Advances in Phage-Assisted Continuous Evolution and Application to Overcoming Bioinsecticide Resistance

Advances in Phage-Assisted Continuous Evolution and Application to Overcoming Bioinsecticide Resistance

A dissertation presented

by

Ahmed Hussein Badran

to

The Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

Cambridge, Massachusetts

April 2016

Dissertation Advisor: David R. Liu Ahmed Hussein Badran

Advances in Phage-Assisted Continuous Evolution and Application to Overcoming Bioinsecticide Resistance

Abstract

The *Bacillus thuringiensis* δ-endotoxins (Bt toxins) are widely used insecticidal proteins in engineered crops that provide agricultural, economic, and environmental benefits, constituting a substantial and increasingly large portion of the total global production of various crops, including cotton, corn, maize and soybeans. Bt toxins are exquisitely selective for targeted pests, typically do not affect off-target insects, and are completely orthogonal to human biology. However, the development of insect resistance to Bt toxins endangers their long-term effectiveness.

In this thesis, I describe the development of methodology for the systematic directed evolution of novel Bt toxins to selectively affect resistant insects. Using Phage-Assisted Continuous Evolution (PACE), a previously developed general platform for the continuous directed evolution of biomolecules, I developed a highly sensitized selection for novel protein-protein interactions. This system robustly reported on interactions spanning affinities from low micromolar to picomolar. However, attempts at using this system for the directed evolution of novel protein-protein interactions were largely unsuccessful, presumably as a consequence of low mutagenesis efficiency.

To increase the utility of the platform, I sought to enhance the mutagenesis levels afforded by PACE, but current *in vivo* methods suffer from a lack of control, genomic instability, low efficiency, and narrow mutational spectra. I used a systematic, mechanism-driven approach to create a potent, inducible, broad-spectrum, and vector-based mutagenesis system in *E. coli* that enhances mutation rates by 322,000-fold over basal levels, surpassing the mutational efficiency and spectra of widely used *in vivo* and *in vitro* mutagenesis methods. This system enabled the high-frequency, broad-spectrum mutagenesis of chromosomal, episomal, and viral nucleic acids *in vivo*, and dramatically enhanced the success of PACE experiments, highlighting the importance of mutagenesis efficiency on selection outcome.

Using this enhanced mutagenesis approach and the previously described sensitized selection platform, I was able to evolve variants of the commonly used Bt toxin Cry1Ac that bind toxin binding region of a cadherin-like receptor from the insect pest *Trichoplusia ni* (TnCAD) that is not targeted by wild-type Cry1Ac. The resulting evolved Cry1Ac variants bind TnCAD with high affinity (K_d = 11-41 nM), kill TnCAD-expressing insect cells that are not susceptible to wild-type Cry1Ac, and kill Cry1Ac-resistant *T. ni* insects up to 335-fold more potently than wild-type Cry1Ac. Our findings establish that the evolution of Bt toxins with novel insect cell receptor affinity can overcome Bt toxin resistance in insects and confer lethality approaching that of the wild-type Bt toxin against non-resistant insects.

In conclusion, these finding offer a novel mechanism of overcoming what is quickly becoming among the largest issue overshadowing the continued success of Bt toxins for pest control and management, and establish a platform for the detection and evolution of a wide array of protein-protein interactions.

Dedicated to my parents, Hussein and Eman.

Acknowledgements

I acknowledge the financial support from the National Science Foundation (NSF) for a three year PhD fellowship. I also acknowledge the support from the GSAS Merit Fellowship from the John Parker Bequest fund during the academic year 2014-2015.

First I would like to thank my advisor Prof. David R. Liu for his continued optimism to counter my pessimism, constant encouragement, and support throughout my graduate career. The intellectual and scientific freedom I was allowed in his lab has undoubtedly contributed to my ability to think critically and troubleshoot ongoing research, allowing me to become the best scientist that I can be.

I would like to thank my Preliminary Qualifying Exam and Dissertation Advisory Committees, composed of Prof. Dan Kahne, Prof. Ann Hochschild, Prof. Greg Verdine, Prof. Emily Balskus and Prof. Alan Saghatelian, for their enthusiastic mentorship from the very beginning of my doctoral training.

I am especially thankful to Prof. Dan Kahne for providing insights into biological phenomena that have enabled the development of various aspects of my system, and his encouragement to pursue important questions and problems. My work with the bacterial 2-hybrid would not have been possible without the guidance of Prof. Ann Hochschild, whose understanding of bacterial biology has illuminated critical improvements to the research throughout the years. I would additionally like to deeply thank Prof. Balskus, Hochschild and Kahne for numerous engaging conversations about science and the pursuit of an academic career.

The research described in this thesis is the culmination of many years of collaboration with researchers from various disciplines and diverse backgrounds. The contributions presented herein would not have been possible without their support and hard work.

I am very thankful to the developers of the Phage-Assisted Continuous Evolution (PACE) system for their dissmination of knowledge and direction throughout the years, as well as fellow PACE team members for their exchange of ideas and collaborations: Dr. Kevin Esvelt, Dr. Jacob Carlson, Jeff Bessen, Dr. David Bryson, Liwei Chen, Kevin Davis, Dr. Bryan Dickinson, Dr. Xue Gao, Johnny Hu, Dr. Basil Hubbard, Dr. Aaron Leconte, Michael Packer, Tim Roth, Dr. Ning Sun, and Dr. David Thompson. I would especially like to thank Dr. Jacob Carlson, whose molecular biology guidance and thoughtful experimental design has largely shaped my scientific approach.

I am grateful to all the contemporary group members of the Liu lab for their friendship and intellectual exchanges throughout these years: Adrian Berliner, Dr. Chihui An, Dr. Travis Blum, Dr. Yevgeny Brudno, Dr. Brian Chaikind, Alix Chan, Dr. Grace Chen, Zhen Chen, Dr. James Cronican, Prof. Ed Curtis, Dennis Dobrovolsky, Dr. Brent Dorr, Dr. Christoph Dumelin, Dr. Nicole Gaudelli, Prof. David Gorin, Chris Green, Dr. Yu He, Dr. John Guilinger, Prof. Ryan Hili, Dr. Adam Kamlet, Bill Kim, Dr. Ralph Kleiner, Dr. Alexis Komor, Dr. Jon Levy, Dr. Margie Li, Dr. Phillip Lichtor, Dr. Juan Pablo Maianti, Dr. Jasmina Marjanovic, Dr. Rick McDonald, Dr. Lynn McGregor, Dr. Jia Niu, Dr. Vikram Pattanayak, Chris Podracky, Holly Rees, Dr. Weixin Tang, Dr. Ben Thuronyi, Dr. Dmitry Usanov, Dr. Tina Wang and Dr. John Zuris. I would especially like to acknowledge Brent Dorr and Tim Roth for being exceptional people and fantastic bay mates. I would like to especially thank Aleksandar Markovic for ensuring that the lab operates smoothly year in and year out.

I would like to thank my collaborators at Monsanto, who played a massive role in our ability to extend the PACE technology to a fantastic application: Thomas Malvar, Victor M. Guzov, Qing Huai, Melissa M. Kemp, Prashanth Vishwanath, Autumn M. Nance, Artem Evdokimov, Farhad Moshiri, Keith H. Turner, Jeff Nageotte, David Rappoli, Jean-Louis Kouadio, Meiying Zheng, Jason Milligan, Mingya Huang, Zijin Du, Xuefeng Zhou, Edward Kraft and Jinling

Wang for their assistance. I would like to specifically acknowledge the contributions of Thomas Malvar, Victor Guzov and Qing Huai to prioritize our collaboration above their ongoing research, and their constant hard work throughout the duration of our collaboration. I would also like especially thank Keith Turner for his beautiful analysis of our Bt evolution sequencing data, and his interest in the intricacies of the PACE technology.

I would like to extend a special thanks to Wendy Kain and Prof. Ping Wang of Cornell University for their willingness to collaborate with us, and providing their expertise and resources to test our evolved Bt toxin variants against their resistant insects.

I am indebted to Prof. David Liu and Prof. Jon Clardy for the opportunity to teach The Molecules of Life SLS11 for 3 consecutive years, as well as Prof. James Bradner and Prof. Ralph Mazitschek for the opportunity to TF the Chemical Biology Program Bootcamp CB2200. I extend my sincerest thanks to all the students I have throughout the years. I am especially grateful to Dr. Marie Spong for her continuous hard work in SLS11 to ensure that all classes, sections, labs and exams proceeded without a hitch.

I am grateful to the staff of the Chemical Biology Program, notably KeyAnna Schmiedl, Samantha Reed and Jason Millberg for all their hard work behind the scenes, keeping the program running smoothly and continuing to attract top students every year. The program has been a great source of enjoyment for me and I thank you for all your hard work.

Finally I would like to thank my friends and family for their support, constant encouragement and suggestions, and dedication to our friendship when times got rough. I would like to specifically thank my best friends and brothers Mounir Koussa and Gregory Boursalian. My relationship with both of you has had a dramatic impact on my personality, my approach to life and research, and has greatly improved the quality of my existence. I am eternally grateful to have met you and hope that you achieve all you set out to accomplish.

I dedicate this thesis to my loving parents, Dr. Hussein Badran and Dr. Eman Basha, whose academic and social sacrifices laid the foundation for immigration to the United States, facilitated a fantastic education and allowed me to pursue research as a career. I remain indebted to you and the opportunities that you have given me, and hope to one day repay that debt.

Thank you all most sincerely.

Table of Contents

List of Sections

List of Figures

List of Tables

Chapter 1

Introduction to *In Vivo* Continuous Directed Evolution

Parts adapted from: Badran & Liu, *Current Opinion in Chemical Biology* **24**, 1-10 (2014).

1.1 – Introduction: *in vitro* and *in vivo* continuous directed evolution

Initially demonstrated as a surrogate for Darwinian evolution by Spiegelman and coworkers almost five decades ago[1], continuous directed evolution has long been envisioned as a potentially highly efficient method to discover novel biomolecules with activities of interest. In Spiegelman's seminal study, Qβ bacteriophage genomic RNA was amplified based on its ability to serve as a substrate for purified Qβ RNA-dependent RNA-replicase. After 74 serial passages of this RNA genome, a variant was produced that was both 83% smaller and could be replicated 15 times faster than the starting Qβ RNA genome. Importantly, this study highlighted the critical nature of selection pressure design, as the evolved Qβ RNA genome could be replicated at a significantly faster rate than the parental genome, but could no longer direct the synthesis of viral particles since this requirement was not implicit in the selection.

Thirty years later, Joyce and coworkers described the first system for the continuous directed evolution of catalytic function. The researchers established a self-sustaining RNA replication cycle for RNA molecules capable of catalyzing their own ligation to an RNA-DNA substrate[2]. Like the Qβ system, the RNA ligase ribozyme continuous evolution system relied on exogenously supplied materials (reverse transcriptase and T7 RNA polymerase). Since these early studies, a number of additional continuous directed evolution systems have been described, the majority of which are limited to either catalytic RNAs or involve the evolution of alternative RNA functions[2-7]. Additionally, these methods have been exclusively carried out *in vitro*, where the comparative ease of manipulation and selection stringency adjustment can be exploited compared to *in vivo* methods. Nevertheless, these early landmark studies vividly

demonstrated the potential of continuous directed evolution and provided a foundation for subsequent advances.

All forms of Darwinian evolution must support four fundamental processes: translation (when the evolving molecule is not identical to the information carrier), selection, replication, and mutation (**Figure 1.1**). Traditional directed evolution methods handle each of these processes discretely, frequently requiring steps in which the researcher must perform experimental manipulations. In contrast, continuous directed evolution systems must seamlessly integrate all of these processes into an uninterrupted cycle. For the purposes of this review, we define continuous directed evolution as a general method capable of evolving a specific phenotype, at the level of an organism or a set of genes, over many cycles of mutation, selection, and replication with minimal researcher intervention. A continuous evolution method therefore must support all four stages of Darwinian evolution and allow the surviving genes from one generation to

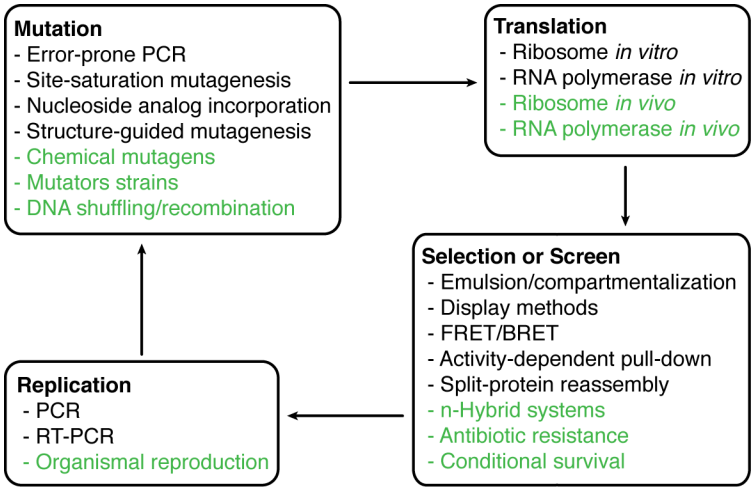


Figure 1.1 | Schematic representation of directed evolution. All complete directed evolution methods must provide four major components: translation, selection or screening, replication, and mutation. Historical examples of each of these components are listed. Techniques that are particularly amenable to *in vivo* continuous evolution are shown in green.

spontaneously enter the translation, selection, replication, and mutation processes of the subsequent generation.

Many key studies on continuous evolution have relied on manual serial dilution of the evolving population as a mechanism for propagating genes and controlling selection stringency. We do not consider this technical manipulation to violate our definition, as such systems have been shown to be amenable to automation, yielding similar results[8]. While *in vivo* continuous evolution methods have exploited the natural occurrence of the three steps of Darwinian evolution in living organisms, these methods have also historically been constrained by modest scope, often focusing on easily selected phenotypes such as antibiotic resistance rather than traits associated with diverse applications and unmet needs. Recently, more general *in vivo* continuous directed evolution methods have been developed using various organisms, ranging from viruses to bacteria to higher eukaryotes. In this review, we highlight a sampling of recent developments in the field of *in vivo* continuous directed evolution over the past two decades and their impact, both realized and prospective, on the directed evolution community.

1.2 – Viral Continuous Evolution

With their high mutation rates[9] and relative ease of manipulation and study, bacteriophages were used early on as model organisms for rapid directed evolution of readily selectable traits[10]. Moreover, the development of methods for continually culturing bacterial strains has greatly reduced the barrier for studying bacteriophage evolution under continuous selection pressure[10,11]. Of historical note is the early and

widespread use of both chemostats, cultures maintained through the continuous inflow of fresh growth medium, and auxostats, cultures regulated by a feedback mechanism that controls the inflow of growth medium based on a monitored culture parameter (**Figure 1.2**). Both methods have proven useful in continuous culturing of bacteriophage-infectable microbes. Methods for bacterial continuous evolution will be covered in the next section.

Among the earliest examples of viral continuous evolution is the 1997 work of Molineux and coworkers using ΦX174, a bacteriophage capable of propagating using *E.*

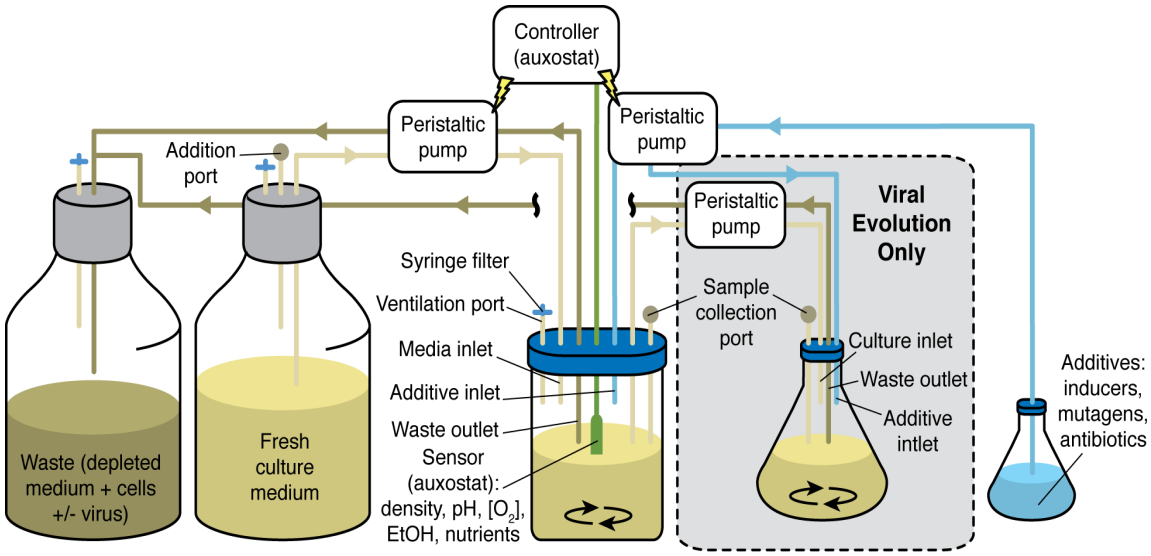


Figure 1.2 | Typical auxostats and chemostats. Fresh culture medium is pumped into a pre-sterilized container containing the microorganism of interest. Medium with cells is pumped out of the culture container. The rates of medium input and output are held constant in a chemostat system, and the growth rate is regulated by the composition of the medium. In an auxostat system, the medium flow rates are dynamically regulated by a controller in response to measurements made in the growing culture, which correspond directly (turbidity meter) or indirectly (other sensors) to the culture density. In schemes for viral continuous evolution, the auxostat or chemostat culture is pumped into a new vessel where the virus of interest is supplied (cellstat). Both cultures can be supplemented with additives supplied through a dedicated inlet. Selection stringency can be regulated by varying flow rates, changing temperature, or adding compounds to the auxostat, chemostat or cellstat. The depleted medium, cells and virus are pumped to a waste container.

coli or *S. typhimurium*[12,13]. These studies highlighted the ability of this phage to rapidly adjust to novel environments such as higher temperatures, as well as an abnormally high degree of mutational convergence depending on selection conditions. Additionally, substantial improvements in phage fitness were observed over relatively short time frames, with an average of 1-2 mutations per 24 hours (corresponding to 0.2-0.5 % of the viral genome, and 0.003-0.005 mutations per generation per kilobase) in the absence of any added mutagens. In an expansion of this work by Bull and coworkers, ΦX174 phage was propagated on bacterial hosts under native conditions for six months, corresponding to ~13,000 phage generations[14]. Mutations continued to accumulate at a constant rate for the majority of the experiment, suggesting a potential arms race within the propagating phage pool. The researchers speculated that this arms race was driven by the high concentration of phage with respect to host cells, resulting in high levels of co-infection and competition between phage. While the mutation rates observed under these conditions were high with respect to the size of the viral genome and sufficient for rapid phage adaptation, higher rates are typically required to access gene-specific evolution goals (1-2 mutations per round per gene)[15].

Similar phage propagation experiments were carried out by Molineux and coworkers using bacteriophage T7 to study organismal phylogenetic histories[16]. To enhance the mutation rate, T7 phage was serially passaged in the presence of the chemical mutagen MNNG (*N*-methyl-*N'*-nitro-*N'*-nitrosoguanidine), resulting in variants with differential restriction enzyme cleavage patterns than the wild-type phage. The correct phylogeny of the diverging phage populations could be determined using these cleavage patterns, supporting their method for phylogenetic reconstruction and

suggesting maximization of parsimony as a promising method for determining ancestral characteristics. While this approach addressed the need for high mutagenesis rates during evolution, the selection was highly qualitative, requiring the accumulation of non-deleterious mutations that enabled restriction enzyme profiling. A similar strategy was later used by the same group to study the effects of successive severe population bottlenecking and expansion on viral fitness, demonstrating the remarkable plasticity of the T7 genome towards mutations and the interdependence of the accumulated mutations in specific phage linages[17]. These studies also highlight the utility of user-defined mutagenesis rates in continuous directed evolution, although chemical mutagens generally offer narrow mutational spectra and are potent carcinogens.

Recently, T7 phage was used by Barrick and coworkers to study organismal adaptation to an expanded genetic code[18]. Two versions of this phage, wild-type T7 and the Δ2 T7 hypermutator, were serially propagated using an *E. coli* strain capable of efficiently incorporating 3-iodotyrosine at amber stop codons. The use of a genetically encodable hypermutator resulted in broad mutational load to the T7 phage without the hazards typically associated with alternative chemical mutagens. Notably, two amber codon substitutions reached high frequencies in the evolving populations: Tyr88Amber in T7 RNA polymerase (T7 RNAP) and Tyr39Amber in T7 type II holin. Whereas the mutation in the RNAP was assumed to be tolerated due to a lack of interaction of this surface-exposed side chain with the remainder of the protein, the mutation in the type II holin was shown to be critical for function in the amber suppressor strain. The findings of this study suggest that an expanded genetic code may increase the evolvability of an organism.

Lambda bacteriophage has been used extensively to study host-virus interactions. Lambda phage natively recognizes LamB as the main point of entry into *E. coli*. Using a strictly lytic lambda phage, Lenski and coworkers studied the ability of this phage to continuously evolve the ability to use alternative outer membrane receptors as the point of entry[19]. Upon co-culturing this phage with host cells in glucose-limited conditions, the bacteria rapidly evolved resistance through downregulation of *lamB* expression. The mutating pool of lambda phage then evolved compensatory mutations in the LamB recognition protein J, resulting in improved fitness on LamB, as well as facilitating the use a new receptor, OmpF. In response, the *E. coli* host reduced fitness of the mutant lambda phage through mutations to *manYZ*, two proteins that form an inner membrane channel required for lambda phage entry. This coevolution highlights the utility of continuous evolution methods in revealing the dynamic interplay between host and virus evolution.

While these systems have enhanced our understanding of fundamental aspects of evolution, they generally are limited by an uncontrolled (usually low) mutagenesis rate intrinsic to the bacteriophage and host, and/or the requirement for broad organismal adaptation to an environmental challenge, rather than enabling the evolution of a specific biomolecular activity of interest. Our group has developed a bacteriophage-based continuous evolution system that attempts to address these two limitations. Phage-assisted continuous evolution (PACE) is a general system for the directed evolution of biomolecules that relies on previously discussed methods for the continued culturing of *E. coli*, can be used to evolve, in principle, any activity that can be coupled to gene transcription in *E. coli*, and allows real-time modulation of selection stringency and

mutagenesis levels[20].

Briefly, PACE takes advantage of the critical role of the minor coat protein pIII from the filamentous bacteriophage M13 during both infection and progeny release[21]. Phage lacking functional pIII have virtually no ability to propagate and are rapidly lost under continuous culturing conditions. However, host cell-provided pIII restores the virulence of the phage and allows robust phage propagation in continuous culture. Accordingly, phage-borne genes encoding the ability to trigger pIII in host cells gain a fitness advantage and propagate at the expense of unfit phage. To increase the mutation rate during PACE, an inducible plasmid driving the expression of error-prone *E. coli* polymerase subunits results in ~100-fold increased mutagenesis with a more-uniform distribution of mutations as compared to traditional chemical mutagens[20]. Due to the continuous flow nature of this system, mutations accumulate only within the phage genome and the gene to be evolved, reducing the likelihood of "cheaters" overtaking the population during selection.

PACE was initially used to evolve novel activity in T7 RNAP[20]. Whereas T7 RNAP efficiently initiates transcription using the native T7 promoter, it lacks detectable activity on the related T3 promoter. Using a combined T7/T3 promoter as an evolutionary stepping-stone, T7 RNAP variants capable of efficiently initiating transcription from the T3 promoter were evolved within a few days. This method has since been expanded to address the effects of varying levels of mutagenesis and selection stringency on enzyme evolution[22], the path dependence of convergent protein evolution[23], the effects of neutral drift, negative selection, and stringency modulation on enzyme evolution[24], and the rapid evolution of protease enzymes from pathogenic viruses that predict how viral proteases

evade clinically relevant antiviral drugs[25]. This system has been further modified to enhance its ability to serve as a basis for alternative selection schemes, and ongoing studies use PACE to mediate the continuous directed evolution of DNA-binding domains, protein-protein interactions, genome modification enzymes, proteases, and other enzymes[25].

1.3 – Bacterial Continuous Evolution

Mechanisms that enable the efficient and unmonitored continuous propagation of bacteria and other microorganisms have served as a key component of many continuous evolution methods. While their utility has been extensively demonstrated for bacteriophage evolution efforts as highlighted in the previous section, these methods have also facilitated the study of continuous directed evolution of bacterial populations, especially in cases where the gene to be evolved is large or cannot be directly coupled to viral fitness. A recent demonstration assesses the evolution of population-wide antibiotic resistance. Using a continuous culture of *E. coli*, Collins and coworkers added increasing concentrations of the antibiotic norfloxacin to study population dynamics during increasing antibiotic selection pressure[26]. Surprisingly, the researchers found that single members of the population were, on average, less resistant to the antibiotic than the population as a whole. Additionally, resistant members were found to produce high levels of indole, a signaling molecule used by the surrounding, less resistant members to induce protective mechanisms against the antibiotic.

In a related study, Kishony and coworkers followed parallel populations of *E. coli* under differing antibiotic selection pressures using a bacterial culture device termed the

morbidostat[27]. Unlike the previous study in which the concentration of antibiotic is adjusted in 24-hour intervals, the morbidostat continuously adjusts the concentration of the antibiotic to maintain a nearly constant selection pressure (as judged by changes in cell density, thereby making it a variation on the prototypical auxostat). The parallel populations were challenged with different antibiotics, and followed using whole-genome sequencing to assess the nature of the mutations endowing the populations with antibiotic resistance. For chloramphenicol and doxycycline, resistance proceeded through multiple pathways, ultimately resulting in similar levels of antibiotic resistance using different sets of mutations. Conversely, resistance to trimethoprim proceeded through a highly sequential series of mutations in *DHFR* and its associated promoter in all the parallel populations. Importantly, many resistance mutations that had been previously identified using plate-based methods were not found in this study, highlighting variable fitness costs for some mutations depending on selection conditions, and the stochasticity of evolution under selection conditions for which many potential pathways to increased fitness exist.

For some applications the basal rate of bacterial mutagenesis may not be high enough to provide access to necessary genetic changes on a practical timescale. Methods for the unbiased, yet targeted, mutagenesis of bacteria would dramatically facilitate these applications. An ideal method would require that the evolving gene(s) be replicated by an orthogonal, error-prone polymerase enzyme. One such system uses the *E. coli* DNA polymerase I (Pol I). Bacteria lacking Pol I are viable, suggesting that manipulation of this polymerase might be tolerated. Loeb and coworkers used previously reported mutations that increase the error rate of Pol I to continuously evolve TEM beta-

lactamases capable of hydrolyzing a third-generation lactam antibiotic, aztreonam[28]. Three serial dilutions of a growing bacterial culture in increasing concentrations of aztreonam strongly enriched clinically relevant TEM mutations. The utility of this technique extends beyond the continuous directed evolution of antibiotic resistance, as it also enables the potent and targeted *in vivo* mutagenesis of a gene of interest. It should be noted, however, that an inverse relationship exists between the length of the gene of interest and the mutagenesis rate using this system and others that depend on Pol I mutants[28,29].

Bacterial continuous evolution has also been used to evolve phenotypes other than antibiotic resistance. The *E. coli* DNA polymerase III epsilon subunit DnaQ is responsible for the proofreading activity of most DNA replication. Dominant-negative variants of DnaQ provided *in trans* are known to have dramatic consequences on replication fidelity[30]. Integrating these principles, Ma and coworkers constructed a library of *dnaQ* mutants using error-prone PCR to be expressed constitutively in continuously growing bacterial cultures[31]. Selection pressures in the form of increasing concentrations of *n*-butanol or acetic acid were applied to the growing cultures. Plasmids carrying mutagenic *dnaQ* variants should be enriched at high concentrations of *n*-butanol or acetic acid because they endow their hosts with the ability to more quickly adapt to the strong selection pressure. Within 24 to 36 days of continuous evolution, the *E. coli* populations became adapted to the high solvent concentrations. As the researchers predicted, all the *dnaQ* variants that evolved in surviving cells were more mutagenic than the wildtype *dnaQ*, increased the mutation rate by up to 2,800-fold, and contained mutations known to dramatically enhance mutagenesis[30,31].

All of these examples of bacterial continuous evolution rely on the manual passaging of the bacteria being studied, or use an automated variant of a chemostat or auxostat. In either case, biofilm formation can dramatically affect the success of the experiment, as bacteria are known to attach to the culture vessel wall upon extended propagation. To ameliorate this issue, an apparatus dubbed the "GM3 cultivation device" was recently developed to allow the maintenance of steady-state microbial growth over extremely long timeframes (**Figure 1.3**). This apparatus uses two growth chambers that are cyclically used to culture the microbe of interest, with alternating rounds of chemical

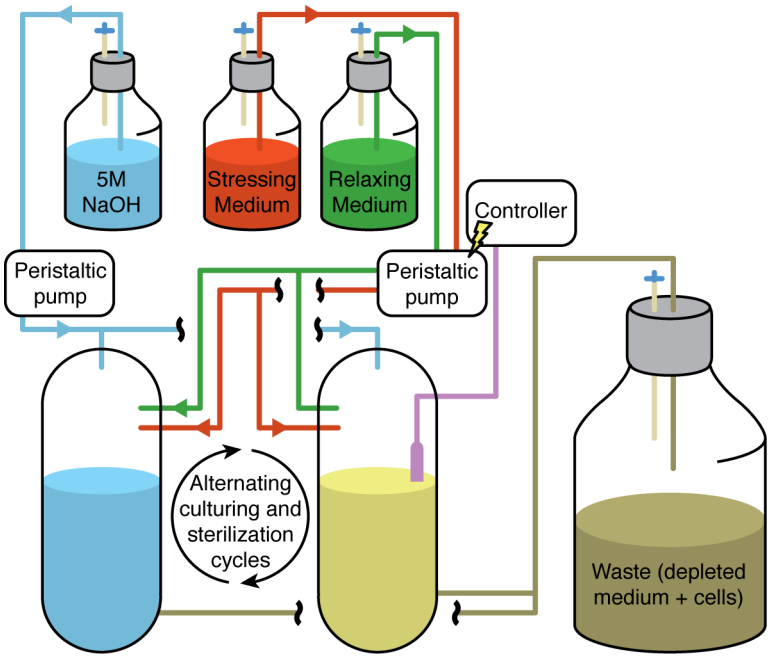


Figure 1.3 | The GM3 cultivation device. A growing culture in the first vessel is supplemented with a defined ratio of relaxing and stressing medium. This ratio is regulated by a controller connected to a turbidity meter measuring cell density. After a defined growth time, the whole culture is transferred to the second vessel to limit selection escape through biofilm formation. Following this transfer, the first vessel is sterilized with 5 M NaOH and washed extensively to limit contamination. After another defined growth period in the second vessel, the growing culture is transferred back to the first vessel, and the cycle is repeated. The selection stringency is modulated through the ratio of the relaxing and stressing media, ultimately reaching up to 100% stressing medium. The depleted medium and cells are pumped to a waste container.

sterilization that ensure that members of the population do not escape selections for faster growth through biofilm formation.

In one application of the GM3 device, Mutzel and coworkers evolved *E. coli* containing the large-scale substitution of chromosomal thymine with 5-chlorouracil[32]. The researchers used a thymine auxotroph grown in continuous culture on two types of media: a permissive media with all essential nutrients including thymine, and a selective media of similar composition with the exception of thymine, which was substituted with 5-chlorouracil. To slowly modulate selection stringency they mixed varying ratios of the two media, starting from 100% thymine-containing media and slowly transitioning to 100% 5-chlorouracil-containing media over six months. By the end of the experiment, strains with a significant number of mutations or chromosomal rearrangements that could grow using 5-chlorouracil were isolated. These strains were capable of incorporating 5-chlorouracil at levels up to 90% of the chromosomal DNA, with the remaining 10% attributed to thymine. After further optimization, the fraction of chromosomal thymine was reduced to less than 2% through removal of a cryptic pathway to thymine deoxynucleotide production.

In another application of the same device, Hilvert and coworkers continuously evolved an artificial chorismate mutase enzyme that had been previously designed using a nine-amino-acid code[33]. Chorismate mutase natively converts chorismate to prephenate in the shikimate pathway, a critical step in the biosynthesis of the aromatic amino acids phenylalanine and tyrosine. A strain of *E. coli* lacking wild-type chorismate mutase and expressing the artificial variant was continuously propagated for more than 200 days, initially in low phenylalanine-containing media and slowly transitioning to media lacking

both phenylalanine and tyrosine. Bacteria carrying improved chorismate mutase mutants grew at a faster rate in the restrictive media, resulting in their enrichment over the course of the experiment. By the end of the experiment, chorismate mutase variants that expanded the nine amino acid code to ten or eleven amino acids could be isolated, and mutants generally had improved activity as compared to the parent. This study suggests that expanded genetic codes can result in improved evolvability, a similar conclusion to those reached by Barrick and coworkers using 3-iodotyrosine incorporation in T7 bacteriophage[18].

1.4 – Eukaryotic Continuous Evolution

Viral and bacterial hosts are not suitable for the directed evolution of biomolecules in certain scenarios. In cases of industrial drug precursor production[34] or whole-cell biocatalysis[35], higher organisms with more extensive or relevant biosynthetic pathways may be necessary. Towards this end, continuous directed evolution techniques using eukaryotic organisms have also been developed.

