in eVOLVER vials. SC-ULH + 7.76 mg/L (50 μM) histidine served as the base media, while SC-ULH served as the full selection media. After inoculation, the eVOLVER PID script was initiated and evolution commenced. During evolution, the only user intervention was media exchange and periodic sampling of cultures. After 715 hrs, all cultures achieved growth rates near wild-type levels in the full selection condition (Figure S3.4), so the experiment was stopped and cultures were frozen in glycerol stocks.

Bulk DNA sequencing and characterization. Final evolution timepoints of *Pf*DHFR and *Tm*HisA were regrown in SC-HW and SC-ULH media, respectively, from glycerol stocks. The orthogonal plasmids encoding evolved *Pf*DHFR or *Tm*HisA were extracted from the bulk cultures as described above, PCR amplified, and sequenced via Sanger sequencing. Mutation frequencies were calculated from Sanger sequencing files with QSVanalyzer as previously described.[2] However, V1 from *Tm*HisA evolution could not be revived from the glycerol stock due to a stocking mistake and was not included for bulk DNA sequencing.

***Tm*HisA isolated mutant cloning.** Final evolution time-points of *Tm*HisA were streaked onto SC-ULH solid media. Individual colonies were regrown in SC-ULH media and the orthogonal plasmid DNA was extracted from the cultures as described above. The evolved *Tm*HisA sequences were sequenced and cloned into a nuclear *CEN6/ARS4* expression vector under control of the pRPL18B promoter and with the *LEU2* selection marker. Since each colony can have different *Tm*HisA mutants due to the multicopy nature of the orthogonal plasmid in OrthoRep, the cloned plasmids were sequenced again to determine the exact mutant of *Tm*HisA being characterized. The resulting plasmids were transformed into ZZ-Y354, which lacks *his6*, for growth rate measurements.

***Tm*HisA growth rate measurements.** Yeast strains containing each *Tm*HisA mutant, WT *Tm*HisA, *S. cerevisiae* HIS6, or none of the above expressed from a nuclear plasmid were grown to saturation in SC-L and diluted 1:100 in SC-LH. Three 100 uL replicates of each strain were placed into a 96 well clear-bottom tray, sealed, and grown at 30 °C. Cultures were continuously shaken and OD_{600} was measured every 30 minutes automatically for 24 hours (Tecan Infinite M200 Pro) according to a previously described protocol.[26] A custom MATLAB script (growthassayV3.m), included in supplemental files, was used to calculate growth rates from raw OD_{600} data. The script carries out a logarithmic transformation of the OD_{600} data. The linear region of the transformed data as a function of time corresponds to log phase growth. A sliding window approach is used to find and fit this linear region in order to calculate the doubling time during log phase growth. This doubling time (T) is converted to the continuous growth rate plotted in Figure 3.3c by the formula $ln(2)/T$.

Statistical analysis. Statistical analysis was done using GraphPad Prism and one-way ANOVA with multiple comparisons versus wild-type *Tm*HisA and corrected for multiple comparisons. Results are reported at $p < 0.05$.

References

(1) Esvelt, K. M., Carlson, J. C., Liu, D. R. (2011) A System for the Continuous Directed Evolution of Biomolecules. *Nature, 472*, 499–503.

(2) Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., Liu, C. C. (2018) Scalable , Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds Resource Scalable , Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell, 175*, 1946–1957.

(3) Wellner, A., Ravikumar, A., Liu, C. C. (accepted) Continuous Evolution of Proteins in vivo. In *Protein Engineering* (Zhao, H., Ed.), Wiley-VCH.

(4) Fact, C., Tan, Z. L., Zheng, X., Wu, Y., Jian, X., Xing, X., Zhang, C. (2019) In Vivo Continuous Evolution of Metabolic Pathways for Chemical Production. *Microb. Cell Fact., 18*, 1–19.

(5) Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., Turner, K. H., et al. (2016) Continuous Evolution of Bacillus Thuringiensis Toxins Overcomes Insect Resistance. *Nature, 533*, 58–63.

(6) Packer, M. S., Rees, H. A., Liu, D. R. (2017) Phage-Assisted Continuous Evolution of Proteases with Altered Substrate Specificity. *Nat. Commun., 8*, 956–966.

(7) Packer, M. S., Liu, D. R. (2015) Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet., 16*, 379–394.

(8) Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V., Dueber, J. E. (2018)

CRISPR-Guided DNA Polymerases Enable Diversification of All Nucleotides in a Tunable Window. *Nature*, *560*, 248–252.

(9) Finney-Manchester, S. P., Maheshri, N. (2013) Harnessing Mutagenic Homologous Recombination for Targeted Mutagenesis in Vivo by TaGTEAM. *Nucleic Acids Res.*, *41*, 1–10.

(10) Crook, N., Abatemarco, J., Sun, J., Wagner, J. M., Schmitz, A., Alper, H. S. (2016) In Vivo Continuous Evolution of Genes and Pathways in Yeast. *Nat. Commun.*, *7*, 13051–13064.

(11) Hess, G. T., Frésard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., Montgomery, S. B., Bassik, M. C. (2016) Directed Evolution Using dCas9-Targeted Somatic Hypermutation in Mammalian Cells. *Nat. Methods*, *13*, 1036–1042.

(12) Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., Chang, X. (2016) Targeted AID-Mediated Mutagenesis (TAM) Enables Efficient Genomic Diversification in Mammalian Cells. *Nat. Methods*, *13*, 1029–1035.

(13) Moore, C. L., Papa, L. J., Shoulders, M. D. (2018) A Processive Protein Chimera Introduces Mutations across Defined DNA Regions in Vivo. *J. Am. Chem. Soc.*, *140*, 11560–11564.

(14) Ravikumar, A., Arrieta, A., Liu, C. C. (2014) An Orthogonal DNA Replication System in Yeast. *Nat. Chem. Biol.*, *10*, 175–177.

(15) Fasan, R., Meharenna, Y. T., Snow, C. D., Poulos, T. L., Arnold, F. H. (2008)

Evolutionary History of a Specialized P450 Propane Monooxygenase. *J. Mol. Biol.*, *383*, 1069–1080.

(16) Steinberg, B., Ostermeier, M. (2016) Environmental Changes Bridge Evolutionary Valleys. *Sci. Adv.*, *2*, 1–9.

(17) Bershtein, S., Goldin, K., Tawfik, D. S. (2008) Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J. Mol. Biol.*, *379*, 1029–1044.

(18) Toprak, E., Veres, A., Michel, J. B., Chait, R., Hartl, D. L., Kishony, R. (2012) Evolutionary Paths to Antibiotic Resistance under Dynamically Sustained Drug Selection. *Nat. Genet.*, *44*, 101–105.

(19) Toprak, E., Veres, A., Yildiz, S., Pedraza, J. M., Chait, R., Paulsson, J., Kishony, R. (2013) Building a Morbidostat: An Automated Continuous-Culture Device for Studying Bacterial Drug Resistance under Dynamically Sustained Drug Inhibition. *Nat. Protoc.*, *8*, 555–567.

(20) DeBenedictis, E. A., Chory, E. J., Gretton, D., Wang, B., Esvelt, K. (2020) A High-Throughput Platform for Feedback-Controlled Directed Evolution. *bioRxiv* Epub Apr 2, 2020. DOI: 10.1101/2020.04.01.021022.

(21) Wong, B. G., Mancuso, C. P., Kiriakov, S., Bashor, C. J., Khalil, A. S. (2018) Precise, Automated Control of Conditions for High-Throughput Growth of Yeast and Bacteria with eVOLVER. *Nat. Biotechnol.*, *36*, 614–623.

(22) Ang, K. H., Chong, G., Li, Y. (2005) PID Control System Analysis, Design, and

Technology. *IEEE Trans. Control Syst. Technol., 13*, 559–576.

(23) Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C. A., Smith, H. O. (2009). Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nature Meth., 6*, 343–345.

(24) Gietz, R. D., Schiestl, R. H. (2018) High-Efficiency Yeast Transformation Using the LiAc / SS Carrier DNA / PEG Method. *Nature Prot., 2*, 31–35.

(25) Mccormack, A. S., Godfrey, K. R. (1998) Rule-Based Autotuning Based on Frequency Domain Identification. *IEEE Transactions on Control Systems Technology, 6*, 43–61.

(26) Jung, P. P., Christian, N., Kay, D. P., Skupin, A., Linster, C. L. (2015) Protocols and Programs for High-Throughput Growth and Aging Phenotyping in Yeast. *PLoS ONE* Epub Mar 30, 2015. DOI: 10.1371/journal.pone.0019807.

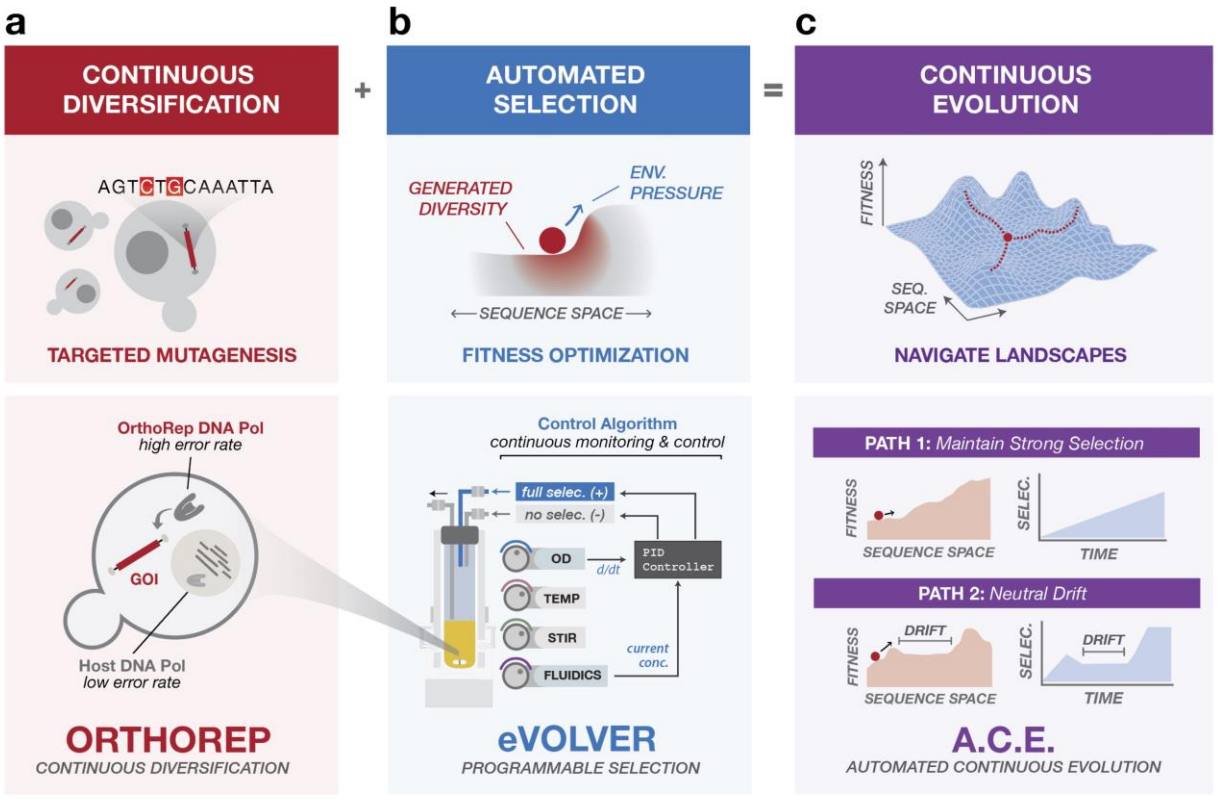


Figure 3.1. Automated Continuous Evolution (ACE). (**a**) OrthoRep enables continuous diversification of genes of interest (GOIs) via *in vivo* targeted mutagenesis in yeast. The basis of OrthoRep is an orthogonal DNA polymerase-plasmid pair that mutates GOIs ~100,000-fold faster than the genome. (**b**) eVOLVER is a continuous culturing platform for programmable, multiparameter control of selection conditions across many independent cultures. A PID control algorithm implemented with eVOLVER dynamically tunes selection pressure of populations as they adapt, precisely challenging them to achieve desired functions. PID control is achieved by tuning the ratio of full selection and no selection media inputs in response to growth rate. (**c**) By running OrthoRep in eVOLVER with PID control, ACE autonomously and rapidly navigates complex fitness landscapes. With a single framework, ACE can guide independent cultures through diverse trajectories.

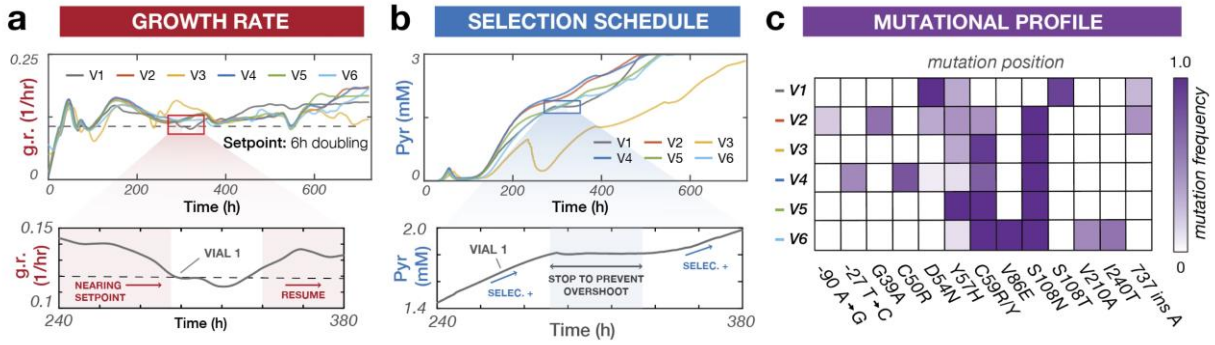


Figure 3.2. Automated continuous evolution of *Pf*DHFR resistance to pyrimethamine. (**a**) Top: Growth rate traces for six independent OrthoRep cultures (V1-6) evolving *Pf*DHFR resistance to pyrimethamine in eVOLVER using PID control. Bottom: A representative time window validating PID control. The growth rate (solid line) is controlled by automated tuning of pyrimethamine concentration (Figure 2b, bottom) to keep cultures constantly challenged at the setpoint growth rate (dashed line). (**b**) Top: Drug selection schedules for OrthoRep cultures evolving *Pf*DHFR. Bottom: A representative time window demonstrating PID-based selection tuning. Pyrimethamine concentration autonomously adjusts in response to growth rate deviation from the setpoint (Figure 2a, bottom). (**c**) Promoter and *Pf*DHFR mutations identified in six evolved populations. Mutation frequencies are estimated from SNP analysis of bulk Sanger sequencing traces.