In one of the earliest examples of eukaryotic continuous evolution, the Evolugator, a proprietary continuous culturing device that reduces biofilm formation and wall growth (**Figure 1.4**), was used to evolve thermotolerant filamentous fungi. During long-term continuous culturing of filamentous fungi, wall growth can be problematic as fungal cells rapidly aggregate, leading to non-uniform culture dilution. To circumvent this issue, de Crécy-Lagard and coworkers devised an alternative method for continuous fungi culturing that dramatically ameliorates fungal wall growth[36,37]. Using this platform over the course of four months, the researchers evolved two thermotolerant variants of

the entomopathogenic fungus *Metarhizium anisopliae* to facilitate the widespread adoption of this fungus for pest control[37]. While successful, this method for continuous directed evolution requires specialized equipment that cannot be readily generalized to

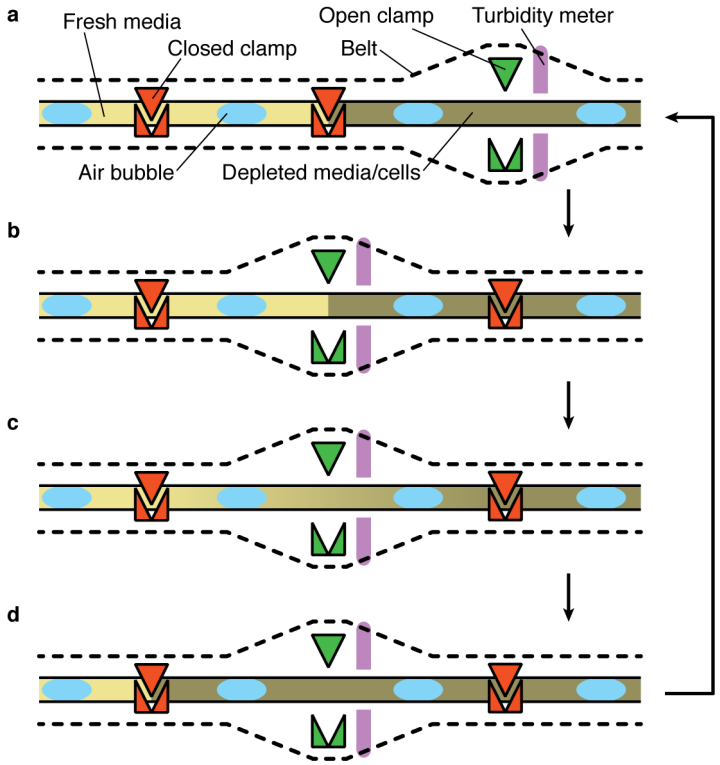


Figure 1.4 | Growth chamber of the Evolugator. a, The growing culture is shown on the right, and is completely contained within a tube clamped at regular intervals to limit unintentional mixing. The density of the culture is measured with a turbidity meter connected to a controller that dynamically regulates the in-flux of new medium. Air bubbles are included at regular intervals to increase the oxygenation level of the medium. **b,** Upon reaching high density, the right clamp closes, the tubing moves along the track, and the next clamp opens allowing for contact between the growing culture and the fresh medium. **c,** The fresh medium mixes with the growing culture through diffusion, and the culture resumes growing. **d,** After sufficient time, the culture again reaches high density. The middle clamp then closes, the tubing shifts on the track once again, the next clamp opens and the new medium is allowed in. This cycle is then continued for the duration of the experiment.

alternative selection methodologies.

Cornish and coworkers recently noted that many *in vivo* selection methods do not take advantage of genetic recombination as a source of diversification, and reasoned that the selective induction of double-stranded breaks (DSBs) with high efficiency could dramatically increase the degree of recombination in *Saccharomyces cerevisiae*[38]. They developed a yeast continuous evolution system made of three parts: a plasmid that tightly controls the expression of the site-specific endonuclease I-*Sce*I, a cassette plasmid that contains regions of homology to the target sequence flanked by I-*Sce*I cleavage sites, and a target plasmid carrying the gene to be mutagenized. Upon induction of the endonuclease, the mutagenic cassette is liberated from its respective plasmid and goes on to mutate the region of interest through homology-directed recombination[38]. Using this system, the authors were able to evolve *hisA* variants capable of complementing *trp1* auxotrophy in yeast in less than one month. This system is accessible and potentially powerful but requires knowledge of the relevant regions of a gene to mutate since genetic diversity is dependent on homology-directed recombination.

One of the most useful features of a continuous directed evolution system is the ability to modulate the mutagenesis of the evolving genes in a manner that does not perturb either the selection or the organism. This feature was one of the major design considerations for the development of the PACE platform[20], in which mutations accumulate only in replicating phage; the error-prone bacterial Pol I variants[28], in which mutations accumulate only in ColE1-type origin vectors; and the I-*Sce*I-based yeast recombination system[38], in which mutations are restricted to regions highly homologous to the mutagenic cassette. All of these systems largely abrogate or avoid host strain

mutagenesis. A similar approach was recently developed by Chang Liu and coworkers using a system that is completely contained in yeast. The authors used an autonomous DNA replication system from *Kluveromyces lactis* to enable the replication of a gene of interest in the *S. cerevisiae* cytosol[39]. Due to the strict requirements for the orthogonal *K. lactis* DNA polymerase initiation, no crosstalk was observed between the *K. lactis* components and the native host machinery. To make this system more amenable to directed evolution, the authors used a homology-guided approach to create error-prone DNA polymerase variants that would only replicate the orthogonal plasmid. The optimized system was able to induce ~300x higher rates of mutagenesis in the orthogonal plasmid as compared to the background *S. cerevisiae* mutagenesis rate, with no reduction in the genomic replication fidelity. This system represents an important methodological advance in continuous directed evolution in eukaryotes.

1.5 – Discussion

Methods for the directed evolution of biomolecules have proven to be important technologies for generating new functional proteins and nucleic acids as research tools and therapeutics, as well as for facilitating the investigation of fundamental biological questions. Some of the limitations of traditional stepwise laboratory evolution approaches can be overcome by continuous directed evolution platforms. Continuous directed evolution intrinsically enables a broader exploration of sequence space in a practical timescale compared to traditional discrete laboratory evolution methods. Indeed, *in vivo* continuous evolution methods can enable > 10^{11} protein variants to be generated and subjected to selection over > 100 generations of evolution in less than a week[20]. The

efficiency of continuous evolution methods can enable long evolutionary trajectories[14,23,32,33] or access to highly evolved biomolecules with new properties that would otherwise require impractical time scales[1,2,18,24,27,31,38]. In addition, some continuous evolution platforms, including several described above, can avoid genetic bottlenecks from modest population sizes, modest screening throughput, or modest mutation rates that commonly constrain some traditional laboratory evolution methods. The minimal reliance on researcher intervention during continuous evolution can also make performing multiple parallel evolution experiments more accessible than using traditional directed evolution platforms[12-14,18-20,22-24,27,31,39].

Continuous directed evolution methods are not without their own drawbacks and should be considered a complement, rather than a replacement, to other laboratory evolution approaches. Establishing a continuous cycle of translation, mutation, selection, and replication that is general for a range of biomolecules is challenging, and therefore most continuous evolution methods perform these key steps inside cells. As a result, while traditional *in vitro* directed evolution techniques enable exquisite control of selection stringency and mutation rate through simple adjustments of additive concentrations, analogous perturbations are more difficult to implement *in vivo* and generally must be empirically determined. Because *in vivo* continuous evolution platforms typically require continuous culture of microorganisms, contamination either by "cheaters" (uninteresting gene variants that bypass selections) or by inoculation with undesired organisms poses additional hurdles for implementing these methods.

Finally, the development of selection methods compatible with continuous evolution is typically more difficult than the development of analogous selections or

screens for traditional stepwise laboratory evolution methods. Evolved biomolecules with desired properties in a continuous evolution system must be linked to gene and organism replication, rather than to the generation of fluorescent or colorimetric signals that have proven to be robust staples of *in vitro* and cell-based screening. While establishing this linkage between desired biomolecule function and cell survival is relatively straightforward for genes that are natively associated with organismal fitness (such as antibiotic resistance, temperature or solvent tolerance, or improved growth rate), the design of selections for continuous evolution is more challenging for genes and gene products that do not intrinsically affect cellular survival. Fortunately, many creative strategies have been described for coupling a wide variety of molecular activities to cell survival, including protein-fragment complementation of enzymatic activities and organelle-specific localization signals, *n*-hybrid systems for binding activity and bond-formation catalysis[40,41], derepression-based methods for bond-cleaving activities[42], gene regulation-based selections for activities associated with transcription or translation[43], and *cis*-acting elements for metabolite sensing[44]. We anticipate that the expansion of such selection methods to address an increasingly broad scope of molecular problems of interest will play a key role in fully realizing the potential and defining the future impact of continuous directed evolution.

1.6 – References

1 Mills, D. R., Peterson, R. L. & Spiegelman, S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 217-224, (1967).

2 Wright, M. C. & Joyce, G. F. Continuous in vitro evolution of catalytic function. *Science* **276**, 614-617, (1997).

3 Breaker, R. R., Banerji, A. & Joyce, G. F. Continuous in vitro evolution of bacteriophage RNA polymerase promoters. *Biochemistry* **33**, 11980-11986, (1994).

4 McGinness, K. E., Wright, M. C. & Joyce, G. F. Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chemistry & biology* **9**, 585-596, (2002).

5 Kuhne, H. & Joyce, G. F. Continuous in vitro evolution of ribozymes that operate under conditions of extreme pH. *Journal of molecular evolution* **57**, 292-298, (2003).

6 Johns, G. C. & Joyce, G. F. The promise and peril of continuous in vitro evolution. *Journal of molecular evolution* **61**, 253-263, (2005).

7 Voytek, S. B. & Joyce, G. F. Emergence of a fast-reacting ribozyme that is capable of undergoing continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 15288-15293, (2007).

8 Paegel, B. M. & Joyce, G. F. Darwinian evolution on a chip. *PLoS biology* **6**, e85, (2008).

9 Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667-1686, (1998).

10 Husimi, Y. Selection and evolution of bacteriophages in cellstat. *Advances in biophysics* **25**, 1-43, (1989).

11 Hoskisson, P. A. & Hobbs, G. Continuous culture--making a comeback?

Microbiology **151**, 3153-3159, (2005).

12 Bull, J. J. *et al.* Exceptional convergent evolution in a virus. *Genetics* **147**, 1497-1507, (1997).

13 Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull, J. J. Different trajectories of parallel evolution during viral adaptation. *Science* **285**, 422-424, (1999).

14 Wichman, H. A., Millstein, J. & Bull, J. J. Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* **170**, 19-31, (2005).

15 Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature reviews. Molecular cell biology* **10**, 866-876, (2009).

16 Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R. & Molineux, I. J. Experimental phylogenetics: generation of a known phylogeny. *Science* **255**, 589-592, (1992).

17 Bull, J. J., Badgett, M. R., Rokyta, D. & Molineux, I. J. Experimental evolution yields hundreds of mutations in a functional viral genome. *Journal of molecular evolution* **57**, 241-248, (2003).

18 Hammerling, M. J. *et al.* Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. *Nature chemical biology* **10**, 178-180, (2014).

19 Meyer, J. R. *et al.* Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428-432, (2012).

20 Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503, (2011).

21 Rakonjac, J., Bennett, N. J., Spagnuolo, J., Gagic, D. & Russel, M. Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Current issues in molecular biology* **13**, 51-76, (2011).

22 Leconte, A. M. *et al.* A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary

outcomes. *Biochemistry* **52**, 1490-1499, (2013).

23 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. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9007-9012, (2013).

24 Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature chemical biology* **10**, 216-222, (2014).

25 Dickinson, B. C., Packer, M. S., Badran, A. H. & Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Submitted*, (2014).

26 Lee, H. H., Molla, M. N., Cantor, C. R. & Collins, J. J. Bacterial charity work leads to population-wide resistance. *Nature* **467**, 82-85, (2010).

27 Toprak, E. *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature genetics* **44**, 101-105, (2012).

28 Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9727-9732, (2003).

29 Fabret, C. *et al.* Efficient gene targeted random mutagenesis in genetically stable Escherichia coli strains. *Nucleic acids research* **28**, E95, (2000).

30 Fijalkowska, I. J. & Schaaper, R. M. Mutants in the Exo I motif of Escherichia coli dnaQ: defective proofreading and inviability due to error catastrophe. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 2856-2861, (1996).

31 Luan, G., Cai, Z., Li, Y. & Ma, Y. Genome replication engineering assisted continuous evolution (GREACE) to improve microbial tolerance for biofuels production. *Biotechnology for biofuels* **6**, 137, (2013).

32 Marliere, P. *et al.* Chemical evolution of a bacterium's genome. *Angewandte*

Chemie **50**, 7109-7114, (2011).

33 Muller, M. M. *et al.* Directed evolution of a model primordial enzyme provides insights into the development of the genetic code. *PLoS genetics* **9**, e1003187, (2013).

34 Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943, (2006).

35 Ishige, T., Honda, K. & Shimizu, S. Whole organism biocatalysis. *Current opinion in chemical biology* **9**, 174-180, (2005).

36 de Crecy, E. *et al.* Development of a novel continuous culture device for experimental evolution of bacterial populations. *Applied microbiology and biotechnology* **77**, 489-496, (2007).

37 de Crecy, E., Jaronski, S., Lyons, B., Lyons, T. J. & Keyhani, N. O. Directed evolution of a filamentous fungus for thermotolerance. *BMC biotechnology* **9**, 74, (2009).

38 Romanini, D. W., Peralta-Yahya, P., Mondol, V. & Cornish, V. W. A Heritable Recombination system for synthetic Darwinian evolution in yeast. *ACS synthetic biology* **1**, 602-609, (2012).

39 Ravikumar, A., Arrieta, A. & Liu, C. C. An orthogonal DNA replication system in yeast. *Nature chemical biology* **10**, 175-177, (2014).

40 Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I. & Stefan, E. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nature reviews. Drug discovery* **6**, 569-582, (2007).

41 Stynen, B., Tournu, H., Tavernier, J. & Van Dijck, P. Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. *Microbiology and molecular biology reviews : MMBR* **76**, 331-382, (2012).

42 van Rossum, T., Kengen, S. W. & van der Oost, J. Reporter-based screening and selection of enzymes. *The FEBS journal* **280**, 2979-2996, (2013).

43 Schaerli, Y. & Isalan, M. Building synthetic gene circuits from combinatorial libraries: screening and selection strategies. *Molecular bioSystems* **9**, 1559-1567, (2013).

44 Wang, Y. H., Wei, K. Y. & Smolke, C. D. Synthetic biology: advancing the design of diverse genetic systems. *Annual review of chemical and biomolecular engineering* **4**, 69-102, (2013).

Chapter 2

Development of a Sensitive Platform for the Detection and Continuous Evolution of Protein-Protein Interactions

Parts adapted from: Badran *et al.*, *Nature* (2016) [**in press**]

2.1 – Introduction: *In vivo* detection platforms for protein-protein interactions

Methods for the *in vivo* detection of protein-protein interactions offer a number of advantages over *in vitro* methods, notably the use of a native cellular environment and lack of biomolecule purification or preparation steps[1]. However, *in vivo* methods can lack the sensitivity afforded by *in vitro* methods, typically require the use of fusion proteins to reporters of the interaction, yield qualitative rather than quantitative measures of the interaction affinity, and have been known to produce a high frequency of false positives[1]. Despite this, methods for the *in vivo* detection of protein-protein interaction continue to be developed, as they enable biomolecule improvements through directed evolution to a greater extent than their *in vitro* counterparts[1-3].

Most widely applied methods for in vivo protein-protein interaction detection fall into one of two categories: split-protein complementation or *n*-hybrid schemes. In the former, a reporter of interest is split into two or more fragments and requires complementation by a protein-protein interaction to yield functional reporter. Notably examples of split reporters include ubiquitin, β-galactosidases, β-lactamases, luciferases, dihydrofolate reductases, and fluorescent proteins. In most cases, these reporters suffer from a high background activity in the absence of a protein-protein interaction, low activity upon complementation, or require the use of specialized techniques (e.g. flow cytometry) to quantify interaction affinity. In the latter, interaction of two or more biomolecules results in the requirement of transcriptional machinery upstream of a reporter of interest, resulting in transcriptional activation. N-hybrid systems have been demonstrated in organisms ranging from *E. coli* to immortalized human cell lines, highlighting the utility of the approach, but all seem to suffer from significant false

positive discovery frequency.

In this chapter, I design and optimize a system for the detection of protein-protein interactions for the ultimate use during Phage-Assisted Continuous Evolution. This system builds upon the work of many others using the bacterial 2-hybird. The finalized system demonstrates a very low false positive discovery frequency, can respond to interactions as low as micromolar affinities, and has a dynamic range rivaling the yeast counterpart.

2.2 – Development of a PACE selection for protein binding

PACE has mediated the rapid laboratory evolution of diverse protein classes including polymerases, proteases, and genome-editing proteins, yielding variants with highly altered activities and specificities[5-11]. While PACE has not been previously used to evolve protein-binding activity, we speculated that the bacterial 2-hybrid system[12] could serve as the basis of a protein-binding PACE selection. The bacterial 2-hybrid consists of a DNA-binding protein domain fused to a target protein (the "bait"), and a prospective protein-binding partner (the "prey") fused to a transcriptional activation domain that recruits *E. coli* RNA polymerase[12] (**Figure 2.1A**). Bait-prey binding results in localization of RNA polymerase upstream of a reporter gene, initiating gene expression. To adapt this system into a protein-binding selection for PACE, we envisioned that protein-target binding could instead activate the expression of the filamentous bacteriophage gene III, which is required for the infectivity of progeny phage[5] (**Figure 2.1B**).

To maximize the sensitivity of the bacterial 2-hybrid, we extensively optimized

selection parameters including (i) choice of activation and DNA-binding domains, (ii) expression level of either protein partner, (iii) interaction binding affinity, (iv) DNA-binding domain multivalency state, (v) reporter gene ribosome-binding site, (vi) operator–promoter distance, (vii) RNA polymerase–promoter affinity, and (viii) DNA-binding domain–bait linker length. While the previously described bacterial 2-hybrid system yielded a 17-fold increase in transcriptional activation using a model high-affinity

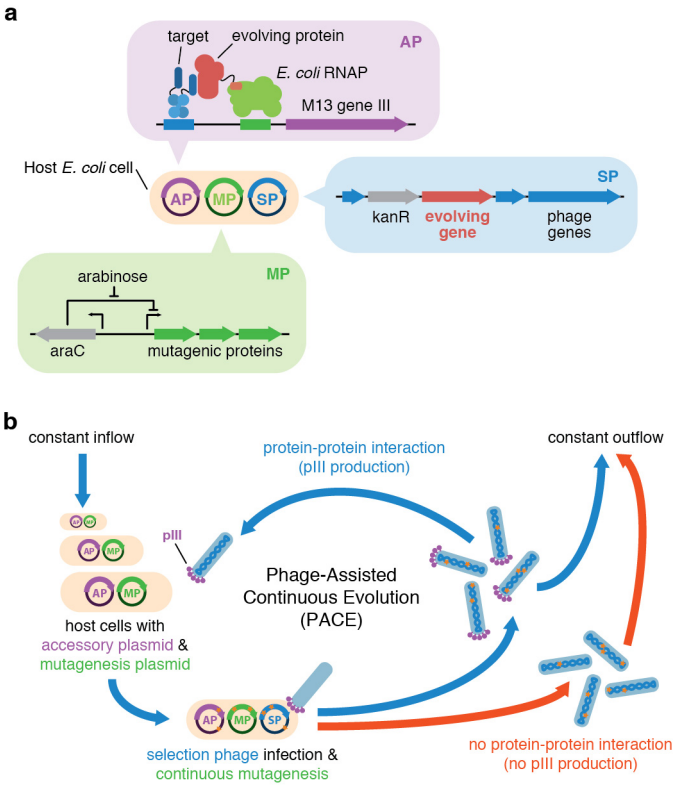


Figure 2.1 | Protein-binding phage-assisted continuous evolution (PACE). a, Anatomy of a phage-infected host cell during PACE. The host *E. coli* cell carries two plasmids: the accessory plasmid (AP), which links protein binding to phage propagation and controls selection stringency, and the mutagenesis plasmid (MP), which enables arabinose-inducible elevated level mutagenesis during PACE. **b,** Following infection, selection phage (SP) that encode an evolving protein capable of binding to the target protein induces expression of gene III from the AP, resulting in the production of pIII, a phage protein required for progeny phage produced by that host cell to infect subsequent host cells. PACE takes place in a fixed-volume vessel (the "lagoon") that is continuously diluted with fresh host cells. Only those SP encoding proteins that bind the target can propagate faster than they are diluted out of the lagoon.

interaction (HA4 monobody binding to the SH2 domain of ABL1 kinase)[13], our optimized system enhanced transcriptional activation > 200-fold using the same interaction (**Figure 2.2 – 2.4**). This system consisted of the *E. coli* RNA polymerase omega subunit (RpoZ) as the activation domain, the 434 phage cI repressor as the DNA-binding domain, and an optimized P_{lacZ}-derived promoter ($P_{lacZ\text{-}opt}$) to drive reporter transcription. Together, these results extend and improve previously described bacterial systems[12] that transduce protein-target binding into gene expression in a manner that can be easily tuned by the researcher.

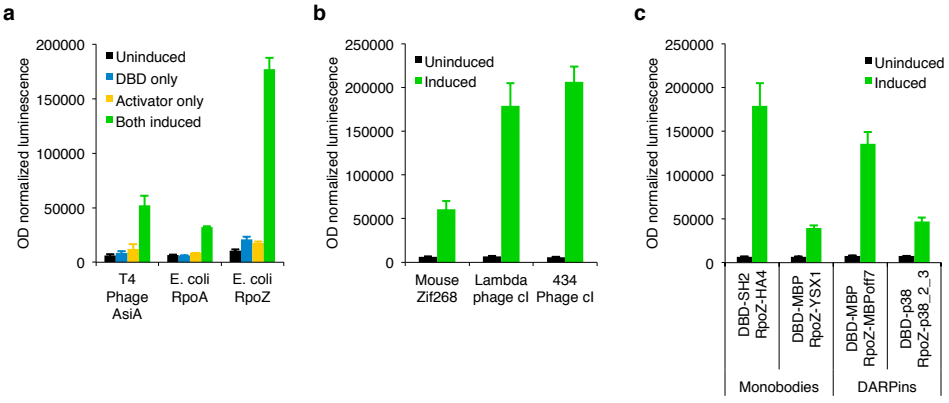


Figure 2.2 | Bacterial 2-hybrid component validation. a, Plasmids encoding an IPTG-inducible *λcI-SH2* cassette ("DBD") and an ATc-inducible activator-HA4 cassette ("activator") were co-transformed into the *E. coli* S1030 host strain and induced using either or both small molecules. T4 AsiA-mediated transcriptional activation required low-level expression of the σ70 (R541C/F563Y/L607P) mutant to alleviate AsiA toxicity[4]. Use of RpoZ as the activation domain showed the greatest degree of transcriptional activation (~17-fold). **b,** DNA-binding domain variation shows that multivalent phage repressors yield a greater degree of transcriptional activation than the monomeric zinc finger Zif268. **c,** Transcriptional activation from a combination of the λcI DNA binding domain and RpoZ transcriptional activator was evaluated using a number of previously evolved protein-protein interactions involving either monobodies or DARPins, showing the generality of binding interaction detection. Error bars reflect the standard deviation of at least three independent biological replicates.

a

DBD	Activator	Promoter	Uninduced	Induced	Fold
Lambda cI	RpoZ	cattaggcaccccgggctttacactttatgcttccggctcgtatgttgtgtcgaccg	525	11689	22
434 cI	RpoZ	cattaggcaccccgggctttacactttatgcttccggctcgtatgttgtgtcgaccg	330	18167	55
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgtgtcgaccg	3252	182235	56
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgGtgtgtcgaccg	4542	467340	103
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtataCgttgtgtcgaccg	767	51636	67
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtaCgGtgtgtcgaccg	945	64088	68
434 cI	RpoZ	cattaggcaccccgggcttGacacGtAatgcttccggcGcgtatgttgtgtcgaccg	432159	712599	2
434 cI	RpoZ	cattaggcaccccgggcttGacacGtAatgcttccggcGcgtatgGtgtgtcgaccg	556230	750804	1
434 cI	RpoZ	cattaggcaccccgggctttacGcGtAatgcttccggcGcgtatgttgtgtcgaccg	1015	97100	96
434 cI	RpoZ	cattaggcaccccgggctGtacGcGtAatgcttccggcGcgtatgttgtgtcgaccg	103	8579	83
434 cI	RpoZ	cattaggcaccccgggctGtacGcGtAatgcttccggcGcgtaCgttgtgtcgaccg	69	899	13
434 cI	RpoZ	cattaggcaccccgggctGtacGcGtAatgcttccggcGcgtaCgGtgtgtcgaccg	52	678	13
434 cI	RpoZ	cattaggcaccccgggctGtacGcGtAatgcttccggcGcgtatgttgCgtcgaccg	202	1280	6
434 cI	RpoZ	cattaggcaccccgggctGtacGcGtAatgcttccggcGcgtatgttgGgtcgaccg	139	2307	17
434 cI	RpoZ	cattaggcaccccgggcttGacGcGtAatgcttccggcGcgtatgttgtgtcgaccg	50526	494779	10
434 cI	RpoZ	cattaggcaccccgggcttGacGcGtAatgcttccggcGcgtatgGtgtgtcgaccg	66275	481567	7
434 cI	RpoZ	cattaggcaccccgggcGttacacGtAatgcttccggcGcgtatgGtgtgtcgaccg	229	1851	8
434 cI	RpoZ	cattaggcaccccgggcAttacacGtAatgcttccggcGcgtatgGtgtgtcgaccg	382	3450	9
434 cI	RpoZ	cattaggcaccccgggcCttacacGtAatgcttccggcGcgtatgGtgtgtcgaccg	131	3062	23
434 cI	RpoZ	cattaggcaccccgggctGtacacGtAatgcttccggcGcgtatgGtgtgtcgaccg	725	76430	105
434 cI	RpoZ	cattaggcaccccgggctCtacacGtAatgcttccggcGcgtatgGtgtgtcgaccg	80	668	8
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgCgtcgaccg	859	86407	101
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgCgtcgaccg	943	97337	103
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgCgCcgaccg	554	21807	39
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgCgGcgaccg	428	34919	82
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgGgtcgaccg	2229	309795	139
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgGgCcgaccg	989	92428	93
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgGgGcgaccg	1523	149661	98
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgAgtcgaccg	2203	296387	135
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgAgCcgaccg	1271	167091	131
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgttgAgGcgaccg	1906	189337	99
434 cI	RpoZ	cattaggcaccccgggctttacGcGtAatgcttccggcGcgtatgttgAgCcgaccg	230	28708	125
434 cI	RpoZ	cattaggcaccccgggctttacGcGtAatgcttccggcGcgtatgttgAgGcgaccg	269	42314	158
434 cI	RpoZ	cattaggcaccccgggctGtacacGtAatgcttccggcGcgtatgttgAgCcgaccg	122	10658	87
434 cI	RpoZ	cattaggcaccccgggctGtacacGtAatgcttccggcGcgtatgttgAgGcgaccg	201	22016	109
434 cI	RpoZ	cattaggcaccccgggctttacGcGtAatgcttccggcGcgtatgGtgAgCcgaccg	566	24652	44
434 cI	RpoZ	cattaggcaccccgggctttacGcGtAatgcttccggcGcgtatgGtgAgGcgaccg	753	78485	104
434 cI	RpoZ	cattaggcaccccgggctGtacacGtAatgcttccggcGcgtatgGtgAgCcgaccg	878	26019	30
434 cI	RpoZ	cattaggcaccccgggctGtacacGtAatgcttccggcGcgtatgGtgAgGcgaccg	1041	52307	50
434 cI	RpoZ	cattaggcaccccgggAtttacacGtAatgcttccggcGcgtatgttgtgtcgaccg	2480	164725	66
434 cI	RpoZ	cattaggcaccccgggctttGcacGtAatgcttccggcGcgtatgttgtgtcgaccg	1015	46198	46
434 cI	RpoZ	cattaggcaccccgggctttacTcGtAatgcttccggcGcgtatgttgtgtcgaccg	1294	151683	117
434 cI	RpoZ	cattaggcaccccgggctttacTcGtCatgcttccggcGcgtatgttgtgtcgaccg	1259	134645	107
434 cI	RpoZ	cattaggcaccccgggctttacTcGtAaAgcttccggcGcgtatgttgtgtcgaccg	412	80809	196
434 cI	RpoZ	cattaggcaccccgggctttacTcGtAatgcttccggTGcgtatgttgtgtcgaccg	613	38195	62
434 cI	RpoZ	cattaggcaccccgggctttacCcGtAatgcttccggcGcgtatgttgtgtcgaccg	601	53127	88
434 cI	RpoZ	cattaggcaccccgggctttacacGtCatgcttccggcGcgtatgttgtgtcgaccg	2975	371487	125
434 cI	RpoZ	cattaggcaccccgggctttacacGtCaAgcttccggcGcgtatgttgtgtcgaccg	1903	255198	134
434 cI	RpoZ	cattaggcaccccgggctttacacGtCatgcttccggTGcgtatgttgtgtcgaccg	1271	2597	2
434 cI	RpoZ	cattaggcaccccgggctttacacGtAaAgcttccggcGcgtatgttgtgtcgaccg	1371	218447	159
434 cI	RpoZ	cattaggcaccccgggctttacacGtAaAgcttccggTGcgtatgttgtgtcgaccg	813	62941	77
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcCtccggcGcgtatgttgtgtcgaccg	33730	218546	6
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggTGcgtatgttgtgtcgaccg	1146	160609	140
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGTgtatgttgtgtcgaccg	39981	777784	19
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGAgtatgttgtgtcgaccg	6010	638749	106
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtaGgttgtgtcgaccg	489	46068	94
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggcGcgtatgtAgtgtcgaccg	291	4553	16
434 cI	RpoZ	cattaggcaccccgggctttacTcGtAatgcttccggcGcgtatgttgAgCcgaccg	685	44500	65
434 cI	RpoZ	cattaggcaccccgggctttacTcGtCatgcttccggcGcgtatgttgAgCcgaccg	638	43688	68
434 cI	RpoZ	cattaggcaccccgggctttacTcGtAaAgcttccggcGcgtatgttgAgCcgaccg	488	18713	38
434 cI	RpoZ	cattaggcaccccgggctttacTcGtAatgcttccggTGcgtatgttgAgCcgaccg	278	9431	34
434 cI	RpoZ	cattaggcaccccgggctttacacGtCatgcttccggcGcgtatgttgAgCcgaccg	1890	176933	94
434 cI	RpoZ	cattaggcaccccgggctttacacGtCaAgcttccggcGcgtatgttgAgCcgaccg	709	68843	97
434 cI	RpoZ	cattaggcaccccgggctttacacGtCatgcttccggTGcgtatgttgAgCcgaccg	908	49205	54
434 cI	RpoZ	cattaggcaccccgggctttacacGtAaAgcttccggcGcgtatgttgAgCcgaccg	1060	70184	66
434 cI	RpoZ	cattaggcaccccgggctttacacGtAaAgcttccggTGcgtatgttgAgCcgaccg	524	17228	33
434 cI	RpoZ	cattaggcaccccgggctttacacGtAatgcttccggTGcgtatgttgAgCcgaccg	696	36530	52

b

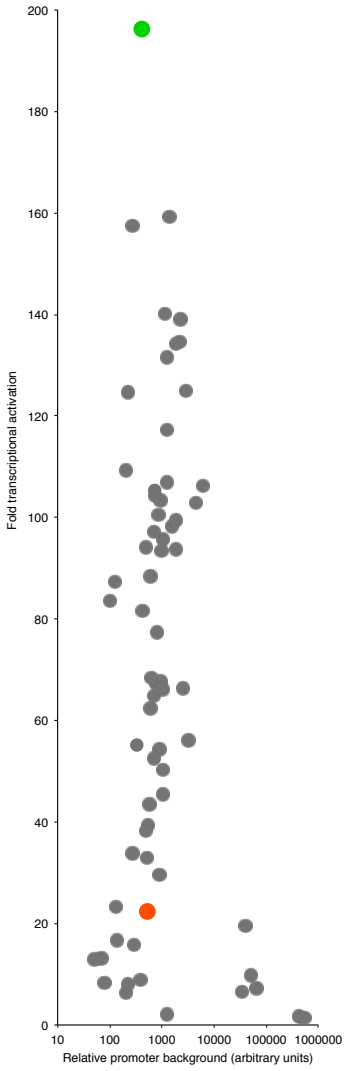


Figure 2.3 | Optimization of the P_{lacZ} promoter for improved sensitivity and dynamic range. a, Promoter and DNA-binding domain combinations tested during P_{lacZ} optimization, showing uninduced and induced levels of OD-normalized luminescence. The SH2/HA4 interaction pair was used in all cases. The fold activation in each case was calculated as the ratio of the induced and uninduced luciferase expression signals. **b,** Graphical representation of the data in part (**a**), showing the wide distribution of promoter background levels and degrees of transcriptional activation. In both (A) and (B), the red and green dots indicate the starting (P_{lac62}) and final ($P_{lacZ\text{-}opt}$) promoter/DNA-binding domain combinations, respectively. Each data point in (**b**) reflects the average of at least three independent biological replicates.

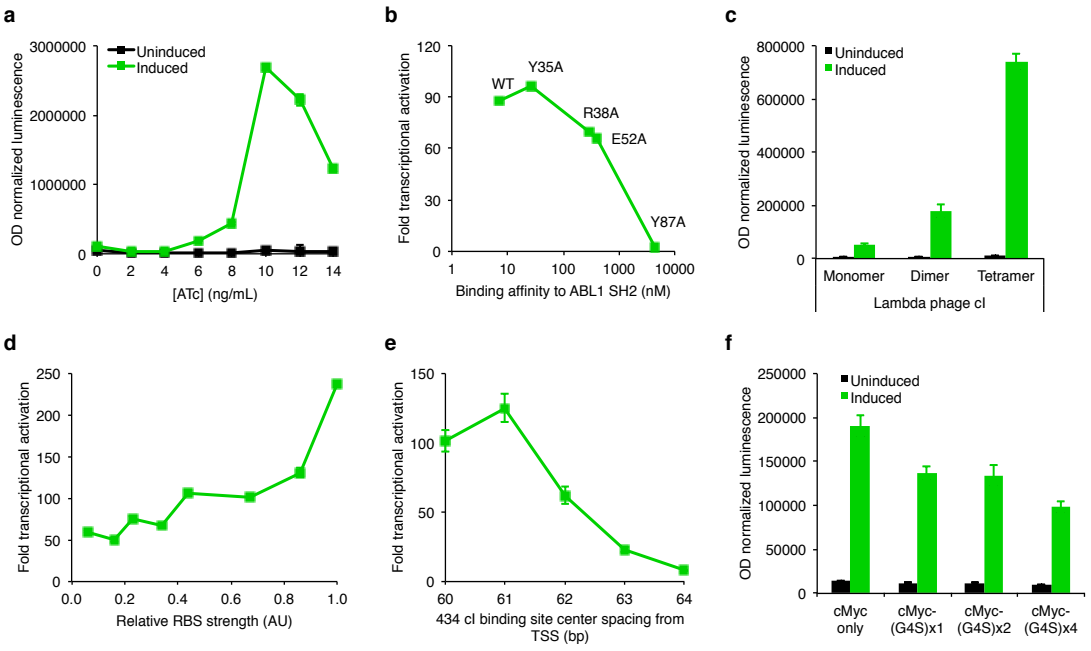


Figure 2.4 | Bacterial 2-hybrid optimization. a, Inducer titration of the interacting fusion proteins driving the 2-hyrbid system. The black and green lines represent the uninduced (0 μM IPTG) and induced (1 μM IPTG) levels of IPTG-inducible *434cI-SH2* expression, while ATc induces expression of the *rpoZ-HA4* cassette. In subsequent graphs and assays, the expression level resulting from the IPTG-inducible P_{lac} promoter was measured by Western blot and approximated using a constitutive promoter to reduce experimental variability. **b,** The degree of transcriptional activation using HA4 monobody mutants correlated with known binding affinities. The highest levels of activation resulted from K_d = low nM affinities, while weak affinities in the K_d = low μM range could still be detected. **c,** The relationship between DNA-binding domain multivalency state (monomeric, dimeric or tetrameric DNA-binding domain fused to the SH2 domain) and transcriptional activation resulting from the SH2/HA4 interaction, with higher multivalency states yielding greater activation levels. **d,** RBS modification enables robust modulation of the relative activation levels from the $P_{lacZ\text{-}opt}$ promoter using the SH2/HA4 interaction. **e,** Operator-promoter binding site spacing strongly affects transcriptional activation levels. 434cI binding at 61 base pairs upstream of the $P_{lacZ\text{-}opt}$ promoter resulted in the most robust activation. **f,** Linker extension to include one, two, or three G_4S motifs result in reduced activation levels using the SH2/HA4 interacting pair. Error bars throughout this figure reflect the standard deviation of at least three independent biological replicates.