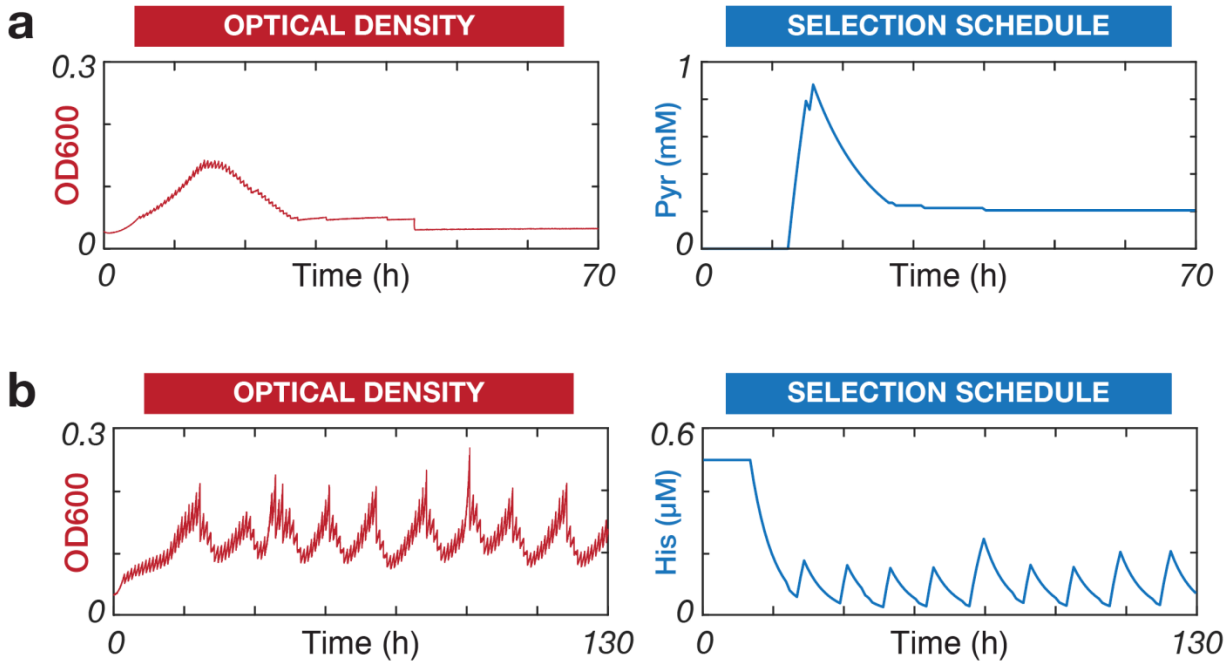


Figure S3.1. Sample evolution experiments in eVOLVER using the control algorithm derived from Toprak et al., 2013.[19] Briefly, this algorithm sampled OD at fixed intervals and if (1) the current OD is greater than a threshold and (2) if the current OD is greater than the previous OD, the growing cultures were diluted with the selection media. Otherwise, the culture would be diluted with the base media **(a)** An example of *Pf*DHFR evolution using the Toprak *et al.* algorithm. ZZ-Y435 was inoculated into eVOLVER and grown as described in **Methods** except the control algorithm was as described in Toprak *et al.*[19] The OD of the adapting culture (red) and the calculated concentration of pyrimethamine (blue) are shown. After the initial increase and subsequent decrease of pyrimethamine, no further growth is seen in 40 hours. **(b)** An example of *Tm*HisA evolution using the Toprak *et al.*[19] algorithm. ZZ-Y323 was inoculated into eVOLVER and grown as described in **Methods** except for the control algorithm. The OD of the adapting culture (red) and the calculated concentration of histidine in the media (blue) are

shown. The concentration of histidine is observed to oscillate during selection and is unable to successfully adapt.

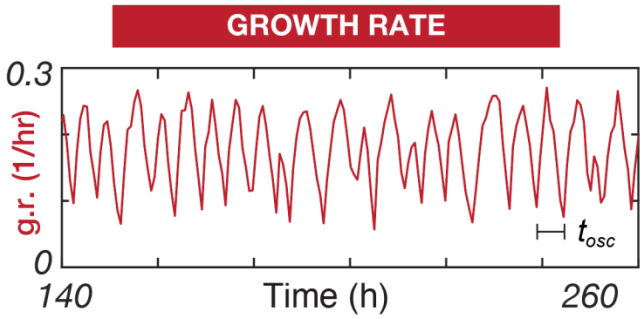


Figure S3.2. Oscillation of growth rate of ZZ-Y323 during the empirical determination of PID settings. ZZ-Y323 was inoculated into eVOLVER and SC-UL was used as the base media and SC-ULH was used as the full adaptation media. K_P was iterated from 0 to 4 over 120 hours while K_I and K_D were set to zero. Oscillations were observed when $K_P = 4$, and the period and amplitude of the oscillation were used to estimate the parameters for the PID control algorithm.

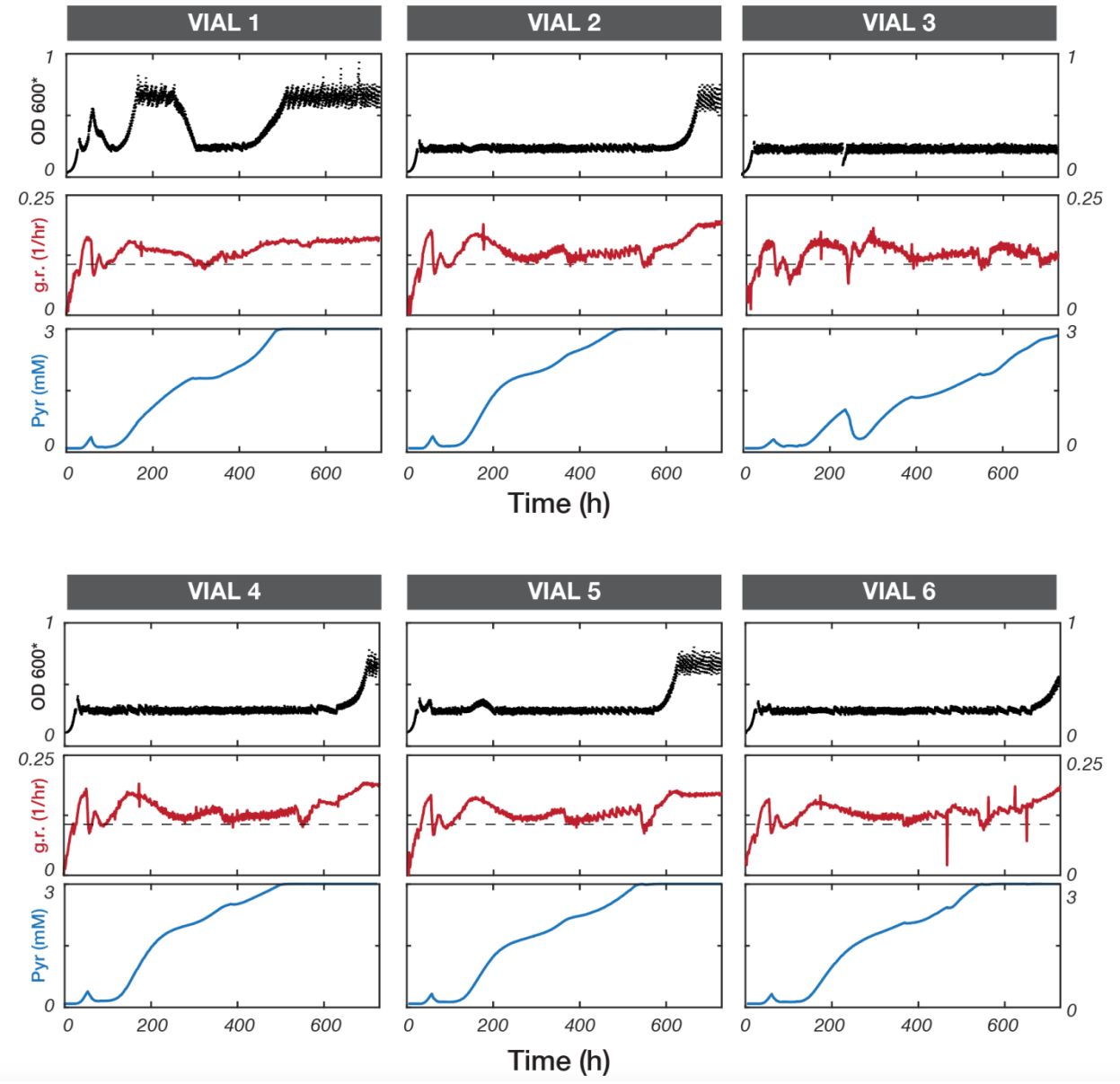


Figure S3.3. Adaptation history for all six replicates during *Pf*DHFR evolution. OD (black dot), growth rate (red line), target growth rate (black dash), and pyrimethamine concentration (blue line) are plotted for all six independent replicates. All six independent cultures were able to adapt to 3 mM pyrimethamine without the need for pre-programmed selection schedules or user intervention except to replenish media stocks.

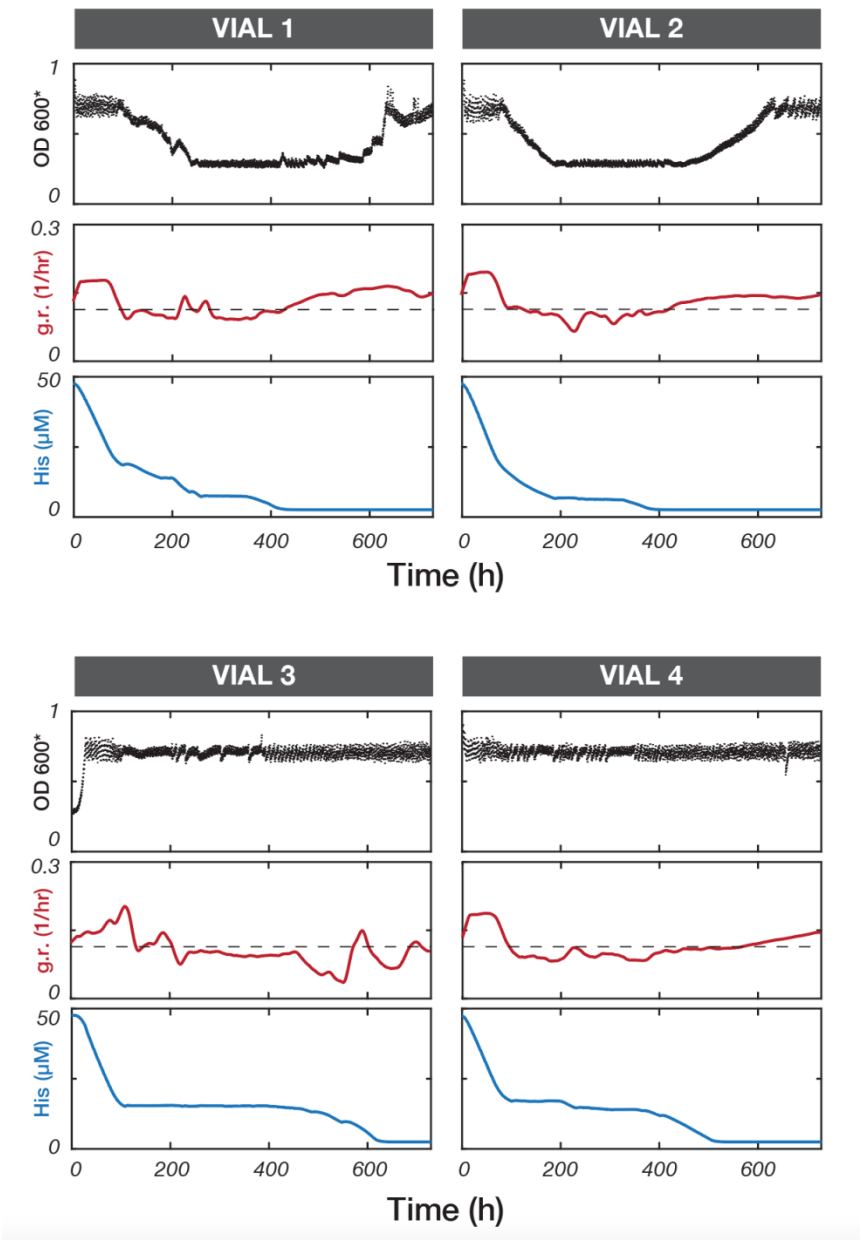


Figure S3.4. Adaptation history for all four replicates during *Tm*HisA evolution. OD (black dot), growth rate (red line), and histidine concentration (blue line) are plotted for all six independent replicates. All four cultures were able to adapt to media lacking histidine without the need for evolution schedules or user intervention except to provide eVOLVER with fresh base and full adaptation media in 700 hours of growth.

Table S3.1. *Tm*HisA mutants characterized. Both nonsynonymous mutations (**bold**) and synonymous are noted. The order of these variants (top to bottom) corresponds to the order of the variants (left to right) in Figure 3c.

Name	Mutations from wt *Tm*HisA
V1-1	**E37G**, N62N, **E71G**, **L85F**, R98R, **F124L**, K168K, **K182R**, E186E, S202S, **V209A**, **F226S**
V1-3	**G12E**, **E29K**, **V36M**, **H75Y**, **I185T**, **R224K**
V1-4	F27F, **I40T**, G43G, **E71G**, **F72L**, L112L, V123V, D127D, **A140T**, E186E, **V209A**
V1-6	**E37G**, **E71G**, **F124L**, L126L, E186E, **S199P**, **V209A**
V2-1	**A15V**, **E37G**, **E71R**, D108D, **F124L**, L172L, E186E, **V209A**
V2-2	**V4A**, **A15V**, **Y28H**, **E37G**, **E71G**, **S102G**, **A134V**, V147V, **D176Y**, **K182E**, **I183T**, E186E, **V209A**, T213T, **E228K**
V2-4	**I18T**, **E71G**, **K181R**, **V209A**, **I221V**
V3-1	**H2Q**, A6A, **E71G**, **F72L**, V123V, **V133A**, E186E, E188E, **V209A**, **E234G**
V3-2	**H2Q**, A6A, D8D, **N24S**, **E71G**, **F72L**, V123V, **V133A**, E186E, E188E, **V209A**, **E234G**
V3-3	**H2Q**, A6A, V14V, **E71G**, **F72L**, V123V, **V133A**, **I161V**, E186E, E188E, **V209A**, **E234G**
V3-5	**H2Q**, A6A, **E71G**, **F72L**, **I82V**, V123V, **V133A**, **G170D**, E186E, E188E, **V209A**, **E234G**
V4-1	**V4A**, **F27S**, **V36M**, E41E, R97R, S125S, **M236T**
V4-2	**E71G**, **F72L**, **V105A**, V123V, **I144T**, E186E, **S199A**, **V209A**, **G223R**
V4-3	D8D, **V14I**, **E37G**, **E71G**, **F135L**, **S148P**, L149L, E186E, **L203S**, **K208R**, **V209A**, **F226L**
V4-4	**V4A**, **N24S**, **V36M**, G60G, R97R, **E107G**, **F124Y**, **I166T**, **I221V**, **M236T**
V4-5	**H2Q**, **I7V**, **A15V**, K22K, E29E, **V36M**, I47I, E67E, **F124L**, **E143G**, **D145N**, **V147A**, K181K, E188E

Table S3.2. List of plasmids used in this study. For plasmids ZZ-Ec903 to ZZ-Ec918, see Table S1 for specific sequences.

Name	Source	Parent Plasmid	Origin of replication (yeast, bacterial)	Selection Marker (yeast, bacterial)	Notes
GA-DNA-*flo1URA3*	This work	n/a	n/a	URA3, n/a	Used to delete flo1 in strains in order to prevent flocculation
GA-Ec51	Ravikumar et al. 2018[2]	See previous work	CEN6/ARS4, ColE1	URA3, AmpR	
GA-Ec64	Ravikumar et al. 2018[2]	See previous work	n/a, ColE1	TRP1, AmpR	Recombination cassette that integrates TRP1, mKate2, *Pf*DHFR, and leu2 (538C>T) in place of wt TP-DNAP1 on the orthogonal plasmid (pGKL1)
AR-Ec633	Ravikumar et al. 2018[2]	See previous work	CEN6/ARS4, ColE1	HIS3, KanR	REV1 promoter > TP-DNAP1 L477V, L640Y, I777K, W814N
ZZ-Ec482	This work	AJ-Ec200[27]	2µ, ColE1	NatMX, AmpR	contains sgRNA targetting *TRP1* (*GTCCATTGGTGAAAGTTTG*)
ZZ-Ec475	This work	GA-Ec64	n/a, ColE1	URA3, AmpR	p1 recombination cassette that integrates *Tm*HisA and URA3 in place of ORFs 1-4 on the orthogonal plasmid (pGKL1)
ZZ-Ec506	This work	AR-Ec633	CEN6/ARS4, ColE1	LEU2, KanR	Plasmid encoding the error-prone orthogonal DNAP (TP-DNAP1 (L477V, L640Y, I777K, W814N)) driven by the REV1 promoter
ZZ-Ec727	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Sc*His6
ZZ-Ec903	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA-V1-1
ZZ-Ec904	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V1-3
ZZ-Ec905	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V1-4
ZZ-Ec906	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V1-6
ZZ-Ec907	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V2-1
ZZ-Ec908	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V2-2
ZZ-Ec909	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V2-4
ZZ-Ec910	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V3-1
ZZ-Ec911	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V3-2

ZZ-Ec912	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V3-3
ZZ-Ec913	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V3-5
ZZ-Ec914	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V4-1
ZZ-Ec915	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V4-2
ZZ-Ec916	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V4-3
ZZ-Ec917	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V4-4
ZZ-Ec918	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > *Tm*HisA -V4-5
ZZ-Ec919	This work	ZZ-Ec506	CEN6/ARS4, ColE1	LEU2, KanR	Encodes RPL18B promoter > wt *Tm*HisA

Table S3.3. List of yeast strains used in this study.