2.3 – PACE evolves monobodies with antigen affinity

The HA4 monobody (antibody mimetic) binds to the SH2 domain of ABL1 kinase (K_d = 7 nM) in a manner highly dependent on HA4 residue Y87[13]. The mutant HA4 Y87A monobody binds the ABL1 SH2 domain with 100- to 1,000-fold weaker affinity[13]. Whereas wild-type HA4 monobody fused to RpoZ in the presence of 434cI-SH2 resulted in potent transcriptional activation in the optimized bacterial 2-hybrid system described above, transcriptional activation using $HA4_{Y87A}$ was negligible (**Figure 2.5a**). Next we tested the ability of this interaction to support phage propagation through activation of gene III expression. Selection phage (SP) expressing the *rpoZ-HA4* fusion robustly propagate in the PACE format using host cell strains carrying accessory plasmids (APs) expressing the *434cI-SH2* fusion, in agreement with our transcriptional activation measurements (**Figure 2.6a**). Importantly, SPs encoding either *rpoZ-HA4*$_{Y87A}$ or *rpoZ* alone did not support phage propagation, demonstrating the inability of the mutant HA4 monobody to support gene III expression (**Figure 2.6a**).

To validate the capability of protein-binding PACE to evolve a known protein-protein interaction, we challenged the system to evolve a functional SH2-binding monobody starting from the inactive $HA4_{Y87A}$ mutant. To revert the $HA4_{Y87A}$ mutant back to a Tyr87 protein requires three adjacent point mutations (GCG to either TAT or TAC). The *rpoZ-HA4*$_{Y87A}$ SP was propagated during PACE for 66 h at a flow rate of 2.0 lagoon vol/h in the absence of selection pressure (i.e., allowing evolutionary drift[8]), before engaging selection pressure by changing the host cell strain to one requiring SH2 binding-dependent phage propagation (**Figure 2.5b**). Under selection pressure, control lagoons that previously experienced neither drift nor mutagenesis, or that experienced

only mutagenesis, quickly lost their SPs encoding the evolving monobody population (phage "wash out"). In contrast, lagoons subjected to both drift and mutagenesis dropped markedly in phage titer for the first 12 h, but recovered over the next 24 h (**Figure 2.5c**). Sequence analysis of eight phage clones surviving 48 h of PACE revealed that all eight evolved either Tyr (three point mutations) or Trp (two point mutations) at HA4 position 87 (**Figure 2.5b,c**). We confirmed potent activation of transcription using either evolved consensus clone (**Figure 2.6c**). These results collectively demonstrate that protein-binding PACE can rapidly evolve proteins with target affinity, even when multiple mutations are required to gain protein-binding activity.

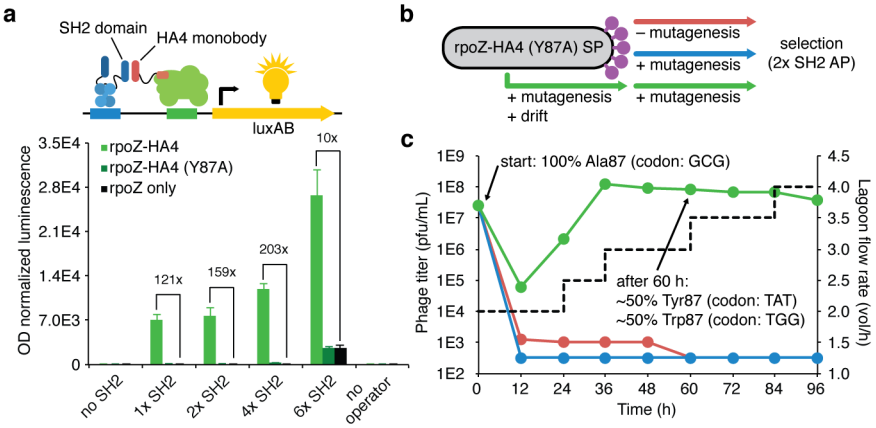


Figure 2.5 | Protein-binding PACE selection development and stringency modulation. a, The relationship between target protein multivalency and transcriptional output measured by luciferase expression. The number of ABL1 SH2 domains available to bind the HA4 monobody was modulated by varying the 434cI DNA-binding domain multivalency state (1x, 2x, 4x, or 6x SH2). "No operator" indicates a scrambled 434cI operator control AP. **b,** During PACE, the inactive monobody mutant $HA4_{Y87A}$ was subjected to no mutagenesis (MP not induced), enhanced mutagenesis (MP induced with arabinose), or enhanced mutagenesis with genetic drift (MP induced with arabinose in addition to an initial period of zero selection stringency), then selected for binding to the ABL1 SH2 target protein. **c,** The combination of drift and enhanced mutagenesis during PACE (green line) resulted in the evolution of Tyr and Trp residues at position 87, either of which restores SH2-binding activity, while no mutagenesis (red line) or enhanced mutagenesis without drift (blue line) resulted in phage washout. Error bars in (**a**) reflect the standard deviation of at least three independent biological replicates.

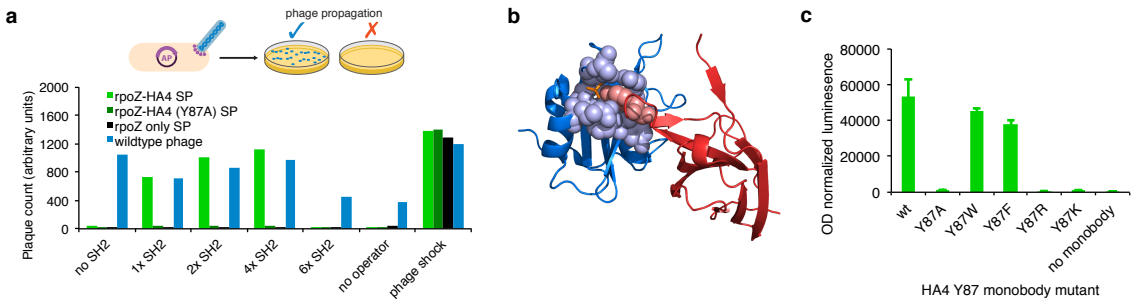


Figure 2.6 | Protein-binding PACE phage propagation and mutant evolution. a, Phage plaque formation as a function of target protein multivalency. "No operator" indicates a scrambled 434cI operator control AP, and "phage control" indicates an AP in which the phage shock promoter (activated by phage infection) drives gene III expression. **b,** Co-crystal structure of the ABL1 SH2 (blue) bound to the HA4 monobody (red), highlighting the interaction of HA4 Y87 (red spheres) with key residues of the phosophotyrosine-binding pocket (blue spheres) of the SH2 domain (PDB 3K2M). The phosphate ion is shown in orange at the interaction interface. **c,** Apparent binding activity of mutants of the HA4 monobody at position 87. Tyrosine, tryptophan and phenylalanine are tolerated at position 87 and enable protein-protein interaction by bacterial 2-hybrid assay. Error bars reflect the standard deviation of at least three independent biological replicates.

2.4 – Discussion

The system presented in this chapter potently reports on protein-protein interactions *in vivo*, and can be used to evolve known antigen-monobody interactions in short timeframes. Importantly, the system enables a substantial degree of control to selection stringency through manipulations to the DNA-binding domain concentration and multimer state, ribosome-binding site strength, operator-promoter spacing, and was shown to enable the general detection of a number of unrelated protein-protein interactions. Further application of this system, however, towards the directed evolution of novel protein-protein interactions from non-functional starting materials were wholly unsuccessful, likely owing to low mutagenesis efficiencies in the system. In the next chapter, I address the issue of low mutagenesis to enable the application of the described

platform for robust protein-protein interaction evolution.

2.5 – Methods

General methods. PCR was performed using PfuTurbo Cx Hotstart DNA polymerase (Agilent Technologies), VeraSeq ULtra DNA polymerase (Enzymatics), or Phusion U Hot Start DNA Polymerase (Life Technologies). Water was purified using a MilliQ water purification system (Millipore). Plasmids and selection phages were constructed using USER cloning (New England Biolabs). Genes were either synthesized as bacterial codon-optimized gBlocks Gene Fragments (Integrated DNA Technologies) or amplified by PCR from native sources. DNA vector amplification was carried out using NEB Turbo or DH5a cells (New England Biolabs).

Electrocompetent strain preparation. The previously described strains S1030[8] and S2060[10] were used in all luciferase and plaque assays, as well as in PACE experiments. The glycerol stock of either strain was used to seed a 2-mL overnight culture using 2xYT media (United States Biological) supplemented with 10 μg/mL tetracycline (Sigma Aldrich), 50 μg/mL streptomycin (Sigma Aldrich), 10 μg/mL fluconazole (TCI America), and 10 μg/mL amphotericin B (TCI America) in a 37 °C shaker at 230 rpm. The saturated culture was diluted 1,000-fold in 50 mL of the same supplemented media and grown under identical conditions until it reached mid log-phase (OD_{600} 0.5-0.8). Once the appropriate OD_{600} was reached, the cells were pelleted in a 50-mL conical tube (VWR) centrifuged at 10,000 *G* for 5 min at 4 °C. The supernatant was immediately decanted and the interior of the tube was wiped with a few Kimwipes (Kimberly-Clark) to remove residual media and salts. The cells were resuspended in 25 mL of pre-chilled, sterile filtered 10% glycerol in MilliQ purified water using a pipette to quickly break up the

pellet. The cells were centrifuged and washed an additional three times. After the last centrifugation step, the interior of the tube was wiped with a few Kimwipes to remove residual glycerol solution. The pellet was resuspended in as little volume as possible, typically ~150 μL, and split into 10 μL aliquots for storage. Cells were flash frozen using a liquid N_2 bath, then quickly transferred to -80°C for extended storage. Electrocompetent S1030 or S2060 cells produced by this method typically yielded 10^7-10^8 colonies/μg plasmid DNA and enable the simultaneous electroporation of up to three plasmids carrying orthogonal origins of replication and antibiotic resistance cassettes to yield transformants containing all plasmids.

General USER cloning. All plasmid and phage materials were constructed via USER cloning[14]. Briefly, primers were designed to include a single internal deoxyuracil base 15-20 bases from the 5' end of the primer, specifying this region as the "USER junction". Criteria for design of the USER junction were: it should contain minimal secondary structure, have 45 °C < T_m < 70 °C, and begin with a deoxyadenosine and end with a deoxythymine (to be replaced by deoxyuridine). The USER junction specifies the homology required for correct assembly. We note that PfuTurbo Cx Hotstart DNA polymerase (Agilent Technologies), VeraSeq ULtra DNA polymerase (Enzymatics), or Phusion U Hot Start DNA Polymerase (Life Technologies) are able to use primers carrying deoxyuracil bases, whereas some other polymerases undergo a phenomenon known as PCR poisoning and do not extend the primer.

All PCR products were purified using MinElute PCR Purification Kit (Qiagen) to 10 μL final volume and quantified using a NanoDrop 1000 Spectrophotometer (Thermo

Scientific). For assembly, PCR products carrying complementary USER junctions were mixed in an equimolar ratio (up to 1 pmol each) in a 10 μL reaction containing 15 units *Dpn*I (New England Biolabs), 0.75 units USER (Uracil-Specific Excision Reagent) enzyme (Endonuclease VIII and Uracil-DNA Glycosylase, NEB), 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 μg/ml BSA at pH 7.9 (1x CutSmart Buffer, New England Biolabs). The reactions were incubated at 37 °C for 45 min, followed by heating to 80 °C and slow cooling to 22 °C at 0.1 °C/s in a temperature-controlled block. The hybridized constructs were directly used for heat-shock transformation of chemically competent NEB Turbo *E. coli* cells according to the manufacturer's instructions. Transformants were selected on 1.8% agar-2xYT plates supplemented with the appropriate antibiotic(s).

For SP cloning, the hybridized constructs were purified using EconoSpin purification columns (Epoch Life Sciences), eluted using 25 μL 10% glycerol, and transformed into electrocompetent S2060 cells carrying the phage-responsive AP pJC175e, which produces functional pIII in response to phage infection (this strain is henceforth referred to as S2208). Following recovery for 3-4 h at 37 °C using 2xYT (United States Biological) media, the culture was centrifuged and the supernatant was purified using a 0.22 μm PVDF Ultrafree centrifugal filter (Millipore). The supernatant was diluted serially in 100-fold increments and used in plaque assays using log-phase S2208 cells. Following overnight at 37 °C, single plaques were picked into 2xYT media and grown for 12-18 h in a 37 °C shaker at 230 rpm. The supernatant was purified again to yield clonal phage stocks. In all cases, cloned plasmids and phages were prepared using the TempliPhi 500 Amplification Kit (GE Life Sciences) according to the

manufacturers protocol and verified by Sanger sequencing.

Plaque assays. S1030[8] or S2060[10] cells transformed with the AP of interest were grown in 2xYT (United States Biological) liquid media supplemented with the appropriate antibiotics to an OD_{600} of 0.6-0.9. The phage supernatant was diluted serially in three, 100-fold increments to yield four total samples (undiluted, 10^2-fold, 10^4-fold, and 10^6-fold diluted) to be used for infections. For each sample, 150 μL of cells were added to 10 μL of phage that had been filtered using a 0.22 μm PVDF Ultrafree centrifugal filter (Millipore). Within 1-2 min of infection, 1 mL of warm (~55°C) top agar (7 g/L bacteriological agar in 2xYT) was added to the phage/cell mixture, mixed by pipetting up and down once, and plated onto quartered plates that had been previously poured with 2 mL of bottom agar (18 g/L bacteriological agar in 2xYT) in each quadrant. The plates were then grown overnight at 37 °C before plaques could be observed.

Phage-Assisted Continuous Evolution (PACE). Host cell cultures, lagoons, media, and the PACE apparatus were as previously described[8]. Recombined selection phage harboring gene III (rSP) will poison a PACE experiment by outcompeting the evolving SP. We have noted that the likelihood of rSP occurrence in an SP stock increases with extended standing culture growth during the initial SP stock preparation. To reduce the likelihood of rSP formation, all SPs are repurified prior to any continuous evolution experiments. Briefly, SPs were plaqued on S2208 cells. A single plaque was picked into 2 mL 2xYT (United States Biological) supplemented with the appropriate antibiotics and grown until the culture reached mid log-phase (OD_{600} 0.5-0.8). The culture was

centrifuged using a tabletop centrifuge for 2 min at 10,000 *G*, followed by supernatant filtration using a 0.22 μm PVDF Ultrafree centrifugal filter (Millipore). This short growth period routinely yields titers of 10^6-10^8 pfu/mL and was found to minimize the occurrence of rSP during PACE experiments.

To prepare the PACE strain, the AP and MP were co-transformed into electrocompetent S1030 cells (see above) and recovered using Davis rich media[8] (DRM) to ensure MP repression. Transformations were plated on 1.8% agar-2xYT containing 50 μg/mL carbenicillin, 40 μg/mL chloramphenicol, 10 μg/mL fluconazole, 10 μg/mL amphotericin B, 100 mM glucose (United States Biological) and grown for 12-18 h in a 37 °C incubator. Following overnight growth, four single colonies were picked and resuspended in DRM, then serially diluted and plated on 1.8% agar-2xYT containing 50 μg/mL carbenicillin, 40 μg/mL chloramphenicol, 10 μg/mL fluconazole, 10 μg/mL amphotericin B, and either 100 mM glucose or 100 mM arabinose (Gold Biotechnology) and grown for 12-18 h in a 37 °C incubator. Concomitant with this plating step, the dilution series was used to inoculate liquid cultures in DRM supplemented with 50 μg/mL carbenicillin, 40 μg/mL chloramphenicol, 10 μg/mL tetracycline, 50 μg/mL streptomycin, 10 μg/mL fluconazole, 10 μg/mL amphotericin B and grown for 12-18 h in a 37 °C shaker at 230 rpm. Following confirmation of arabinose sensitivity using the plate assay, cultures of the serially diluted colonies still in log-phase growth were used to seed a 25-mL starter culture for the PACE chemostat.

Once the starter culture had reached log-phase density (OD_{600} 0.5-0.8), the 25-mL culture was added directly to 175 mL of fresh DRM in the chemostat. The chemostat culture was maintained at 200 mL and grown at a dilution rate of 1.5-1.6 vol/hr as

previously described[8]. Lagoons flowing from the chemostats were maintained at 40 mL, and diluted as described for each experiment. Lagoons were supplemented with 25 mM arabinose to induce the MP for 8-16 h prior to infection with packaged SP. Samples were taken at the indicated time points, centrifuged at 10,000 *G* for 2 min, then filtered with a 0.2 μm filter and stored overnight at 4°C. Phage aliquots were titered by plaque assay on S2208 cells (total phage titer) and S1030 or S2060 cells (rSP titer) for all time points.

Luciferase assays. Complementary plasmids (CPs) were co-transformed with an AP of interest into electrocompetent S1030[8] or S2060[10] cells and plated onto 1.8% agar-2xYT plates with 50 μg/mL carbenicillin and 100 μg/mL spectinomycin. After overnight growth at 37 °C, single colonies were each picked into 2 mL DRM supplemented with 50 μg/mL carbenicillin, 100 μg/mL spectinomycin, 10 μg/mL tetracycline, 50 μg/mL streptomycin, 10 μg/mL fluconazole, 10 μg/mL amphotericin B and grown for 12-18 h in a 37 °C shaker at 230 rpm. Following overnight growth, cultures were diluted 1000-fold in a 96-well deep well plate containing 500 μL DRM with 50 μg/mL carbenicillin, 100 μg/mL spectinomycin and the indicated arabinose, IPTG or anhydrotetracycline (ATc) concentration to induce protein expression from either the AP or CP. Constitutive APs and CPs were used where no inducer concentration is given. After growth with shaking at 37 °C for 4-5 hours, 150 μL of each culture was transferred to a 96-well black wall, clear bottom plate (Costar), and the OD_{600} and luminescence for each well was measured on an Infinite M1000 Pro microplate reader (Tecan). The OD_{600} of a well containing only media was subtracted from all sample wells to obtain a corrected OD_{600} value for each well. The raw luminescence value for each well was then divided by that well's corrected

OD_{600} value to obtain the luminescence value normalized to cell density. Each variant was assayed in at least biological triplicate, and the error bars shown reflect the standard deviations of the independent measurements.

2.6 – References

1 Stynen, B., Tournu, H., Tavernier, J. & Van Dijck, P. Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. *Microbiology and molecular biology reviews : MMBR* **76**, 331-382, (2012).

2 Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I. & Stefan, E. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nature reviews. Drug discovery* **6**, 569-582, (2007).

3 Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. Protein-protein interaction detection: methods and analysis. *International journal of proteomics* **2014**, 147648, (2014).

4 Gregory, B. D., Deighan, P. & Hochschild, A. An artificial activator that contacts a normally occluded surface of the RNA polymerase holoenzyme. *Journal of molecular biology* **353**, 497-506, (2005).

5 Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503, (2011).

6 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. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9007-9012, (2013).

7 Leconte, A. M. *et al.* A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* **52**, 1490-1499, (2013).

8 Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature chemical biology* **10**, 216-222, (2014).

9 Dickinson, B. C., Packer, M. S., Badran, A. H. & Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nature communications* **5**, 5352, (2014).

10 Hubbard, B. P. *et al.* Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat Methods*, (2015).

11 Badran, A. H. & Liu, D. R. Development of Potent In Vivo Mutagenesis Plasmids with Broad Mutational Spectra. *Nat Communications* **6**, 8425, (2015).

12 Dove, S. L. & Hochschild, A. Conversion of the omega subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes & development* **12**, 745-754, (1998).

13 Wojcik, J. *et al.* A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nature structural & molecular biology* **17**, 519-527, (2010).

14 Lund, A. M. *et al.* A versatile system for USER cloning-based assembly of expression vectors for mammalian cell engineering. *PloS one* **9**, e96693, (2014).

Chapter 3

Potent *In Vivo* Mutagenesis Plasmids with Broad Mutational Spectra for Improved Outcome During Phage-Assisted Continuous Evolution

Parts adapted from: Badran & Liu, *Nature Communications* **6**, 8425 (2015).

3.1 – Introduction: unbiased and untargeted *in vivo* mutagenesis

Access to new mutations drives both natural and laboratory evolution. Native biological mutation rates are modest, occurring at frequencies of approximately one mutation per billion replicated DNA bases in most eukaryotes and prokaryotes[1]. Laboratory evolution methods must increase this mutation rate to accelerate the discovery of biomolecules with desired properties on a practical time scale. In addition to mutation rate, mutational spectrum is also a crucial component of the genetic diversity that fuels evolution. Access to diverse amino acid substitutions during evolution is enhanced by more complete coverage of the 12 possible mutation types.

Several systems to enhance mutational efficiency and broaden mutational spectra have been developed for use in laboratory evolution efforts[15]. *In vitro* mutagenesis through error-prone PCR, site-saturation mutagenesis, or DNA shuffling has become the standard approach to introduce diversity into genes of interest. Whereas *in vitro* mutagenesis methods enable control of mutation rate and mutational spectrum, *in vivo* mutagenesis methods enable mutation and selection cycles to be coupled, bypass transformation efficiency bottlenecks that frequently limit the size of gene populations that can be accessed following *in vitro* mutagenesis, and avoid labor-intensive library creation, cloning, and manipulation steps[16]. The difficulty of tuning mutagenic load and spectrum in live cells, however, has prevented the development of safe, effective *in vivo* mutagenesis methods that can rival or exceed the efficiency and mutational spectra of state-of-the-art *in vitro* methods.

The most commonly used *in vivo* mutagenesis methods include chemical mutagens, nucleobase analogs, UV light, and hypermutator strains[17]. Chemical mutagens

yield narrow mutagenic spectra and are potent human carcinogens. UV radiation is known to generate a wide mutation spectrum with little sequence preference, but with low potency that is limited by cellular toxicity. Perhaps the most widespread method for *in vivo* mutagenesis has been the use of hypermutator strains such as XL-1 Red[18] that have been engineered to have higher mutation rates through the deletion or modification of genes involved in DNA replication and repair. These strains suffer from numerous drawbacks, however, including modest mutational potency (*vide infra*), moderately biased mutational preference[19], poor transformation efficiency, slow growth rate, and difficulty of modification when additional genetic manipulations are required. In addition, the rate of mutagenesis is not controllable, and mutagenesis generally must be separated from other steps such as selection or screening due to the instability or poor growth of the strain. Loeb and coworkers previously described an elegant system that uses a low-fidelity *E. coli* DNA Pol I (LF-Pol I) to increase *in vivo* mutagenesis efficiency with a wide scope of mutational types[20]. Unfortunately, this method is restricted to Pol I temperature-sensitive strains and mutates only vectors carrying ColE1-related origins of replication, with mutation rate being highly dependent on distance from the ColE1 origin[2,20].

An ideal *in vivo* mutagenesis method should offer (*i*) a broad mutagenesis spectrum, (*ii*) a mutation rate that can be very high but is easily modulated by the researcher, and (*iii*) entirely episomal encoding so that it can be applied to virtually any *E. coli* strain. Here we describe the development of a general *in vivo* mutagenesis method that meets these criteria. We generate a series of vectors that enable robust chromosomal, episomal, and viral DNA mutagenesis in a strain-independent manner. Importantly, this

system outperforms commonly used *in vivo* mutator and chemical mutagenesis methods in mutational efficiency and spectrum, and can even achieve mutational potency and spectrum comparable to that of state-of-the-art *in vitro* mutagenesis methods. We apply this system to two directed evolution case studies in bacteria and bacteriophage, validating their ability to access biomolecule variants with novel properties with greater effectiveness than previously described methods. Collectively, this system substantially advances *in vivo* mutagenesis capabilities and increases the effectiveness of laboratory evolution efforts.

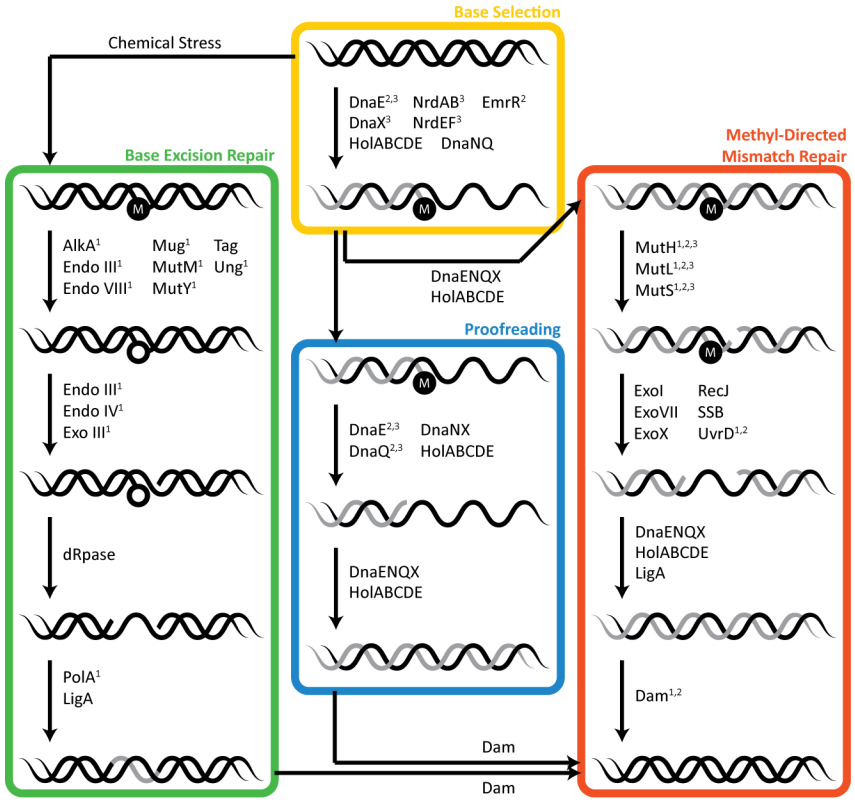


Figure 3.1 | Summary of major pathways that influence *E. coli* DNA replication fidelity. Steps during replication and mutation correction are grouped according to their mechanisms of action. Methylated DNA is shown in black, unmethylated DNA is shown in grey, and the mutation to be corrected is depicted as "M". Gene superscripts denote if a mutator phenotype results upon gene deletion[4] (1), gene overexpression[5-9] (2) or modification of the chromosomal allele to circumvent potential knockout lethality[5-10] (3).

Strain	Genotype
MG1655 Δ*recA*	*F⁻ ΔrecA1918::apra, rph-1 λ⁻*
CSH101	*F' lacI373 lacZ571 / ara-600 Δ(gpt-lac)5 relA1 spoT1 thiE1 λ⁻*
CSH102	*F' lacI373 lacZ572 / ara-600 Δ(gpt-lac)5 relA1 spoT1 thiE1 λ⁻*
CSH103	*F' lacI373 lacZ573 / ara-600 Δ(gpt-lac)5 relA1 spoT1 thiE1 λ⁻*
CSH104	*F' lacI373 lacZ574 / ara-600 Δ(gpt-lac)5 relA1 spoT1 thiE1 λ⁻*
CSH105	*F' lacI373 lacZ575 / ara-600 Δ(gpt-lac)5 relA1 spoT1 thiE1 λ⁻*
CSH106	*F' lacI373 lacZ576 / ara-600 Δ(gpt-lac)5 relA1 spoT1 thiE1 λ⁻*
S1030	*F' proA+B+ Δ(lacIZY) zzf::Tn10 lacI^{Q1} P$_{N25}$-tetR luxCDE / endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC λ⁻*
S1021	*F⁻ endA1 recA1 galE15 galK16 nupG rpsL ΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 araE201 ΔrpoZ Δflu ΔcsgABCDEFG ΔpgaC λ⁻*

Table 3.1 | Summary of all strains used in this chapter. MG1655 and CSH strains were obtained from the Yale University *E. coli* Genetic Stock Center.

3.2 – Mutagenesis plasmid minimization

Bacteria control the rate of chromosomal substitutions through a series of overlapping mechanisms that can be subdivided into three main pathways: proofreading (reduces mutation rate by a factor of ~10^2 substitutions bp^{-1} generation^{-1}), mismatch repair (reduces mutation rate by ~10^3 substitutions bp^{-1} generation^{-1}), and base selection (reduces mutation rate by ~10^5 substitutions bp^{-1} generation^{-1})[21] (**Figure 3.1**). These redundant replication maintenance mechanisms collectively account for the basal substitution rate of bacterial chromosomal DNA of ~10^{-9} to 10^{-10} substitutions bp^{-1} generation^{-1} [21]. Based on prior knowledge of dominant mutators alleles that interfere with DNA replication fidelity, we sought to design a series of small-molecule inducible mutagenesis plasmids (MPs) that offer broad mutational spectra and high levels of mutagenesis in bacterial cells.

We recently reported the development and application of an MP for *in vivo* mutagenesis during phage-assisted continuous evolution (PACE)[22]. This MP increases

the mutation rate ~100-fold above the basal *E. coli* mutation rate through the arabinose-induced expression of *dnaQ926*, a dominant negative variant of the *E. coli* DNA Pol III proofreading domain. This plasmid additionally provides *umuD'*, *umuC* and *recA730*, which together enable *in vivo* translesion mutagenesis employing UV light or chemical mutagens. This MP (designated MP1) results in a substitution rate of 7.2×10^{-5} and 5.4×10^{-8} substitutions bp^{-1} $generation^{-1}$ for M13 phage and *E. coli*, respectively[22,23]. This mutation rate, however, is still several orders of magnitude below the mutation rates provided by conventional *in vitro* mutagenesis techniques[24].

Since we sought to avoid the use of exogenous mutagens, we first minimized MP1 by removing *umuD'*, *umuC*, and *recA730* from MP1 to yield MP2 carrying only

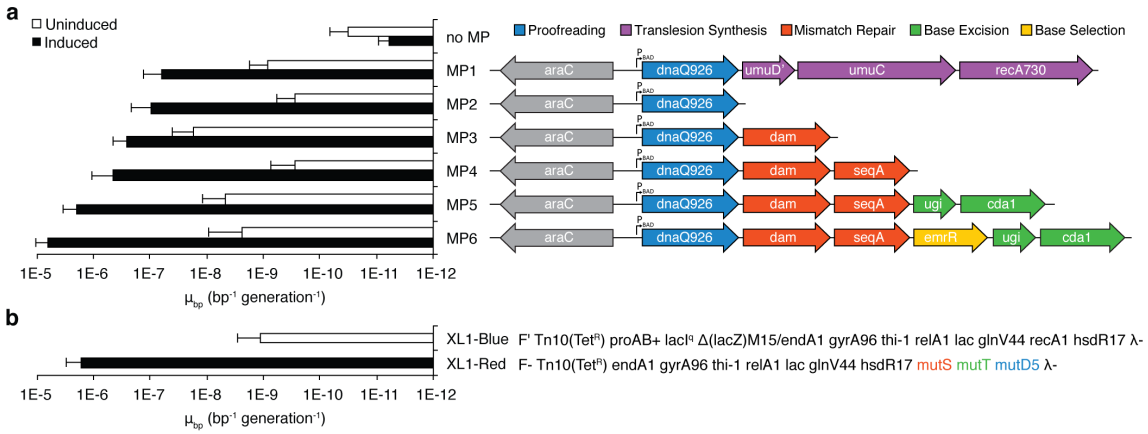


Figure 3.2 | Mutagenesis plasmid (MP) design and effect on mutation rate in bacteria. The mutator genes tested are color-coded to indicate the canonical pathway being disrupted through the overexpression (A) or deletion (B) of that gene. All MPs express mutator genes from the *E. coli* arabinose-inducible P_{BAD} promoter, with each gene preceded by a ribosome-binding site to enable translation from a single transcript. For (A), the mutagenesis rate μ_{bp} (substitutions per base pair of the *E. coli* genome per generation) was calculated using rifampin resistance under uninduced (25 mM glucose, white bars) and induced (25 mM glucose + 25 mM arabinose, black bars). For (B), the rifampin resistance of XL1-Blue (white bar) and XL1-Red (black bar) was used to calculate μ_{bp}. In all cases, all known 77 rifampin-resistant rpoB alleles were used to calculate the mutation rate. Error bars denote the standard deviation of at least three biological replicates.

dnaQ926, and observed mutation rates in the absence of mutagens to be modestly improved compared to MP1 through a rifampin-resistance assay using the nearly wild-type *E. coli* MG1655 Δ*recA::apra* (**Table 3.1**), enabling an average of 9.9×10^{-8} substitutions bp^{-1} $generation^{-1}$ (**Figure 3.2**). Since *dnaQ926* abrogates the proofreading component of DNA replication, we began assessing additional genes that when expressed from the MP can further enhance mutation rate.

3.3 – DNA methylation state manipulation enhances mutagenesis

Mismatch repair reduces the error rate of bacterial DNA replication by a factor of up to ~10^3 [21]. Specialized proteins monitor (MutSL) and excise (MutH) mismatched nucleotides following DNA replication, using DNA methylation (Dam) to indicate the parental strand and thus the basis for correction. Not surprisingly, the deletion of *mutS*, *mutL*, or *mutH*, and the overexpression of *dam* are known to have a strong mutator effect due to impaired mismatch repair[4].

We added wild-type *dam* downstream of *dnaQ926* on the same arabinose-inducible expression cassette to yield MP3, and observed increased mutagenesis potency but also increased uninduced background mutagenesis, likely due to a strong cryptic σ^{70}

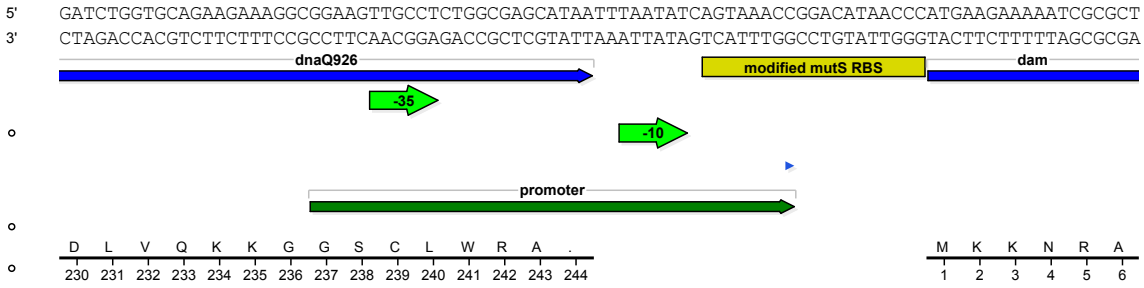


Figure 3.3 | Cryptic σ^{70} promoter at the 3' end of the *dnaQ926* ORF. Annotated sequence of the predicted σ^{70} promoter in MP3 bridging the 3' end of the *dnaQ926* ORF and RBS driving the *dam* ORF.

promoter at the 3' end of *dnaQ926* (**Figure 3.3**). Overall, MP3 resulted in a 3-fold mutagenesis potency increase in the presence of arabinose relative to MP2 but > 60-fold increase in background mutagenesis, greatly reducing the dynamic range of induced mutagenesis to only 11-fold (**Figure 3.2**).

To restore dynamic range, we installed the gene encoding the hemimethylated GATC-binding domain SeqA downstream of *dam*. Low-level expression of *seqA* is known to delay Dam methylation of hemimethylated GATC sequences[25] and induce positive supercoils in chromosomal and episomal DNA[26], which may reduce MP gene transcription in the absence of arabinose. Conversely, high-level *seqA* expression results in a potent mutator response[8] and induction of negative supercoils in chromosomal and episomal DNA[26], potentially increasing global gene transcription, including that of the MP, upon induction with arabinose. Indeed, the MP carrying *dnaQ926*, *dam*, and *seqA* (MP4) resulted in > 60-fold reduced background and 2-fold improved mutagenesis potency in the presence of arabinose, representing a cumulative > 100-fold improvement in dynamic range, relative to MP3 (**Figure 3.2**). In agreement with previous reports showing that dominant-negative *dnaQ* mutants partially saturate the mismatch repair response[27], dominant-negative variants of *mutS*[8,9], *mutL*[5,6], or *mutH*[6] with *dnaQ926* had no additional effect on mutagenesis (**Tables 3.2 and 3.3**). Cumulatively, expression of *dam* and *seqA* from MP4 is sufficient to disrupt the mismatch repair response, enabling an average of 4.4×10^{-7} substitutions bp^{-1} $generation^{-1}$.

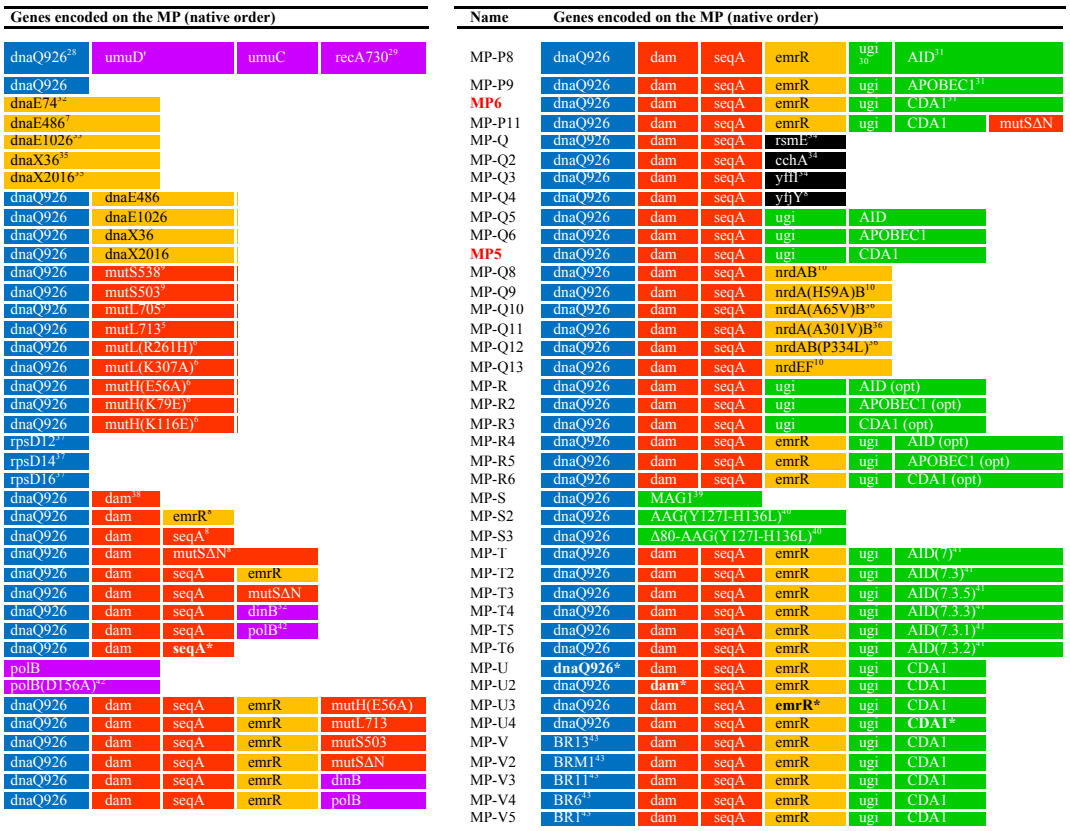


Table 3.2 | Summary of all ORFs carried by the MPs. All MPs use the identical vector backbone: a cloDF13 origin of replication (20-40 copies/cell), a chloramphenicol resistance cassette, the arabinose responsive promoter P_{BAD} driving the mutator genes, and the weak promoter P_C driving *araC*. Genes carried by each MP are arranged in the order found in the table and are highlighted according to their mechanism of action and/or the canonical repair pathway that they disrupt: proofreading (blue), translesion synthesis (purple), methyl-directed mismatch repair (red), base excision repair (green), base selection (yellow) and unknown (black). Additional optimizations included: codon usage optimization (opt) and increased ribosome-binding site strength (*). Boxes are not drawn to scale.

Name	Uninduced μ_{bp}	Induced μ_{bp}	Viability (%)	Name	Uninduced μ_{bp}	Induced μ_{bp}	Viability (%)
None	3.26E-11	6.07E-12	100	MP-P8	5.49E-10	1.31E-07	32.8
MP1	8.20E-10	6.42E-08	87.5	MP-P9	3.28E-08	7.04E-06	2.9
MP2	2.77E-10	9.85E-08	136.8	**MP6**	2.44E-09	6.24E-06	1.7
MP-B2	0.00E+00	0.00E+00	25.4	MP-P11	1.66E-09	5.30E-06	1.4
MP-B4	0.00E+00	0.00E+00	45.8	MP-Q	4.49E-10	1.35E-07	14.7
MP-B5	0.00E+00	0.00E+00	55.9	MP-Q2	1.55E-09	9.24E-08	20.6
MP-C2	0.00E+00	0.00E+00	63.6	MP-Q3	2.37E-10	9.36E-08	16.8
MP-C3	0.00E+00	0.00E+00	53.4	MP-Q4	1.94E-09	2.64E-07	15
MP-D3	0.00E+00	2.31E-09	24.2	MP-Q5	1.92E-09	1.09E-07	22.5
MP-D4	0.00E+00	2.59E-09	68.6	MP-Q6	9.86E-10	1.83E-07	16.3
MP-E	0.00E+00	2.13E-10	61	**MP5**	4.79E-09	2.01E-06	3.8
MP-E2	3.98E-10	6.95E-10	40.7	MP-Q8	1.70E-10	4.82E-08	37.4
MP-F2	9.83E-10	2.68E-09	48.3	MP-Q9	2.50E-10	1.88E-08	73.7
MP-F3	0.00E+00	4.59E-09	73.7	MP-Q10	2.26E-09	1.04E-08	81.4
MP-H	0.00E+00	7.07E-09	94.1	MP-Q11	1.16E-10	1.15E-08	66.1
MP-H2	7.15E-10	9.61E-09	35.6	MP-Q12	2.40E-10	5.48E-09	144.9
MP-H3	1.27E-10	2.75E-09	71.2	MP-Q13	3.98E-10	4.24E-09	101.7
MP-H4	0.00E+00	2.03E-09	63.6	MP-R	9.47E-10	1.64E-07	13.9
MP-I	0.00E+00	4.26E-09	61	MP-R2	7.52E-09	2.37E-07	13.1
MP-I2	0.00E+00	1.75E-09	83.9	MP-R3	1.12E-08	1.17E-06	4.3
MP-I3	0.00E+00	2.72E-09	89	MP-R4	2.34E-09	6.94E-07	3.7
MP-J	0.00E+00	0.00E+00	223.7	MP-R5	4.90E-08	4.35E-06	5.8
MP-J2	0.00E+00	0.00E+00	142.4	MP-R6	5.80E-09	1.71E-05	0.7
MP-J3	0.00E+00	0.00E+00	137.3	MP-S	3.23E-11	6.29E-09	115.3
MP3	1.74E-08	2.66E-07	31.2	MP-S2	1.73E-11	5.76E-09	313
MP-K7	6.53E-10	8.99E-08	9.4	MP-S3	6.85E-11	2.68E-08	428.4
MP4	2.73E-10	4.38E-07	22.9	MP-T	1.92E-10	2.77E-07	13.6
MP-K9	1.55E-09	8.99E-08	23.4	MP-T2	1.93E-09	9.59E-08	23.1
MP-K10	1.12E-11	3.07E-08	80.6	MP-T3	8.70E-09	1.55E-07	12.6
MP-K11	9.87E-11	1.85E-08	63.6	MP-T4	6.09E-09	2.42E-07	13.7
MP-K12	2.68E-10	1.48E-08	86.4	MP-T5	6.28E-09	1.65E-07	21.7
MP-K13	0.00E+00	1.83E-08	63.6	MP-T6	2.50E-09	2.29E-07	16
MP-K14	1.12E-09	0.00E+00	0.8	MP-U	1.90E-09	2.66E-05	0.6
MP-L	0.00E+00	0.00E+00	129.7	MP-U2	8.02E-10	2.88E-06	1.7
MP-L2	0.00E+00	0.00E+00	93.8	MP-U3	6.66E-09	1.49E-06	2.8
MP-P	5.32E-10	3.84E-09	144.9	MP-U4	1.36E-09	2.51E-06	3.7
MP-P3	1.14E-09	2.51E-07	33.3	MP-V	4.59E-10	1.49E-07	5.7
MP-P4	4.58E-09	1.57E-07	81.4	MP-V2	7.79E-10	2.88E-07	7.9
MP-P5	2.83E-10	6.06E-07	26.7	MP-V3	2.95E-09	6.44E-07	3.4
MP-P6	4.90E-10	1.13E-07	10.7	MP-V4	1.35E-09	5.98E-07	4.6
MP-P7	3.54E-10	7.58E-08	83.9	MP-V5	2.38E-10	5.09E-07	6.2

Table 3.3 | Summary of induced and uninduced mutagenesis levels for all designed MPs. All MPs were tested using the rifampin resistance assay to assess their relative mutagenic load under uninduced (glucose) and induced (arabinose) conditions. The viability of the MP-carrying strains under the induced conditions (as a percentage of the viability of the strain without an MP) is also shown. Ideal MPs show low background and high induced mutagenesis, with only moderate reductions in viability.

3.4 – Cytosine deamination and reduced base excision repair

Overexpression of the catalytic domains of several cytidine deaminases in *E. coli* has been shown to have a mutagenic effect, resulting in primarily C→T transitions through a deoxyuracil intermediate[28]. The cytidine deaminase CDA1 from *Petromyzon marinus* is reported to promote the mutation of prokaryotic and eukaryotic genomic DNA[28]. Impairing uracil-DNA glycosylase (Ung) synergizes with the effect of deaminases and can enhance mutagenesis through disruption of the native uracil-excision repair pathway[28] (**Figure 3.1**). Two natural protein inhibitors of Ung, Ugi and p56[29], inhibit Ung through mimicry of structural and electronic features of uracil-containing DNA[29]. As the p56:Ung interaction has been shown to be less stable than the Ugi:Ung interaction[29], we proceeded with Ugi.

We placed *ugi* and *cda1* downstream of *dnaQ926*, *dam* and *seqA* to yield MP5, and observed a 5-fold mutagenesis potency increase under induced conditions, representing an 18-fold increase in background mutagenesis compared to MP4 (**Figure**

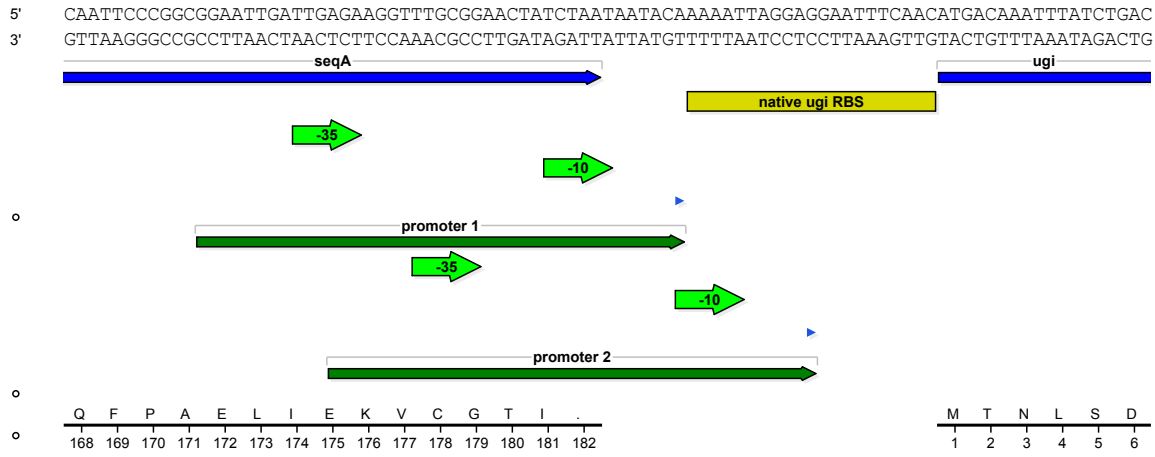


Figure 3.4 | Cryptic σ^{70} promoters at the 3' end of the *seqA* ORF. Annotated sequence of the predicted σ^{70} promoters in MP5 bridging the 3' end of the *seqA* ORF and native RBS driving the *ugi* ORF.

3.2). This background increase was caused by two predicted σ^{70} promoters at the 3' end of the *seqA* open reading frame, resulting in constitutive *ugi* and *cda1* transcription (**Figure 3.4**). We considered this background mutation rate for MP5 to be acceptable, as it was only ~5-fold higher than the starting MP1. Alternative cytosine deaminases, including rat APOBEC1 and human AID, generally resulted in weaker effects on mutation rate than CDA1, in agreement with previous reports (CDA1 >> AID ≈ APOBEC1)[28] (**Tables 3.2 and 3.3**). Overall, MP5 yielded 2.0×10^{-6} substitutions bp^{-1} $generation^{-1}$, a 31-fold increase in mutation rate relative to MP1 (**Tables 3.2 and 3.3**).

3.5 – Impairing mutagenic nucleobase export

Two major determinants of base selection during DNA replication are the catalytic alpha subunit of *E. coli* DNA Pol III, and the intracellular concentration of dNTPs (**Figure 3.1**). Mutations affecting the former are generally not viable or exert a mutator effect through reduced affinity to the proofreading domain, DnaQ[7], whereas perturbations affecting the latter are generally more tolerated and can be modified to affect the mutational spectrum[10]. We screened a number of proteins that are known to compromise intracellular dNTP pools, and found expression of the *emrR* transcriptional repressor to be the most promising (**Figure 3.2**; **Tables 3.2 and 3.3**).

Derepression by EmrR results in upregulation of *emrAB*, which produces a multidrug efflux pump responsible for antibiotic resistance and the putative export of mutagenic nucleobase intermediates[30]. In an unbiased screen, *emrR* overexpression was found to have a potent mutator effect[8], presumably as a consequence of retaining these mutagenic nucleobases within the cell. To decrease background mutagenesis compared to

MP5, we placed the *emrR* cassette between *seqA* and *ugi*, thereby disrupting the strong predicted σ^{70} promoter, yielding MP6. MP6 exhibited 2-fold lower background mutagenesis, while improving the overall mutator effect under induced conditions by 3-fold (**Figure 3.1**).

3.6 – Features of the MP6 mutagenesis system

We chose MP6 for in-depth characterization because it offered the highest mutagenic potency with acceptable levels of toxicity (< 100-fold reduction in cell viability under maximal MP induction compared to the uninduced control). We retained native bacterial ribosome-binding sites (RBSs) upstream of the ORFs for four of the six MP6 genes. The exceptions are *dam*, which natively lacks a canonical RBS in the *E. coli* genome, and *cda1*, which derives from the eukaryote *P. marinus* and thus does not use a bacterial Shine-Dalgarno sequence. In an attempt to further enhance mutational potency by modulating the expression of *dnaQ926*, *dam*, *seqA*, *emrR*, *ugi*, and *cda1*, we varied RBS upstream of each of these six genes by individually mutating them to fully complement the 16S rRNA, and resulting in optimal transcript translation. Interestingly, strengthening the RBSs upstream of each of the six genes generally reduced the potency of the MP, with the exception of the *seqA* and *dnaQ926* RBSs (**Tables 3.2 and 3.3**). Strengthening the *seqA* RBS proved highly lethal under induced conditions, likely as a consequence of impeded genomic replication (**Tables 3.2 and 3.3**). Increasing the strength of the *dnaQ926* RBS enhanced the mutagenic potency of the MP by 4-fold under induced conditions, concomitant with a minor increase in background mutagenesis, but was more toxic to bacteria as evidenced by a greater loss of viability under induced

conditions (**Tables 3.2 and 3.3**). Additionally, measurement of the mutation rate of the resulting MP became irreproducible, consistent with MP instability under these conditions. These findings together suggest that additional mutagenic potency gains beyond that of MP6 may result in error catastrophe and reduced MP stability.

When induced, MP6 results in a 322,000-fold increase in mutation rate of chromosomal DNA over that of wild-type *E. coli*, and a 100-fold increase in mutation rate relative to that of MP1. Induced MP6 results in an average of 6.2×10^{-6} substitutions bp^{-1} $generation^{-1}$, representing to our knowledge the most potent inducible and genetically encodable mutagenesis method in bacteria reported to date (**Table 3.4**). This potency compares favorably to overexpression of *PolB (D156A)* ($\mu_{bp} = 3.2 \times 10^{-6}$), *dnaQ926* ($\mu_{bp} = 2.4 \times 10^{-6}$), or *mutD5* ($\mu_{bp} = 3.9 \times 10^{-7}$) (**Table 3.4**). We note that most previously published mutagenic constructs are not inducible, negatively affect host viability, and require overexpression of the mutagenic protein. In contrast, the MPs described here rely on low-level expression of multiple genes, thereby affecting multiple pathways and enabling broad mutagenic spectra.

Additional permutations of this design or the inclusion of alternative mutators impaired overall mutation rate, or strongly decreased host viability, as evidenced by the characteristics of 80 candidate MPs with mutation rates spanning five orders of magnitude (**Figures 3.5 – 3.7**; **Tables 3.2 and 3.3**). We observed that a loss of bacterial viability proportional to mutagenic potency beyond a mutation rate of $\sim 4 \times 10^{-7}$ substitutions bp^{-1} $generation^{-1}$, corresponding to an average of ~1.9 substitutions $genome^{-1}$ $generation^{-1}$. Given that ~10-30% of the *E. coli* genome has been estimated to be essential[31], this rate corresponds to ~0.2-0.6 mutations in an essential gene/generation.

Source organism	Gene(s)	Fraction Rif^R		μ_{bp} (bp^{-1} generation^{-1})			vs MP6 (%)	Dynamic range*	Ref.
		- MP	+ MP	- MP	+ MP	Fold			
Ec	dam	9.00E-09	3.00E-07	6.16E-11	2.05E-09	0	0	-	[38]
Ec	dnaE173	5.00E-09	7.20E-06	3.42E-11	4.93E-08	1440	1	24	[7]
Ec	dnaQ926	3.00E-08	3.47E-04	2.05E-10	2.38E-06	11567	38	-	[28]
Ec	mutD5	3.00E-08	5.70E-05	2.05E-10	3.90E-07	1900	6	-	[28]
Sc	mag1	5.13E-08	3.61E-06	3.51E-10	2.47E-08	70	0	-	[39]
Sc	mag1	1.00E-08	2.00E-06	6.85E-11	1.37E-08	200	0	200	[47]
Ec	dinB	4.00E-08	4.55E-06	2.74E-10	3.11E-08	114	0	114	[48]
Ec	dnaE (K655Y)	1.00E-08	8.45E-06	6.85E-11	5.78E-08	845	1	-	[32]
Hs	AID	1.30E-08	1.03E-07	8.90E-11	7.05E-10	8	0	8	[49]
Rn	APOBEC1	2.53E-08	1.23E-05	1.73E-10	8.42E-08	486	1	14	[50]
Rn	APOBEC2	2.53E-08	2.50E-08	1.73E-10	1.71E-10	1	0	-	[50]
Hs	AID	2.53E-08	1.66E-07	1.73E-10	1.14E-09	7	0	-	[50]
Hs	APOBEC3C	2.53E-08	2.93E-07	1.73E-10	2.01E-09	12	0	-	[50]
Hs	APOBEC3G	2.53E-08	2.70E-07	1.73E-10	1.85E-09	11	0	-	[50]
Ec	mutH (E56A)	4.40E-08	8.74E-06	3.01E-10	5.98E-08	199	1	-	[6]
Ec	mutH (K116E)	4.40E-08	8.40E-06	3.01E-10	5.75E-08	191	1	-	[6]
Ec	mutH (K79E)	4.40E-08	7.00E-06	3.01E-10	4.79E-08	159	1	-	[6]
Ec	mutH (E77A)	4.40E-08	6.94E-06	3.01E-10	4.75E-08	158	1	-	[6]
Ec	mutH (D70A)	4.40E-08	6.88E-06	3.01E-10	4.71E-08	156	1	-	[6]
Ec	mutH CΔ5	4.40E-08	9.10E-07	3.01E-10	6.23E-09	21	0	-	[6]
Ec	mutL (R95F/N302A)	4.40E-08	1.17E-05	3.01E-10	8.03E-08	267	1	-	[6]
Ec	mutL (R261H)	4.40E-08	1.04E-05	3.01E-10	7.13E-08	237	1	-	[6]
Ec	mutL (E29A)	4.40E-08	8.32E-06	3.01E-10	5.70E-08	189	1	-	[6]
Ec	mutL (K307A)	4.40E-08	6.75E-06	3.01E-10	4.62E-08	153	1	-	[6]
Ec	mutL (N302A)	4.40E-08	3.50E-06	3.01E-10	2.40E-08	80	0	-	[6]
Ec	mutL (R95F)	4.40E-08	1.27E-07	3.01E-10	8.69E-10	3	0	-	[6]
Ec	mutL (K159E)	4.40E-08	7.26E-06	3.01E-10	4.97E-08	165	1	-	[6]
Ec	mutL (R266E)	4.40E-08	6.05E-06	3.01E-10	4.14E-08	138	1	-	[6]
Ec	mutL (R177E)	4.40E-08	1.77E-06	3.01E-10	1.21E-08	40	0	-	[6]
Ec	mutL (I90R)	4.40E-08	1.56E-07	3.01E-10	1.07E-09	4	0	-	[6]
Ec	mutL (R237E)	4.40E-08	1.40E-07	3.01E-10	9.58E-10	3	0	-	[6]
Ec	mutL (G238A)	4.40E-08	5.40E-08	3.01E-10	3.70E-10	1	0	-	[6]
Ec	mutL (G238D)	4.40E-08	1.03E-05	3.01E-10	7.05E-08	234	1	-	[6]
Ec	mutL (I90E)	4.40E-08	1.70E-08	3.01E-10	1.16E-10	0	0	-	[6]
Ec	mutS (S668A/T669V)	4.40E-08	1.33E-05	3.01E-10	9.12E-08	303	1	-	[6]
Ec	mutS (K620M)	4.40E-08	1.21E-05	3.01E-10	8.31E-08	276	1	-	[6]
Ec	mutS (D693A)	4.40E-08	1.15E-05	3.01E-10	7.90E-08	262	1	-	[6]
Ec	mutS (E694Q)	4.40E-08	1.13E-05	3.01E-10	7.75E-08	257	1	-	[6]
Ec	mutS (E694A)	4.40E-08	8.84E-06	3.01E-10	6.05E-08	201	1	-	[6]
Ec	mutS (E694A/S668A/T669V)	4.40E-08	8.02E-06	3.01E-10	5.49E-08	182	1	-	[6]
Ec	mutS (D693N)	4.40E-08	5.58E-06	3.01E-10	3.82E-08	127	1	-	[6]
Ec	mutS (H760A)	4.40E-08	8.01E-07	3.01E-10	5.48E-09	18	0	-	[6]
Ec	mutS (H728A)	4.40E-08	4.48E-07	3.01E-10	3.07E-09	10	0	-	[6]
Ec	mutS (F596A)	4.40E-08	3.14E-07	3.01E-10	2.15E-09	7	0	-	[6]
Ec	mutS (S612A)	4.40E-08	3.70E-08	3.01E-10	2.53E-10	1	0	-	[6]
Ec	mutS (F36A)	4.40E-08	1.06E-05	3.01E-10	7.26E-08	241	1	-	[6]
Ec	mutS (E38Q)	4.40E-08	4.19E-06	3.01E-10	2.87E-08	95	0	-	[6]
Ec	mutS (D162R/E164R)	4.40E-08	4.30E-06	3.01E-10	2.94E-08	98	0	-	[6]
Ec	mutS (R163E)	4.40E-08	3.90E-08	3.01E-10	2.67E-10	1	0	-	[6]
Ec	mutS (R197E/R198E)	4.40E-08	1.11E-05	3.01E-10	7.61E-08	253	1	-	[6]
Ec	mutS (R197E/R198E/R199E)	4.40E-08	6.15E-06	3.01E-10	4.21E-08	140	1	-	[6]
Ec	mutS (E177A)	4.40E-08	1.10E-05	3.01E-10	7.50E-08	249	1	-	[6]
Ec	mutS (T115A)	4.40E-08	7.99E-06	3.01E-10	5.47E-08	182	1	-	[6]
Ec	dam	3.00E-08	2.00E-06	2.05E-10	1.37E-08	67	0	-	[8]
Ec	emrR	3.00E-08	5.00E-06	2.05E-10	3.42E-08	167	1	-	[8]
Ec	mutS*	3.00E-08	3.00E-07	2.05E-10	2.05E-09	10	0	-	[8]
Ec	seqA	3.00E-08	8.00E-07	2.05E-10	5.48E-09	27	0	-	[8]
Ec	dinB	3.00E-08	3.00E-07	2.05E-10	2.05E-09	10	0	-	[8]
Pa	nfxB	8.00E-08	3.50E-06	5.48E-10	2.40E-08	44	0	-	[45]
Ll	dnaN	8.00E-08	9.80E-08	5.48E-10	6.71E-10	1	0	-	[45]
Ll	dnaA	8.00E-08	1.90E-07	5.48E-10	1.30E-09	2	0	-	[45]
Ll	uvrA, ysjE	8.00E-08	3.90E-08	5.48E-10	2.67E-10	0	0	-	[45]
Ll	uvrA	8.00E-08	1.10E-07	5.48E-10	7.53E-10	1	0	-	[45]
Ll	rnhA, sipL, purR	8.00E-08	4.80E-08	5.48E-10	3.29E-10	1	0	-	[45]
Ll	rnhA, sipL	8.00E-08	6.70E-08	5.48E-10	4.59E-10	1	0	-	[45]
Ll	rnhA	8.00E-08	1.40E-08	5.48E-10	9.58E-11	0	0	-	[45]

Table 3.4 | Comparison of MP1-MP6 with previously described mutator plasmids.

Ec	**polB**	**2.20E-08**	**2.32E-06**	**1.51E-10**	**1.59E-08**	**105**	**0**	**116**	[42]
Ec	polB Q779V	2.20E-08	6.14E-08	1.51E-10	4.20E-10	3	0	-	[42]
Ec	polB Δ780-783	2.20E-08	5.93E-08	1.51E-10	4.06E-10	3	0	-	[42]
Ec	polB (D156A)	2.20E-08	4.73E-04	1.51E-10	3.23E-06	21478	52	-	[42]
Ec	uvrAB	7.90E-09	5.40E-08	5.41E-11	3.70E-10	7	0	-	[51]
Ec	uvrABC	7.90E-09	1.90E-07	5.41E-11	1.30E-09	24	0	-	[51]
Hs	AID	1.00E-08	1.40E-07	6.85E-11	9.58E-10	14	0	-	[41]
Hs	AID (K10E/E156G)	1.00E-08	6.30E-07	6.85E-11	4.31E-09	63	0	-	[41]
Hs	AID (K34E/K160E)	1.00E-08	4.20E-07	6.85E-11	2.88E-09	42	0	-	[41]
Hs	AID	2.40E-08	4.90E-07	1.64E-10	3.35E-09	20	0	-	[52]
Hs	AID-3FL	2.40E-08	3.70E-07	1.64E-10	2.53E-09	15	0	-	[52]
Hs	AID-3GL	2.40E-08	1.60E-07	1.64E-10	1.10E-09	7	0	-	[52]
Hs	AAG	1.00E-09	2.00E-07	6.85E-12	1.37E-09	200	0	1	[40]
Hs	AAG(Y127I/H136L)	1.00E-09	1.15E-05	6.85E-12	7.87E-08	11500	1	4	[40]
Ec	nrdAB	1.00E-08	8.00E-08	6.85E-11	5.48E-10	8	0	-	[10]
Ec	nrdEF	1.00E-08	3.50E-07	6.85E-11	2.40E-09	35	0	-	[10]
Ec	nrdA(H59A)B	1.00E-08	3.80E-07	6.85E-11	2.60E-09	38	0	-	[10]
Hs	APOBEC3G	2.60E-07	2.40E-06	1.78E-09	1.64E-08	9	0	-	[53]
Hs	APOBEC3G (E259Q)	2.60E-07	3.60E-07	1.78E-09	2.46E-09	1	0	-	[53]
Hs	APOBEC3G (E254R)	2.60E-07	3.50E-06	1.78E-09	2.40E-08	13	0	-	[53]
Hs	APOBEC3G (R313E)	2.60E-07	2.60E-07	1.78E-09	1.78E-09	1	0	-	[53]
Hs	APOBEC3G (R320E)	2.60E-07	5.80E-07	1.78E-09	3.97E-09	2	0	-	[53]
Hs	APOBEC3G (R313E/R320E)	2.60E-07	3.10E-07	1.78E-09	2.12E-09	1	0	-	[53]
Hs	APOBEC3G (R374E)	2.60E-07	5.50E-07	1.78E-09	3.77E-09	2	0	-	[53]
Hs	APOBEC3G (R376E)	2.60E-07	1.90E-06	1.78E-09	1.30E-08	7	0	-	[53]
Hs	APOBEC3G (R374E/R376E)	2.60E-07	6.90E-07	1.78E-09	4.72E-09	3	0	-	[53]
Hs	APOBEC3G (R213E)	2.60E-07	4.40E-07	1.78E-09	3.01E-09	2	0	-	[53]
Hs	APOBEC3G (R215E)	2.60E-07	3.20E-07	1.78E-09	2.19E-09	1	0	-	[53]
Hs	APOBEC3G (R213E/R215E)	2.60E-07	2.90E-07	1.78E-09	1.99E-09	1	0	-	[53]
Ec	dnaQ926, umuD', umuC, recA730	2.83E-09	9.38E-06	1.94E-11	6.42E-08	3320	1	78	[23]
Ec	dnaQ926	2.83E-09	1.44E-05	1.94E-11	9.85E-08	5087	2	355	This work
Ec	dnaQ926, dam	2.83E-09	3.88E-05	1.94E-11	2.66E-07	13723	4	15	This work
Ec	dnaQ926, dam, seqA	2.83E-09	6.39E-05	1.94E-11	4.38E-07	22610	7	1604	This work
Ec, PBS2, Pm	dnaQ926, dam, seqA, ugi, cda1	2.83E-09	2.93E-04	1.94E-11	2.01E-06	103749	32	419	This work
Ec, PBS2, Pm	dnaQ926, dam, seqA, emrR, ugi, cda1	2.83E-09	9.11E-04	1.94E-11	6.24E-06	322414	100	34941	This work

Table 3.4 (Continued) | Comparison of MP1-MP6 with previously described mutator plasmids. In each case, the mutator genes are listed with the source organism(s): *Ec = Escherichia coli; Sc = Saccharomyces cerevisiae; Hs = Homo sapiens; Rn = Rattus norvegicus; Pa = Pseudomonas aeruginosa; Ll = Lactococcus lactis; PBS2 = Bacillus subtilis phage PBS2; Pm = Petromyzon marinus*. In all cases, the fraction of cells showing rifampin resistance (Rif^R) was used to calculate μ_{bp} as described in the methods section, using R=77 sites, N=10^8, and N_0 = 1.5x10^7 to approximate the levels as compared to the MP1-6 series. The fold increase in mutagenesis is shown for each MP (defined as the ratio of the mutagenesis in the strain without the MP vs. in the strain with the MP). All MPs are compared to MP6 in total mutagenesis efficiency. *In cases where the MP was inducible, the dynamic range represents the fold increase between the uninduced and induced states for strains carrying the MP.