Name	Genotype	Source	Parent Strain	Notes
F102-2u	F102-2u *MATa can1 his4-519 leu2-3,112* ρ^0+ pGKL1 + pGKL2	ATCC #200585	n/a	
AR-Y383	F102-2u *MATa can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4* ρ^0+ pGKL1 + pGKL2	Ravikumar et al., 2018[2]	AR-Y292	
ZZ-Y454	F102-2u *MATa flo1::URA3 can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4* ρ^0+ pGKL1 + pGKL2	This work	AR-Y383	AR-Y383 transformed with GA-DNA-*flo1URA3*
ZZ-Y455	F102-2u *MATa flo1 can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4* ρ^0+ pGKL1 + pGKL2	This work	ZZ-Y454	ZZ-Y454 streaked on solid media containing 5-FOA and confirmed with a negative flocculation phenotype
GA-Y235	F102-2u *MATa flo1 can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0+ pGKL1 + pGKL2+GA-Ec51	This work	ZZ-Y455	ZZ-Y455 transformed with GA-Ec51
ZZ-Y431	F102-2u *MATa flo1 can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0+ p1- FulldelPol-W-mK-PfDHFR-l*(TAA) + pGKL1 + pGKL2+GA-Ec51	This work	GA-Y235	GA-Y235 transformed with GA-Ec64
ZZ-Y435	F102-2u *MATa flo1 can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0+ p1- FulldelPol-W-mK-PfDHFR-l*(TAA) + pGKL1 + pGKL2 + AR-Ec633	This work	ZZ-Y431	GA-Y235 transformed with AR-Ec633
ZZ-Y292	F102-2u *MATa can1 leu2Δ0 ura3Δ0 his6::KanMX HIS4* ρ^0+ pGKL1 +	This work	AR-Y288	AR-Y288 transformed with PCR product from Yeast Knockout Collection strain with *his6* deleted with primers TCATCATCAAGGGTCATCTTTTTAT

	pGKL2			and GAAAAAGGTTGCCTCAATATTGTTA
ZZ-Y299	F102-2u *MATa can1 leu2Δ0 ura3Δ0 trp1Δ0 his6::KanMX HIS4* ρ^0+ pGKL1 + pGKL2	This work	ZZ-Y292	ZZ-Y292 transformed with ZZ-Ec482 and linear DNA corresponding to 40 basepairs upstream and downstream of *TRP1*. The strain was restreaked x2 on YPD solid media to remove ZZ-Ec482
ZZ-YT17-A3	F102-2u *MATa can1 leu2Δ0 ura3Δ0 trp1Δ0 his6::KanMX HIS4* ρ^0+ p1-FullDelPol-*Tm*HisA-URA3 + pGKL2 + ZZ-Ec506	This work	ZZ-Y299	ZZ-Y299 transformed with ZZ-Ec506 and ScaI-digested ZZ-Ec475
ZZ-Y323	F102-2u *MATa can1 leu2Δ0 ura3Δ0 trp1Δ0 his6::KanMX HIS4 flo1::NatMX* ρ^0+ p1-FullDelPol-*Tm*HisA-URA3 + pGKL2 + pZZ-Ec506	This work	ZZ-YT17-A3	ZZ-Y299 transformed with linear DNA containing 500 basepairs upstream of *FLO1*, *NatMX*, and 500 basepairs downstream of *FLO1*
BY4741	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+	ATCC #201388		
ZZ-Y332	BY4741 *MAT***a** *HIS3 leu2Δ0 met15Δ0 ura3Δ0* ρ^+	This work	BY4741	BY4741 transformed with HIS3 PCR from F102-2u
ZZ-Y336	BY4741 *MAT***a** *HIS3 leu2Δ0 met15Δ0 ura3Δ0 his6::KanMX* ρ^+	This work	ZZ-Y332	ZZ-Y332 transformed with PCR product from Yeast Knockout Collection strain with *his6* deleted with primers TCATCATCAAGGGTCATCTTTTTAT and GAAAAAGGTTGCCTCAATATTGTTA
ZZ-Y354	BY4741 *MAT***a** *HIS3 leu2Δ0 met15Δ0 ura3Δ0 trp1Δ0 his6::KanMX* ρ^+	This work	ZZ-Y336	ZZ-Y292 transformed with ZZ-Ec482 and linear DNA corresponding to 40 basepairs upstream and downstream of *TRP1*. The strain was restreaked x2 on YPD solid media to remove ZZ-Ec482

Supplement References

(27) Javanpour, A. A., Liu, C. C. (2019) Genetic Compatibility and Extensibility of Orthogonal Replication. *ACS Synth. Biol.*, *8*, 1249–1256.

Chapter 4. Evolving enzymes for novel catalytic activity using OrthoRep

Ziwei Zhong[1] and Chang C. Liu[1,2,3,*]

[1] Department of Biomedical Engineering, University of California, Irvine, California 92697, United States

[2] Department of Chemistry, University of California, Irvine, California 92697, United States

[3] Department of Molecular Biology & Biochemistry, University of California, Irvine, California 92697, United States

Introduction

Understanding how enzymes develop new catalytic activities and recognize non-native substrates is important for drug discovery[1,2], commodity chemical synthesis[3], biodegradation[4,5], and many other industrial functions[6,7]. These enzymes are often from an organism which can be difficult to culture, slow growing, and/or lack of genetic tools. Therefore, the enzymes can be transplanted into a host organism, such as *Saccharomyces cerevisiae* or *Escherichia coli*, which grows easily in lab conditions and has a wide panel of genetic tools available for use[8,9]. However, repurposing natural enzymes is a difficult task, as often the enzyme is optimized for a natural task and/or requires protein associations or cofactors that are not present in the new host[9]. Further, the catalytic efficiencies of new or secondary functions are often poor[10,11], requiring time and labor-intensive directed evolution that often does not result in substantial gains in catalytic activity[12].

In light of the challenges encountered when evolving novel catalytic abilities or higher catalytic efficiencies, it is not well understood how nature has evolved a stunning number of enzymes with extremely high catalytic efficiencies. *In vitro* evolution experiments have sought to determine how enzymes evolve and how evolutionary pressures have guided exploration of sequence space to develop a diverse range of substrates and catalyze different reactions[12,13]. Experimental evolution studies have already offered insight as to how enzyme diversity has developed. For instance, it is possible to improve a protein function, such as catalysis[14,15] or binding[16] via step-wise improvements even if the catalytic efficiency is initially low, given that the function is already present in the protein. Additionally, many proteins have promiscuous functions that can be subjected to selection and improvement in terms of increased turnover, tighter binding, or both[11,17,18]. Not only can these secondary functions can be selected for with

little or no detriment to the primary function[19], random mutations can also result in the development of novel secondary functions[11]. Another particularly striking case was demonstrated by Fasan et al., where they demonstrated that the turnover number of cytochrome P450 for short-chain alkanes could only be further improved when they performed an evolutionary step for stability[13,20]. Despite these successes, not all enzymes can be evolved to a level suitable for industrial applications, as generally, the improvement seen in turnover number of evolved enzymes largely remains below a couple orders of magnitude while the turnover number of natural enzymes span over six orders of magnitude. Further, the increase catalytic efficiency largely stems from tighter binding of substrates as opposed to improvements in enzyme turnover number[21].

To gain further insight into how enzymes develop and improve novel catalytic activities, we have developed a model enzyme evolution system that evolves the enzyme HisA to catalyze the Trp1 function in yeast. HisA and Trp1 are enzymes in the biosynthesis pathways of histidine and tryptophan which show remarkable similarity in their mechanism and enzyme structure. The enzymes N′-[(5′-Phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase (HisA in bacteria, His6 in eukaryotes) and phosphoribosylanthranilate (PRA) isomerase (TrpF in bacteria, Trp1 in eukaryotes) catalyze the similar Amadori rearrangements of the sugar backbones (Figure 4.1). Interestingly, depending on the organism, these two reactions can either be catalyzed by the same enzyme, as demonstrated in *Mycobacterium tuberculosis* and *Streptomyces coelicolor*[22], or by two distinct enzymes as in found in *Esherichia coli*, *Thermotoga maritima*, and *Saccharomyces cerevisiae*[23]. In the case where there are two distinct enzymes, these enzymes often have poor sequence alignment and similarity as demonstrated by a 10% similarity in *T. maritima* and no sequence similarity in *S.*

cerevisiae[23]. Despite the absence of similarity, it has been previously shown that HisA from *T. maritima* (*Tm*HisA) can catalyze the reaction of TrpF with a single amino acid change, and further, these two enzymatic activities are opposing and that improvement in the kinetics of one reaction results the deterioration of the other[23,24].

Using the scalability of OrthoRep[25], we sought to establish *Tm*HisA as a model for enzyme evolution to explore the landscape of *Tm*HisA for His6 and Trp1 activity as well as to understand the tradeoffs, if any, in the selection for one enzymatic activity over the other. We first demonstrate that *Tm*HisA possesses no baseline activity of either His6 or Trp1 in *S. cerevisiae*. Further, we show that while His6 activity can be readily established on *Tm*HisA, Trp1 activity cannot be as readily evolved. Rather, an intermediate selection for His6 activity is required but not sufficient to obtain Trp1 activity. Finally, we demonstrate that the fitness landscape of *Tm*HisA for Trp1 is complex, and that certain trends may be important in the context of establishing Trp1 activity on the *Tm*HisA backbone.

Materials and Methods

Strain creation

Strains and plasmids used are noted in supplemental tables S1. Detailed procedure for making these strains can be obtained from Chapter 3 or from Ref 26[26].

HisA evolution

ZZ-Y323 was used as the base strain for evolution experiments and SC-UL from US Biological with standard 76 mg/L histidine and 76 mg/L tryptophan was used as the media unless otherwise stated. A starter 100 mL culture was grown from a glycerol stock which was subsequently diluted 1:100 into eight replicates of 30 mL of SC-UL media containing 7.6 mg/L tryptophan in a 50 mL conical tube, loosely capped to facilitate gas exchange. The replicates

were then placed at a 20° angle in an orbital shaker at 200 RPM and 30°C. Once the cultures had reached saturation, usually after two to three days, they were passaged with media containing decreasing concentrations of tryptophan as shown in Figure 4.4. After 20 passages, the tryptophan concentration was 0.76 mg/L and additional tryptophan could not be dropped out from the media as cultures failed to saturate, and these eight replicates were designated as plateau points (Figure 4.4).

Each plateau culture was then diluted 1:100 in three conditions of increasingly histidine limiting, tryptophan limiting, or no limiting conditions in six replicates for a total of 18 replicates per plateau culture (Figure 4.4). For the histidine limiting condition, cultures were serially passaged 1:100 twenty times in SC-UL media initially containing 3.8 mg/L histidine at the first passage, down to 1.52 mg/L histidine at the twentieth passage. For tryptophan limiting conditions, cultures were passaged 1:100 twenty times in SC-UL media containing 0.76 mg/L tryptophan. For no limiting conditions, cultures were serially passaged 1:100 twenty times in SC-UL media containing 76 mg/L of histidine and tryptophan.

To reselect *Tm*HisA for Trp1 activity, each of the 144 cultures underwent twenty 1:100 passaging in SC-UL media initially containing 2 mg/L tryptophan and decreasing the concentration of tryptophan with subsequent passages to 0.2 mg/L tryptophan at the end of twenty passages. Successful evolution of *Tm*HisA for Trp1 activity was determined when the OD after 3 days of culture was greater than 2 while cultures that did not successfully evolve Trp1 activity had OD values < 0.2.

HisA growth rate measurement

Wild-type *Tm*HisA, His6, and Trp1 from *S. cerevisiae* were PCR amplified and cloned into a CEN/ARS plasmid[26]. ZZ-Y354 was transformed with these plasmids and grown to saturation. Eight biological replicates were picked for each clone and their growth rates were measured on a Tecan Infinite 2000 by measuring OD600 every 30 minutes for 20 hours. A previously documented custom MATLAB script was used to calculate growth rates and doubling times from the OD600 data[26].

***Tm*HisA NGS sequencing**

The eight plateau point cultures, and the 144 cultures just prior to and after twenty passages selecting for Trp1 activity (Figure 4.4) were grown from glycerol stocks in SC-UL media and miniprepped as previously described[27]. *Tm*HisA was PCR amplified from these minipreps and barcoded (Table S2) on the forward and reverse primers to allow for deconvolution after NGS sequencing in a PacBio Sequel II. Prior to submission for sequencing, DNA concentrations were measured on a NanoDrop 2000 and the 1 ng from each sample was combined and sent to the UCI Genomics High Throughput Facility for sequencing. A custom Python script was used to deconvolute the sequences to each condition as well as to call mutations from each sequence read. Only cultures where greater than 10 reads could be obtained were used in subsequent mutation analysis.

Statistics

Statistical analysis was done with GraphPad Prism 8. Analyses were either one way ANOVA with Tukey's correction for multiple comparisons or two sample t-tests. Error bars reported herein represent mean ± standard deviation.

Results and Discussion

His6 activity can be readily evolved on *Tm*HisA

Contrary to expectations, wild-type (wt) *Tm*HisA does not exhibit sufficient activity to complement a *his6* or *trp1* deletion in *S. cerevisiae* (Figure 4.2). In ZZ-Y354, both *his6* and *trp1* were knocked out as described previously[26], allowing for the assaying of His6 or Trp1 activity of *Tm*HisA and its evolved variants by measuring growth rate in media limited by histidine or tryptophan, respectively. When wt *Tm*HisA was expressed with an RPL18b promoter (medium strength promoter in yeast and similar to levels that can be expected on OrthoRep[28]), there was no significant difference in growth rate compared with a plasmid control in media limited by histidine or tryptophan (Figure 4.3). This demonstrates that wt *Tm*HisA exhibits minimal His6 activity and below the limit of detection of our growth rate assay. Therefore, we explored using OrthoRep to first establish His6 activity on the *Tm*HisA backbone.

We used OrthoRep to successfully evolve *Tm*HisA for His6 activity in 24 of 24 independent 30mL cultures. We encoded wt *Tm*HisA on the p1 plasmid of OrthoRep along with the mutagenic polymerase TP-DNAP-**4-2** *in trans* on a nuclear plasmid and knocked out *his6* and *trp1* from the strain to allow for selection of His6 or Trp1 activity simply by growth in media limited or lacking histidine or tryptophan (Figure 4.3). Additionally, we knocked out *flo1* with NatMX as described previously[26] to prevent flocculation. Pilot experiments have shown that this was essential for successful evolution of *Tm*HisA as flocculation, a known phenomenon of yeast during times of stress and nutrient limitation[29–32], may enable survival of individuals that have not adapted *Tm*HisA for His6 or Trp1 activity. We used this strain, ZZ-Y323, to seed 24 independent replicates of 30 mL cultures with media limited by histidine. Once cultures were saturated, generally after 48-72 hours, they were diluted 1:100 into fresh media containing

decreasing amounts of histidine. After 14-20 passages, all 24 out of 24 cultures were saturating fully in media lacking histidine, indicating that His6 activity can be readily evolved on the *Tm*HisA backbone.

Establishment of Trp1 activity rapidly reaches a plateau

Using the same strain, ZZ-Y323, we were unable to successfully evolve wt *Tm*HisA for significant Trp1 activity in any of 24 independent 30 mL cultures after 100 generations. We seeded 24 replicate 30 mL cultures with media limited by tryptophan and similarly passaged the cultures at a 1:100 dilution into media containing decreasing concentrations of tryptophan (Figure 4.4, left). After 20 passages, all 24 out of 24 cultures grew minimally in media containing 0.76 mg/L of tryptophan (1% of tryptophan in regular SC media) and did not demonstrate significant growth in media completely lacking in tryptophan, indicating that cultures may have reached a plateau in Trp1 activity. From these 24 cultures that were unsuccessful in evolving Trp1 activity, we selected the first eight cultures, subsequently denoted as "plateaus", to conduct further analysis and evolution.

We performed deep sequencing on all cultures on the eight plateau points (Figure 4.4, left), and we observed a complex mutational landscape with few consensus mutations (Table 4.1). At these plateau points, there is significant diversity among the cultures with no consensus mutation present across all eight cultures (Table 4.1). While S53L and G60S were the most prevalent mutations, observed in 6 and 5 out of 8 plateau cultures, respectively, other mutations that observed at high frequencies were only observed in at most 3 plateau cultures. Further, these most common mutations did not reflect the mutations seen in prior evolution experiments with *Tm*HisA. Indeed, the only previously observed mutation that is present in the plateau cultures is D127G, and it is only present in one of eight cultures.

Interestingly, in contrast with previous reports, it was difficult to establish Trp1 activity on the *Tm*HisA backbone. In prior studies with *E. coli* and *S. cerevisiae* where *Tm*HisA was used as the backbone to evolving Trp1 activity[23,33], at most five non-synonymous mutations were seen with only 1-2 key mutations sufficient to impart sufficient Trp1 activity to enable survival and propagation without exogenous tryptophan supplementation. However, we observe 8-10 consensus mutations in cultures that able to evolve Trp1 activity, with individual sequences containing as high as 18 non-synonymous mutations and only a minority of mutations reflecting those seen previously. This discrepancy between our results and those prior may be due in part to differences in OrthoRep compared to traditional directed evolution, which includes a biased mutational supply and decreased expression of proteins encoded on OrthoRep.

One difference is that the mutational preferences of OrthoRep may not allow sufficient sampling of favorable mutations as the mutational supply for OrthoRep is biased towards transitions over transversions. Calculations suggest that with 30 mL culture volumes for evolution and OrthoRep's current mutation rate of 10^{-5} substitutions per base, all single mutants and a significant percentage of double mutants are sampled as each generation should generate upwards of one million mutations. However, this may be skewed by the mutational preferences of OrthoRep's TP-DNAP-**4-2** polymerase for transitions, as Ravikumar *et al.* previously report the transition to transversion ratio as 50:1[25]. When accounting for this bias for transitions, the number of mutations supplied per passage may only offer sufficient sampling of single mutants and very limited sampling of double mutants that require transversions, and even then, may be diluted out by passaging. Further, though mutations observed in previous *E. coli* selection of *Tm*HisA for Trp1 activity were primarily transition mutations[23], mutations previously observed

in *S. cerevisiae* are often transversions or require two base pair changes[33], conceivably indicating under-sampling with OrthoRep.

Another difference is that *Tm*HisA may need to be expressed at higher levels on OrthoRep to have an observable phenotype. For OrthoRep polymerases with high mutation rates, genes encoded on OrthoRep can maximally be expressed at medium levels[28]. While the *S. cerevisiae* versions of His6 and Trp1 are naturally expressed at low levels[34], wt *Tm*HisA, as a thermophilic and xenologous version of these enzymes, may have reduced or minimal function in yeast. As a result, mutations that impart a small increase in enzymatic activity may not exhibit sufficient activity to enable survival when encoded on OrthoRep. While not explicitly stated, previous work likely used high expression systems, thus both allowing isolation of weakly active mutants and allowing greater discrimination from non-functional mutants. This in turn allowed for the identification of mutants with low but survivable Trp1 activity[23,33], and therefore, may not have higher activity than the mutants identified at our plateau points. Further, when we expressed those previously reported mutants at high levels on nuclear plasmids, we could not replicate complementation in *S. cerevisiae*. However, mutants identified with OrthoRep can sufficiently complement a *trp1* deletion in *S. cerevisiae* even when expressed at medium levels, reflecting the expression levels at which they were evolved. By encoding *Tm*HisA on OrthoRep, a greater increase in enzymatic activity is required to achieve similar discriminating power and to enable complementation in media lacking tryptophan. As a result, the mutants identified via OrthoRep should consequently have higher Trp1 activity compared to previously identified variants of *Tm*HisA.

An intermediate selection for His6 activity is required, but not sufficient for evolution of higher Trp1 activity

We next explored alternative selections that may enable evolution of *Tm*HisA for higher Trp1 activity, demonstrating that only through an alternative selection for His6 activity can *Tm*HisA be evolved for sufficient Trp1 activity to complement a *trp1* deletion. From the eight plateau cultures, we further subjected each plateau in six replicate cultures to alternative selection conditions to determine if they can escape the plateau to higher Trp1 activity. The first of these conditions, denoted **W**, the plateau cultures continued to passaged in media containing 0.76 mg/L of tryptophan (1% of regular SC media, Figure 4.4, center top). In the second condition, denoted **H**, there was partial selection for His6 activity, as tryptophan concentration was restored to the stock concentration of 76 mg/L and rather, increasing amounts of histidine was removed from the media down to 1.52 mg/L (2% of regular SC media, Figure 4.4, center middle). In the third condition, denoted **UL**, plateau cultures were allowed to drift with stock concentrations of both tryptophan and histidine (both 76 mg/L, Figure 4.4, center bottom). During the drift or alternative selection periods, cultures were similarly passaged in 30 mL volumes and diluted 1:100 in media with varying concentrations of tryptophan and histidine as shown in Figure 4.4. After twenty 1:100 passages in alternative selection or drift conditions, evolution of Trp1 activity was once again attempted on all 144 cultures in a similar manner to the initial selection for Trp1 activity (Figure 4.4, right). After an additional twenty passages reselecting for Trp1 activity, only nine cultures demonstrated the ability to grow and fully saturate in media lacking tryptophan. All nine of these cultures came from replicates that underwent alternative selection for His6 activity, with three cultures descending from the Plateau 4 replicate and six cultures descending from the Plateau 6 replicate (Table 5 and 7). Conversely, only two out of eight plateaus that underwent alternative selection for His6 activity were eventually able to successfully evolve Trp1 activity, indicating that this alternative selection for

His6 activity was necessary, but not sufficient for successful evolution of Trp1 activity on the *Tm*HisA backbone.

This result is a slight variation compared to previous studies. Most prior work that have sought to bypass fitness plateaus has focused on the effect that neutral or purifying drift has on protein evolvability[11,19,35,36]. In these cases of neutral drift, Bershtein *et al.* demonstrated that the stability of the protein was increased, and thus enabled destabilizing mutations that may allow for greater ornovel functions[35]. However, in our preliminary experiments, we attempted neutral drift by drifting plateaus in 7.6 mg/L tryptophan (10% of normal SC media), but did not achieve any successes in bypassing the plateaus. Additionally, other studies have employed a band-pass selection, which subjected TEM-15 β-lactamase to an intermediate selection prior to strong selection and compared it to a continuously strong selection and demonstrated that TEM-15 can achieve higher activity through an intermediate selection step than via strong selection alone[37]. Both neutral drift and a band-pass selection can be selection schemes that can be further tested by our model in future experiments to compare with the alternative selection for His6 activity.

Deep sequencing offers few insights into the difficulty of evolving *Tm*HisA for Trp1 activity

To investigate the factors that contributed to the difficulty in establishing Trp1 activity on *Tm*HisA, we deep sequenced cultures at the end of alternative selection and drift, as well as at the end of reselection for Trp1 activity (Figure 4.4, right). Mutations that were present at the plateau generally carried through even after continued selection or drift (Tables 4.2-9). With the exception of plateaus 5 and 8 (Tables 4.6 and 4.9), the mutations present in the cultures that underwent drift or continued selection for Trp1 activity showed similar mutations as the plateau points with the occasional development of additional mutations that were observed at high frequencies, such as D145V, I185T, and E188K in Plateau 1 (Table 4.2) and V14A and I221T in

Plateau 2 (Table 3), among others. Additionally, after further passaging, S53L and G60S, the mutations most prevalent at the plateaus, also appeared in most cultures. In the cultures where subsequent selection caused divergence from the plateaus, there were two trends. The first of these was predominantly present in replicates that were able to successfully evolve Trp1 activity (Plateau 4 and 6). In this case, certain mutations present in the plateau became undetectable only after alternative selection for His6 activity and did not reappear during reselection for Trp1 activity (Table 4: V33I, D127G, G170D; Table 7: I7M, K154E, K168E, D169N, G195D). Instead, these cultures developed mutations not seen in other replicate cultures of the same plateau that had undergone continued selection for Trp1 activity or drift with no selection for activity, such as I144T, L149P, L158F, I185V/T, and E188G in Plateau 4 and V50A, D127V/H, and I197V in Plateau 6 (Tables 4.5 and 4.7). However, there exists a second case of mutations that were present at high frequency at the plateaus but were not carried through in any subsequent replicates (Table 6: I144N, K154E, K168E, G195D; Table 9: V3I, G60S, S70P, S148P, G157S, A225V).

One potential difference we investigated is differential mutational supply for different conditions of *Tm*HisA during evolution, but the data suggest that this is not likely. Through deep sequencing, we observe 8-10 consensus mutations in cultures that were able to evolve Trp1 activity and an average of 12.7 mutations per sequences (mps) in these cultures. This is similar to the average of 12.6 mps of all cultures at the end of ~400 generations of evolution. Further, we observe that the average mps increases as the number of passages increase. In the plateau cultures, there is an average of 9.9 mps (Figure 4.5). While this decreases to 9.3 mps in the replicates that are alternatively selected for His6 activity, there is still an increase to 11.4 and 10.9 mps in the cultures that underwent either continued selection for Trp1 activity or was

allowed to drift in non-selective conditions, respectively (Figure 4.5). Mutations continue to accumulate as the mutations per sequence increases to 12.2, 13.2, and 13.5 after reselection for Trp1 activity after selection for His6 activity, Trp1 activity, or drift in non-selective conditions, respectively (Figure 4.5). This indicates that mutation accumulation is still present during later generations of evolution and at similar rates across all conditions. While mutation accumulation is not sufficient to explain the differences between successful and unsuccessful cultures, future evolution experiments using clonal plateau points in fresh evolution strains can limit the possibility of decreasing mutational load affecting evolution.

Thermophilic origin of *Tm*HisA may contribute to difficulty establishing Trp1 activity

Another explanation for the difficulty in establishing Trp1 activity on *Tm*HisA could be its extreme thermophile origin, as *T. maritima* grows optimally at 80° C[38]. Previous work done by Näsvall *et al*. demonstrated considerable ease in evolving HisA from *Salmonella enterica* for TrpF activity in its native host[24,39]. Therefore in evolving *Tm*HisA, a significant number of mutations are needed to enable catalysis in *S. cerevisiae*, which grows at the mesophile temperature of 30° C, as both an adaptation for a new environment and substrate are required. Consistent with the need for mutations that enable activity at mesophile temperatures is that at plateau points for Trp1 activity, there are mutations consistent with previous data in which *Tm*HisA was selected for His6 activity (V4A, Y28H, V36M, E71G, A134V, S148P, G170D, and M236T from ref. 26[26]). Yet, these mutations are neither widespread nor consistent: plateaus have 0-3 of the aforementioned mutations, and each mutation is only present in at most 2 plateau cultures. This suggests that the mutations for mesophile activity are varied, generally weak, and/or have to compete with mutations that impart Trp1 activity. However, the fact that these mutations are not outcompeted by mutations that impart Trp1 activity suggests that Trp1

enabling mutations are similarly weak and/or varied. Therefore, this adaptation to mesophile temperatures coupled with the evolution for Trp1 activity may have been too large of an evolutionary step to adequately sample in direct selection for Trp1 activity.

Further, there may be tradeoffs between temperature and substrate as the mutations that enable activity at mesophile temperatures may have negative or even reciprocal sign epistasis with mutations that enable Trp1 activity. Consistent with this hypothesis, previous reports demonstrated that when enzymes from thermophiles are adapted for lower temperatures, the K_m, a measure of affinity of the enzyme for the substrate, is worsened and k_{cat}, the enzyme turnover number, is increased[40,41]. Therefore, while selecting *Tm*HisA for Trp1 activity in *S. cerevisiae*, there is a simultaneous pressure to worsen K_m for general activity and a pressure to improve K_m for the new substrate of Trp1, PRA. In the data, this is suggested by the replicates that were eventually able to develop Trp1 activity from Plateau-4 and Plateau-6. In Plateau-4's alternative selection for His6 activity, sequences lost V33I and G170D mutations that were present at the plateau point and eventual reselection for Trp1 activity (Table 5). Similarly, in Plateau-6's alternative selection for His6 activity, I7M, I56T, K154E, D169N, G195D, and Y239H were lost from the plateau sequence during selection for His6 activity and did not reappear in the successful reselection for Trp1 activity. In contrast, these mutations persisted in replicate cultures of the same plateaus that underwent continued Trp1 selection or non-selective drift, none of which were able to give rise to mutants that had significant Trp1 activity. Additionally, none of the other replicates display any reversal of mutations present at the plateau in alternative selection for His6 activity (Tables 2-4,6,8,9). Therefore, the alternative selection for His6 activity may have enabled simply acted as a "stepping stone" for the evolution for Trp1 activity by decoupling the opposing pressure for K_m values. Future experiments examining the mutations

that enable Trp1 activity can determine if there are mutational paths that enable monotonically increasing Trp1 activity with increasing mutations, or if only an alternative selection for His6 activity enables steadily increasing fitness. In addition, further analysis of the aforementioned mutations that disappeared in the successful evolutions for Trp1 activity can examine if there are epistatic interactions among mutations that limit step-wise gains in fitness.

However, the possibility of the thermophilic origin of *Tm*HisA is actually detrimental to establishing Trp1 activity comes unexpectedly. Prior work by Bloom and colleagues demonstrated that stability is crucial for development of higher activities, as they were only able to achieve higher k_{cat} on a Cytochrome P450 propane monooxygenase via a selection for stability that resulted in an consequent decrease in activity[42]. Additional studies have further demonstrated the importance of stability in the evolvability of a protein[35,43], the presence of a tradeoff between stability and function[44,45], and that mutations that impart activity are generally destabilizing[46]. Therefore, the difficulty of *Tm*HisA to readily evolve Trp1 function, despite its origin from a thermophile with an excess of stability in a mesophilic organism, may indicate that there are additional driving forces beyond protein stability.

Conclusion

Despite the challenges in evolving *Tm*HisA for Trp1 activity, we demonstrate OrthoRep's ability to explore complex fitness landscapes. Through high-replicate evolution experiments, we not only observed a plethora of novel mutations that impart Trp1 activity, but also to probe multiple evolutionary conditions in search of a successful solution. Specifically, we were able to evolve highly successful mutations of *Tm*HisA with high Trp1 activity, albeit only with an alternative selection step for *HIS6* activity. To conduct such an experiment using traditional mutagenesis techniques would have been a significant undertaking, requiring multiple rounds of

mutagenesis on multiple culture conditions. While *Tm*HisA may simply be a model evolution system, the data suggests that in previous directed evolution experiments where the gains in function were observed to non-existent or limited, alternative selections may be a possible strategy to overcome the barrier to higher activity. This adds another dimension for experimental evolution and questions whether traditional selections with a singular pressure is truly the most efficient way of achieving high or novel activities.

References

(1) Williams, G. J.; Goff, R. D.; Zhang, C.; Thorson, J. S. Optimizing Glycosyltransferase Specificity via “Hot Spot” Saturation Mutagenesis Presents a Catalyst for Novobiocin Glycorandomization. *Chem. Biol.* **2008**, *15* (4), 393–401. https://doi.org/10.1016/j.chembiol.2008.02.017.

(2) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic Asymmetric Synthesis of Sitagliptin Manufacture. *Science (80-.).* **2010**, *329* (July), 305–310. https://doi.org/10.1126/science.1188934.

(3) Zhang, K.; Sawaya, M. R.; Eisenberg, D. S.; Liao, J. C. Expanding Metabolism for Biosynthesis of Nonnatural Alcohols. *Proc. Natl. Acad. Sci.* **2008**, *105* (52), 20653–20658. https://doi.org/10.1073/pnas.0807157106.

(4) Bosma, T.; Damborský, J.; Stucki, G.; Janssen, D. B. Biodegradation of 1,2,3-Trichloropropane through Directed Evolution and Heterologous Expression of a Haloalkane Dehalogenase Gene. *Appl. Environ. Microbiol.* **2002**, *68* (7), 3582–3587. https://doi.org/10.1128/AEM.68.7.3582-3587.2002.