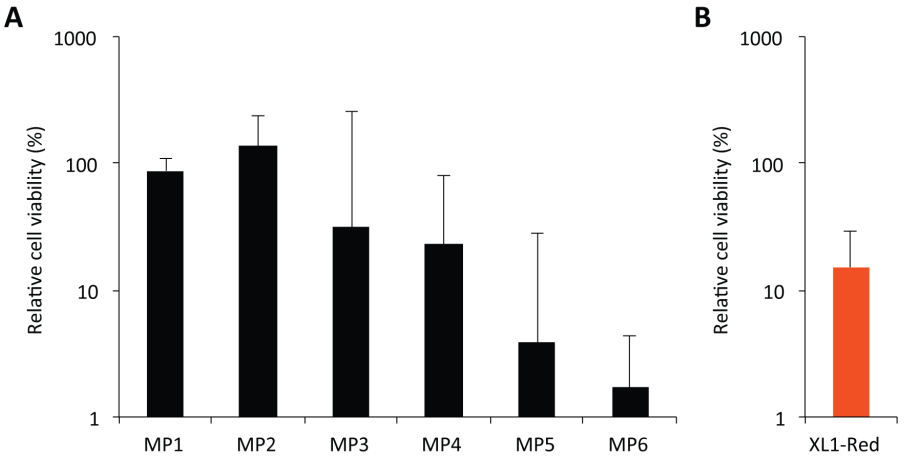


Figure 3.5 | Effect of MPs on host viability under induced conditions. A) Relative cell viability was calculated as the fractional cell titer following arabinose induction as compared to the uninduced control for each MP. Viability is anti-correlated with mutagenic potency at high levels. B) The XL1-Red strain shows the expected level of viability (compared to the control, XL1-Blue) given its mutagenic potency as compared to the designed MPs.

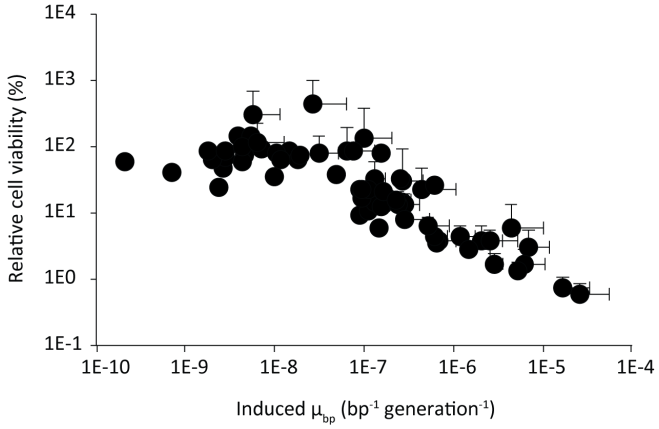


Figure 3.6 | Relationship between host viability and induced levels of mutagenesis for all MPs. Low potency MPs which induce up to ~$4x10^{-7}$ substitutions/bp/generation were well tolerated by the *E. coli* MG1655 Δ*recA::apra* host, while higher levels of mutagenesis generally resulted in a reduced host viability, as expected. This inflection point corresponds to ~1.9 substitutions/genome/generation for wild-type *E. coli MG1566* (genome size = 4.64 x 10^6 bp).

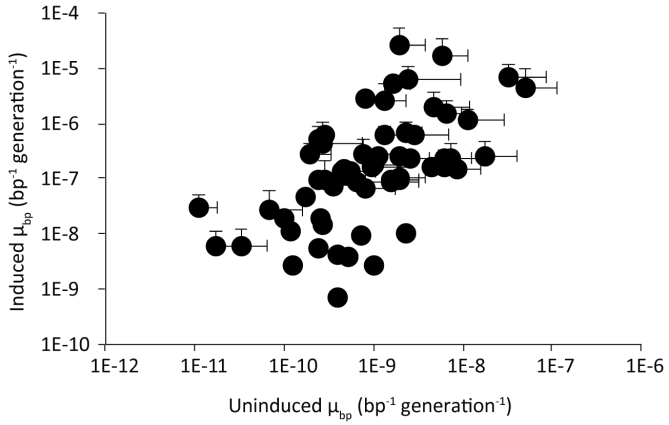


Figure 3.7 | Relationship between uninduced and induced levels of mutagenesis for all MPs. Higher levels of background (uninduced mutagenesis) were generally accompanied by an increase in overall MP mutagenesis upon induction. The full data set is provided in **Table 3.2**.

To test the inducibility of MP6, we increased the concentration of glucose from 25 mM to 200 mM during the transformation and growth stages prior to induction to maximize catabolite repression of the arabinose-inducible promoter. Under these modified conditions, induction of log-phase cultures of MG1655 Δ*recA::apra* carrying MP6 with increasing concentrations of arabinose resulted in a 35,000-fold range of mutational potency between 10 μM and 100 mM arabinose (**Figure 3.8**). Despite this strong induction effect, MP6-carrying cultures maintained low levels of mutagenesis when suppressed with 200 mM glucose (**Figure 3.8**).

To further evaluate MP6, we compared its performance with that of the most commonly used *in vivo* mutagenesis strain, XL1-Red (Agilent Technologies). XL1-Red is deficient in proofreading (*mutD5*), mismatch-repair (*mutS*) and base-excision (*mutT*) activities, resulting in high levels of substitutions in chromosomal and episomal DNA[18].

However, the strain grows very slowly, is difficult to manipulate genetically, has poor transformation efficiency, and produces a fairly narrow mutagenic spectrum[19]. Using the rifampin resistance assay, we found that XL1-Red results in 8 substitutions genome^{-1} generation^{-1}, while MP6 in the induced state produces an average of 29 substitutions genome^{-1} generation^{-1}, a ~4-fold higher mutation rate (**Figure 3.8**). The uninduced state of MP6 yields similar background mutagenesis levels (0.01 substitutions genome^{-1} generation^{-1}) as the non-mutagenic related strain XL1-Blue (0.005 substitutions genome^{-1} generation^{-1}) (**Figure 3.2 and 3.8**). Together, these results establish that MP6 enables mutagenesis levels exceeding that of the most commonly used mutator *E. coli* strain, and offers the ability to control mutation rate with an exogenous small molecule.

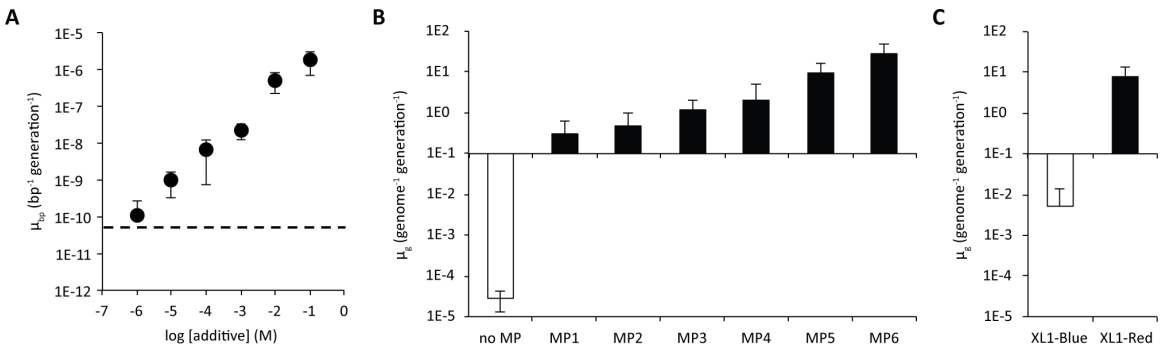


Figure 3.8 | Features of the MP system. (A) The dynamic range of MP6 was evaluated using increasing concentrations of arabinose (black circles) in the presence of 25 mM glucose in all cases. Treatment with 200 mM glucose only (dotted line) showed low mutagenesis under identical conditions. Comparison of the mutation rate under induced and uninduced conditions reveals a 35,000-fold dynamic range. Using the number of known unique mutations found in rifampin-resistant *rpoB* alleles (77 sites), the average number of substitutions per *E. coli* genome per generation (μ_g) was calculated for (B) MG1655 Δ*recA::apra* without an MP (white bar) vs. carrying MP1-6 under induced conditions (black bars), and (C) XL1-Blue (white bar) vs. XL1-Red (black bar). Error bars denote the standard deviation of at least three biological replicates.

3.7 – MP6 augments M13 bacteriophage mutagenesis

To further characterize MP1, MP4, and MP6, we assessed their impact on the mutagenesis of bacteriophage DNA, a common component of laboratory screening and evolution efforts. We measured the mutagenesis rate of M13 phage in host cells harboring a variety of MP variants or in XL1-Red using a *lacZ* inactivation assay. Briefly, a *lacZ* cassette was integrated downstream of *geneIII* in the wild-type M13 genome to yield SP063, resulting in the production of blue plaques in the presence of an X-Gal analog (**Figure 3.9**). Disruption of the *lacZ* cassette due to high mutagenesis reduces or ablates this blue-plaque phenotype, enabling the estimation of phage mutagenesis rates. We compared the ratio of white:blue plaques (*lacZ*$^-$ phenotype) using MP1, MP4, MP6, and XL1-Red.

Under conditions supporting phage inoculum expansion in overnight culture using S1030 cells (**Figures 3.10 and 3.11; Table 3.1**), we observed that the phage-borne *lacZ* inactivation rates scaled with MP potency, reaching up to 27% white plaques

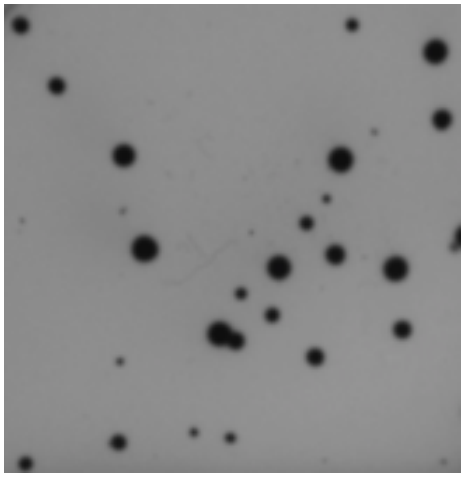


Figure 3.9 | Plaque assay of the *lacZ*-carrying M13 phage SP063. SP063 carries the wild-type E. coli β-galactosidase gene with a consensus ribosome-binding site directly downstream of *geneIII*. Plating using soft agar containing S1030 cells in the presence of the X-Gal analog Bluo-Gal (Life Technologies) results in a strong, deep blue plaques (shown here as dark circles).

(representing mutant, inactive *lacZ* genes) after 18 to 24 h of culture with MP6 (**Figure 3.12**). As MP6 increases the rate of mutagenesis by ~100-fold as compared to MP1, the expected M13 bacteriophage mutation rate is elevated to $7.2x10^{-3}$ substitutions bp^{-1} $generation^{-1}$, resulting in an average of 22 substitutions $genome^{-1}$ $generation^{-1}$ in the SP063 phage. This mutagenesis potency of 2.3 substitutions kbp^{-1} (0.23%) approaches the potency of the most commonly used *in vitro* mutagenesis methods, Mutazyme II (0.3%-1.6%, Agilent Technologies).

To enable comparison to XL1-Red, which lacks an *F'* episome and thus cannot be infected by M13 phage particles but allows for the production of fully functional phage, purified SP063 dsDNA was transformed into S1021 (the F^- variant of S1030, **Table 3.1**) cells carrying the aforementioned MPs (induced prior to or after SP063 transformation, or both; **Figure 3.11**) or XL1-Red. Whereas phage produced from transformed XL1-Red cells only yielded an average of 5% white plaques on S1030 cells, phage produced from the S1021 strain carrying MP6 yielded 15% white plaques and MP1 and MP4 yielded 7%

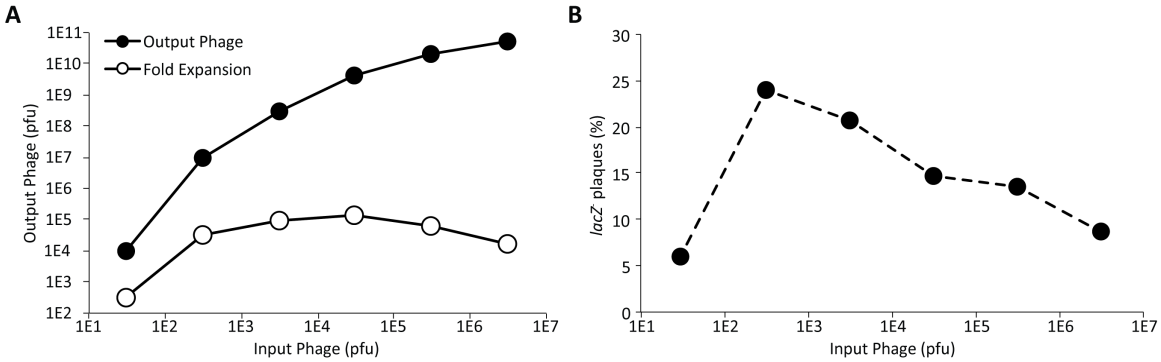


Figure 3.10 | Optimization of phage inoculant for optimal expansion and mutagenesis. A) *E. coli* S1030 carrying MP6 were induced with arabinose during log-phase growth and concomitantly infected with serially diluted SP063 phage. Phage were titered after overnight propagation and the fold expansion of the phage population was determined. B) The percentage of white and light blue ($lacZ^-$) plaques in the presence of Bluo-Gal suggests a correlation between phage population expansion and mutagenesis efficiency.

and 10% white plaques, respectively, on S1030 cells (**Figure 3.12**). These results further demonstrate the greater mutational potency of MP6 compared to that of XL1-Red, and also highlight the strain flexibility enabled by using inducible, genetically encodable mutagenesis systems.

3.8 – Mutational spectra of designed MPs

In addition to mutagenic potency, the distribution of mutation types is also important, as a narrow mutational spectrum limits the diversity of changes that can be accessed. To analyze the spectrum of produced mutations, we took advantage of previously reported distributions of rifampin-resistant *rpoB* mutants. Importantly, mutations covering each of the 12 possible transitions and transversions in the *rpoB* gene are known to endow *E. coli* with resistance to high levels of rifampin[32]. We assessed the spectra of MP1, MP4, and MP6 by sequencing rifampin-resistant *rpoB* alleles in mutated MG1655 Δ*recA::apra*, and compared the spectra of each to the *rpoB* mutation spectrum afforded by XL1-Red (**Figure 3.13**). MP1 yielded a narrow mutagenic spectrum strongly

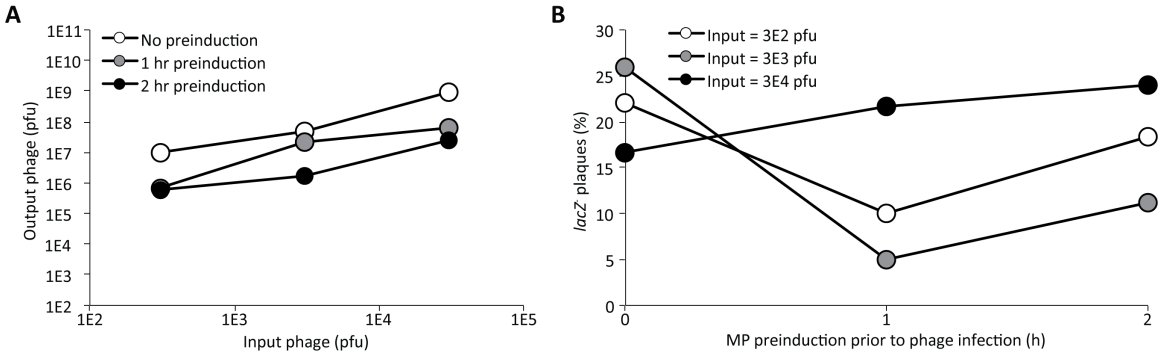


Figure 3.11 | Effect of MP pre-induction on phage mutagenesis. *E. coli* S1030 carrying MP6 were induced with arabinose during log-phase growth and concomitantly infected with SP063 phage at defined titers. Alternatively, infection was delayed for 1 h or 2 h. Phage were titered after overnight propagation and the percentage of white plaques in the presence of Bluo-Gal was counted.

biased towards A:T → T:A transversions, a known side-effect of using *recA730*-based mutators on genomic DNA[33]. Comparatively, the intermediate MP4 had a more uniform distribution, covering more transitions and a marked increase in G:C → A:T and A:T → G:C, a hallmark of mutagenesis methods that perturb the mismatch repair response[34]. MP6 exhibited a wider spectrum still, with a more uniform distribution of transitions and transversions. The identities of the observed rifampin-resistant mutations are in agreement with previous studies (**Figure 3.13**)[32]. Notably, XL1-Red exclusively displayed two types of mutations, C:G → T:A and T:A → C:G transitions (**Figure 3.13**). This observation is consistent with previous reports describing the narrow mutational spectrum of XL1-Red[19]. Additional MP characterization the β-galactosidase (*lacZ)* reversion strains developed by Miller and coworkers[11] revealed similar mutational spectra

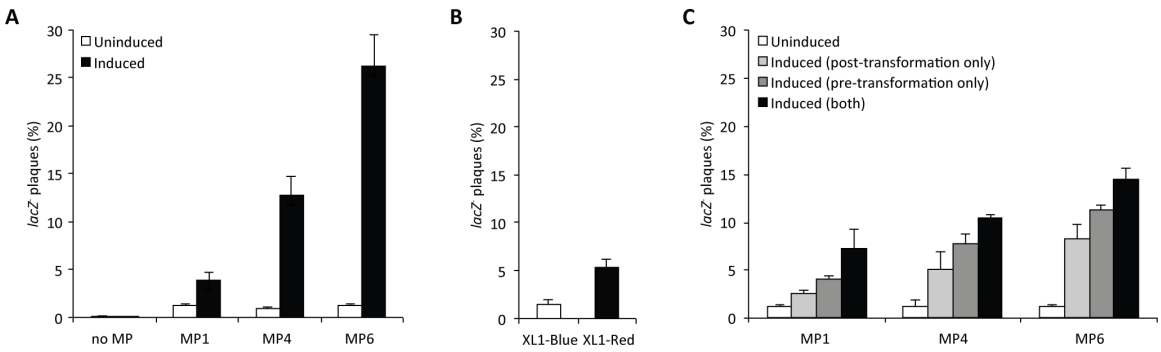


Figure 3.12 | *In vivo* mutagenesis of M13 bacteriophage DNA. (A) S1030 cells (F^+) carrying the indicated MPs were infected with M13 bacteriophage carrying a constitutive *lacZ* expression cassette (SP063), accompanied by the induction of the MP using arabinose, or suppression of the MP using glucose. (B) XL1-Blue or XL1-Red cells were transformed with purified SP063 DNA and recovered after overnight growth. (C) S1021 cells (identical to S1030, but lacking *F'*) carrying the indicated MPs were induced with arabinose or repressed with glucose for 2 h, transformed with purified SP063 DNA, and again induced with arabinose or repressed with glucose during recovery. For (A)-(C), progeny phage from the overnight propagation were plaqued on S1030 cells and stained using the X-Gal analog Bluo-Gal. The fraction of plaques that showed a white or light-blue *lacZ*$^-$ phenotype reflecting loss-of-function mutation(s) is shown. Error bars denote the standard deviation of at least three biological replicates.

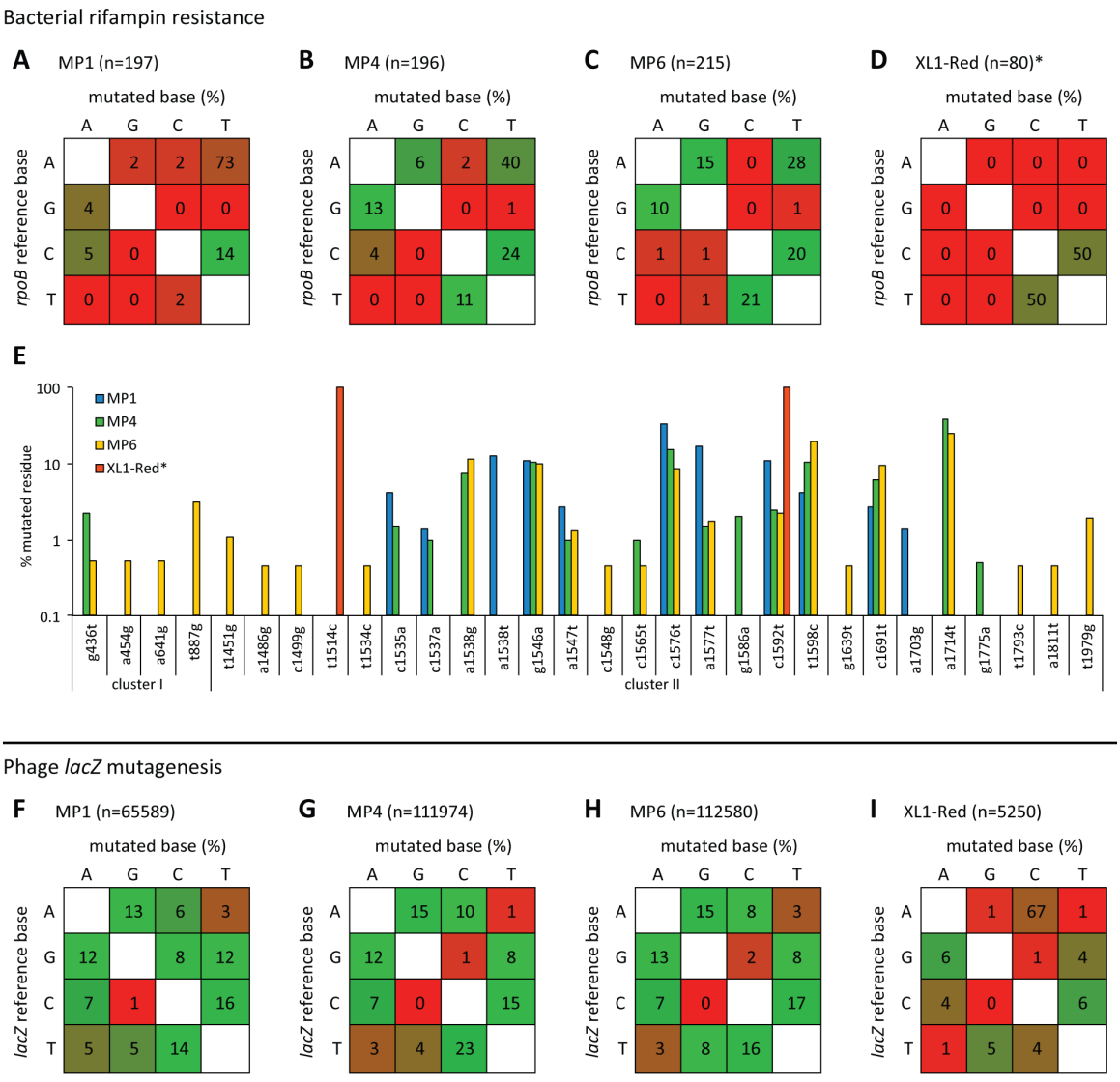


Figure 3.13 | Mutagenic spectra of the MPs. (A-D) The *rpoB* locus of single rifampin-resistant colonies was amplified by PCR and sequenced in both clusters I (aa 451-754) and II (aa 84-401). (E) Identities of the rifampin-resistant *rpoB* alleles from MP1, MP4, or MP6 mutagenesis, or using XL1-Red. The MPs yielded a wide distribution of mutation types, with the diversity of alleles strongly correlating with MP potency. (F-I) SP063 phage containing a constitutive *lacZ* expression cassette was propagated on the indicated mutator strain under induced conditions, and mutations were analyzed by high-throughput sequencing. In all cases, the number of sequenced mutations (n) is indicated for the MP and XL1-Red assays. * For (D) and (E), all sequenced XL1-Red rifampin-resistant colonies carried the identical F505S/S531F *rpoB* genotype.

(**Figure 3.14**). Strains CSH101 to CSH106 each carry different nonfunctional missense mutants of *lacZ* at codon 461 (natively encoding glutamic acid) on the *F'* episome. If a mutation reverts the nonfunctional codon to a glutamic acid, the strain can synthesize functional LacZ and survives using lactose as the only carbon source. These six strains are designed to report on all 12 possible mutations using this codon reversion. Using these strains, we observed that MP1 had the most narrow episomal mutational spectrum, with a moderate bias towards G:C → A:T (CSH102) and A:T → G:C (CSH106) substitutions (**Figure 3.14**). MP4 showed an improved distribution of mutations, with a still moderate preference for A:T → G:C (CSH106) substitutions (**Figure 3.14**). MP6 showed a near equal distribution of mutations in all six strains, with the exception of A:T → C:G (CSH101) substitutions which were detected at 10- to 100-fold lower levels than the other substitutions (**Figure 3.14**). Taken together, these results suggest that MP6 generally outperforms the other MPs and XL1-Red based on both the frequency of *lacZ* reversion as well as the breadth of mutation types detected by these strains. Importantly, the mutational potency and spectra using the episomal *lacZ* reversion assays were in agreement with those from the rifampin resistance assays (**Figure 3.13 and 3.14**). These results establish the ability of the MPs to affect a wide variety of mutations in both genomic and episomal DNA.

To further characterize the mutagenic spectra, we propagated the *lacZ*-encoding phage SP063 using S1030 strains carrying MP1, MP4 and MP6, or transformed SP063 DNA directly into XL1-Red to produce progeny phage, and subjected the *lacZ* ORF from all the resultant phage to high-throughput DNA sequencing (**Figure 3.13**). The mutagenesis efficiency for all conditions was in agreement with that revealed by the other

assays (**Figure 3.13 and 3.14**). The MPs generated broad mutagenic spectra in progeny phage, consistent with both *rpoB* single clone sequencing and *lacZ* reversion assays, with the exception of MP1, which showed a more uniform distribution of mutation types using the phage assay and *lacZ* reversion assays than using *rpoB* sequencing. This discrepancy is likely a result of the MP1-encoded *recA730* allele, which selectively enhances the rate of A:T → T:A transversions in DNA of strains lacking *recA* (MG1655 Δ*recA::apra*), as it no longer competes for substrates with wild-type (CSH strains[11], **Table 3.1**) or reduced-activity *recA* mutants (S1021 and S1030, **Table 3.1**) for function. Interestingly, the MP6 mutagenic spectrum shows an enhancement in C→T transitions with a concomitant decrease in T→C transitions, a hallmark of cytidine deaminase-mediated mutagenesis. Phage sequences produced from XL1-Red revealed a much narrower mutational spectrum, with a bias for A:T → C:G mutations. We further note that the number of mutations among the tested MPs does not agree with differences in mutagenesis potency, likely due to mutation accumulation in the bacteriophage genome,

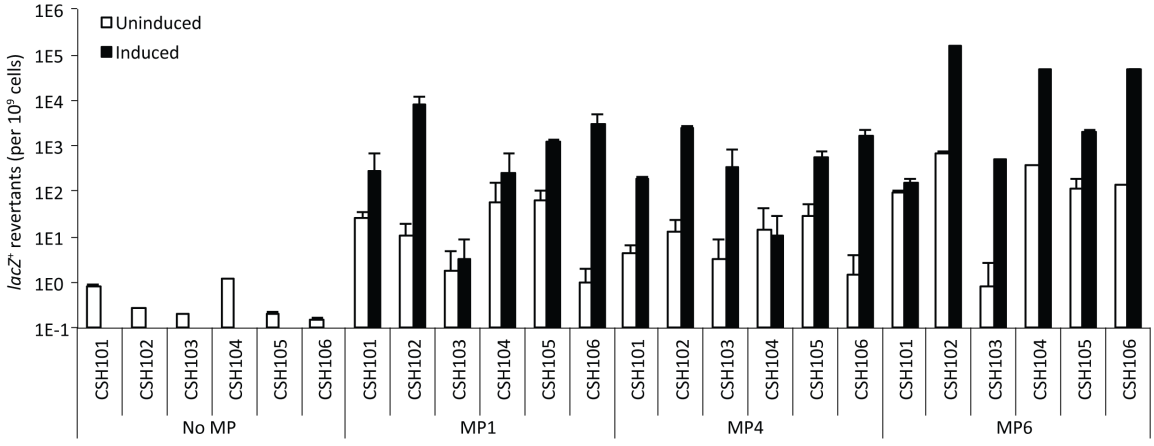


Figure 3.14 | Analysis of *F'* episomal mutations rates using various MPs. The frequency of *lacZ*+ revertants is the fraction of colonies surviving on lactose as the sole carbon source as compared to the total colony count (colonies that survive on glucose as sole carbon source). Each strain reports the MP's ability to increase the frequency of a specific mutation type.

which would negatively affect propagation and thus mutation accumulation. Taken together, these results reveal that the MPs developed in this study can outperform the most widely used *in vivo* and *in vitro* mutagenesis techniques (**Figure 3.15**) both in mutagenic potency and mutational breadth.

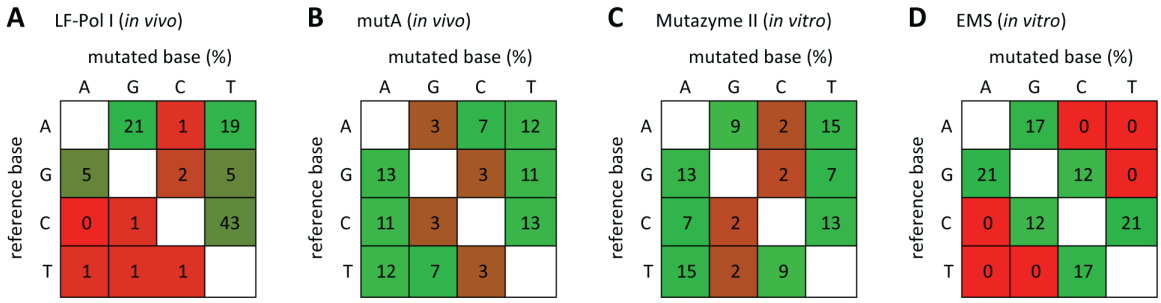


Figure 3.15 | Mutagenic spectra of commonly used mutagenesis techniques. (A-D) Previously reported mutagenic spectra of four commonly used mutagenesis methods: LF Pol I[2], *mutA*[3], Mutazyme II[13] and EMS[14].

3.9 – Evolution of antibiotic resistance using the designed MPs

To evaluate the impact of these MPs on laboratory evolutionary outcomes, we evolved antibiotic resistance in *E. coli*, as well as RNA polymerase substrate specificity changes in bacteriophage, using cells harboring different MPs. First we tested the ability of MP1, MP4, or MP6 to evolve resistance of *E. coli* strain MG1655 Δ*recA::apra* to a number of commonly used antibiotics. Mid-log-phase cultures of MG1655 Δ*recA::apra* carrying MP1, MP4, or MP6 were induced for 18-21 h, then serially diluted and plated on agar plates without antibiotics, or with 5-100 μg/mL of carbenicillin, cephotaxime, fosfomycin, kanamycin, metronidazole, norfloxacin, rifampin, spectinomycin, streptomycin, or tetracycline. The antibiotic concentrations used in all ten cases were well above known MIC values (**Table 3.5**).

Antibiotic	MIC (ug/mL)
carbenicillin	2 - 25
cefotaxime	0.016 - 0.25
fosfomycin	0.125 - 8
kanamycin	0.25 - 8
metronidazole	8 - 32
norfloxacin	0.016 - 0.125
rifampin	0.5 - 16
spectinomycin	8 - 64
streptomycin	1 - 16
tetracycline	0.5 - 8

Table 3.5 | Minimum inhibitory concentrations (MICs) for selected antibiotics. All data regarding *E. coli* MICs was tabulated from the Antimicrobial Index Knowledgebase (http://antibiotics.toku-e.com).

After only 18 h of growth on antibiotic-containing solid medium, large fractions of the bacterial population showed resistance to high concentrations of carbenicillin, fosfomycin, kanamycin, metronidazole, rifampin, spectinomycin, streptomycin, and tetracycline (**Figure 3.16 and Table 3.5**). No resistant colonies were detected for cefotaxime or norfloxacin. The frequency of antibiotic resistance strongly correlated with MP potency (**Figure 3.16**). For example, we observed the evolution of resistance to high levels of kanamycin with no intermediate selection step using MP6, which enabled up to 11% of the total population to grow in the presence of 30 μg/mL kanamycin. In contrast, only 0.02% or 0.66% of the population survived this level of kanamycin when using MP1 or MP4, respectively (**Figure 3.16**).

We repeated these antibiotic resistance evolution experiments to compare the performance of these MPs to those of the commonly used chemical mutagens[11] ethyl methanesulfonate (EMS), methylnitronitrosoguanidine (MNNG), and 2-aminopurine

(2AP), as well as UV irradiation, XL1-Red, and XL1-Blue (**Figure 3.16 and Table 3.5**). Importantly, UV irradiation can only be used for short durations due to toxicity from random crosslinking of cellular components. As such, mutagenesis with UV irradiation typically constitutes a single round of mutagenesis, while chemical mutagens, hypermutator strains and mutagenesis plasmids enable continuous mutagenesis. MP6 outperformed all six of the other mutagenesis treatments or strains for all but one of the antibiotics. Cephotaxime resistance was not observed from use of any of the MPs, but was weakly detected from the chemical mutagens (**Figure 3.16**). We speculate that the known ability of all three of these chemical mutagens to greatly enhance G:C → A:T transitions may contribute to the evolution of cephotaxime resistance. This finding further suggests that a period of neutral drift or intermediate selection pressure may be needed to evolve complex phenotypes requiring mutations to multiple cellular components.

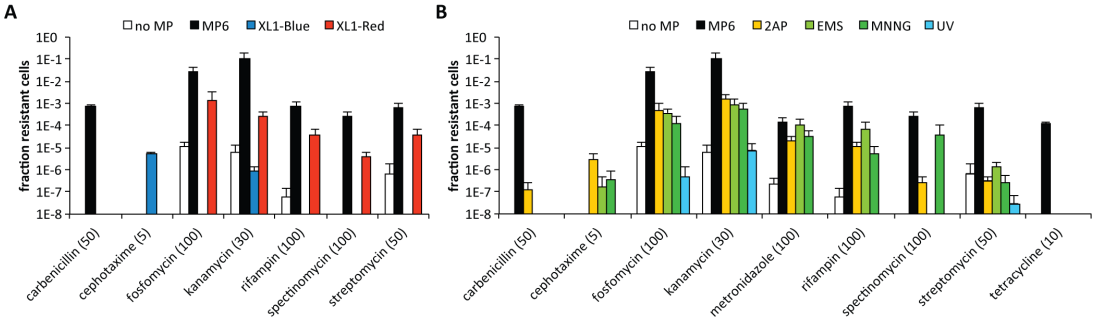


Figure 3.16 | Comparison of MP6 and other mutagenesis approaches for the evolution of antibiotic resistance. (A) MG1655 Δ*recA::apra* cells with or without MP6 were grown for 18-21 h under induced conditions. XL1-Blue and XL1-Red were grown for 18-21 h. (B) MG1655 Δ*recA::apra* cells were treated using chemical mutagens as previously described [11]. In all cases, cultures were plated on the indicated antibiotics following overnight culture in the absence of any selection. The numbers in parenthesis indicate the antibiotic concentrations used in μg/mL. The fraction of cells resistant to each antibiotic was calculated relative to the total number of cells on plates without antibiotics. Resistance to norfloxacin was not detected for any of the tested strains or treatments. XL1-Blue and XL1-Red are both inherently resistant to tetracycline and metronidazole, so no comparison is shown for either antibiotic. See **Table 3.6** for full antibiotic resistance data. Error bars denote the standard deviation of at least three biological replicates.