(5) Kumamaru, T.; Suenaga, H.; Mitsuoka, M.; Watanabe, T.; Furukawa, K. Enhanced Degradation of Polychlorinated Biphenyls by Directed Evolution of Biphenyl Dioxygenase. *Nat. Biotechnol.* **1998**, *16* (7), 663–666. https://doi.org/10.1038/nbt0798-663.

(6) Ling, H.; Teo, W.; Chen, B.; Leong, S. S. J.; Chang, M. W. Microbial Tolerance Engineering toward Biochemical Production: From Lignocellulose to Products. *Curr. Opin. Biotechnol.* **2014**, *29* (1), 99–106. https://doi.org/10.1016/j.copbio.2014.03.005.

(7) Abatemarco, J.; Hill, A.; Alper, H. S. Expanding the Metabolic Engineering Toolbox with Directed Evolution. *Biotechnol. J.* **2013**, *8* (12), 1397–1410. https://doi.org/10.1002/biot.201300021.

(8) Anderson, C. Registry of Standard Biological Parts.

(9) Lee, M. E.; DeLoache, W. C.; Cervantes, B.; Dueber, J. E. A Highly-Characterized Yeast Toolkit for Modular, Multi-Part Assembly. *ACS Synth. Biol.* **2015**, 150414151809002. https://doi.org/10.1021/sb500366v.

(10) Obexer, R.; Pott, M.; Zeymer, C.; Griffiths, A. D.; Hilvert, D.; Grif, D.; Hilvert, D. Efficient Laboratory Evolution of Computationally Designed Enzymes with Low Starting Activities Using Fluorescence-Activated Droplet Sorting. *Protein Eng. Des. Sel.* **2016**, *29* (9), 355–356. https://doi.org/10.1093/protein/gzw032.

(11) Bloom, J. D.; Romero, P. A.; Lu, Z.; Arnold, F. H. Neutral Genetic Drift Can Alter Promiscuous Protein Functions, Potentially Aiding Functional Evolution. *Biol. Direct* **2007**, *2*, 7–10. https://doi.org/10.1186/1745-6150-2-17.

(12) Goldsmith, M.; Tawfik, D. S. Enzyme Engineering: Reaching the Maximal Catalytic Efficiency Peak. *Curr. Opin. Struct. Biol.* **2017**, *47*, 140–150. https://doi.org/10.1016/j.sbi.2017.09.002.

(13) Bloom, J. D.; Arnold, F. H. In the Light of Directed Evolution: Pathways of Adaptive Protein Evolution. *Proc. Natl. Acad. Sci.* **2009**, *106*, 9995–10000. https://doi.org/10.17226/12692.

(14) Glieder, A.; Farinas, E. T.; Arnold, F. H. Laboratory Evolution of a Soluble, Self-Sufficient, Highly Active Alkane Hydroxylase. *Nat. Biotechnol.* **2002**, *20* (11), 1135–1139. https://doi.org/10.1038/nbt744.

(15) Peters, M. W.; Meinhold, P.; Glieder, A.; Arnold, F. H. Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. **2003**, No. 3, 13442–13450. https://doi.org/10.1021/ja0303790.

(16) Boder, E. T.; Midelfort, K. S.; Wittrup, K. D. Directed Evolution of Antibody Fragments with Monovalent Femtomolar Antigen-Binding Affinity. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (20), 10701–10705. https://doi.org/10.1073/pnas.170297297.

(17) Khersonsky, O.; Tawfik, D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annual Review of Biochemistry*. 2010, pp 471–505. https://doi.org/10.1146/annurev-biochem-030409-143718.

(18) Kaltenbach, M.; Emond, S.; Hollfelder, F.; Tokuriki, N. Functional Trade-Offs in Promiscuous Enzymes Cannot Be Explained by Intrinsic Mutational Robustness of the Native Activity. *PLoS Genet.* **2016**, *12* (10), 1–18. https://doi.org/10.1371/journal.pgen.1006305.

(19) Amitai, G.; Gupta, R. D.; Tawfik, D. S. Latent Evolutionary Potentials under the Neutral Mutational Drift of an Enzyme. *HFSP J.* **2007**, *1* (1), 67–78. https://doi.org/10.2976/1.2739115/10.2976/1.

(20) Fasan, R.; Meharenna, Y. T.; Snow, C. D.; Poulos, T. L.; Arnold, F. H. Evolutionary History of a Specialized P450 Propane Monooxygenase. *J. Mol. Biol.* **2008**, *383* (5), 1069–1080. https://doi.org/10.1016/j.jmb.2008.06.060.

(21) Nannermann, D. P.; Birmingham, W. R.; Scism, R. A.; Bachmann, B. O.; Nannemann, D. P.; Birmingham, W. R.; Scism, R. A.; Bachmann, B. O. *Assessing Directed Evolution Methods for the Generation of Biosynthetic Enzymes with Potential in Drug Biosynthesis*; 2011; Vol. 3, pp 803–819. https://doi.org/10.4155/fmc.11.48.

(22) Barona-Gómez, F.; Hodgson, D. A. Occurrence of a Putative Ancient-like Isomerase Involved in Histidine and Tryptophan Biosynthesis. *EMBO Rep.* **2003**, *4* (3), 296–300. https://doi.org/10.1038/sj.embor.embor771.

(23) Jürgens, C.; Strom, A.; Wegener, D.; Hettwer, S.; Wilmanns, M.; Sterner, R. Directed Evolution of a (Bα)8-Barrel Enzyme to Catalyze Related Reactions in Two Different Metabolic Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (18), 9925–9930. https://doi.org/10.1073/pnas.160255397.

(24) Lundin, E.; Näsvall, J.; Andersson, D. I. Mutational Pathways and Trade-Offs Between HisA and TrpF Functions: Implications for Evolution via Gene Duplication and Divergence. *Front. Microbiol.* **2020**, *11* (October), 1–12. https://doi.org/10.3389/fmicb.2020.588235.

(25) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **2018**, *175* (7), 1946-1957.e13. https://doi.org/10.1016/j.cell.2018.10.021.

(26) Zhong, Z.; Wong, B. G.; Ravikumar, A.; Arzumanyan, G. A.; Khalil, A. S.; Liu, C. C. Automated Continuous Evolution of Proteins in Vivo. *ACS Synth. Biol.* **2020**, *9* (6), 1270–1276. https://doi.org/10.1021/acssynbio.0c00135.

(27) Ravikumar, A.; Arrieta, A.; Liu, C. C. An Orthogonal DNA Replication System in Yeast. *Nat. Chem. Biol.* **2014**, *10* (3), 175–177. https://doi.org/10.1038/nchembio.1439.

(28) Zhong, Z.; Ravikumar, A.; Liu, C. C. Tunable Expression Systems for Orthogonal DNA Replication. *ACS Synth. Biol.* **2018**, *7* (12), 2930–2934. https://doi.org/10.1021/acssynbio.8b00400.

(29) Zhao, X. Q.; Bai, F. W. Yeast Flocculation: New Story in Fuel Ethanol Production.

Biotechnol. Adv. **2009**, *27* (6), 849–856. https://doi.org/10.1016/j.biotechadv.2009.06.006.

(30) Fidalgo, M.; Barrales, R. R.; Ibeas, J. I.; Jimenez, J. Adaptive Evolution by Mutations in the FLO11 Gene. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (30), 11228–11233. https://doi.org/10.1073/pnas.0601713103.

(31) Watanabe, J.; Uehara, K.; Mogi, Y. Adaptation of the Osmotolerant Yeast Zygosaccharomyces Rouxii to an Osmotic Environment through Copy Number Amplification of FLO11D. *Genetics* **2013**, *195* (2), 393–405. https://doi.org/10.1534/genetics.113.154690.

(32) Smukalla, S.; Caldara, M.; Pochet, N.; Beauvais, A.; Guadagnini, S.; Yan, C.; Vinces, M. D.; Jansen, A.; Prevost, M. C.; Latgé, J. P.; Fink, G. R.; Foster, K. R.; Verstrepen, K. J. FLO1 Is a Variable Green Beard Gene That Drives Biofilm-like Cooperation in Budding Yeast. *Cell* **2008**, *135* (4), 726–737. https://doi.org/10.1016/j.cell.2008.09.037.

(33) Romanini, D. W.; Peralta-yahya, P.; Mondol, V.; Cornish, V. W. A Heritable Recombination System for Synthetic Darwinian Evolution in Yeast. *ACS Synth. Biol.* **2012**, *1*, 602–609.

(34) Howson, R.; Huh, W. K.; Ghaemmaghami, S.; Falvo, J. V.; Bower, K.; Belle, A.; Dephoure, N.; Wykoff, D. D.; Weissman, J. S.; O'Shea, E. K. Construction, Verification and Experimental Use of Two Epitope-Tagged Collections of Budding Yeast Strains. *Comp. Funct. Genomics* **2005**, *6* (1–2), 2–16. https://doi.org/10.1002/cfg.449.

(35) Bershtein, S.; Goldin, K.; Tawfik, D. S. Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J. Mol. Biol.* **2008**, *379* (5), 1029–1044. https://doi.org/10.1016/j.jmb.2008.04.024.

(36) Smith, W. S.; Hale, J. R.; Neylon, C. Applying Neutral Drift to the Directed Molecular

Evolution of a B-Glucuronidase into a b-Galactosidase: Two Different Evolutionary Pathways Lead to the Same Variant. *BMC Res. Notes* **2011**, *4* (1), 138. https://doi.org/10.1186/1756-0500-4-138.

(37) Steinberg, B.; Ostermeier, M. Environmental Changes Bridge Evolutionary Valleys. *Sci. Adv.* **2016**, *2* (1). https://doi.org/10.1126/sciadv.1500921.

(38) Huber, R.; Langworthy, T. A.; König, H.; Thomm, M.; Woese, C. R.; Sleytr, U. B.; Stetter, K. O. Thermotoga Maritima Sp. Nov. Represents a New Genus of Unique Extremely Thermophilic Euhacteria Growing up to 90 ° C. *Arch. Microbiol.* **1986**, *144*, 324–333.

(39) Näsvall, J.; Sun, L.; Roth, J. R.; Andersson, D. I. Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. *Science (80-.).* **2012**, *338* (6105), 384–387. https://doi.org/10.1126/SCIENCE.1226521.

(40) Siddiqui, K. S.; Cavicchioli, R. Cold-Adapted Enzymes. *Annu. Rev. Biochem.* **2006**, *75*, 403–433. https://doi.org/10.1146/annurev.biochem.75.103004.142723.

(41) Smal S, A. O.; Leiros, H. K. S.; Os, V.; Willassen, N. P. Cold Adapted Enzymes. *Biotechnol. Annu. Rev.* **2000**, *6*, 1–57. https://doi.org/10.1016/S1387-2656(00)06018-X.

(42) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci.* **2006**, *103* (15), 5869–5874. https://doi.org/10.1073/PNAS.0510098103.

(43) Goldsmith, M.; Aggarwal, N.; Ashani, Y.; Jubran, H.; Greisen, P. J.; Ovchinnikov, S.; Leader, H.; Baker, D.; Sussman, J. L.; Goldenzweig, A.; Fleishman, S. J.; Tawfik, D. S. Overcoming an Optimization Plateau in the Directed Evolution of Highly Efficient Nerve Agent Bioscavengers. *Protein Eng. Des. Sel.* **2017**, *30* (4), 333–345.

https://doi.org/10.1093/PROTEIN/GZX003.

(44) Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. A Relationship between Protein Stability and Protein Function. *Proc. Natl. Acad. Sci.* **1995**, *92* (2), 452–456. https://doi.org/10.1073/PNAS.92.2.452.

(45) Beadle, B. M.; Shoichet, B. K. Structural Bases of Stability–Function Tradeoffs in Enzymes. *J. Mol. Biol.* **2002**, *321* (2), 285–296. https://doi.org/10.1016/S0022-2836(02)00599-5.

(46) Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How Protein Stability and New Functions Trade Off. *PLoS Comput. Biol.* **2008**, *4* (2), 35–37. https://doi.org/10.1371/journal.pcbi.1000002.

Table 4.1. Summary of mutations identified in NGS sequencing of cultures at the Plateau point. Letters indicate amino acid changes found at their respective positions, and multiple letters indicate multiple mutations observed in the same culture. Color intensity denotes frequency of mutations present in the culture.

Position	3	4	6	7	13	14	15	28	33	36	53	56	58	59	60	62	63	64	68	70	71	74	84	92	96	108	114	122	127	133	134	141	144	148	154	157	166	168	169	170	187	195	204	211	219	221	225	236	239
wt AA	V	V	A	I	K	V	A	Y	V	V	S	I	N	S	G	N	L	P	K	S	E	E	S	R	Y	D	S	P	D	V	A	E	I	S	K	G	I	K	D	G	A	G	K	T	G	I	A	M	Y
Plateau-1		A			R						L				S	S						AV			C	G	P				V										T			I					
Plateau-2											L				DS	S																G				S		E										
Plateau-3											L							L															NS							V							T	
Plateau-4							T		I															Q					G										D									
Plateau-5				M		A				M		T	S															A					DNS		ER			E	N			D						
Plateau-6			V	M						M	L	T	S		DS		H						P							I			N		E	S	V	E	GN			DV	E			T	V		H
Plateau-7	I								A		L				NS					P													P		S										T	V		
Plateau-8	I							CH	A		L			GN	INS				R	P	G												P		S									D		V	IL	

Mutational Frequency

0 0.5 1.0

Table 4.2. Summary of mutations identified in NGS sequencing of descendents of Plateau-1 after alternative selection or drift (Timepoint 1) and after reselection for Trp1 activity (Timepoint 2). Cultures are named by the Timepoint-condition-replicate scheme. H denotes cultures that have underwent alternative selection for His6 activity. W denotes cultures that underwent continued strong selection for Trp1 activity. UL denotes cultures that underwent selection in conditions not limited by tryptophan or histidine.

Position	4	13	18	27	28	33	36	38	47	53	54	55	60	62	72	73	74	96	108	111	113	114	122	134	135	138	145	157	170	185	187	188	191	211	217
wt AA	V	K	I	F	Y	V	V	K	I	S	N	A	G	N	F	A	E	Y	D	F	K	S	P	A	F	W	D	G	G	I	A	E	V	T	L
Plateau-1	A	R						R		L			S	S			AV	C	G			P		V								T			I	
1-H-1		R						R	V	L			S	S			V	C	G			P		V	LS			D				T			I	P
1-H-2		R						R		L			S	S	L		V	C	G			P		V								T			I	
1-H-3	A	R	M		H			R		L			S	S				C				P				R								A	I	
1-H-4		R			C		M	R	F	L			S	S			V	C	G			P		V					V		T			I	
1-H-5	A	R						R		L			NS	S			DV	C				P							D					I	
1-H-6		R				AI		R		L			NS	S			V	C	G			P		V							T			I	
2-H-1		R					IM	R	V	L			S	S			V	C	G			P		V	L			D			T			I	P
2-H-2		R		SV				R		L			S	S	L		V	C	G			P		V							T		AD	I	
2-H-3	A	R	M					R		L			S	S				C				P											A	I	
2-H-4		R		S	C		M	R	F	L			S	S			V	C	G			P		V						T	T			I	
2-H-5	A	R						R		L			S	S			V	C	G			P		V					D					I	
2-H-6	A	R	MT	V		AI	M	R		L			S	S	L		V	C	G			P		V					SV		T			I	
1-W-1		R	T			I	AM	R		L			S	S	L		V	C	G			P		V	L			DS			T			I	
1-W-2		R						R		L			S	S		V	V	C				P						D			T			I	
1-W-3	A	R				AI		R	V	L			S	S			V	C	G			P		V							T			I	
1-W-4		R						R		L			S	S			V	C	G			P		V					DV		T			I	
1-W-5	A	R				AIT		R		L			S	S				C				P		V					AV			DGK		I	
1-W-6	A	R			C			R		L			S	S			V	C	G			P		V							T			I	
2-W-4		R			C			R		L			S	S			V	C	G			P		V					D		T			I	
2-W-5	A	R				A		R		L			S	S			G	C				P												I	
2-W-6	A	R						R		L			S	S		V	V	C	G			P		V							T			I	
1-UL-1	A	R	M					R	V	L			S	S			V	C	G			AP		V					D				A	I	
1-UL-2	A	R	M					R		L			S	S				C				P						D					A	I	
1-UL-3	A	R						R		L			S	S				C				P	S					D						I	
1-UL-4	A	R						R		L			S	S			V	C	G			P		V	Y						T		A	I	
1-UL-5	A	R						R		L			S	S				C		CS		P	S							TV				I	
1-UL-6	A	R			C		M	R		L			S	S		V	Q	C	G			P		V				D	DV					I	
2-UL-1	A	R			H			R		L		T	S	S				C			R	P					V	D		T		K		I	P
2-UL-2	A	R						R		L	D	T	S	S				C			R	P			LS		IV			T		K		I	P
2-UL-4		R						R		L	Y		S	S		P	A	C	G			P		V							T		A		
2-UL-6	A	R				I		R		L		T	S	S	L			C		L		P					V			T		K		I	P

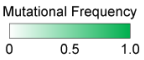


Table 4.3. Summary of mutations identified in NGS sequencing of descendents of Plateau-2 after alternative selection or drift (Timepoint 1) and after reselection for Trp1 activity (Timepoint 2). Cultures are named by the Timepoint-condition-replicate scheme. H denotes cultures that have underwent alternative selection for His6 activity. W denotes cultures that underwent continued strong selection for Trp1 activity. UL denotes cultures that underwent selection in conditions not limited by tryptophan or histidine.

Position	4	14	18	22	33	36	53	56	60	62	105	135	141	144	156	168	170	204	221	226
wt AA	V	V	I	K	V	V	S	I	G	N	V	F	E	I	Y	K	G	K	I	F
Plateau-2							L		DS	S			G			E				
1-H-1		A					L		AS			LS	G		CH	E				S
1-H-2			M				L		S				G		CH	E				S
1-H-3		A					L		S	KS			G		H	EV		E		
1-H-4		ALT					L		S	KS			G		H	EG				
1-H-5	A		M				L		S			LS	G		CH	E				S
1-H-6			MT	R			L		NS	DS		LY	GR		H	EV		E		
2-H-1						M	L		S	S		S	G	N	C	E				S
2-H-2	A	A	M			M	L	T	S	S	A	LS	G	N	C	E	V			S
2-H-3		A	T		A	M	L	VT	S	S		LS	G	N	CH	EG	D			
2-H-4	A				I		L		S	S		V	G			E				
2-H-5		A	M				L	N	S			LS	G		CH	ER		R		S
2-H-6						M	L	T	DNS			L	G	N	H	E	V	E		
1-W-1					I		L		S				G			E				
1-W-2		A			AI	M	L		S	S			G		CH	E				
1-W-3	A	I	M		I		L		S				G		H	EG		E		
1-W-4	A	A	M		I		L		S	S			G			E	D			
1-W-5	AL	A					L	T	DRS	DS			G		H	E		ER		
1-W-6		AI			AI	M	LM		S				G	NT	CH	EG				
2-W-1					I		L		S							E				
2-W-2					I		L		NS	S	A					EG	A			
2-W-5					I		L		S		A		G	N		E				
1-UL-1		A					L		S	S			G		C	E			T	
1-UL-2		A					L		S				G			E				
1-UL-3	A	A		R			L		S	S			G		C	E			T	
1-UL-4		A		R		M	L	T	S				G	N	C	E	V	ER	T	
1-UL-5		A				M	L	TM	S			S	G		C	E				
1-UL-6		A		R			L		DS			LS	G		C	E			T	S
2-UL-3		A		R			L		S		A		G		C	E			T	
2-UL-4		A		R			L	M	S	S			G	V	CH	E			T	
2-UL-6		T	M				L	T	S				G	N		E				

Mutational Frequency

0 0.5 1.0

Table 4.4. Summary of mutations identified in NGS sequencing of descendents of Plateau-3 after alternative selection or drift (Timepoint 1) and after reselection for Trp1 activity (Timepoint 2). Cultures are named by the Timepoint-condition-replicate scheme. H denotes cultures that have underwent alternative selection for His6 activity. W denotes cultures that underwent continued strong selection for Trp1 activity. UL denotes cultures that underwent selection in conditions not limited by tryptophan or histidine.

Position	4	10	15	53	60	64	72	75	87	92	127	144	157	170	187	194	236
wt AA	V	F	A	S	G	P	F	H	Y	R	D	I	G	G	A	A	M
Plateau-3				L		L						NS		V			T
1-H-1	A	SY		L		L		Y				N		AV		V	T
1-H-2				L		LR						N		V			T
1-H-3			TV	L		L						N		V			T
1-H-4				L	DS	L						N		V			T
1-H-5			T						H	Q							
1-H-6		Y		L		L					GH	N		V			T
2-H-1		S		L	SV	LT		DY				N		ADV		V	LT
2-H-2		S		L		L						N	ADS	AV		GV	T
2-H-3	A		TV	L	NS	L	LY					NS		V	T		LRT
2-H-4			T	L	DS	LT			H	Q	G	NT					
2-H-5			PT	L	CDS	LST	LS		H	Q			DS	DV	TV		T
2-H-6	A	SY	T	LV	NS	LR	CS	DY				NY	S	V		V	T
1-W-1				L		LT						NSV		DIV			ALT
1-W-2	A			L	DS	L			CH			N		AV	T		T
1-W-3				L		L						N		V			T
1-W-4		LY	T	LW	SV	L			H			N	S	DV	T		T
1-W-5				L	NST	LT						NST		DIV			TV
1-W-6		Y		L		LT						N		DV	T		LTV
2-W-2		LS		L		LT		Y	C			NS		DV		EV	T
2-W-4			ST	L		LT	LS					NS		DV		V	T
2-W-5			T		NS					PQ	G			D			
1-UL-1				FLM	S	L	L					NY		DV			T
1-UL-2				L		L						N		V			T
1-UL-3				L		L	LS					N		AV			T
1-UL-4				L		LT						NS		DIV		V	LT
1-UL-5			TV		S	L					G	NT		V			T
1-UL-6				L		LT						DN		DV			T
2-UL-6				L		L						MNS		ADV			T

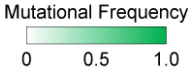


Table 4.5. Summary of mutations identified in NGS sequencing of descendents of Plateau-4 after alternative selection or drift (Timepoint 1) and after reselection for Trp1 activity (Timepoint 2). Cultures are named by the Timepoint-condition-replicate scheme. H denotes cultures that have underwent alternative selection for His6 activity. W denotes cultures that underwent continued strong selection for Trp1 activity. UL denotes cultures that underwent selection in conditions not limited by tryptophan or histidine.

Position	4	15	18	30	33	53	57	60	73	82	87	92	120	127	138	144	145	149	158	160	162	164	168	170	185	188	228
wt AA	V	A	I	K	V	S	E	G	A	I	Y	R	V	D	W	I	D	L	L	E	V	T	K	G	I	E	E
Plateau-4		T			I							Q		G										D			
1-H-1		T									H	Q															
1-H-2		T									H	Q														G	
1-H-3		T									H	Q									I						
1-H-4		T									H	HQ															
1-H-5		T									H	Q											R				
1-H-6		T			I							Q															
● 2-H-1		T	T							V	H	Q		G		T		P	F			R			V	G	
● 2-H-2		T	T							AV	H	Q		G		T		P	F			R			V	G	
● 2-H-3		T									H	Q		G			V			G	I				T	G	
2-H-4		T				L		S			H	Q				N							ER	DV			
2-H-5	A	T	V			L				V	H	Q			R								R				
2-H-6	A	T			I	L		S				Q		G			G							D			
1-W-1		MT		E	I							Q		G										DV			
1-W-2		T		E	I	L						Q		G	R									D			
1-W-3		T	V	E	IM	L		SV				Q		G	R									D			
1-W-4		T		E	IM							Q		G										D			
1-W-5		T		E	IM							Q		G										D			
1-W-6		T		E	I	L						Q		G	R									DV			
2-W-1		T	V	E	I	L		S			H	Q		G		T		P	F					D			
2-W-2	A	T		E	I	L						Q		G					P						D			
1-UL-1		T			I	L	G	S	T		H	Q		G	R									D			
1-UL-2		T					G		T			Q		G										D			V
1-UL-3	A	T			I	L	G	S	T	V		Q		G										D			V
1-UL-4	A	T			IL		G	DS	T			Q		G										D			DV
1-UL-5		T				L	G		T		H	Q		G	R								R	D			V
1-UL-6		T			I		G		T			Q		G	R									D			V
2-UL-1		T		E	I							Q		G		V								D			
2-UL-3	A	T			I							Q	M	G										D			
2-UL-4	A	T			I		G		T			Q	M	G										D			

● Denotes culture that has evolved to complement a Trp1 deletion

Mutational Frequency

0 0.5 1.0

Table 4.6. Summary of mutations identified in NGS sequencing of descendents of Plateau-5 after alternative selection or drift (Timepoint 1) and after reselection for Trp1 activity (Timepoint 2). Cultures are named by the Timepoint-condition-replicate scheme. H denotes cultures that have underwent alternative selection for His6 activity. W denotes cultures that underwent continued strong selection for Trp1 activity. UL denotes cultures that underwent selection in conditions not limited by tryptophan or histidine.

Position	4	7	13	14	18	24	27	28	33	36	38	53	56	58	60	73	82	122	144	151	154	156	168	169	185	195	201
wt AA	V	I	K	V	I	N	F	Y	V	V	K	S	I	N	G	A	I	P	I	K	K	Y	K	D	I	G	N
Plateau-5		M		A				C		M			T	S				A	DNS		ER		E	N		D	
1-H-1								C		M																	S
1-H-2	I					D		C		M															T		KS
1-H-3						D		C		M												H			T		S
1-H-4								C		M																	KS
1-H-5						D		C		M												H			T		S
1-H-6								C		M																	KS
2-H-1								C		M					S	T											S
2-H-2						D		C		M												H					KS
2-H-3	A					D	S	C		M												H					KS
2-H-4					V	D	L	C		IM		L			S							H					KS
2-H-5						D		C		M				S					N			H					S
2-H-6								C		M	R	L			S												S
1-W-1							LS	C		M					S							H					S
1-W-2							LSV	C		M					S												S
1-W-3								CD		M																	S
1-W-4							SV	C		M		L				V							ER				KS
1-W-5								C		M					DS							CH					S
1-W-6	A						LV	C		M		L			DS							H					S
2-W-5								C		M												H					S
1-UL-1	A		R					C	A	M	R	L			S	T	V			R							S
1-UL-2			E		V			C	A	M					S	TV						H			V		S
1-UL-4								C		M							V			R							S
1-UL-5	A		R	A			L	C		M	R	L			S	TV	V			R			E	G			S
1-UL-6								C		M	R						V			R							S
2-UL-2					V			C	A	M						V						H			V		S
2-UL-4								C		M												H					S
2-UL-5			R		T		L	C		M		L			S	V	N					H					S

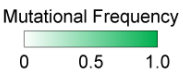


Table 4.7. Summary of mutations identified in NGS sequencing of descendents of Plateau-6 after alternative selection or drift (Timepoint 1) and after reselection for Trp1 activity (Timepoint 2). Cultures are named by the Timepoint-condition-replicate scheme. H denotes cultures that have underwent alternative selection for His6 activity. W denotes cultures that underwent continued strong selection for Trp1 activity. UL denotes cultures that underwent selection in conditions not limited by tryptophan or histidine.

Position	4	6	7	15	26	36	50	53	56	58	60	63	73	75	81	82	84	100	112	121	127	133	135	144	154	156	157	166	168	169	182	185	195	197	204	220	221	225	239
wt AA	V	A	I	A	I	V	V	S	I	N	G	L	A	H	G	I	S	I	L	E	D	V	F	I	K	Y	G	I	K	D	K	I	G	I	K	V	I	A	Y
Plateau-6		V	M			M		L	T	S	DS	H					P					I		N	E		S	V	E	GN			DV		E		T	V	H
1-H-1				V		M				S										D		I	L	N		H									M				
1-H-2	AT					M				S								V				AIT		N		H													
1-H-3				V		M				S												I		N		H													
1-H-4				V		M				S	DS											I	L	N				V			R	T							
1-H-5						M				S										D		I	L	N		H									M				
1-H-6					M	M				S				Y								T		N		H													H
● 2-H-1				V		M	M			S	DS									D		I	L	N		H									M				
● 2-H-2						M				S								V			V	T		N		H								V					
● 2-H-4			V	V			A			S	D								M			I	L	N				V			R	T		V					
● 2-H-5						M	A			S	C			Y		V				D	H	ILT	L	N		H									M				
● 2-H-6				V		M	A			S	CD		T	Y		V				D	H	IT	L	N		H									I				H
1-W-1			MT	TV		M			T		DS													N	E				E	N			CD						H
1-W-2			M			M			T	S												I		DN	E				E	N			D						H
1-W-3		T	M	T		M			T						R									N	E				E	N			D						H
1-W-4			M	T		M	AM		T						R									N	E				E	N			D			M			H
1-W-5			M	T		M			T	DS														N	E			TV	E	N			D						H
1-W-6		V	M	T		M		L	T	S	DS	H			R							I		N	E				E	N			D						H
2-W-2			MT		M	M			T	S														N	E	H			E	N			D						H
2-W-3	A					M		L	T	S	S											I		N	E			T	E	N	E	F	D						
2-W-6		V	V	V		M			T	S	DS	H					P			G		I		N	E			V	E	N			DV						
1-UL-1						M		L	T	S	S		T									I	L	N	E			V	E	N		T	D				V		
1-UL-2	A		M			M			T	S			V				P			G		I		N	E			AV	E	N			DV						
1-UL-3	A		M			LM			T															N	E				E	N			D						
1-UL-4	A				M	M		LT	T		S													N	E				E	N			D				T		
1-UL-5			M		M	M		L		S	S		T									I		N	E				E	N			D						
1-UL-6	A				M	M				S									P					DN						KN		ST		V					

● Denotes culture that has evolved to complement a Trp1 deletion

Mutational Frequency

0 0.5 1.0

Table 4.8. Summary of mutations identified in NGS sequencing of descendents of Plateau-7 after alternative selection or drift (Timepoint 1) and after reselection for Trp1 activity (Timepoint 2). Cultures are named by the Timepoint-condition-replicate scheme. H denotes cultures that have underwent alternative selection for His6 activity. W denotes cultures that underwent continued strong selection for Trp1 activity. UL denotes cultures that underwent selection in conditions not limited by tryptophan or histidine.

Position	3	26	29	33	36	53	56	59	60	64	67	70	71	72	73	89	101	111	138	140	144	148	157	170	177	190	197	199	201	221	225	236
wt AA	V	I	E	V	V	S	I	S	G	P	E	S	E	F	A	E	V	F	W	A	I	S	G	G	F	K	I	S	N	I	A	M
Plateau-7	I			A		L			NS			P										P	S							T	V	
1-H-1									S		GR			L	T	K																
1-H-2	I					L			S		G			L	T											R		T				
1-H-3	I	T							S						T	K																
1-H-4	I								NS		G			L	T																	
1-H-5	I								S		G			L	T						T					R						
1-H-6	I								S		G			L	T											R						
2-H-1						L			NS			P		L	T	K						P	DS		S					N	V	
2-H-2					AM	L			NS	L		P									N	P	DS	D							V	
2-H-3	I					L			N	L		P			T			S				P	S		S						V	
2-H-4	I		G		M	L			NS	L		P		L				Y			N	P	S	D				L	S			T
2-H-5	I				M	L			N	L		P	G	S				L			N	P	S	V	S			P	S		V	IT
2-H-6	I	M			M	L			S	L				L	T						N		S	D					S			T
1-W-1	I	MT		A		L			NS			P	G						R			P	S								V	
1-W-2	I					L		G	NS			P										P	S								V	A
1-W-3	I	MTV		A	M	L			NS			P							R			P	S								V	AIT
1-W-4	I	M		A		L			NS			P	G								T	P	S								V	
1-W-5	I			AI	LM	L			NS			P	G						R		NT	P	S								V	AILT
1-W-6	I	M		A	M	L			NS	L		P		L	TV				R		NS	P	S								V	T
1-W-2	I				M	L	TV	G	NS			P		L					R			P	S					P			V	A
1-W-3	I			I	M	L		G	S	L		P			V	G			R		N	P	S					P	S		V	AET
1-UL-1	I	M	K	I	M	L			S						T		M			T	N						AV					
1-UL-2	I		K			L			S	L				L	T		M			T							V					T
1-UL-3			K			L	V		S						T		M			T	T						SV					
1-UL-4	I				M	L			S	L					T		M			T	N											T
1-UL-5	I		K			L			S	L					T		M		R	T	N											T
1-UL-6	I					L	V		S	LT					T					T												
2-UL-2						L	T		S				G		T																	

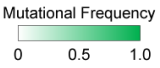
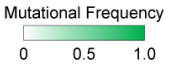


Table 4.9. Summary of mutations identified in NGS sequencing of descendents of Plateau-8 after alternative selection or drift (Timepoint 1) and after reselection for Trp1 activity (Timepoint 2). Cultures are named by the Timepoint-condition-replicate scheme. H denotes cultures that have underwent alternative selection for His6 activity. W denotes cultures that underwent continued strong selection for Trp1 activity. UL denotes cultures that underwent selection in conditions not limited by tryptophan or histidine.