Cumulatively, these results suggest that MP6 can rapidly generate strains with novel properties, outperforming several commonly using chemical mutagens, UV irradiation, and XL-1 Red.

	CRB	CTX	FOS	KAN	MTX	RIF	SPC	STR	TET
no MP	0.00E+00	0.00E+00	1.10E-05	5.78E-06	2.22E-07	5.56E-08	0.00E+00	6.67E-07	0.00E+00
MP1	9.57E-06	0.00E+00	9.17E-04	1.84E-04	6.47E-01	1.26E-05	1.91E-07	5.68E-06	0.00E+00
MP4	1.37E-05	0.00E+00	6.97E-03	6.56E-03	1.52E-06	6.17E-05	1.14E-05	1.56E-05	5.97E-06
MP6	7.39E-04	0.00E+00	2.78E-02	1.11E-01	1.50E-04	7.17E-04	2.55E-04	6.22E-04	1.15E-04
2AP	1.17E-07	2.80E-06	4.83E-04	1.53E-03	2.03E-05	1.03E-05	2.50E-07	3.17E-07	0.00E+00
EMS	0.00E+00	1.67E-07	3.75E-04	8.17E-04	1.10E-04	6.68E-05	0.00E+00	1.33E-06	0.00E+00
MNNG	0.00E+00	3.67E-07	1.22E-04	5.33E-04	3.22E-05	5.56E-06	3.67E-05	2.44E-07	0.00E+00
UV	0.00E+00	0.00E+00	5.00E-07	7.49E-06	0.00E+00	0.00E+00	0.00E+00	2.89E-08	0.00E+00
XL1-Blue	0.00E+00	5.40E-06	0.00E+00	8.66E-07	4.82E-01	0.00E+00	0.00E+00	0.00E+00	9.44E-01
XL1-Red	0.00E+00	0.00E+00	1.42E-03	2.62E-04	9.84E-01	3.62E-05	3.81E-06	3.60E-05	1.10E-04

Table 3.6 | Comparison of developed MPs to chemical mutagens, UV light and XL1-Red. The fraction of cells resistant to each antibiotic upon mutagenic treatment is shown for all tested antibiotics. No resistance was observed for norfloxacin using any MP, chemical mutagen, or strain. All MPs, chemical mutagens, and UV light treatments used *E. coli* MG1655 Δ*recA::apra*. CRB, 50 μg/mL carbenicillin; CTX, 5 μg/mL cefotaxime; FOS, 100 μg/mL fosfomycin; KAN, 30 μg/mL kanamycin, MTX, 100 μg/mL metronidazole; RIF, 100 μg/mL rifampin; SPC, 100 μg/mL spectinomycin; STR, 50 μg/mL streptomycin; TET, 10 μg/mL tetracycline. We note that MP1 (*recA730*), XL1-Blue (*recA1*), and XL1-Red (wt *recA*) are all proficient at recombination, a known requirement for high-level resistance to metronidazole[12]. Additionally, XL1-Blue and XL1-Red are both inherently resistant to tetracycline, explaining the observed high incidence of resistance.

3.10 – MP6 enables direct RNA polymerase evolution during PACE

Next we compared the performance of MP1 and MP6 during phage-assisted continuous evolution (PACE)[22,23,35-37] of T7 RNA polymerase. We previously showed that PACE can evolve T7 RNA polymerase to recognize the distant T3 promoter (P_{T3}) (**Figure 3.17**), but only using either an intermediate evolutionary stepping-stone (a T7/T3 hybrid promoter)[22,35,36], or an initial period of evolutionary drift in the absence of

selection pressure[23]. Without either an evolutionary stepping-stone or initial evolutionary drift, PACE cannot evolve T7 RNAP variants that recognize P_{T3}, leading to rapid phage washout[22,23,35,36]. We compared the ability of MP1 and MP6 to rapidly evolve P_{T3}-active T7 RNAP variants in the absence of evolutionary drift, or under conditions in which the selection stringency was slightly reduced. Importantly, previous attempts of T7 RNAP evolution towards P_{T3} activity using MP1 required a drift period of virtually no selection pressure over 18 h to yield P_{T3}-active variants, whereas the high or intermediate selection pressures resulted in rapid phage washout[23].

In agreement with our previous results, T7 RNAP phage added to lagoons fed by host cells harboring MP1 under high or intermediate selection stringency rapidly washed out in < 20 h of PACE when required to recognize P_{T3} (**Figure 3.18**). In contrast, both lagoons tested in which host cells harbored MP6 enabled the propagation of T7 RNAP

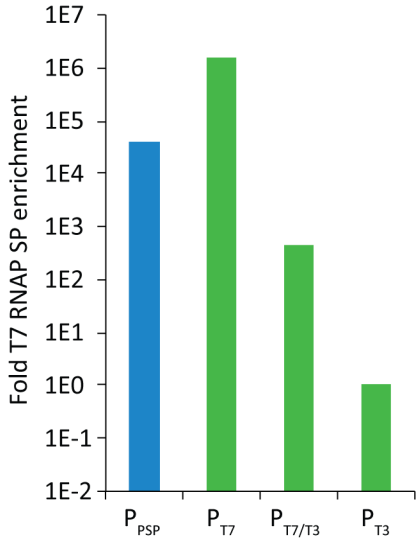


Figure 3.17 | Activity of T7 RNAP on cognate and non-cognate promoters. Log-phase S1030 cells carrying accessory plasmids (APs) with a *geneIII* cassette with an upstream phage shock protein (PSP), T7, hybrid T7/T3, or T3 promoter were infected with selection phage (SPs) carrying the wild-type T7 RNAP. The fraction of output phage vs. input phage indirectly reports on the activity of the T7 RNAP on the various promoters. Enrichment factors of ~100 or less indicate extremely weak to non-existent activity.

phage on host cells requiring P_{T3} recognition under high or intermediate selection stringency, with nearly 10% of both populations after only 10 h of PACE exhibiting activity on P_{T3} by activity-dependent plaque assays (**Figure 3.18**). Sequencing confirmed mutations M219K/E222K/E222Q together with N748D in all surviving clones (**Figure 3.19**), in agreement with previous findings of our group and others[22,23,35,36,38]. The ability of MP6 to enable the discovery of P_{T3}-active T7 RNAP variants highlights the ability of this MP to access mutations more efficiently, and thus mediate a more thorough and larger sampling of sequence space. Collectively, these results establish that the enhanced mutagenesis mediated by the new MPs can enable accelerated access to evolved proteins that are difficult to access directly using previous methods.

3.11 – Discussion

Using a systematic, mechanism-guided approach, we developed a series of

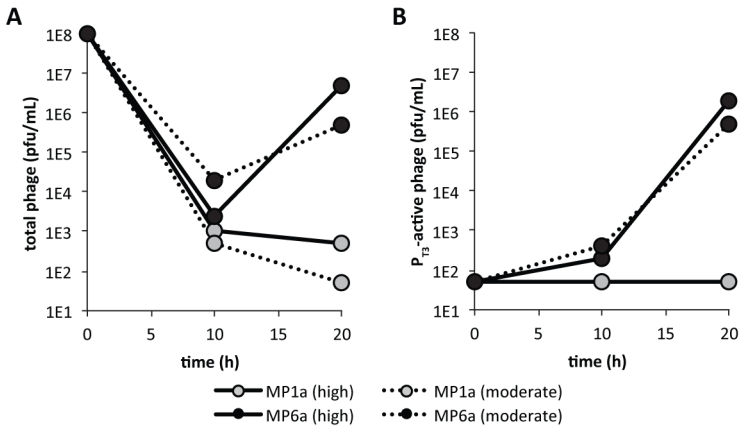


Figure 3.18 | Continuous evolution of P_{T3}-active RNAP variants. S1030 cells carrying an accessory plasmid (AP) encoding P_{T3} upstream of M13 bacteriophage *geneIII* and either MP1a or MP6a were infected with selection phage (SP) carrying wild-type T7 RNAP under conditions in which selection stringency was high (0 ng/mL ATc) or moderate (30 ng/mL ATc). (A) Total phage titers at 10 h and 20 h after lagoon inoculation with the SP encoding T7 RNAP. (B) Titers of phage encoding evolved RNAP variants active on P_{T3}. The limit of detection is 50 pfu/mL.

vectors that express a variety of genes known to adversely affect DNA replication fidelity. In total we generated and assayed 80 candidate MPs with mutation rates spanning five orders of magnitude (**Figures 3.6 and 3.7; Tables 3.2 and 3.3**). The resulting MPs enable highly potent, broad-spectrum, inducible, vector-based *in vivo* mutagenesis that rival the performance characteristics of popular *in vitro* methods such as error-prone PCR, while offering key advantages of *in vivo* mutagenesis. These advantages include enabling mutation and selection cycles to be coupled, and bypassing

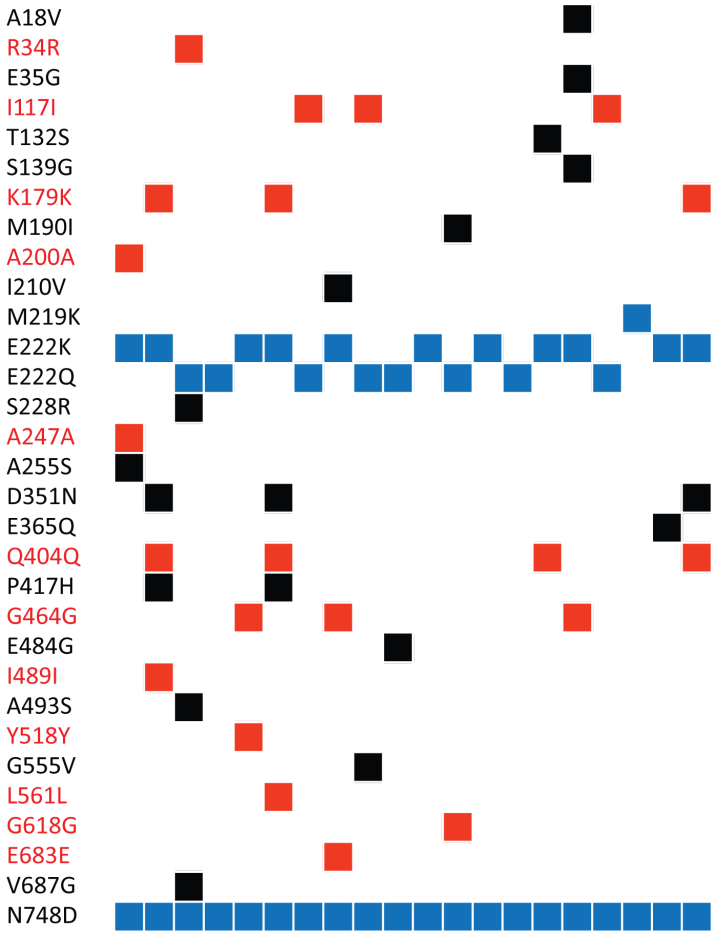


Figure 3.19 | Single-phage plaque sequencing of P_{T3}-active SPs. Single phage plaques at 10 h of PACE using the P_{T3} AP were isolated and subjected to Sanger sequencing. All clones carried T7 RNAP variants with conserved mutations known to confer activity on P_{T3} (blue), as well as additional mutations that may further enhance activity (black). Silent mutations were also detected (red).

transformation efficiency bottlenecks that limit the size of populations that can be generated from DNA diversified *in vitro*. The MPs developed here offer major advantages over current *in vivo* mutagenesis methods such as chemical mutagens or base analogs, UV irradiation, or constitutive hypermutator strains. Hypermutator strains, for example, generally suffer from poor transformation efficiency (XL1-Red, ~1 x 10^6 cfu/μg plasmid DNA), high instability, and narrow mutagenic spectra. MP6 increases the mutation rate of *E. coli* by 322,000-fold, and substantially exceeds both the mutation rate and the mutagenic spectra of XL1-Red. Importantly, MP6 can enable approximately ~2.3 substitutions kbp^{-1} in a gene of interest using phage vectors in a single generation, with additional increases in mutagenesis efficiency concomitant with longer propagation times.

To demonstrate the utility of these vectors, we used a whole-genome mutagenesis approach to evolve high-level antibiotic resistance in *E. coli*. In the absence of any prior selection or mutagenesis step, MP6 rapidly mediates the evolution of antibiotic resistance to many commonly used antibiotics within 18 h. The efficiency and effectiveness of antibiotic resistance mediated by MP6 compares favorably with that of a number of potent chemical mutagens (2AP, EMS, MNNG), UV irradiation, and the hypermutator strain XL1-Red. In addition, we observed that MP6 enables the continuous evolution of T7 RNA polymerase variants capable of initiating transcription at the non-cognate T3 promoter in less than 10 h, without requiring evolutionary stepping-stones or an initial period of evolutionary drift. We anticipate that these MPs will broadly enable the use of *in vivo* mutagenesis to provide efficient access to rare solutions in sequence space that would otherwise be much more difficult to reach using current *in vitro* or *in vivo*

methods. Importantly, the MPs offer non-targeted mutagenesis enhancements that complement sequence-targeted mutagenesis approaches such as MAGE[39] or using CRISPR-Cas9[40]. Furthermore, the MPs operate largely independently of host recombination pathways and bias mutagenesis towards substitutions rather than insertions or deletions, both of which are desirable characteristics in library generation during directed evolution. The properties of MP6, which include very high mutagenesis efficiency, broad mutational spectrum, small-molecule inducibility, and compatibility with a variety of bacterial strains, together represent a substantial advance in *in vivo* mutagenesis methodology for the laboratory evolution community.

3.12 – Methods

General methods. All PCR reactions were performed using *Pfu*Turbo Cx polymerase (Agilent Technologies) or VeraSeq ULtra polymerase (Enzymatics). Water was purified using a MilliQ water purification system (Millipore, Billerica MA). All MPs were constructed using USER cloning (New England Biolabs). Native *E. coli* genes were amplified by PCR directly from genomic DNA, and non-bacterial genes were synthesized as bacterial codon-optimized gBlocks Gene Fragments (Integrated DNA Technologies). All DNA cloning was carried out using NEB Turbo cells (New England Biolabs).

General MP strain preparation. Mid log-phase (OD_{600} = ~0.5-0.8) cells of the strain of interest grown in 2xYT (United States Biological) were transformed with the desired MP, and recovered for 45 min in Davis rich media[23] to suppress MP induction. All transformations were plated on 2xYT in 1.8% agar (United States Biological) containing 40 μg/mL chloramphenicol (Sigma Aldrich), 10 μg/mL fluconazole (TCI America), 10 μg/mL amphotericin B (TCI America), 25 mM glucose (United States Biological) and grown for 12-18 h in a 37 °C incubator. Colonies transformed with the appropriate MP were picked the following day and grown in Davis rich media containing 40 μg/mL chloramphenicol, 10 μg/mL fluconazole, and 10 μg/mL amphotericin B for 12-18 h. Following overnight growth of the MP-carrying strains, cultures were diluted 1,000-fold into fresh Davis rich media containing 40 μg/mL chloramphenicol, 10 μg/mL fluconazole, and 10 μg/mL amphotericin B. The remainder of each experiment is described in each of the following sections.

Rifampin resistance assay. Upon reaching mid log-phase, cultures were induced with 25 mM arabinose (Davis rich media + arabinose) or suppressed with 25 mM glucose (Davis rich media only) and allowed to continue growth for an additional ~18-24 h in a 37 °C shaker. The high arabinose concentration ensures sufficient induction of the plasmid-borne mutators MG1655 Δ*recA::apra* despite arabinose catabolism by this strain upon glucose depletion. For XL1-Blue and XL1-Red strains, cultures were started directly from glycerol stocks according to the manufacturer's instructions and incubated for an identical amount of time as the MP-carrying strains. After overnight growth, cultures were serially diluted in 10-fold increments and plated on 2xYT-agar containing 10 μg/mL fluconazole, 10 μg/mL amphotericin B, and 100 mM glucose +/- 100 μg/mL rifampin. After 18-24 h, the number of colonies on the glucose +/- rifampin plates was counted for each culture. The mutation efficiency induced by the MP (t_{bp}, substitutions bp^{-1} $generation^{-1}$) was calculated using the equation: $\mu_{bp} = f/[R \times ln(N/N_0)]$, where f is the frequency of rifampin-resistant mutants (as compared to the glucose control), R is the number of unique sites yielding rifampin resistance (77 previously identified sites; we identified only 21 sites across both *rpoB* clusters in our experiments), N is the final population size, and N_0 is the population size at which resistance is first observed (empirically determined to be ~1.5×10^7). To calculate μ_G, μ_{bp} was multiplied by the genome size, which for MG1655 was 4.64×10^6 bp.

Episomal *lacZ* reversion assay. Upon reaching mid log-phase, the cultures were induced with 25 mM arabinose or suppressed with 25 mM glucose, and allowed to continue growth for an additional ~18-24 h. After overnight growth, the cultures were centrifuged

for 2 min at 10,000x rcf and resuspended in an equal volume of 10% glycerol. This procedure was carried out twice to remove trace glucose or other carbon sources from the supernatant prior to plating. Washed cells were serially diluted in 10-fold increments using 10% glycerol and plated on M9 minimal media agar supplemented with 5 mM $MgSO_4$, 0.01% thiamine, 335 phase, the cultures were induced with 25 mM arabinose or suppressed with 25 mM glucose, and allowed to continue growth for an additional ~18-24 h. After overnight *lacZ* reversion, and not purely due to extracellular lactose hydrolysis. After extended growth (~24-36 h), the fraction of lactose-catabolizing colonies was calculated using the number of blue colonies on the lactose plates vs. the total number of colonies on the glucose plates.

Phage *lacZ* inactivation assay. Upon reaching mid log-phase, the cultures were induced with 25 mM arabinose or suppressed with 25 mM glucose, allowed to grow for an additional 0-2 h, then, in the case of strain S1030, infected with SP063 phage, and allowed to grow for an additional ~18-24 h. For S1021 and XL1-Red (Agilent Technologies), SP063 DNA was miniprepped from infected S1030 cells and electroporated into these strains as they both lack F' episomes. For F^- cells, cultures were either induced for 2 h prior to being made electrocompotent, induced immediately following transformation, induced both prior to and following electroporation, or not induced at all. After overnight growth and phage propagation, the cultures were centrifuged 2 min at 10,000x rcf and the supernatant was filtered through a 0.2 μm PVDF filter (Millipore). The supernatant was serially diluted in 10-fold increments using Davis rich media and plaqued on S1030 cells using 1.8% 2xYT-agar for the bottom layer and

0.6% 2xYT-agar supplemented with 400 μg/mL Bluo-Gal (Life Technologies) for the top layer. The fraction of white or light blue plaques plaques (lacZ^- phenotype) was counted as a function of all plaques (blue + light blue + white), and used as a measure of mutation frequency for the *lacZ* cassette.

Sanger sequencing of *rpoB* mutations. Rifampin-resistant colonies were picked into 96 well plates and grown overnight in Davis rich media supplemented with 100 μg/mL rifampin. Following overnight growth, 10 μL aliquots were heated at 100 °C for 10 min, followed by PCR using primers AB1678 (5'-AATGTCAAATCCGTGGCGTGAC) and AB1682 (5'-TTCACCCGGATACATCTCGTCTTC) to amplify an *rpoB* fragment containing both clusters I and II. Each fragment was sequenced twice using primers AB1680 (5'-CGGAAGGCACCGTAAAAGACAT) and AB1683 (5'-CGTGTAGAGCGTGCGGTGAAA).

High-throughput sequencing of *lacZ* mutations. SP063 phage that was propagated using S1030 carrying MP1, MP4 or MP6, produced by XL1-Red following SP063 DNA electroporation, or the unmutated stock phage was amplified by PCR using primers AB437 (5'-GGCGCTGGTAAACCATATG) and DB213 (5'-GGAAACCGAGGAAACGCAA) to yield a ~3,400 bp fragment containing the *lacZ* gene. SP063 phage that was propagated under similar conditions on S1030 cells was used as the negative control. Three biological replicates were carried out for each of the aforementioned samples. The resulting PCR products were purified by gel electrophoresis using a 1% agarose gel and prepared for HTS using a Nextera kit

(Illumina) and a previously described procedure[37]. Briefly, 4 μL of DNA (2.5 ng/μL), 5 μL TD buffer, and 1 μL TDE1 were mixed together and then heated to 55 °C for 5 min. After purification (Zymo DNA purification kit), the resultant "tagmented" DNA samples were amplified with Illumina-supplied primers using the manufacturer's protocol. The resulting PCR products were then purified using AMPure XP beads and the final concentration of DNA was quantified using PicoGreen (Invitrogen) and qPCR. The samples were sequenced on a MiSeq Sequencer (Illumina) in 2x300 paired-end runs using the manufacturer's reagents following the manufacturer's protocols.

High-throughput sequencing data analysis. A previously described custom MATLAB script[37] (available upon request) was used to align MiSeq reads with Q score ≥ 30 to the wild-type sequence and count the nucleotide positions from which the experimental sample deviates from the wild-type sequence, yielding called mutations with ≥ 99.9% accuracy, corresponding to > 3 standard deviations above the mean error rate of the MiSeq high-throughput sequencing reads. To compensate for systemic sample preparation and sequencing errors, the observed fraction of mutations at each nucleotide position of the wild-type *lacZ* reference gene was subtracted from the fraction of mutations in a given experimental sample to result in the "corrected fraction mutated". Mutations were defined as nucleotide positions with a corrected fraction mutation that is both greater than the average corrected fraction mutated of the treatment of interest and at least one standard deviation higher than the corrected fraction mutation of the wild-type reference sequence. Duplicates belonging to set of paired-end reads were treated as a single sample, while duplicate reads of the same region with alternative adaptor/index

sequences were not removed so as not to introduce bias into the sequencing analysis. This process yielded an average of ~50,000 reads per position for each of the sequenced samples.

Evolution of novel antibiotic resistance. MG1655 Δ*recA::apra* cells without an MP or carrying MP1, MP4, or MP6 were grown for 18-21 hr in Davis rich media containing 40 μg/mL chloramphenicol, 10 μg/mL fluconazole, 10 μg/mL Amphotericin B, and supplemented with 200 mM arabinose to induce the MPs. Small molecule and UV mutagenesis was carried out as previously described[11]. For 2AP treatment, log-phase MG1655 Δ*recA::apra* cells were diluted to ~1000 cells, the media was supplemented with 700 μg/mL 2AP (TCI America), and the culture was allowed for grow at 37 °C for an additional 18-21 hr . For EMS treatment, 2 mL of a log-phase MG1655 Δ*recA::apra* culture (~$1x10^8$-$1x10^9$ cells) was centrifuged, washed twice with 1 mL A buffer on ice, then supplemented with 14 μL EMS (TCI America). Cells were lightly vortexed, and allowed to shake at 200 rpm at 37 °C for 45 min. After this time, the culture was centrifuged, washed twice with 1 mL A buffer on ice, diluted by 20-fold into Davis rich media without antibiotics, and allowed to grow for 18-21 hr. For MNNG treatment, 2 mL of a log-phase MG1655 Δ*recA::apra* culture (~$1x10^8$-$1x10^9$ cells) was centrifuged, washed twice with 1 mL citrate buffer (pH 5.5) on ice, supplemented with 111 μL of 1mg/mL MNNG (TCI America) and placed in a 37 °C water bath for 30 min. Following treatment, the cells were centrifuged, washed twice with 1mL 0.1 M potassium phosphate buffer (pH 7.0), diluted by 4-fold into Davis rich media without antibiotics, and allowed to grow for 18-21 hr. For UV irradiation, 2 mL of a log-phase MG1655 Δ*recA::apra*

culture (~1x10^8-1x10^9 cells) was centrifuged, resuspended in 1 mL 0.1 M MgSO4 and placed on ice for 10 min. Cells were placed in a petri dish and exposed to UV light from a SM-36-2GR UV lamp (American Air & Water) for 1 min, uncovered, at a distance of ~10 cm. Immediately following UV exposure, cells were diluted by 20-fold into Davis rich media without antibiotics, and allowed to grow for 18-21 hr. For XL1-Blue and XL1-Red strains, cultures were started directly from glycerol stocks according to the manufacturer's instructions and allowed to grow for 18-21 hr in Davis rich media. Following overnight growth, all cultures were serially diluted in Davis rich media and plated on 2xYT-agar containing 10 μcell fluconazole, 10 μfluc amphotericin B, and 100 mM glucose +/- the appropriate antibiotic. After overnight growth (~18-24 h), the numbers of colonies on the glucose +/- antibiotics plates were counted.

Continuous evolution of P_{T3}-active T7 RNAP variants. Two modified versions of MP1 and MP6 (DP1 and DP6, respectively) were generated to enable robust phage propagation during PACE. These MPs carry all of the components of their respective MPs, in addition to the previously described anhydrotetracycline (ATc)-dependent drift promoter driving *geneIII*[23]. S1030 strains carrying either MP in addition to the P_{T3} accessory plasmid (AP) were inoculated into host-cell cultures (chemostats) and grown at a dilution rate of 1.6 vol/hr as previously described[23]. Lagoons flowing from the respective chemostats were maintained at 40 mL, diluted at 0.75 vol/hr, and supplemented with either 25 mM arabinose only (high stringency) or 25 mM arabinose with 30 ng/mL ATc (intermediate stringency) for 8 h prior to infection with packaged T7 RNAP SP. We note that concentrations exceeding 30 ng/mL for extended timeframes

during PACE (> 24 hr) result in excision of the evolving gene from the selection phage. As continuous flow conditions effectively enrich for SPs capable of rapid replication, selection phage with smaller genomes are rapidly enriched to totally dominate the evolving pool. Each lagoon was infected with $4x10^9$ pfu, resulting in an initial titer of 10^8 pfu/mL of the lagoon. Samples were taken 10 h and 20 h after infection, centrifuged at 10,000 rcf for 2 min, then sterile filtered with a 0.2 μm filter and stored overnight at 4°C. Phage aliquots were titered on S1030 cells carrying either the PSP-*geneIII* AP (total phage) or the P_{T3}-*geneIII* AP (P_{T3}-active phage).

3.13 – References

1 Lynch, M. Evolution of the mutation rate. *Trends Genet* **26**, 345-352, (2010).

2 Troll, C. *et al.* The mutagenic footprint of low-fidelity Pol I ColE1 plasmid replication in E. coli reveals an extensive interplay between Pol I and Pol III. *Current genetics* **60**, 123-134, (2014).

3 Balashov, S. & Humayun, M. Z. Specificity of spontaneous mutations induced in mutA mutator cells. *Mutation research* **548**, 9-18, (2004).

4 Horst, J. P., Wu, T. H. & Marinus, M. G. Escherichia coli mutator genes. *Trends in microbiology* **7**, 29-36, (1999).

5 Aronshtam, A. & Marinus, M. G. Dominant negative mutator mutations in the mutL gene of Escherichia coli. *Nucleic acids research* **24**, 2498-2504, (1996).

6 Junop, M. S., Yang, W., Funchain, P., Clendenin, W. & Miller, J. H. In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. *DNA repair* **2**, 387-405, (2003).

7 Maki, H., Mo, J. Y. & Sekiguchi, M. A strong mutator effect caused by an amino acid change in the alpha subunit of DNA polymerase III of Escherichia coli. *The Journal of biological chemistry* **266**, 5055-5061, (1991).

8 Yang, H., Wolff, E., Kim, M., Diep, A. & Miller, J. H. Identification of mutator genes and mutational pathways in Escherichia coli using a multicopy cloning approach. *Molecular microbiology* **53**, 283-295, (2004).

9 Wu, T. H. & Marinus, M. G. Dominant negative mutator mutations in the mutS gene of Escherichia coli. *Journal of bacteriology* **176**, 5393-5400, (1994).

10 Gon, S., Napolitano, R., Rocha, W., Coulon, S. & Fuchs, R. P. Increase in dNTP pool size during the DNA damage response plays a key role in spontaneous and induced-mutagenesis in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 19311-19316, (2011).

11 Cupples, C. G. & Miller, J. H. A set of lacZ mutations in Escherichia coli that allow rapid detection of each of the six base substitutions. *Proceedings of the*

National Academy of Sciences of the United States of America **86**, 5345-5349, (1989).

12 Yeung, T. C., Beaulieu, B. B., Jr., McLafferty, M. A. & Goldman, P. Interaction of metronidazole with DNA repair mutants of Escherichia coli. *Antimicrobial agents and chemotherapy* **25**, 65-70, (1984).

13 Stratagene. Overcome mutational bias. *Strategies* **17**, 20–21, (2004).

14 Lai, Y. P., Huang, J., Wang, L. F., Li, J. & Wu, Z. R. A new approach to random mutagenesis in vitro. *Biotechnology and bioengineering* **86**, 622-627, (2004).

15 Tee, K. L. & Wong, T. S. Polishing the craft of genetic diversity creation in directed evolution. *Biotechnology advances* **31**, 1707-1721, (2013).

16 Badran, A. H. & Liu, D. R. In vivo continuous directed evolution. *Current opinion in chemical biology* **24C**, 1-10, (2014).

17 Wong, T. S., Zhurina, D. & Schwaneberg, U. The diversity challenge in directed protein evolution. *Combinatorial chemistry & high throughput screening* **9**, 271-288, (2006).

18 Greener, A., Callahan, M. & Jerpseth, B. An efficient random mutagenesis technique using an E. coli mutator strain. *Methods in molecular biology* **57**, 375-385, (1996).

19 Rasila, T. S., Pajunen, M. I. & Savilahti, H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. *Analytical biochemistry* **388**, 71-80, (2009).

20 Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9727-9732, (2003).

21 Schaaper, R. M. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. *The Journal of biological chemistry* **268**, 23762-23765, (1993).

22 Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503, (2011).

23 Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature chemical biology* **10**, 216-222, (2014).

24 Wong, T. S., Roccatano, D., Zacharias, M. & Schwaneberg, U. A statistical analysis of random mutagenesis methods used for directed protein evolution. *Journal of molecular biology* **355**, 858-871, (2006).

25 Kang, S., Lee, H., Han, J. S. & Hwang, D. S. Interaction of SeqA and Dam methylase on the hemimethylated origin of Escherichia coli chromosomal DNA replication. *The Journal of biological chemistry* **274**, 11463-11468, (1999).

26 Odsbu, I., Klungsoyr, H. K., Fossum, S. & Skarstad, K. Specific N-terminal interactions of the Escherichia coli SeqA protein are required to form multimers that restrain negative supercoils and form foci. *Genes to cells : devoted to molecular & cellular mechanisms* **10**, 1039-1049, (2005).

27 Schaaper, R. M. & Radman, M. The extreme mutator effect of Escherichia coli mutD5 results from saturation of mismatch repair by excessive DNA replication errors. *The EMBO journal* **8**, 3511-3516, (1989).

28 Lada, A. G. *et al.* Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry. Biokhimiia* **76**, 131-146, (2011).

29 Serrano-Heras, G. *et al.* Protein p56 from the Bacillus subtilis phage phi29 inhibits DNA-binding ability of uracil-DNA glycosylase. *Nucleic acids research* **35**, 5393-5401, (2007).

30 Gabrovsky, V., Yamamoto, M. L. & Miller, J. H. Mutator effects in Escherichia coli caused by the expression of specific foreign genes. *Journal of bacteriology* **187**, 5044-5048, (2005).

31 Mackie, A. *et al.* Addition of Escherichia coli K-12 growth observation and gene essentiality data to the EcoCyc database. *Journal of bacteriology* **196**, 982-988, (2014).

32 Garibyan, L. *et al.* Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome. *DNA repair* **2**, 593-608, (2003).

33 Fijalkowska, I. J., Dunn, R. L. & Schaaper, R. M. Genetic requirements and mutational specificity of the Escherichia coli SOS mutator activity. *Journal of bacteriology* **179**, 7435-7445, (1997).

34 Schaaper, R. M. & Dunn, R. L. Spectra of spontaneous mutations in Escherichia coli strains defective in mismatch correction: the nature of in vivo DNA replication errors. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 6220-6224, (1987).

35 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. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9007-9012, (2013).

36 Leconte, A. M. *et al.* A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* **52**, 1490-1499, (2013).

37 Dickinson, B. C., Packer, M. S., Badran, A. H. & Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nature communications* **5**, 5352, (2014).

38 Raskin, C. A., Diaz, G., Joho, K. & McAllister, W. T. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *Journal of molecular biology* **228**, 506-515, (1992).

39 Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898, (2009).

40 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239, (2013).

Chapter 4

Continuous evolution of *B. thuringiensis* toxins overcomes insect resistance

Parts adapted from: Badran *et al.*, *Nature* (2016) [**in press**]

4.1 – Introduction: The importance of *Bacillus thuringiensis* toxins

The expression of insecticidal toxins derived from *Bacillus thuringiensis* (Bt) has proven to be a valuable strategy for agricultural pest management[3]. Bt toxins are multi-domain proteins that kill insects by selectively binding midgut cell proteins to trigger a series of events leading to cell permeabilization and death. Bt toxin-producing crops have been widely adopted in agriculture with substantial economic and environmental benefits[4], and have increased global agricultural productivity by an estimated $78 billion from 1996 to 2013[5]. Bt toxin resistance has rapidly evolved among insect pests both in the laboratory and in the field, and threatens the continued success of this important strategy for pest control[6]. While researchers have developed approaches to manage insect resistance to Bt toxins, such as using multiple Bt toxins that recognize different binding sites in the insect midgut, or preserving susceptible alleles in insect populations, the development of insect resistance to Bt toxins remains the most serious current threat to sustaining the gains offered by transgenic crops[6].

Bt toxins interact with protein receptors on the surface of insect midgut cells, leading to pore formation in the cell membrane and cell death[7]. Bt toxin resistance is commonly associated with the mutation, down-regulation, or deletion of these receptors[4]. Among the known Bt toxin receptors are the cadherin-like proteins[8], and mutations in the midgut cadherin gene are associated with high-level resistance in major agricultural pests including *H. virescens*[9], *P. gossypiella*[10], and *H. armigera*[11]. We hypothesized that it might be possible to overcome Bt toxin resistance by evolving novel Bt toxins that bind with high affinity to non-cognate receptors in insects. If successful, such an approach has the potential to alter toxin specificity, improve toxin potency, and bypass resistance

mechanisms in insects.

In this chapter we used phage-assisted continuous evolution (PACE) to rapidly evolve Bt toxins through > 500 generations of mutation, selection, and replication to bind a non-cognate receptor expressed on the surface of insect midgut cells. The PACE-derived Bt toxins bind the new receptor with high affinity and specificity, induce lysis of insect cells in a manner dependent on the expression of the new target receptor, and enhance the insecticidal activity against both sensitive and Bt-resistant insect larvae up to 335-fold. Collectively, these results establish an approach to overcoming Bt toxin resistance and also provide a new platform for the rapid evolution of other protein-binding biomolecules.