Position	3	4	15	18	20	26	33	36	37	53	59	60	68	70	71	72	73	78	87	138	142	148	157	191	197	215	219	225	236
wt AA	V	V	A	I	G	I	V	V	E	S	S	G	K	S	E	F	A	I	Y	W	E	S	G	V	I	G	G	A	M
Plateau-8	I						A			L	GN	INS	R	P	G							P	S				D	V	IL
1-H-1		A	T	MV	E	MV				L						L			CH	R		P		D	V	EV			
1-H-2		A	RT	V	E	V				L										R				D	V	V	D		
1-H-3		AT	T	ADV	EKR	V				L										R	G			DG	V	V			
1-H-4		A	T	V	E	V				L								T		R			DS	D	AV	V			
1-H-5		A	T	LV	E	V				L									CH	R				D	V	V			
1-H-6		A	T	V	E	V				L										R				D	V	V			
2-H-1		A	T	V	E	V		M		L						L			C	R	G			D	V	V			
2-H-2		A	T	V	E	V		M		L									H					D		V			
2-H-3	I	A	T	V	E	V		A		L		S				L		V		R	G			D	V	VW			
2-H-4		A			E	V			K	L		S					T	T		R				D	V	V			
2-H-5		A	T	V	E	V				L						L			CH	R				D	V	V			
2-H-6		A	T	V	E	V		I		L			E				V			R				D	V	V			
1-W-1		A	T	MV	E	V				L										R				D	V	V			
1-W-2		A	T	V	E	AV		IM		L							TV	T		R				D	V	MV			
1-W-3		A	T	V	E	V		AI	K	L									H	R				D	V	V			
1-W-4		A	T	V	E	AV	AI		K	L										R				D	V	V			
1-W-5		A	T	ALV	E	ALV	IL			L										R				D	V	V			
1-W-6		A	T	LV	E	V				L										R			D	D	V	V			
2-W-3		A	T	V	E	V		M		L										R				D	V	V			
2-W-5		A		AV	E	V			K	L						L			H	R				D	V	V			
2-W-6		A	T	V	E	V	I			L						S				R			D	D	V	V			
1-UL-1		A	T	V	E	V			K	L		S					T			R	G			D	V	V			
1-UL-2		A		V	E	V			K	L										R				D	V	V			
1-UL-3		A	T	V	E	V		I	G	L									H	R				D	V	V			
1-UL-4		A	T	V	E	V	AI			L										R				D	V	V	D		
1-UL-5		A	T	V	E	V			K	L										R				D	V	V			
1-UL-6		A	T	V	E	AV	I		K	L							V			R				D	V	V			
2-UL-5		A	T	V	E	V	I	I		L										R				D	V	V			



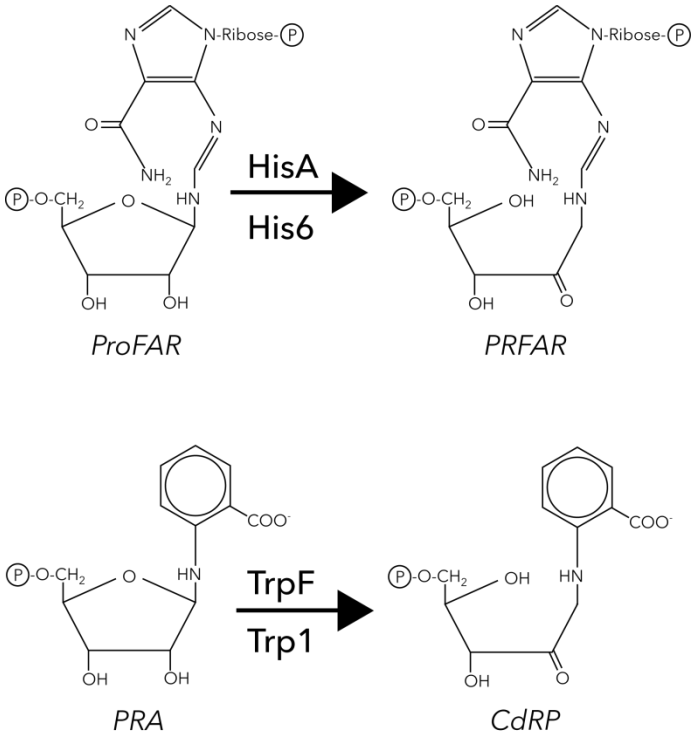


Figure 4.1. HisA (or designated His6 in eukaryotes) catalyzes the isomerization of ProFAR to PRFAR. Similarly, TrpF (or designated Trp1 in eukaryotes) catalyzes the isomerization of PRA to CdRP. The two reactions are the same isomerization reaction on two similar substrates.

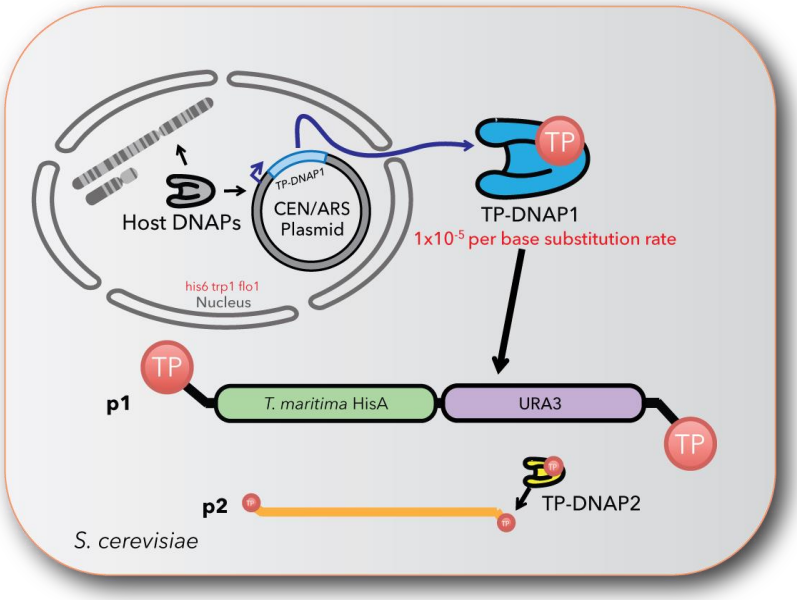


Figure 4.2. OrthoRep strain containing a mutagenic TP-DNAP1 and encoding *Tm*HisA and a selection marker on p1. A mutagenic TP-DNAP1 that only replicates p1 is expressed from a nuclear expression plasmid. This allow mutations to be specifically directed to genes encoded on p1.

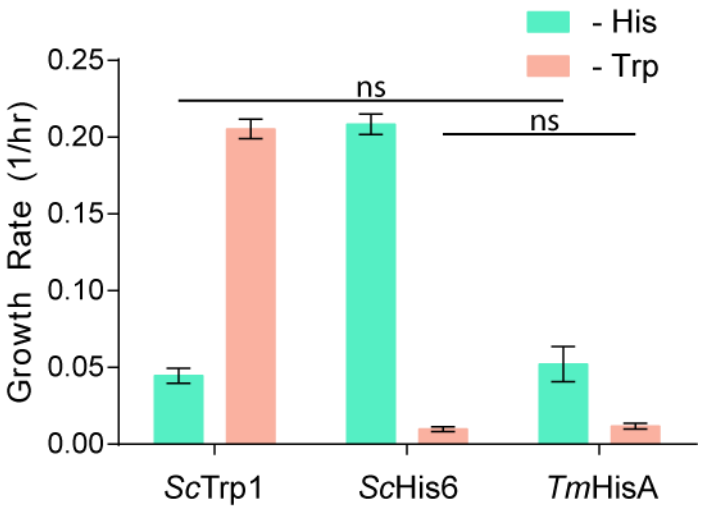


Figure 4.3. Comparisons of growth rates with either *Sc*Trp1, *Sc*His6, or *Tm*HisA expressed from a nuclear plasmid at medium expression levels in media lacking histidine (green) or tryptophan (orange). *Tm*HisA neither complements a His6 nor Trp1 deletion, as shown by the lack of difference in growth rate between HisA and a non-complementing enzyme.

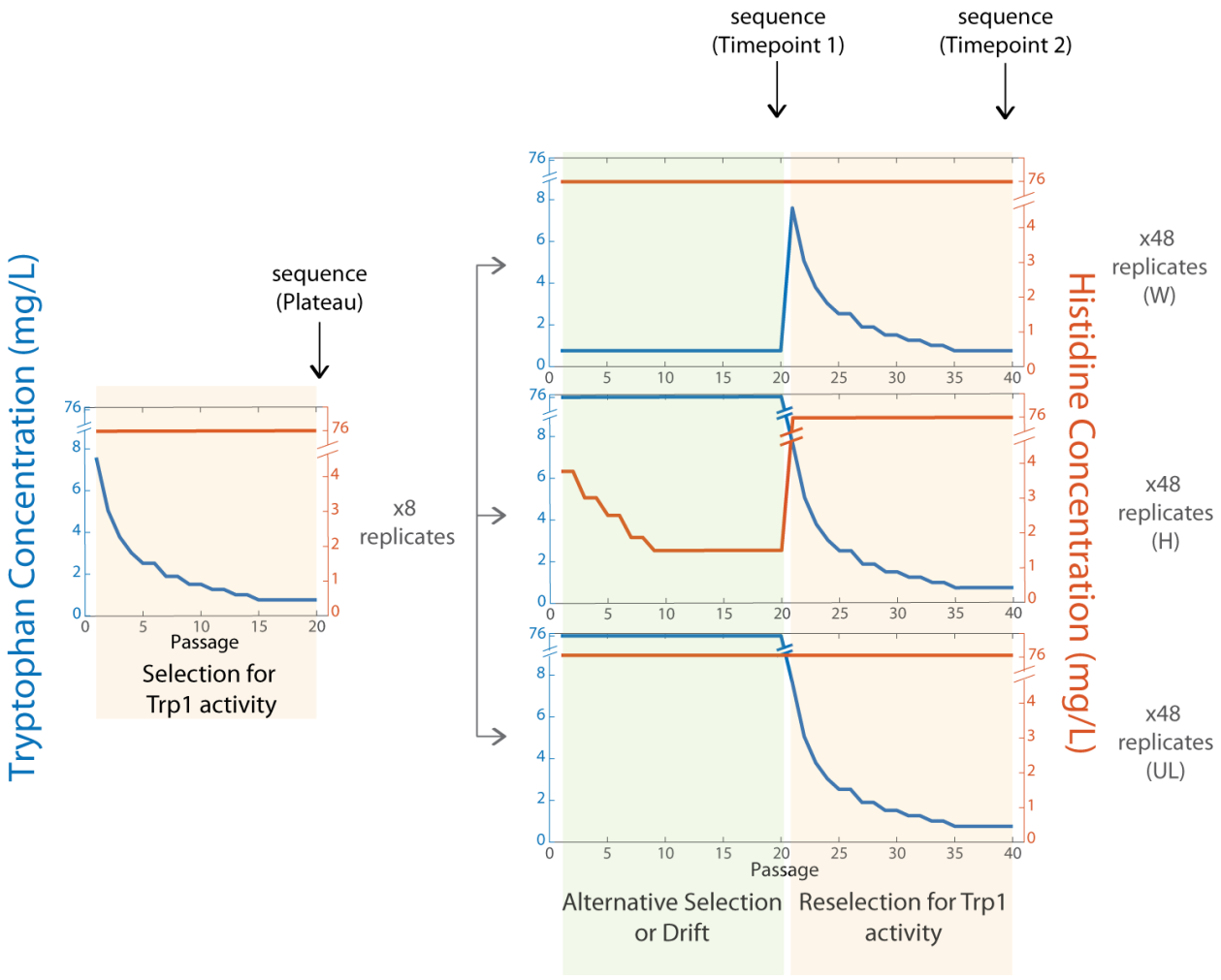


Figure 4.4. Experimental design of the evolution of *Tm*HisA for Trp1 activity with two modes of alternative selection. TmHisA is first selected for Trp1 activity by passaging in media with continually decreasing concentrations of tryptophan in eight replicates. After 20 1:100 passages, each replicate is further divided into 6 replicates for each alternative selection condition. The top condition demonstrates alternative selection for His6 by passaging in decreasing concentrations of histidine (H). The middle condition demonstrates continued selection for Trp1 activity by continued passaging in media limited by tryptophan (W). The bottom condition demonstrates

drift with no selection pressures by passaging in media that contains the full concentration of tryptophan or histidine (UL). After 20 1:100 passages in alternative selections or drift, cultures are reselected for Trp1 activity in a scheme similar to the initial selection. Samples for NGS are taken from cultures at the end of the initial selection for Trp1 activity, after alternative selection or drift, and after reselection for Trp1 activity.

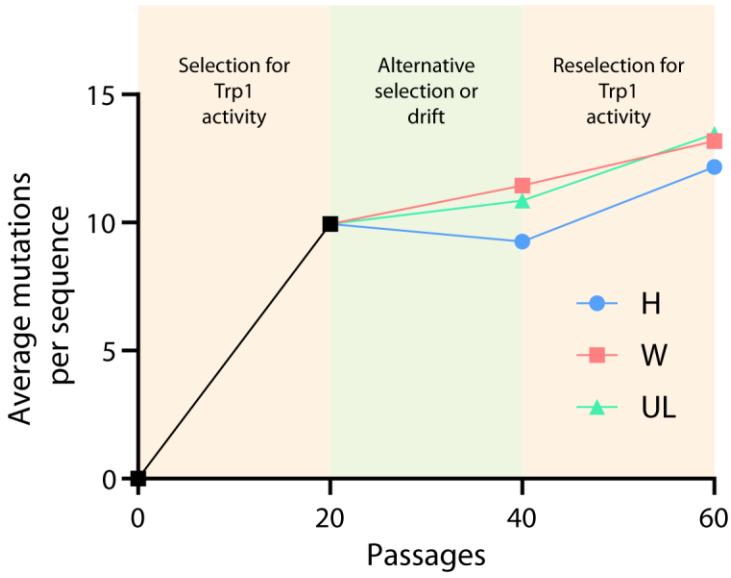


Figure 4.5. Trend of average mutations per sequence identified by NGS. The eight plateau cultures average 9.9 mutations per sequence. This increased in replicates undergoing continued selection for Trp1 activity (W) and for non-selective drift conditions (UL) but decreased in replicates that were alternatively selected for His6 activity (H). During reselection for Trp1 activity, all replicates demonstrated an increase in the average number of mutations per sequence.

Table S4.1. List of yeast strains used in this study.