4.2 – Bt toxin target receptor design

Bt toxins exhibit potent insecticidal activity through disruption of the osmotic balance in insect midgut cells as a consequence of pore formation[7]. The events ultimately leading to cell death include a series of protein-protein binding events that are critical for Bt toxin activation and membrane insertion[7,8]. To illuminate the requirements for Bt toxin-mediated insecticidal activity, as well as to develop a strategy to overcome Bt toxin resistance, we sought to evolve Cry1Ac, the most widely used Bt toxin, to bind a non-cognate insect cell membrane receptor from *Trichoplusia ni* (cabbage looper) Among the insect pests that have developed Bt resistance in agricultural settings, *T. ni* has been widely studied, with high-level resistance known to occur following long-term toxin exposure[4].

Prior studies of Cry1Ac binding to cadherin-like receptor proteins from

lepidopteran insects identified highly conserved epitopes proximal to the plasma membrane that are required for toxin binding[19,20]. Sequence alignment of known Cry1Ac-binding receptors suggests a short ~50-amino acid motif, the toxin-binding region (TBR), in cadherin-like proteins that is necessary for Cry1Ac binding and insecticidal activity (**Figure 4.1a**). Importantly, Cry1Ac toxicity in *T. ni* is not mediated through binding a cadherin-like receptor, but instead is genetically associated with the *ABCC2* gene and down-regulation of the aminopeptidase *APN1*[21-23]. Further, Cry1Ac resistance in *T. ni* is not dependent on the *T. ni* cadherin-like receptor TnCAD[23], as we confirm below. We hypothesized that evolved Cry1Ac mutants capable of binding TnCAD with high affinity would create a new mechanism of action that might restore insecticidal activity against Cry1Ac-resistant *T. ni*.

The TBR of TnCAD differs from that of cadherin-like proteins from susceptible lepidopteran species at seven amino acid positions (**Figure 4.1a,b**). To create an evolutionary stepping-stone from cadherin-like proteins that bind Cry1Ac to TnCAD, three residues (F1433, S1436, and A1437) from the TBR of susceptible lepidopteran species[20,24-26] were introduced into TnCAD, resulting in an artificial receptor fragment designated TnTBR3 (**Figure 4.1a,b**). We constructed APs expressing various TnTBR3 fragments fused to 434cI and assessed transcriptional activation levels in the presence of various domains of Cry1Ac fused to RpoZ (**Figure 4.1c**). Only Cry1Ac containing all three domains of the active toxin (residues 1-609) showed binding activity for TnTBR3 fragments, with TnTBR3 fragment 3 (TnTBR3-F3) showing the greatest degree of transcriptional activation at ~8-fold (**Figure 4.1c**). To assess

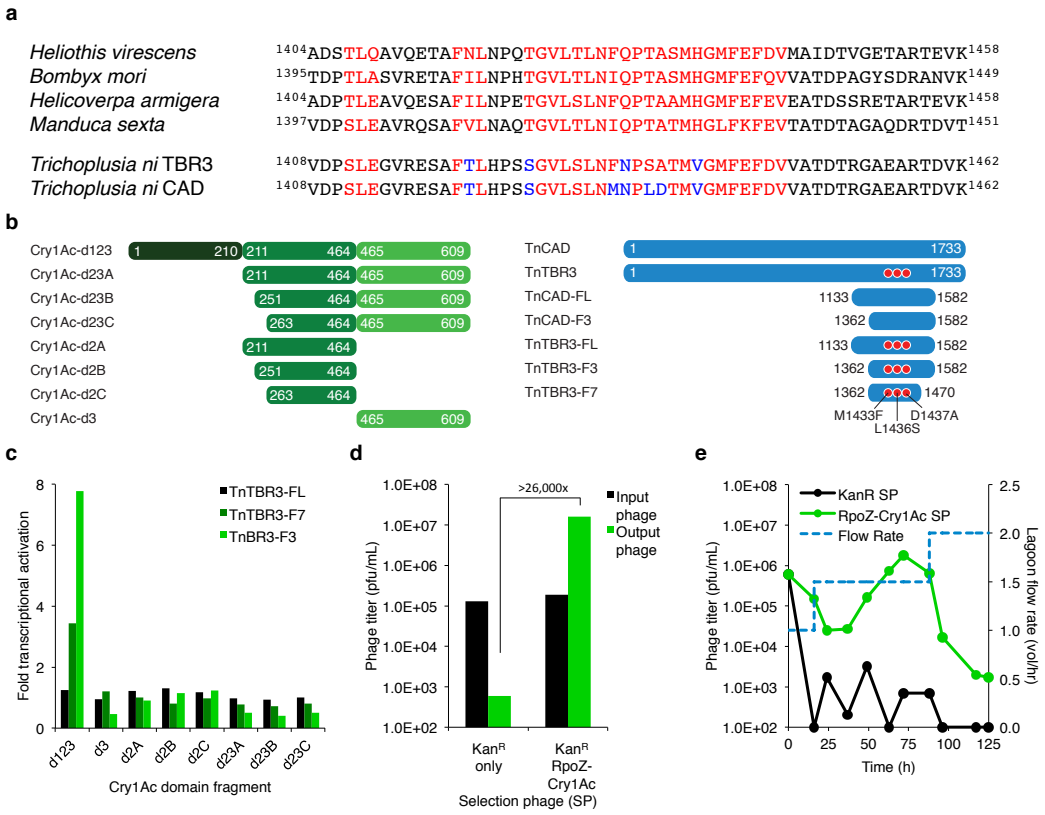


Figure 4.1 | Choice of Cry1Ac and TnTBR3 fragments used in PACE. a, Protein sequence alignment of known Cry1Ac-binding motifs from cadherin receptors in a number of lepidopteran species, as well as the cadherin receptor from *Trichoplusia ni* (TnCAD). The toxin-binding region (TBR; shown in red) of the known Cry1Ac-binding motifs differs from TnCAD at seven positions (shown in blue). Mutation of three residues in the TnCAD TBR (M1433F, L1436S, and D1437A) to resemble the corresponding positions of the cadherin-receptor TBRs yielded the evolutionary stepping-stone target TnTBR3. **b,** Schematic representations of the Cry1Ac and *T. ni* TBR3/CAD full-length receptors and fragments tested in this study. The red stars in the TnTBR3 variants represent the three mutations introduced into TnCAD to generate TnTBR3. **c,** Transcriptional activation assay using Cry1Ac and TnTBR3 fragments shows that the greatest degree of transcriptional activation resulted from full-length Cry1Ac together with TBR3 fragment 3 (TnTBR3-F3). RpoZ-Cry1Ac and 434cI-TnTBR3 fusions were used in all cases. **d,** Overnight phage enrichment assays using selection phages (SPs) that encode either kanamycin resistance (Kan^R) only or Kan^R together with RpoZ-Cry1Ac. Compared to the Kan^R-only SP, the RpoZ-Cry1Ac SP enriches > 26,000-fold overnight. **e,** Continuous propagation assays in the PACE format using either the Kan^R-only SP or the RpoZ-Cry1Ac SP show that the moderate affinity of Cry1Ac for TnTBR3 allows phage propagation at low flow rates (≤ 1.5 lagoon vol/h).

if this modest level of transcriptional activation was sufficient to support PACE, we constructed an SP carrying the *rpoZ-Cry1Ac* fusion gene and observed ~100-fold phage enrichment using a strain carrying the cognate TnTBR3-F3 AP after propagation overnight in a culture without continuous flow, whereas a control SP lacking the *rpoZ-Cry1Ac* fusion did not replicate (**Figure 4.1d,e**). These findings identified TnTBR3-F3 as a promising evolutionary stepping-stone to serve as a starting target for continuous evolution in PACE.

4.3 – Continuous evolution of Cry1Ac to bind TnCAD

Encouraged by these results, we performed 528 h of PACE on Cry1Ac in four segments while varying mutagenesis levels and selection stringency (**Figure 4.2**). For the first two segments (0–144 h, and 144–276 h), the AP in the host cells expressed the TnTBR3-F3 stepping-stone target fused to 434cI. For the final two segments of PACE (276–396 h, and 396–528 h), the AP expressed the TnCAD-F3 final target fused to 434cI. To enhance mutagenesis, we used our recently described MP4[2] mutagenesis plasmid during PACE for binding to TnTBR3-F3 (PACE segments 1 and 2). MP4 offers moderate mutagenesis rates that we speculated could help avoid creating early mutations that destroy the ability of Cry1Ac to kill insects, which is not explicitly selected during PACE. During the final two PACE segments for binding to TnCAD-F3 (PACE segments 3 and 4) we used MP6, which induces a greater mutation rate and broader mutational spectrum than MP4[2]. This change was motivated by consistent phage washout during TnCAD-F3 PACE attempts with MP4, suggesting that higher levels of mutagenesis were required to access rare Cry1Ac mutational combinations that conferred binding to the

final TnCAD-F3 target.

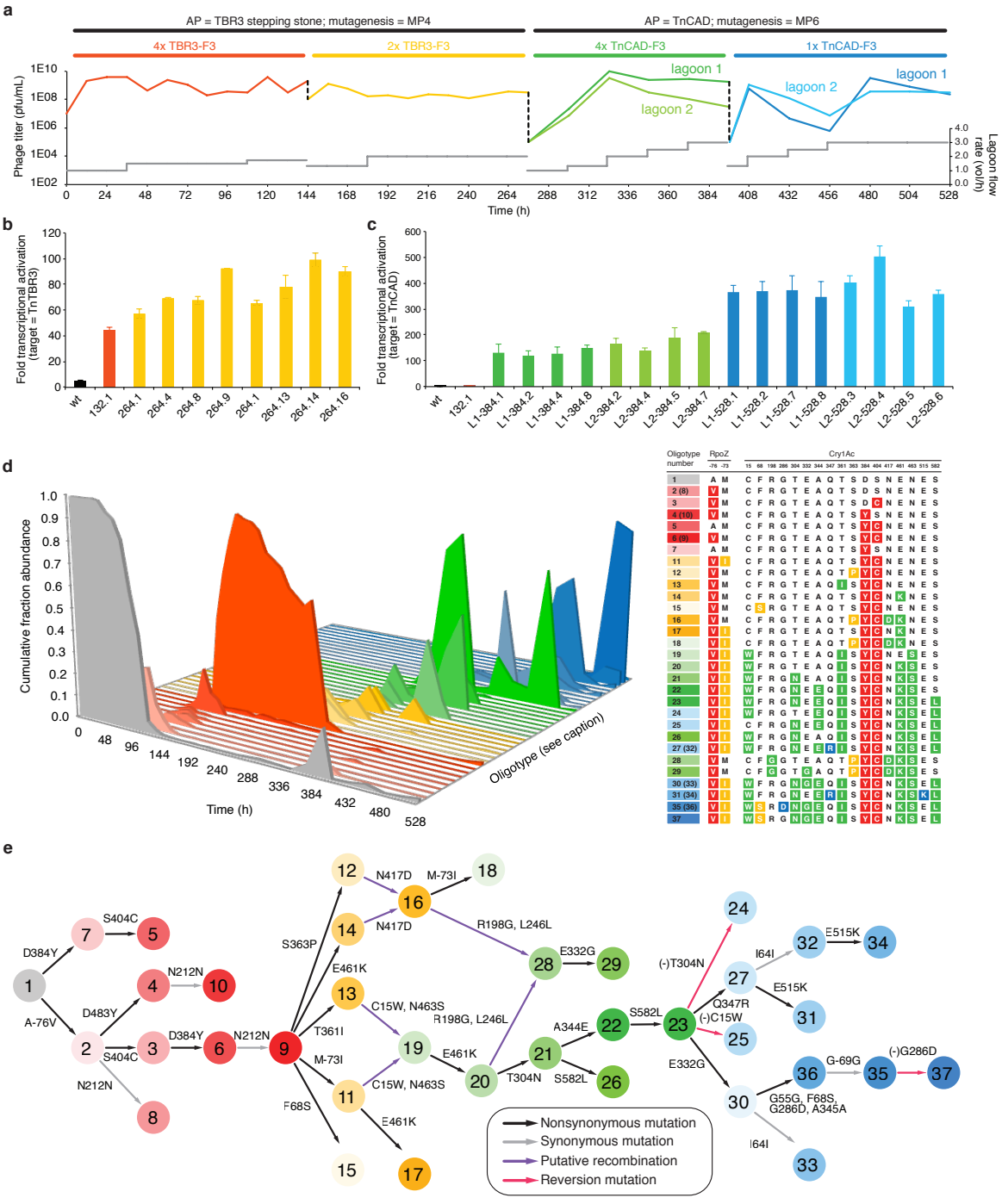


Figure 4.2 | Continuous evolution of Cry1Ac variants that bind the *Trichoplusia ni* cadherin receptor.

Figure 4.2 | (Continued) Continuous evolution of Cry1Ac variants that bind the *Trichoplusia ni* cadherin receptor. a, PACE was executed in four segments. The first two segments implemented the designed TnTBR3-F3 "stepping-stone" target under intermediate levels of mutagenesis (MP4). The final two segments implemented the final TnCAD-F3 target under high levels of mutagenesis (MP6). Phage titer (colored lines) and lagoon flow rate (grey lines) are shown at all sampled time points. The dotted lines indicate transfer of evolving phage to a new lagoon fed by the host cell culture corresponding to the next segment of PACE. **b, c,** Transcriptional activation assays using 434cI-TnTBR3-F3 (**b**) or 434cI-TnCAD-F3 (**c**) and individual RpoZ-Cry1Ac variants evolved during PACE, compared to wild-type RpoZ-Cry1Ac (wt). **d,** Oligotyping analysis of lagoon samples during PACE based on high-throughout DNA sequencing data. Oligotypes containing high frequency mutations (≥ 1%) are represented by different polygons, colored based on the stage in which they first became abundant in the evolving Cry1Ac gene pool. Mutations in Cry1Ac for each oligotype are shown in the table. Numbers in parentheses indicate the oligotype number assigned to that mutant following a synonymous (silent) mutation. **e,** Plausible evolution trajectories over the entire PACE experiment derived from oligotyping analysis strongly suggests instances of recombination during PACE, and also reveals the influence of mutation rate, selection stringency, and target protein on evolutionary outcomes. The colors and numbers in each circle correspond to those in (**d**).

We increased selection stringency during PACE by increasing lagoon flow rates and reducing the number of TnTBR3-F3 or TnCAD-F3 fragments participating in Cry1Ac variant recognition (**Figure 2.4, 2.5a**). At the end of every segment, we assayed the ability of evolved Cry1Ac clones to activate transcription in host *E. coli* cells expressing either the TnTBR3-F3 or the TnCAD-F3 fragment. Phage surviving 528 h of PACE experienced on average 511 generations of mutagenic replication under selection conditions[1].

Single-clone sequencing at the end of the first segment (144 h PACE on cells containing four copies of TnTBR3-F3 associated with each $P_{lacZ\text{-}opt}$ promoter) showed a strong consensus of two coding mutations in *Cry1Ac*, and one coding mutation in *rpoZ* (**Figure 4.3**). The three mutations together resulted in 11-fold higher transcriptional activation than that of the wild-type *rpoZ-Cry1Ac* fusion. At the end of the second

segment (276 h PACE with two copies of TnTBR3-F3 per $P_{lacZ\text{-}opt}$ promoter), even greater degrees of transcriptional activation were observed, up to 20-fold higher than the level resulting from the starting fusion protein (**Figure 4.2b, 4.3a**). At the end of third segment (396 h PACE with four copies of TnCAD-F3 per $P_{lacZ\text{-}opt}$

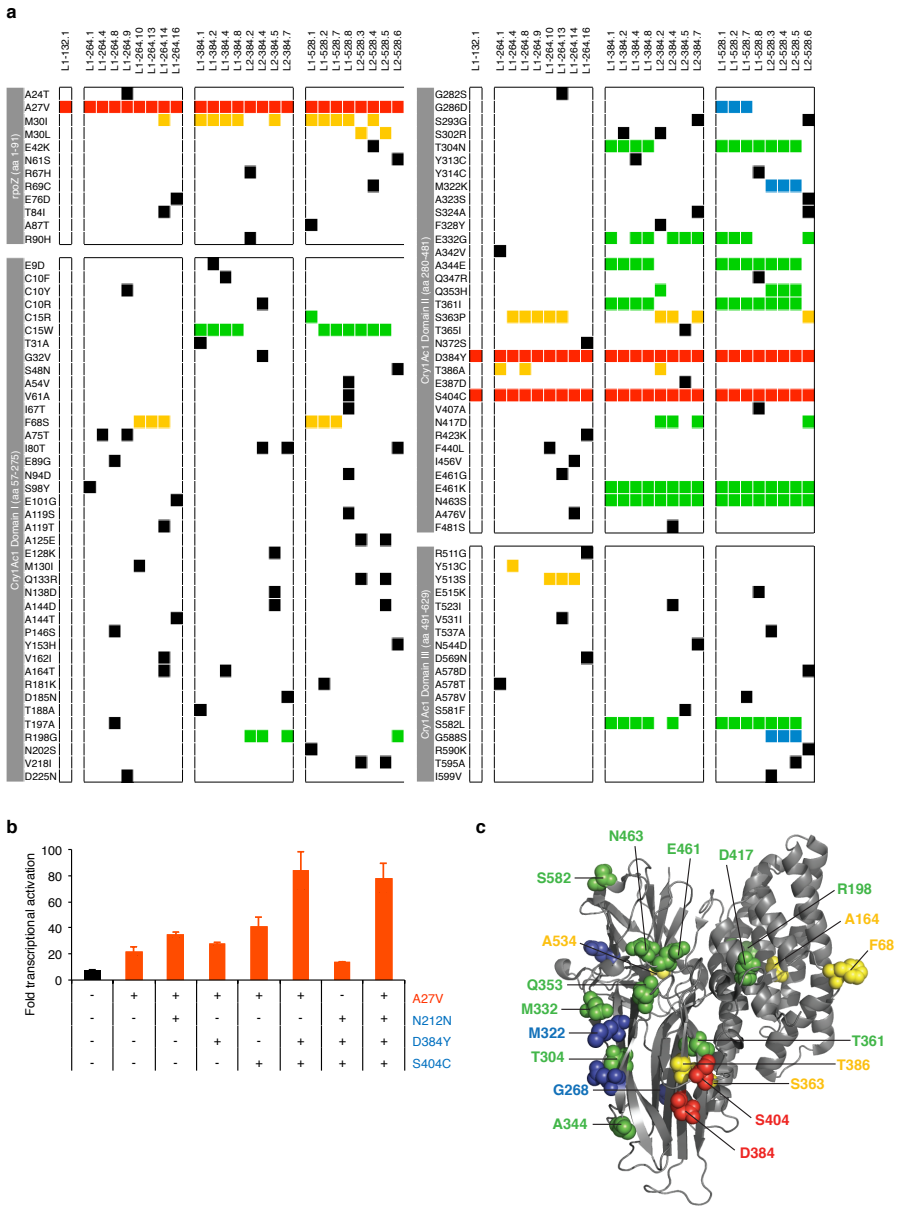


Figure 4.3 | Single-clone sequencing and evolved Cry1Ac characterization following PACE using the bacterial 2-hybrid luminescence reporter.

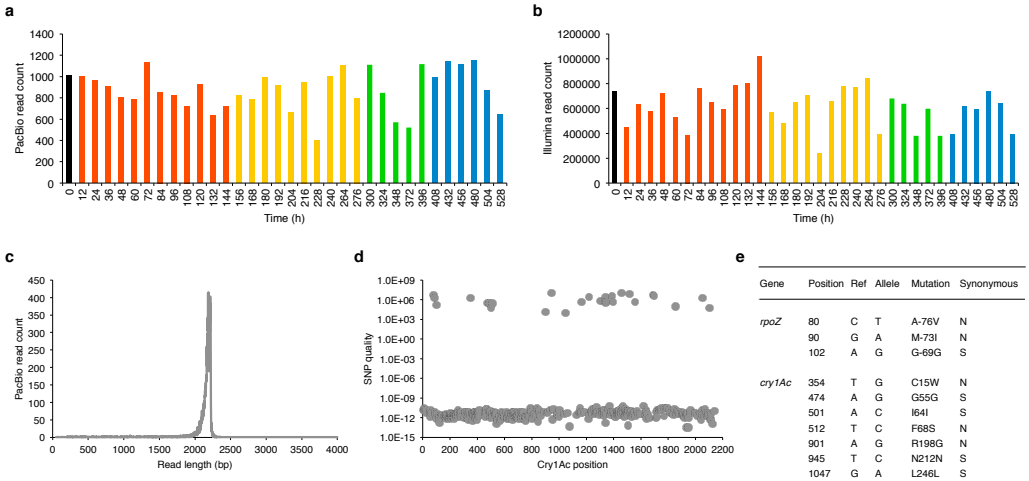
Figure 4.3 | (Continued) Single-clone sequencing and evolved Cry1Ac characterization following PACE using the bacterial 2-hybrid luminescence reporter. a, Coding mutations of the tested RpoZ-Cry1Ac clones at the end of each of the four segments of PACE. Consensus mutations are colored according to the segment in which they became highly enriched in the population (Fig. 3**a**). Mutations colored in black were observed at low abundance (≤ 5% of sequenced clones). **b,** Mutational dissection of the consensus mutations from the first segment of PACE reveals the requirement for both D384Y and S404C to achieve high-level transcriptional activation using the TnTBR3-F3 target. Mutations listed in red occurred in the RpoZ activation domain, whereas mutations listed in blue occurred in the Cry1Ac domain. Error bars reflect the standard deviation of at least three independent biological replicates. **c,** Structure of wild-type Cry1Ac (PDB: 4ARX) showing the positions of the evolved consensus mutations. The colors correspond to the PACE segments shown in Fig. 3 during which the mutations became highly abundant.

promoter), Cry1Ac variants evolved with greatly enhanced apparent affinity for TnCAD-F3 (**Figure 4.2c, 4.3a**). Whereas wild-type Cry1Ac could not detectably activate transcription when challenged to bind TnCAD-F3, single Cry1Ac variants emerging from a total of 384 h of PACE robustly activated transcription up to 210-fold above background in the absence of Cry1Ac variant expression. The end of the fourth segment (528 h PACE with one copy of TnCAD-F3 per $P_{lacZ\text{-}opt}$ promoter) yielded Cry1Ac mutants that could activate transcription when challenged to bind TnCAD-F3 by up to 500-fold (**Figure 4.2c, 4.3a**), suggesting strong binding to the TnCAD-F3 final target.

4.4 – Characterization of evolved Cry1Ac variants

DNA sequencing of individual clones surviving 528 h of PACE revealed a number of consensus genotypes carrying up to 16 mutations per clone out of 22 consensus mutations (C15W, F68S, A164S, R198G, G286D, T304N, M322K, E332G, A344E, Q353H, T361I, S363P, D384Y, T386A, S404C, N417D, E461K, N463S, A534V, E515K, A534V, and S582L), most of which localize to domain II, the predicted

cadherin-binding domain of Cry1Ac (**Figure 4.1, 4.3**). To illuminate the evolution trajectories en route to TnCAD-F3 binding activity, we analyzed all lagoon samples, collected every 12-24 h, by high-throughput DNA sequencing using both shorter-read (Illumina) and longer-read (Pacific Biosciences) methods (**Figure 4.4**). These efforts identified 25 mutations commonly occurring over the 528 h of PACE (**Figure 4.1, 4.3**). Oligotyping analysis[27] of the long-read data revealed plausible evolutionary trajectories over the entire course of the experiment (**Figure 4.2d,e**). While PACE does not explicitly promote recombination as a mechanism of gene diversification, we observed multiple



e

Gene	Position	Ref	Allele	Mutation	Synonymous
rpoZ	80	C	T	A-76V	N
	90	G	A	M-73I	N
	102	A	G	G-69G	S
cry1Ac	354	T	G	C15W	N
	474	A	G	G55G	S
	501	A	C	I64I	S
	512	T	C	F68S	N
	901	A	G	R198G	N
	945	T	C	N212N	S
	1047	G	A	L246L	S
	1166	G	A	G286D	N
	1220	C	A	T304N	N
	1304	A	G	E332G	N
	1340	C	A	A344E	N
	1344	T	C	A345A	S
	1349	A	G	Q347R	N
	1391	C	T	T361I	N
	1396	T	C	S363P	N
	1459	G	T	D384Y	N
	1519	A	T	S404C	N
	1558	A	G	N417D	N
	1690	G	A	E461K	N
	1697	A	G	N463S	N
	1852	G	A	E515K	N
	2054	C	T	S582L	N

Figure 4.4 | High-throughput DNA sequencing of PACE Cry1Ac selection phage libraries. The number of reads mapped to the wt *rpoZ-Cry1Ac* reference sequence using (**a**) Pacific Biosciences (PacBio) or (**b**) Illumina sequencing. Time points are colored according to the corresponding segment of the PACE experiment (Fig. 3A). **c,** In general, the majority of PacBio reads aligned to the wt *rpoZ-Cry1Ac* reference sequence were found to cluster around ~2,200 base pairs, corresponding to the size of the full-length fusion gene and indicating high-quality sequencing reads. **d,** Illumina high-throughput sequencing yielded a number of high quality SNPs across all time points. The corresponding mutations are shown in (**e**).

putative recombination events during the course of Cry1Ac evolution (**Figure 4.2e**). These recombination events, which we presume arose from multiple phage occasionally infecting the same host cell, yielded seminal, highly functional new variants.

Based on our mutational analysis, we designed and synthesized consensus Cry1Ac variants containing the most commonly observed mutations (**Figure 4.5a**). Purified activated Cry1Ac variants encoding PACE-derived consensus mutations bind very strongly to a TnCAD fragment containing the TBR (TnCAD-FL; **Figure 4.1b**) by ForteBio bio-layer interferometry analysis (K_d = 11-41 nM), with evolved

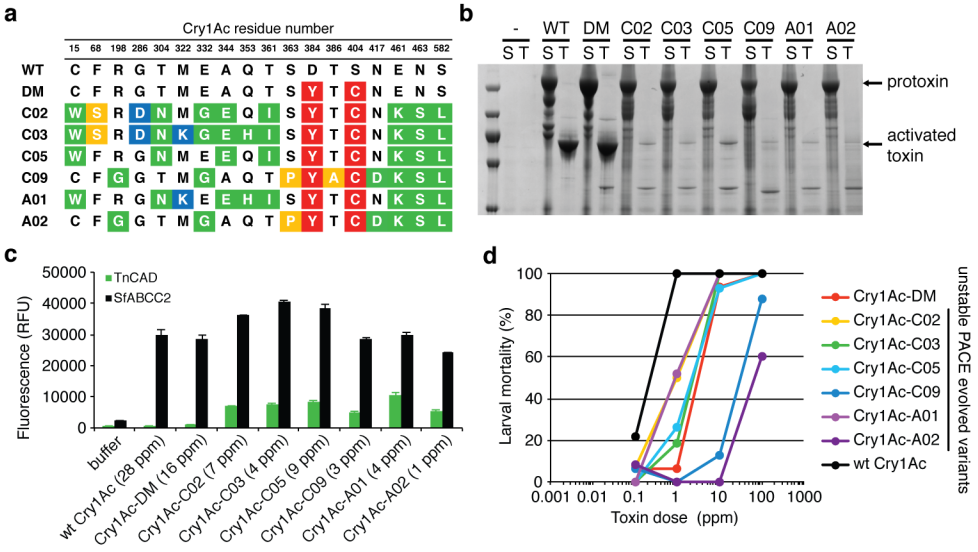


Figure 4.5 | Characterization of consensus evolved Cry1Ac variants. a, Consensus evolved Cry1Ac mutant sequences, including the D384Y/S404C double mutant (DM) that enabled TnTBR3-F3 recognition during the first segment of PACE. **b,** SDS-PAGE analysis of Cry1Ac variants following trypsin digestion, revealing proteolytic instability of consensus evolved variants. S = solubilized crude Bt crystals; T = trypsin-treated. **c,** Toxicity assays using Sf9 cells expressing the ABCC2 (black) or TnCAD receptor (green). Cells were incubated with Cry1Ac variants following trypsin digestion at the concentrations of activated toxin shown. Cry1Ac-induced cell permeabilization causes a fluorescent dye to enter cells, resulting in an increase in fluorescence. The evolved Cry1Ac variants, but not wild-type Cry1Ac, induce permeabilization of cells expressing TnCAD. Error bars reflect the standard deviation of at least three independent biological replicates. **d,** Insect larvae diet bioassays using wild-type and evolved consensus Cry1Ac variants, showing the loss of evolved Cry1Ac potency in insect larvae arising from impaired stability.

Cry1Ac variants C03 and C05 exhibiting the highest binding affinity (**Figure 4.6**). In contrast, wild-type Cry1Ac exhibited no detectable affinity for TnCAD under the same conditions. These results together establish the ability of protein-binding PACE to rapidly evolve extensively mutated proteins with high target affinity.

Cry1Ac requires activation by proteases in the insect midgut[4]. The evolved consensus mutants, however, exhibited extensive proteolysis in the presence of trypsin under conditions in which the wild-type Cry1Ac was cleanly cleaved into its active form (**Figure 4.5b**). Thermal melting studies confirmed this reduced stability (consensus variants: T_m = ~45 °C; wild-type Cry1Ac: T_m = 71 °C; **Figure 4.6**). Despite this lower stability, trypsin-activated consensus variants robustly killed Sf9 cells expressing the TnCAD receptor, whereas wild-type Cry1Ac did not exhibit toxicity at the tested concentration (**Figure 4.5c**). Moreover, these consensus evolved Cry1Ac mutants showed insecticidal activity in *T. ni* larvae, although they were less potent than wild-type Cry1Ac (**Figure 4.5d**).