Name	Genotype	Source	Parent Strain	Notes
F102-2u	F102-2u *MATa can1 his4-519 leu2-3,112* ρ^0+ pGKL1 + pGKL2	ATCC #200585	n/a	
ZZ-Y292	F102-2u *MATa can1 leu2Δ0 ura3Δ0 his6::KanMX HIS4* ρ^0+ pGKL1 + pGKL2	eVOLVER paper	AR-Y288	AR-Y288 transformed with PCR product from Yeast Knockout Collection strain with *his6* deleted with primers TCATCATCAAGGGTCATCTTTTTAT and GAAAAAGGTTGCCTCAATATTGTTA
ZZ-Y299	F102-2u *MATa can1 leu2Δ0 ura3Δ0 trp1Δ0 his6::KanMX HIS4* ρ^0+ pGKL1 + pGKL2	eVOLVER paper	ZZ-Y292	ZZ-Y292 transformed with ZZ-Ec482 and linear DNA corresponding to 40 basepairs upstream and downstream of *TRP1*. The strain was restreaked x2 on YPD solid media to remove ZZ-Ec482
ZZ-YT17-A3	F102-2u *MATa can1 leu2Δ0 ura3Δ0 trp1Δ0 his6::KanMX HIS4* ρ^0+ p1-FullDelPol-*Tm*HisA-URA3 + pGKL2 + ZZ-Ec506	eVOLVER paper	ZZ-Y299	ZZ-Y299 transformed with ZZ-Ec506 and ScaI-digested ZZ-Ec475
ZZ-Y323	F102-2u *MATa can1 leu2Δ0 ura3Δ0 trp1Δ0 his6::KanMX HIS4 flo1::NatMX* ρ^0+ p1-FullDelPol-*Tm*HisA-URA3 + pGKL2 +	eVOLVER paper	ZZ-YT17-A3	ZZ-Y299 transformed with linear DNA containing 500 basepairs upstream of *FLO1*, *NatMX*, and 500 basepairs downstream of *FLO1*

	pZZ-Ec506			
BY4741	*MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+	ATCC #201388		
ZZ-Y332	BY4741 *MAT***a** *HIS3 leu2Δ0 met15Δ0 ura3Δ0* ρ^+	eVOLVER paper	BY4741	BY4741 transformed with HIS3 PCR from F102-2u
ZZ-Y336	BY4741 *MAT***a** *HIS3 leu2Δ0 met15Δ0 ura3Δ0 his6::KanMX* ρ^+	eVOLVER paper	ZZ-Y332	ZZ-Y332 transformed with PCR product from Yeast Knockout Collection strain with *his6* deleted with primers TCATCATCAAGGGTCATCTTTTTAT and GAAAAAGGTTGCCTCAATATTGTTA
ZZ-Y354	BY4741 *MAT***a** *HIS3 leu2Δ0 met15Δ0 ura3Δ0 trp1Δ0 his6::KanMX* ρ^+	eVOLVER paper	ZZ-Y336	ZZ-Y292 transformed with ZZ-Ec482 and linear DNA corresponding to 40 basepairs upstream and downstream of *TRP1*. The strain was restreaked x2 on YPD solid media to remove ZZ-Ec482

Table S4.2. List of barcodes used in NGS sequence to deconvolute sequences.

Forward barcode primers	Reverse barcode primers
AACGGACT	AAACAAAC
ATGAAGGC	GCTACCTG
GGCATTTG	AGACTCGT
GGAGCGAT	ATGGCATT
GACAACGC	CCTTATTT
TGTTGATT	CTGTGGTA
GATCTTAA	GGGATTCC
CCGACATC	GGGTAACG
CTATCTAT	TAACTGTT
TATGTCCC	GTAAGCCT
GTCTGAGG	TCCTTGCG
CCACTGAA	CGAGGTTC
ACCGTTTA	ACATAAGT
AGCTTGCC	GAAGGCAG
CACCTCCA	ATTCCGGA
GTAAGCCT	ACTAACCA
TGGACCAG	CAGTCTTG
TTCTATCA	CGTGACTG
	AAAGTTGG

Chapter 5. Concluding Remarks

Summary

Directed evolution is a powerful tool for both developing technologies to improve human lives as well as gaining insight into how the protein world developed. Numerous successes has been noted in medical technologies, including therapeutic antibodies[1,2], technologies for medical imaging[3–5], and for vectors for gene therapy[6]. Additionally, the ability to conduct prospective evolution experiments has allowed for insight into how proteins have evolved and expanded their functions[7–9]. However, the evolutionary methods employed in these experiments have been with *ex vivo* mutagenesis coupled with *in vivo* selection, leading to arduous experiments when multiple rounds of evolution are required[10]. Newer technologies developed by our lab[11,12] and others[13,14] seek to reduce the intensity and investment needed for laboratory scale evolution efforts by enabling mutagenesis *in vivo*, thus allowing for the continuous evolution of proteins.

OrthoRep, a continuous evolution system based in *S. cerevisiae*, has been developed by our lab and enables continuous diversification of genes *in vivo*[11,12]. Previous attempts to enable *in vivo* diversification subjected the entire genome to high mutation rates, resulting in decreased fitness over time as mutations accumulate on essential genes. Instead, OrthoRep is based on a system of two linear, cytoplasmic plasmids with their own dedicated DNA polymerases and previous work demonstrated that increased mutation rates of genes encoded on OrthoRep does not increase genomic mutation rates[11]. Further work by our lab has developed higher mutagenesis rates on OrthoRep, achieving rates of 10^{-5} substitutions per base during

replication[12]. This power and simplicity has led to my work to further develop OrthoRep as a platform for directed evolution.

I have engineered a panel of genomic constructs that allow expression of proteins on OrthoRep to span a range comparable to traditional nuclear expression in *S. cerevisiae*. OrthoRep's distinct origin from a pair high-copy cytoplasmic plasmids that encode a toxin-antitoxin system[15] also necessitates novel tools for manipulation and expression. Previously developed genomic promoters and expression cassettes that have been used in *S. cerevisiae*[16] are non-functional on OrthoRep[17]. However, expression of genes encoded on OrthoRep using native upstream conserved sequences (UCS) have been low[11,18]. My work screened sequences identified in preliminary evolution studies to identify UCSs that enable higher expression of genes encoded on OrthoRep. Further, I developed a genomically encoded poly-A tail that mimics a eukaryotic poly-A tail, which further increases expression the type of genes encoded on OrthoRep. This has enabled a panel of expression cassettes for OrthoRep that span a range comparable to genomic expression.

I have also established the paring of OrthoRep with eVOLVER, a continuous culture device, to enable hands-free evolution of proteins. Due to the stochastic nature of mutation accumulation of genes of interest encoded on OrthoRep, different replicate cultures undergo divergent rates of adaptation during evolution. Therefore, in high-replicate evolution experiments, cultures that lag behind in adaptation are driven to extinction when the stringency of selection outpaces their ability to adapt[12], as tuning conditions for each individual culture is not feasible. I coupled OrthoRep with eVOLVER, a programmable continuous culture platform

and developed the algorithm necessary to track the adaptation of individual cultures and to automatically increase or decrease selection stringency, allowing for hands-free continuous evolution in a system I termed Automated Continuous Evolution (ACE). I demonstrate the feasibility of ACE with the evolution of *Plasmodium falciparum* dihydrofolate reductase (*Pf*DHFR) to resist the competitive inhibitor, pyrimethamine, on time scales faster than manual serial passaging. I further validate ACE with the adaptation of thermophilic *Thermotoga maritima* HisA (*Tm*HisA) for activity at mesophile temperatures in *S. cerevisiae*. This enables a platform that offers unparalleled simplicity and speed for the evolution of proteins.

I used the unique capabilities or OrthoRep to demonstrate that alternative selection may be superior to continuous strong selection in evolving enzymes for higher activity. Traditional directed evolution studies use a monotonically increasing selection stringency that can be slowed down or sped up based on how fast cultures are adapting. However, this has mostly led to only modest improvements in the directed evolution of enzymes for novel or higher catalytic activities[19,20]. Using a model evolution system that evolves *Tm*HisA for Trp1 activity in *S. cerevisiae*, I demonstrated that continued strong selection results in a plateauing of Trp1 activity, which does not allow for growth in media lacking tryptophan. Moreover, continued strong selection over an additional 200 generations for Trp1 activity does not result in any noticeable improvement in enzymatic activity. However, when an alternative selection for His6 activity is implemented in series with selection for Trp1 activity, *Tm*HisA is able to develop mutations that allows for full complementation of Trp1 activity that allows for growth in media lacking

tryptophan. This suggests that in the directed evolution of enzymes, schemes incorporating selection for alternative activities may allow for more successful variants to emerge.

Future Directions

Finer expression control and additional tools can still be developed for OrthoRep. The use of the mutagenic OrthoRep polymerase TP-DNAP-**4-2**[12] results in a significant decrease in expression of genes encoded on OrthoRep[21]. This limits the maximal expression that can be achieved during continuous mutagenesis. Further developments to OrthoRep that allows for a range of expression comparable to genomic constructs during mutagenesis will expand the potential evolution targets with OrthoRep and allow for identification of small increases in fitness that are magnified with higher expression.

Automated Continuous Evolution remains to be demonstrated as a stand-alone technology. While I previously evolved *Pf*DHFR for pyrimethamine resistance and *Tm*HisA for His6 activity in *S. cerevisiae*, both of these experiments were done after successful evolution through serial passaging. It would be a strong demonstration of ACE if this technology was used to conduct a directed evolution experiment without prior evidence of success, or if it was able to achieve success despite prior failures through serial passaging.

The mechanisms of how alternative selection enables evolution remains to be elucidated. The contributions of individual mutations on *Tm*HisA that facilitates Trp1 activity are still unknown. Therefore, it is difficult to draw definitive conclusions on the mechanism of alternative selection in the development of higher catalytic activities. Elucidating such mechanisms would be useful not only in experimental design during directed evolution

experiments, but also to give insight into how nature has evolved a stunning number of enzymes with a broad range of activities and catalytic rates.

Conclusions

In this dissertation, I developed three technologies that directly affects the continuous evolution system OrthoRep, or aids in experimental design during experimental evolution. These technologies have already been used in evolution by subsequent experiments using OrthoRep[22]. Ultimately, the technologies and applications discussed herein should further the field of directed evolution and evolutionary biology through the discovery or development of novel or highly efficient enzymes and offer insight into the driving forces of natural evolution.

References

(1) Lu, R. M.; Hwang, Y. C.; Liu, I. J.; Lee, C. C.; Tsai, H. Z.; Li, H. J.; Wu, H. C. Development of Therapeutic Antibodies for the Treatment of Diseases. *J. Biomed. Sci.* **2020**, *27* (1), 1–30. https://doi.org/10.1186/s12929-019-0592-z.

(2) Chan, D. T. Y.; Groves, M. A. T. Affinity Maturation: Highlights in the Application of in Vitro Strategies for the Directed Evolution of Antibodies . *Emerg. Top. Life Sci.* **2021**, *0*, 1–8. https://doi.org/10.1042/etls20200331.

(3) Lee, T.; Cai, L. X.; Lelyveld, V. S.; Hai, A.; Jasanoff, A. Molecular-Level Functional Magnetic Resonance Imaging of Dopaminergic Signaling. *Science (80-.).* **2014**, *344* (6183), 533–535. https://doi.org/10.1126/science.1249380.

(4) Shapiro, M. G.; Westmeyer, G. G.; Romero, P. A.; Szablowski, J. O.; Küster, B.; Shah, A.; Otey, C. R.; Langer, R.; Arnold, F. H.; Jasanoff, A. Directed Evolution of a Magnetic Resonance Imaging Contrast Agent for Noninvasive Imaging of Dopamine. *Nat. Biotechnol.* **2010**, *28* (3), 264–270. https://doi.org/10.1038/nbt.1609.

(5) Willmann, J. K.; Kimura, R. H.; Deshpande, N.; Lutz, A. M.; Cochran, J. R.; Gambhir, S. S. Targeted Contrast-Enhanced Ultrasound Imaging of Tumor Angiogenesis with Contrast Microbubbles Conjugated to Integrin-Binding Knottin Peptides. *J. Nucl. Med.* **2010**, *51* (3), 433–440. https://doi.org/10.2967/jnumed.109.068007.

(6) Dalkara, D.; Byrne, L. C.; Klimczak, R. R.; Visel, M.; Yin, L.; Merigan, W. H.; Flannery, J. G.; Schaffer, D. V. In Vivo-Directed Evolution of a New Adeno-Associated Virus for Therapeutic Outer Retinal Gene Delivery from the Vitreous. *Sci. Transl. Med.* **2013**, *5*

(189), 1–12. https://doi.org/10.1126/scitranslmed.3005708.

(7) Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (12), 866–876. https://doi.org/10.1038/nrm2805.

(8) Bloom, J. D.; Romero, P. A.; Lu, Z.; Arnold, F. H. Neutral Genetic Drift Can Alter Promiscuous Protein Functions, Potentially Aiding Functional Evolution. *Biol. Direct* **2007**, *2*, 7–10. https://doi.org/10.1186/1745-6150-2-17.

(9) Bershtein, S.; Goldin, K.; Tawfik, D. S. Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J. Mol. Biol.* **2008**, *379* (5), 1029–1044. https://doi.org/10.1016/j.jmb.2008.04.024.

(10) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet.* **2015**, *16* (7), 379–394. https://doi.org/10.1038/nrg3927.

(11) Ravikumar, A.; Arrieta, A.; Liu, C. C. An Orthogonal DNA Replication System in Yeast. *Nat. Chem. Biol.* **2014**, *10* (3), 175–177. https://doi.org/10.1038/nchembio.1439.

(12) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **2018**, *175* (7), 1946-1957.e13. https://doi.org/10.1016/j.cell.2018.10.021.

(13) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **2011**, *472* (7344), 499–503. https://doi.org/10.1038/nature09929.

(14) Halperin, S. O.; Tou, C. J.; Wong, E. B.; Modavi, C.; Schaffer, D. V.; Dueber, J. E. CRISPR-Guided DNA Polymerases Enable Diversification of All Nucleotides in a

Tunable Window. *Nature* **2018**, *560* (7717), 248–252. https://doi.org/10.1038/s41586-018-0384-8.

(15) Gunge, N.; Sakaguchi, K. Intergeneric Transfer of Deoxyribonucleic Acid Killer Plasmids, PGKl1 and PGKl2, from Kluyveromyces Lactis into Saccharomyces Cerevisiae by Cell Fusion. *J. Bacteriol.* **1981**, *147* (1), 155–160.

(16) Lee, M. E.; DeLoache, W. C.; Cervantes, B.; Dueber, J. E. A Highly-Characterized Yeast Toolkit for Modular, Multi-Part Assembly. *ACS Synth. Biol.* **2015**, 150414151809002. https://doi.org/10.1021/sb500366v.

(17) Kamper, J.; Meinhardt, F.; Gunge, N.; Esser, K. New Recombinant Linear DNA-Elements Derived from Kluyveromyces Lactis Killer Plasmids. *Nucleic Acids Res.* **1989**, *17* (4), 1781. https://doi.org/10.1093/nar/17.4.1781.

(18) Schickel, J.; Helmig, C.; Meinhardt, F. Kluyveromyces Lactis Killer System: Analysis of Cytoplasmic Promoters of the Linear Plasmids. *Nucleic Acids Res.* **1996**, *24* (10), 1879–1886. https://doi.org/10.1093/nar/24.10.1879.

(19) Nannermann, D. P.; Birmingham, W. R.; Scism, R. A.; Bachmann, B. O.; Nannemann, D. P.; Birmingham, W. R.; Scism, R. A.; Bachmann, B. O. *Assessing Directed Evolution Methods for the Generation of Biosynthetic Enzymes with Potential in Drug Biosynthesis*; 2011; Vol. 3, pp 803–819. https://doi.org/10.4155/fmc.11.48.

(20) Goldsmith, M.; Tawfik, D. S. Enzyme Engineering: Reaching the Maximal Catalytic Efficiency Peak. *Curr. Opin. Struct. Biol.* **2017**, *47*, 140–150. https://doi.org/10.1016/j.sbi.2017.09.002.

(21) Zhong, Z.; Ravikumar, A.; Liu, C. C. Tunable Expression Systems for Orthogonal DNA Replication. *ACS Synth. Biol.* **2018**, *7* (12), 2930–2934. https://doi.org/10.1021/acssynbio.8b00400.

(22) Wellner, A.; McMahon, C.; Gilman, M. S. A.; Clements, J. R.; Clark, S.; Nguyen, K. M.; Ho, M. H.; Hu, V. J.; Shin, J.-E.; Feldman, J.; Hauser, B. M.; Caradonna, T. M.; Wingler, L. M.; Schmidt, A. G.; Marks, D. S.; Abraham, J.; Kruse, A. C.; Liu, C. C. Rapid Generation of Potent Antibodies by Autonomous Hypermutation in Yeast. *Nat. Chem. Biol. 2021* **2021**, 1–8. https://doi.org/10.1038/s41589-021-00832-4.