We hypothesized that a subset of the consensus mutations were impairing apparent toxin potency against insect larvae by decreasing Cry1Ac stability and thus promoting degradation in insect gut. We generated Cry1Ac variants containing reversions of the identified consensus mutations, with a focus on mutations that arose during early-stage evolution on the stepping-stone target (**Figure 4.2e**), and evaluated their ability to survive trypsin treatment and to kill Sf9 insect cells and *T. ni* insect larvae (**Figure 4.6**). Using this approach, we identified mutations D384Y and S404C, two mutations that arose early during PACE against the TBR3 stepping-stone target (**Figures 4.2d, 4.2e, 4.5a**), as the main source of reduced protein stability. Variants lacking these two

Construct Name		Amino acid																				Cell growth?	Stable to trypsin?	T_m @ pH 8.5		Crude insect cell-based assay (TnCAD)		Purified (10 ppm) insect cell-based assay	
	# mut	15	68	164	198	286	304	322	332	344	353	361	363	384	386	404	417	461	463	534	582			Crude	Purified	Average	Std. Dev.	TnCAD	SfABCC2
WT Cry1Ac	0	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	S	N	E	N	A	S	Y	Y	-	71	3197	3250	4370	172950
Cry1Ac1_WT_S582L	1	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	S	N	E	N	A	L	Y	Y	72	-	7206	1242	-	-
Cry1Ac1_WT_E461K_N463S	2	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	S	N	K	S	A	S	Y	Y	72	-	10877	4822	-	-
Cry1Ac1_WT_S404C_E461K_S582L	3	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	C	N	K	N	A	L	Y	Y	64	-	8315	1393	-	-
Cry1Ac1_WT_S404C_E461K_N463S	3	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	C	N	K	S	A	S	Y	Y	66	-	23032	312	-	-
Cry1Ac1_WT_D384Y_N463S_S582L	3	C	F	A	R	G	T	M	E	A	Q	T	S	Y	T	S	N	E	S	A	L	Y	Y	62	-	15449	5700	-	-
Cry1Ac1_WT_D384Y_E461K_S582L	3	C	F	A	R	G	T	M	E	A	Q	T	S	Y	T	S	N	K	N	A	L	Y	Y	62	-	5535	4068	-	-
Cry1Ac1_WT_D384Y_S404C	2	C	F	A	R	G	T	M	E	A	Q	T	S	Y	T	C	N	E	N	A	S	Y	N	48	49	-	-	4303	198886
Cry1Ac1_WT_D384Y_N463S	2	C	F	A	R	G	T	M	E	A	Q	T	S	Y	T	S	N	E	S	A	S	Y	Y	61	-	13097	3552	-	-
Cry1Ac1_WT_N463S_S582L	2	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	S	N	E	S	A	L	Y	Y	70	-	16132	952	-	-
Cry1Ac1_WT_D384Y_S582L	3	C	F	A	R	G	T	M	E	A	Q	T	S	Y	T	S	N	E	N	A	L	Y	N	62	-	3730	1503	-	-
Cry1Ac1_WT_S404C_N463S_S582L	3	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	C	N	E	S	A	L	Y	Y	64	-	12371	11966	-	-
Cry1Ac1_WT_D384Y_E461K	2	C	F	A	R	G	T	M	E	A	Q	T	S	Y	T	S	N	K	N	A	S	Y	Y	62	-	6465	1294	-	-
Cry1Ac1_WT_S404C_E461K	2	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	C	N	K	N	A	S	Y	Y	66	-	18011	2542	-	-
Cry1Ac1_WT_E461K_S582L	2	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	S	N	K	N	A	L	Y	Y	72	-	10195	1128	-	-
Cry1Ac1_WT_D384Y_E461K_N463S	3	C	F	A	R	G	T	M	E	A	Q	T	S	Y	T	S	N	K	S	A	S	Y	N	-	-	28813	5220	-	-
Cry1Ac1_WT_E461K_N463S_S582L	3	C	F	A	R	G	T	M	E	A	Q	T	S	D	T	S	N	K	S	A	L	Y	Y	68	-	21756	392	-	-
Cry1Ac1_WT_T361I	1	C	F	A	R	G	T	M	E	A	Q	I	S	D	T	S	N	E	N	A	S	Y	Y	68	-	15149	1727	-	-
Cry1Ac1_WT_T361I	1	C	F	A	R	G	T	M	E	A	Q	I	S	D	T	S	N	E	N	A	S					8121	1522	-	-
Cry1Ac1_WT_S404C_T361I_S582L	3	C	F	A	R	G	T	M	E	A	Q	I	S	D	T	C	N	E	N	A	L	Y	Y	62	-	5546	1629	-	-
Cry1Ac1_WT_S404C_T361I_N463S	3	C	F	A	R	G	T	M	E	A	Q	I	S	D	T	C	N	E	S	A	S	Y	Y	63	-	10177	2023	-	-
Cry1Ac1_WT_T361I_S582L	2	C	F	A	R	G	T	M	E	A	Q	I	S	D	T	S	N	E	N	A	L	Y	Y	68	-	6316	2631	-	-
Cry1Ac1_WT_D384Y_T361I_S582L	3	C	F	A	R	G	T	M	E	A	Q	I	S	Y	T	S	N	E	N	A	L	Y	Y	58	-	15817	6297	-	-
Cry1Ac1_WT_T361I_N463S	2	C	F	A	R	G	T	M	E	A	Q	I	S	D	T	S	N	E	S	A	S	Y	Y	68	-	10608	1445	-	-
Cry1Ac1_WT_D384Y_T361I_N463S	3	C	F	A	R	G	T	M	E	A	Q	I	S	Y	T	S	N	E	S	A	S	Y	Y	60	-	11332	3053	-	-
Cry1Ac1_WT_D384Y_S404C_T361I_N463S_S582L	5	C	F	A	R	G	T	M	E	A	Q	I	S	Y	T	C	N	E	S	A	L	Y	Y	51	-	9799	909	-	-
Cry1Ac1_WT_T361I_E461K_S582L	3	C	F	A	R	G	T	M	E	A	Q	I	S	D	T	S	N	K	N	A	L	Y	Y	68	67	35119	5795	10761	159851
Cry1Ac1_WT_T361I_E461K_N463S	3	C	F	A	R	G	T	M	E	A	Q	I	S	D	T	S	N	K	S	A	S	Y	Y	68	-	9695	2551	-	-
Cry1Ac1_WT_A344E	1	C	F	A	R	G	T	M	E	E	Q	T	S	D	T	S	N	E	N	A	S					6445	2659	-	-
Cry1Ac1_WT_A344E_E461K_S582L	3	C	F	A	R	G	T	M	E	E	Q	T	S	D	T	S	N	K	N	A	L	Y	Y	68	-	11950	495	-	-
Cry1Ac1_WT_A344E_E461K_N463S	3	C	F	A	R	G	T	M	E	E	Q	T	S	D	T	S	N	K	S	A	S	Y	Y	68	-	9773	1734	-	-
Cry1Ac1_WT_D384Y_A344E	2	C	F	A	R	G	T	M	E	E	Q	T	S	Y	T	S	N	E	N	A	S	Y	Y	52	-	12400	3208	-	-
Cry1Ac1_WT_S404C_A344E	2	C	F	A	R	G	T	M	E	E	Q	T	S	D	T	C	N	E	N	A	S	Y	Y	62	58	6692	1909	4977	235310
Cry1Ac1_WT_D384Y_A344E_S582L	3	C	F	A	R	G	T	M	E	E	Q	T	S	Y	T	S	N	E	N	A	L	N	-	-	-	1152	705	-	-
Cry1Ac1_WT_A344E_N463S	2	C	F	A	R	G	T	M	E	E	Q	T	S	D	T	S	N	E	S	A	S	Y	Y	67	64	7214	3289	5515	235626
Cry1Ac1_WT_D384Y_A344E_N463S	3	C	F	A	R	G	T	M	E	E	Q	T	S	Y	T	S	N	E	S	A	S	Y	Y	54	-	13121	4174	-	-
Cry1Ac1_WT_A344E_N463S_S582L	3	C	F	A	R	G	T	M	E	E	Q	T	S	D	T	S	N	E	S	A	L	Y	Y	67	64	10298	2115	7129	234448
Cry1Ac1_WT_D384Y_S404C_A344E_N463S_S582L	5	C	F	A	R	G	T	M	E	E	Q	T	S	Y	T	C	N	E	S	A	L	Y	N	-	-	19256	1657	-	-
Cry1Ac1_WT_S404C_A344E_E461K	3	C	F	A	R	G	T	M	E	E	Q	T	S	D	T	C	N	K	N	A	S	Y	Y	61	-	15248	3302	-	-
Cry1Ac1_WT_D384Y_S404C_A344E_E461K_S582L	5	C	F	A	R	G	T	M	E	E	Q	T	S	Y	T	C	N	K	N	A	L	Y	Y	60	-	24482	2772	-	-
Cry1Ac1_WT_A344E_T361I_S582L	3	C	F	A	R	G	T	M	E	E	Q	I	S	D	T	S	N	E	N	A	L	Y	Y	63	-	3014	110	-	-
Cry1Ac1_WT_A344E_T361I_N463S	3	C	F	A	R	G	T	M	E	E	Q	I	S	D	T	S	N	E	S	A	S	Y	Y	64	-	10784	949	-	-
Cry1Ac1_WT_A344E_T361I_E461K	3	C	F	A	R	G	T	M	E	E	Q	I	S	D	T	S	N	K	N	A	S	Y	Y	63	-	14922	3673	-	-
Cry1Ac1_WT_D384Y_A344E_T361I	3	C	F	A	R	G	T	M	E	E	Q	I	S	Y	T	S	N	E	N	A	S	Y	Y	51	-	13277	3595	-	-
Cry1Ac1_WT_S404C_A344E_T361I	3	C	F	A	R	G	T	M	E	E	Q	I	S	D	T	C	N	E	N	A	S	Y	Y	55	-	5968	1815	-	-
Cry1Ac1_WT_T304N_N463S_S582L	3	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	S	N	E	S	A	L	Y	Y	72	70	31456	2151	8184	158265
Cry1Ac1_WT_S404C_T304N_S582L	3	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	C	N	E	N	A	L	Y	Y	66	-	24062	4140	-	-
Cry1Ac1_WT_S404C_T304N_N463S	3	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	C	N	E	S	A	S	Y	Y	67	-	8061	253	-	-
Cry1Ac1_WT_T304N	1	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	S	N	E	N	A	S	Y	Y	70.5	68	4685	2817	7099	234335
Cry1Ac1_WT_D384Y_T304N	2	C	F	A	R	G	N	M	E	A	Q	T	S	Y	T	S	N	E	N	A	S	Y	Y	60	-	13788	3133	-	-
Cry1Ac1_WT_S404C_T304N	2	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	C	N	E	N	A	S	Y	Y	64	-	20864	6917	-	-
Cry1Ac1_WT_T304N_S582L	2	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	S	N	E	N	A	L	Y	Y	70	-	8725	5932	-	-
Cry1Ac1_WT_T304N_N463S	2	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	S	N	E	S	A	S	Y	Y	72	70	31701	59	11031	147276
Cry1Ac1_WT_D384Y_T304N_N463S	3	C	F	A	R	G	N	M	E	A	Q	T	S	Y	T	S	N	E	S	A	S	Y	N	53	-	3236	2465	-	-
Cry1Ac1_WT_D384Y_S404C_T304N_N463S_S582L	5	C	F	A	R	G	N	M	E	A	Q	T	S	Y	T	C	N	E	S	A	L	Y	Y	48	-	6815	2210	-	-
Cry1Ac1_WT_T304N_E461K	2	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	S	N	K	N	A	S	Y	Y	71	70	33874	3286	47427	158635
Cry1Ac1_WT_D384Y_T304N_E461K	3	C	F	A	R	G	N	M	E	A	Q	T	S	Y	T	S	N	K	N	A	S	Y	Y	60	-	5677	3566	-	-
Cry1Ac1_WT_T304N_E461K_N463S	3	C	F	A	R	G	N	M	E	A	Q	T	S	D	T	S	N	K	S	A	S	Y	Y	73	-	28167	2906	-	-
Cry1Ac1_WT_D384Y_S404C_T304N_E461K_N463S_S582L	6	C	F	A	R	G	N	M	E	A	Q	T	S	Y	T	C	N	K	S	A	L	Y	Y	51	-	8124	919	-	-
Cry1Ac1_WT_T304N_T361I	2	C	F	A	R	G	N	M	E	A	Q	I	S	D	T	S	N	E	N	A	S	Y	Y	68	-	11735	743	-	-
Cry1Ac1_WT_D384Y_S404C_T304N_T361I	4	C	F	A	R	G	N	M	E	A	Q	I	S	Y	T	C	N	E	N	A	S	Y	Y	48	50	34891	7503	9129	169831
Cry1Ac1_WT_D384Y_S404C_T304N_T361I_N463S	5	C	F	A	R	G	N	M	E	A	Q	I	S	Y	T	C	N	E	S	A	S	Y	Y	48	-	6624	789	-	-
Cry1Ac1_WT_D384Y_S404C_T304N_T361I_E461K_S582L	6	C	F	A	R	G	N	M	E	A	Q	I	S	Y	T	C	N	K	N	A	L	N	-	-	-	8802	605	-	-
Cry1Ac1_WT_D384Y_S404C_T304N_T361I_E461K_N463S	6	C	F	A	R	G	N	M	E	A	Q	I	S	Y	T	C	N	K	S	A	S	Y	Y	-	48	11389	949	35201	196720
Cry1Ac1_WT_T304N_A344E_S582L	3	C	F	A	R	G	N	M	E	E	Q	T	S	D	T	S	N	E	N	A	L	Y	Y	59	-	8132	656	-	-
Cry1Ac1_WT_T304N_A344E	2	C	F	A	R	G	N	M	E	E	Q	T	S	D	T	S	N	E	N	A	S	Y	Y	68	-	20168	1459	-	-

Table 4.1 | Mutations and *in vitro* properties of Cry1Ac variants derived from PACE.

Table 4.1 (Continued) | Mutations and *in vitro* properties of Cry1Ac variants derived from PACE.

mutations, but containing the other seven consensus C05 mutations, exhibited greatly improved, wild-type-like stability (T_m = ~60°C). Variants lacking D384Y and S404C also exhibited proteolytic resistance similar to that of wild-type Cry1Ac, while retaining high binding affinity to TnCAD-FL (up to K_d = 11 nM) (**Figures 4.6a, 4.6b, 4.5**).

We assayed the toxicity of two evolved consensus Cry1Ac variants (C05 and C03) and three stabilized evolved consensus Cry1Ac variants (C05s, C03s, and A01s) lacking D384Y and S404C to cultured Sf9 insect cells expressing an ABCC2 receptor (positive control) or TnCAD. The stabilized evolved Cry1Ac variants retain their ability to bind to the cognate ABCC2 receptor, while acquiring the ability to potently kill Sf9 cells expressing TnCAD, in contrast to the ability of wild-type Cry1Ac to only kill cells expressing the cognate ABCC2 receptor, but not cells expressing TnCAD (**Figure 4.6c**).

4.5 – *In vivo* activity of evolved Cry1Ac variants

Finally, we assayed the insecticidal activity of the stabilized evolved Cry1Ac variants against Cry1Ac-sensitive *T. ni* larvae when added to their diet. Consistent with the above *in vitro* results, the stabilized evolved Cry1Ac variants exhibited substantially increased toxicity to *T. ni* larvae compared with that of the consensus evolved Cry1Ac mutants prior to stabilization (**Figure 4.6d**). Interestingly, the stabilized evolved Cry1Ac variants also exhibited up to 4-fold more potent insecticidal activity against Cry1Ac-sensitive *T. ni* than that of wild-type Cry1Ac, suggesting that the evolved ability of the toxins to recognize a non-cognate receptor with high affinity may augment their insecticidal potency, even against insects susceptible to wild-type Cry1Ac. These results also suggest that the evolution of Bt toxins that recognize non-cognate or novel receptors

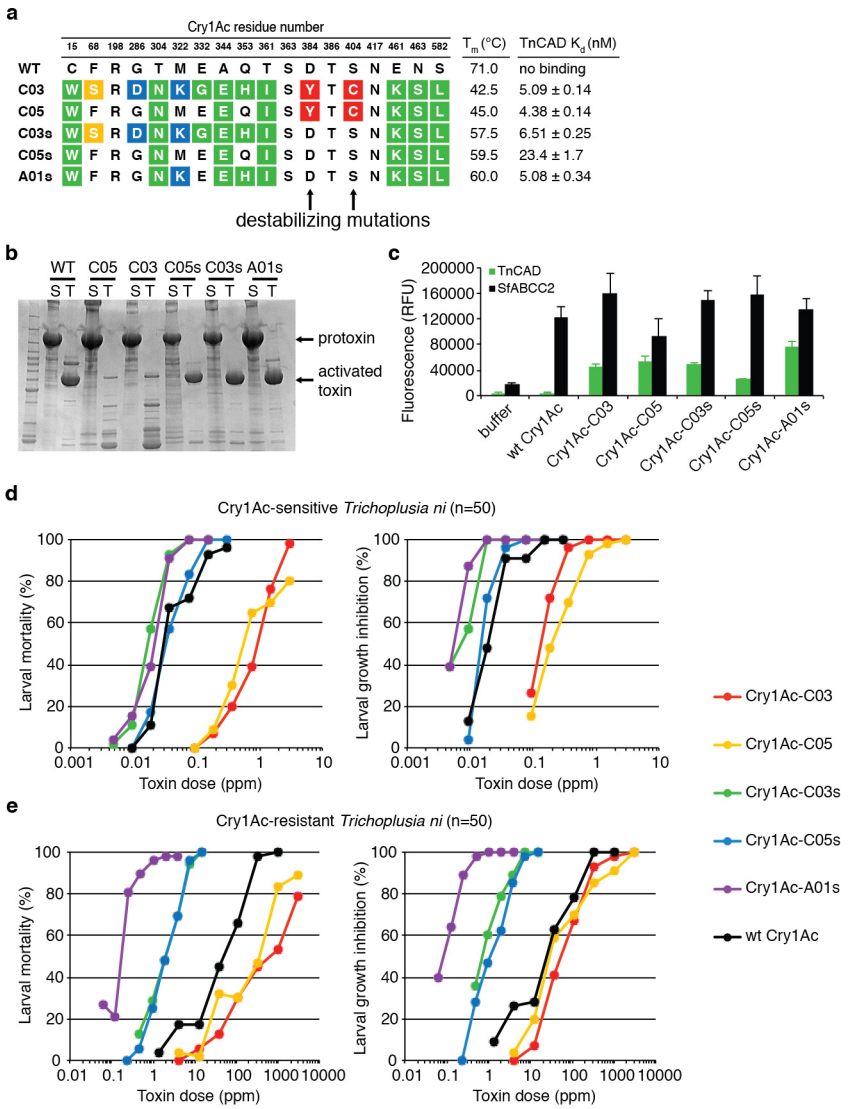


Figure 4.6 | Characterization of stabilized evolved Cry1Ac variants reveals potently enhanced activity. a, Sequence, thermal stability, and TnCAD target-binding affinity of unstable and stabilized PACE evolved consensus mutants. **b,** SDS-PAGE analysis of trypsin digestion reactions showing dramatically enhanced stability upon D384Y and S404C reversion. S = solubilized crude Bt crystals; T = trypsin-treated. **c,** Toxicity assays using Sf9 cells overexpressing the ABCC2 (black) or TnCAD receptor (green), demonstrating maintained activity of stabilized variants against both ABCC2 or TnCAD. All variants were used at 10 ppm. Error bars reflect the standard deviation of at least three independent biological replicates. **d, e,** Highly purified wild-type Cry1Ac, evolved consensus variants, or stabilized evolved variants were added to the diets of Cry1Ac-susceptible (**d**) or Cry1Ac-resistant (**e**) *T. ni* larvae at the indicated doses. Stabilized evolved variants moderately enhance mortality and growth inhibition in Cry1Ac-susceptible larvae compared to wild-type Cry1Ac. Stabilized evolved variants greatly outperform wild-type Cry1Ac toxin in killing and inhibiting the growth of Bt toxin-resistant *T. ni* larvae.

could expand the range of insects that can be targeted by Bt toxins, consistent with previous *in vitro* studies[28,29] using designed Bt toxin derivatives.

Next we evaluated the insecticidal activity of the stabilized evolved Cry1Ac variants against Cry1Ac-resistant *T. ni* larvae. *T. ni* resistance to Cry1Ac has been genetically mapped to the *ABCC2* transporter gene and down-regulation of expression of *APN1*[21,22], and is not dependent on alteration of the cadherin-like receptor[23]. Indeed, we observed ~1,000-fold lower potency of wild-type Cry1Ac against a Cry1Ac-resistant *T. ni* strain, compared to the potency of wild-type Cry1Ac against susceptible *T. ni*[30] (**Figure 4.6e**). Compared to wild-type Cry1Ac, stabilized evolved Cry1Ac variants C05s, C03s, and A01s showed dramatically improved activity against Cry1Ac-resistant *T. ni*, inducing larval mortality up to 335-fold more potently and larval growth inhibition up to 282-fold more potently than wild-type Cry1Ac (**Figure 4.6e**). Importantly, these evolved and stabilized Cry1Ac variants showed similar toxicity in Bt-resistant *T. ni* (as potent as LC_{50} = 0.15 ppm) as that of wild-type Cry1Ac in susceptible larvae (LC_{50} = 0.04 ppm) (**Figure 4.6e**). Taken together, these results establish that the evolution of novel receptor affinity among Bt toxins can provide a solution to overcome Bt toxin resistance in an agricultural pest.

To characterize the species profile of their insecticidal activity, we tested the evolved Cry1Ac variants in diet bioassays on 11 additional agricultural pests: a lepidopteran related to *T. ni* (*Chrysodeixis includes*, soybean looper) that encodes a cadherin-like receptor highly homologous to TnCAD, eight more distantly related lepidopteran pests (*Heliothis virescens*, tobacco budworm; *Helicoverpa zea*, corn earworm; *Plutella xylostella*, diamondback moth; *Agrotis ipsilon*, black cutworm;

Spodoptera frugiperda, fall armyworm*; Anticarsia gemmatalis*, velvetbean caterpillar; *Diatraea saccharalis*, sugarcane borer; and *Spodoptera eridania*, southern armyworm), and three non-lepidopteran pests (*Leptinotarsa decemlineata*, Colorado potato beetle; *Lygus lineolaris*, tarnished plant bug; and *Diabrotica virgifera virgifera*, western corn rootworm) (**Figures 4.7, 4.8**). As expected, the stabilized evolved Cry1Ac variants were up to 20-fold more potent than wild-type Cry1Ac against *C. includes*, and comparably potent as wild-type Cry1Ac against the other lepidopteran pests assayed (**Figure 4.7a**). Neither the evolved nor wild-type Cry1Ac exhibited insecticidal activity against the lepidopteran *S. eridania* or the three non-lepidopterans tested (**Figure 4.7b-h**). These results further support the mechanism of action of the PACE-evolved Bt toxins as binding to non-cognate cadherin receptor in *T. ni* and the related cadherin receptor in *C. includes*, while retaining binding to cognate receptors in all tested lepidopteran species. These data also reveal that the evolved Bt toxins did not acquire activity against species lacking a receptor homologous to TnCAD. Taken together, these findings demonstrate that an evolved Bt toxin that binds a novel target can potently kill closely related insect pest species, while maintaining a similar overall insect spectrum specificity as the parental Bt toxin.

4.6 – Discussion

Protein-binding PACE was able to rapidly discover variants of Cry1Ac that bind with high affinity to the non-cognate receptor TnCAD. Perhaps unsurprisingly, we observed a moderate reduction in stability of the evolved variants as compared to wild-type Cry1Ac, as this was not an implicit requirement of the selection. Interestingly, the

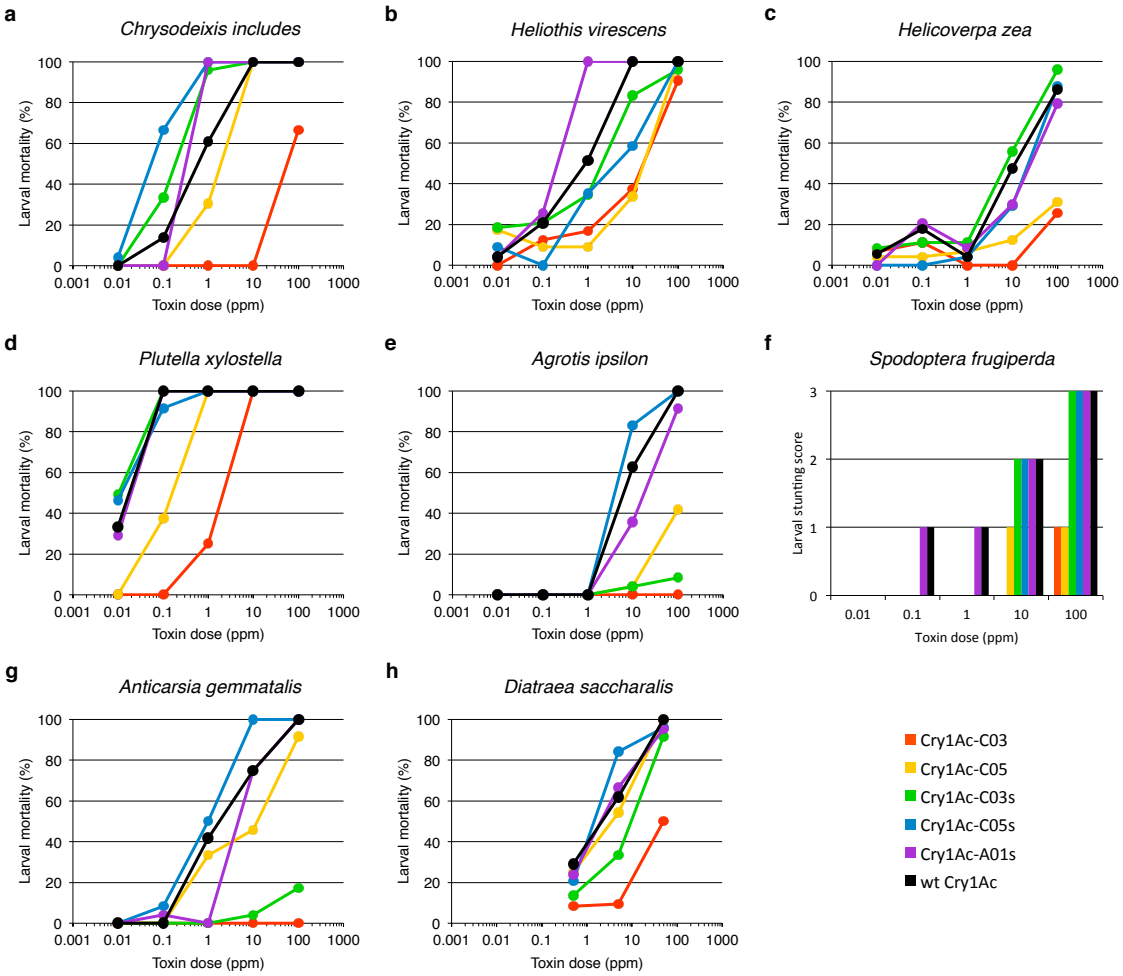


Figure 4.7 | Insect diet bioassay activity of PACE-evolved Cry1Ac variants against various agricultural pests. Two consensus and three stabilized PACE-evolved Cry1Ac variants were tested for activity in eleven pests: a, C*hrysodeixis includes* (soybean looper); b, *Heliothis virescens* (tobacco budworm); c, *Helicoverpa zea* (corn earworm); d, *Plutella* xylostella (diamondback moth); e, *Agrotis ipsilon* (black cutworm); f, *Spodoptera frugiperda* (fall armyworm); g, *Anticarsia gemmatalis* (velvetbean caterpillar); h, *Diatraea saccharalis* (sugarcane borer); *Spodoptera eridania* (southern armyworm); *Leptinotarsa decemlineata* (Colorado potato beetle); and *Lygus lineolaris* (tarnished plant bug). Stabilized variants showed enhanced activity in *C. includens* and *H. virescens* as compared to wild-type Cry1Ac, and comparable activity to wild-type Cry1Ac in *H. zea*, *P. xylostella*, *A. ipsilon*, *S. frugiperda*, *A. gemmatalis*, and *D. saccharalis*. No activity was observed for any of the Cry1Ac variants at any tested dose for *S. eridania*, *L. decemlineata* or *L. lineolaris*. No insect larvae mortality was observed for *S. frugiperda*, although high toxin doses greatly stunted growth.

two mutations that reduced Cry1Ac stability (D384Y and S404C) arose within the first few days of PACE on the stepping-stone target TnTBR3-F3 and were inherited by virtually all subsequent evolved variants (**Figure 4.2e**). It is tempting to speculate that these mutations broadened the substrate scope of Cry1Ac binding to enable downstream protein evolution, at the expense of stability, but were not required once affinity for the TnCAD-F3 receptor fragment evolved. Comparing the TnCAD binding affinities of Cry1Ac variants with and without D384Y and S404C suggests that they do not impair, and in some contexts may enhance, TnCAD binding *in vitro* and during PACE, two contexts in which their negative impact on stability likely does not limit potency (**Figure 4.6**).

Additional affinity measurements of reverted consensus mutations reveal the key roles of E461K and N463S to achieve high-affinity binding to TnCAD (**Figure 4.6**). Both of these mutations evolved in quick succession early during the third PACE

a

	T. ni	*C. includens*	*H. virescens*	*H. zea*	*P. xylostella*	*A. ipsilon*	*S. frugiperda*	*L. decemlineata*	*D. virgifera*
T. ni	-	78.6 (1362)	57.9 (1004)	57.3 (993)	53.1 (920)	51.8 (897)	51.2 (888)	25.6 (443)	26.1 (453)
C. includens	76.8 (1362)	-	55.7 (988)	54 (958)	51.8 (918)	49.9 (884)	49.8 (883)	24.6 (437)	25.5 (452)
H. virescens	58 (1004)	57 (988)	-	84.5 (1464)	55.4 (959)	58.9 (1021)	57.3 (992)	26 (451)	25.7 (445)
H. zea	57.4 (993)	55.4 (958)	84.6 (1464)	-	54.6 (945)	58.7 (1016)	57.3 (992)	26.3 (455)	25.4 (439)
P. xylostella	53.6 (920)	53.5 (918)	55.9 (959)	55 (945)	-	49.4 (848)	50.5 (867)	24.6 (422)	25.3 (435)
A. ipsilon	51 (897)	50.2 (884)	58 (1021)	57.7 (1016)	48.2 (848)	-	53.1 (935)	24 (422)	25 (440)
S. frugiperda	51.2 (888)	50.9 (883)	57.2 (992)	57.2 (992)	50 (867)	53.9 (935)	-	25.7 (445)	25.1 (436)
L. decemlineata	26.3 (443)	25.9 (437)	26.7 (451)	27 (455)	25 (422)	25 (422)	26.4 (445)	-	39.6 (668)
D. virgifera	26.8 (453)	26.8 (452)	26.4 (445)	26 (439)	25.8 (435)	26.1 (440)	25.8 (436)	39.6 (668)	-

b

	T. ni	*C. includens*	*H. virescens*	*H. zea*	*P. xylostella*	*A. ipsilon*	*S. frugiperda*	*L. decemlineata*	*D. virgifera*
T. ni	-	76.9 (170)	55.2 (122)	54.3 (120)	52 (115)	47.5 (105)	47.1 (104)	24 (53)	22.2 (49)
C. includens	75.2 (170)	-	51.3 (116)	51.3 (116)	47.8 (108)	42.5 (96)	42.5 (96)	22.6 (51)	22.6 (51)
H. virescens	55 (122)	52.3 (116)	-	81.5 (181)	50 (111)	49.5 (110)	47.7 (106)	23.4 (52)	19.8 (44)
H. zea	53.8 (120)	52 (116)	81.2 (181)	-	49.3 (110)	50.2 (112)	47.5 (106)	24.2 (54)	19.3 (43)
P. xylostella	52 (115)	48.9 (108)	50.2 (111)	49.8 (110)	-	43.9 (97)	48 (106)	22.2 (49)	19 (42)
A. ipsilon	47.3 (105)	43.2 (96)	49.5 (110)	50.5 (112)	43.7 (97)	-	42.8 (95)	18.5 (41)	18.9 (42)
S. frugiperda	46.8 (104)	43.2 (96)	47.7 (106)	47.7 (106)	47.7 (106)	42.8 (95)	-	24.8 (55)	23.4 (52)
L. decemlineata	22.3 (53)	21.4 (51)	21.8 (52)	22.7 (54)	20.6 (49)	17.2 (41)	23.1 (55)	-	38.2 (91)
D. virgifera	20.2 (49)	21 (51)	18.1 (44)	17.7 (43)	17.3 (42)	17.3 (42)	21.4 (52)	37.4 (91)	-

< 40%
< 70%
> 70%

Figure 4.8 | Comparison of cadherin receptor sequence identity. The % sequence identity using the full-length cadherin receptor (a) or fragment used for directed evolution experiments (b) for insects tested in Extended Data Figure 8. Numbers in parenthesis denote the number of identical amino acids between the two receptors. In general, mortality and stunting data from diet bioassays correlates with cadherin receptor sequence identity.

segment (**Figure 4.2e**), consistent with their contribution to TnCAD binding. Similarly, reversion analysis also shows that S582L can enhance affinity to TnCAD (**Figure 4.6**), in agreement with its evolution near the end of the third PACE segment and its predominance among subsequent Cry1Ac variants. Importantly, all three mutations lie on the same face of Cry1Ac (albeit in different domains), consistent with direct interaction with the cadherin receptor fragments.

Collectively, our findings establish for the first time that the laboratory evolution of novel or enhanced Bt toxin-receptor interactions can overcome insect resistance to Bt toxins. This strategy complements other approaches to limit the incidence of Bt toxin resistance. The "gene pyramiding" strategy for resistance management[31], for example, has been shown to be effective to delay resistance development in insect populations, but requires the availability of multiple effective toxins with distinctively different binding sites in target insects. The refuge strategy[6] is effective to slow the development of resistance, but necessitates that the resistance is a recessive trait, depends on a low initial frequency of resistance alleles in the field, requires compliance by growers, and delays but does not overcome resistance. The engineering of pre-oligomerized Bt toxins to reduce the reliance on cadherin receptor interaction[32] has also been shown to enhance toxicity in resistant insects, but elimination of the integral recognition event with cadherin greatly reduces the target specificity of the toxin, creating a risk to non-target organisms.

The approach presented herein enables targeting of Bt-resistant pests through the evolution of high-affinity receptor-binding Bt toxin variants designed to bind specific proteins in target insects. In principle, this strategy should be applicable to target a variety

of insect pests for which midgut cell-membrane proteins have been identified. While the evolution of insect resistance to an evolved Bt toxin is a likely possibility, this work has the potential to provide access to many new Bt toxins that individually or in combination could substantially extend the effectiveness of this important approach to pest control. We also note that the ability of protein-binding PACE to rapidly evolve novel protein-protein interactions as demonstrated in this study may prove useful in the discovery or improvement of protein therapeutics.

4.7 – Methods

General methods. PCR was performed using PfuTurbo Cx Hotstart DNA polymerase (Agilent Technologies), VeraSeq ULtra DNA polymerase (Enzymatics), or Phusion U Hot Start DNA Polymerase (Life Technologies). Water was purified using a MilliQ water purification system (Millipore). Plasmids and selection phages were constructed using USER cloning (New England Biolabs). Genes were either synthesized as bacterial codon-optimized gBlocks Gene Fragments (Integrated DNA Technologies) or amplified by PCR from native sources. *Cry1ac* was amplified by PCR from the *B. thuringiensis* strain *Bt*_B107284 and cloned into the *Bt* expression vector pMON101647 using Hot Fusion[33] to generate the expression plasmid pMON133051, which served as a template for amplifying *Cry1ac* fragments for constructing PACE vectors. The toxin-binding region from *T. ni* cadherin (A1133-T1582, AEA29692.10), referred to as TnCAD-FL, was synthesized using 45-60-mer oligonucleotides (Integrated DNA Technologies) by overlap extension PCR using KOD Hot Start DNA polymerase (EMD Millipore). The synthetic wild-type TnCAD-FL template was used to generate the TnTBR3-FL fragment via site-directed mutagenesis using the QuikChange II kit according to the manufacturers' instructions. (Agilent Technologies). DNA vector amplification was carried out using NEB Turbo or DH5a cells (New England Biolabs).

Electrocompetent strain preparation. The previously described strains S1030[14] and S2060[16] were used in all luciferase and plaque assays, as well as in PACE experiments. The glycerol stock of either strain was used to seed a 2-mL overnight culture using 2xYT media (United States Biological) supplemented with 10 μg/mL tetracycline (Sigma

Aldrich), 50 μg/mL streptomycin (Sigma Aldrich), 10 μg/mL fluconazole (TCI America), and 10 μg/mL amphotericin B (TCI America) in a 37 °C shaker at 230 rpm. The saturated culture was diluted 1,000-fold in 50 mL of the same supplemented media and grown under identical conditions until it reached mid log-phase (OD_{600} 0.5-0.8). Once the appropriate OD_{600} was reached, the cells were pelleted in a 50-mL conical tube (VWR) centrifuged at 10,000 *G* for 5 min at 4 °C. The supernatant was immediately decanted and the interior of the tube was wiped with a few Kimwipes (Kimberly-Clark) to remove residual media and salts. The cells were resuspended in 25 mL of pre-chilled, sterile filtered 10% glycerol in MilliQ purified water using a pipette to quickly break up the pellet. The cells were centrifuged and washed an additional three times. After the last centrifugation step, the interior of the tube was wiped with a few Kimwipes to remove residual glycerol solution. The pellet was resuspended in as little volume as possible, typically ~150 μL, and split into 10 μL aliquots for storage. Cells were flash frozen using a liquid N_2 bath, then quickly transferred to -80°C for extended storage. Electrocompetent S1030 or S2060 cells produced by this method typically yielded 10^7-10^8 colonies/μg plasmid DNA and enable the simultaneous electroporation of up to three plasmids carrying orthogonal origins of replication and antibiotic resistance cassettes to yield transformants containing all plasmids.

General USER cloning. All plasmid and phage materials were constructed via USER cloning[34]. Briefly, primers were designed to include a single internal deoxyuracil base 15-20 bases from the 5' end of the primer, specifying this region as the "USER junction". Criteria for design of the USER junction were: it should contain minimal secondary

structure, have 45 °C < T_m < 70 °C, and begin with a deoxyadenosine and end with a deoxythymine (to be replaced by deoxyuridine). The USER junction specifies the homology required for correct assembly. We note that PfuTurbo Cx Hotstart DNA polymerase (Agilent Technologies), VeraSeq ULtra DNA polymerase (Enzymatics), or Phusion U Hot Start DNA Polymerase (Life Technologies) are able to use primers carrying deoxyuracil bases, whereas some other polymerases undergo a phenomenon known as PCR poisoning and do not extend the primer.

All PCR products were purified using MinElute PCR Purification Kit (Qiagen) to 10 µL final volume and quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific). For assembly, PCR products carrying complementary USER junctions were mixed in an equimolar ratio (up to 1 pmol each) in a 10 µL reaction containing 15 units *Dpn*I (New England Biolabs), 0.75 units USER (Uracil-Specific Excision Reagent) enzyme (Endonuclease VIII and Uracil-DNA Glycosylase, NEB), 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 µg/ml BSA at pH 7.9 (1x CutSmart Buffer, New England Biolabs). The reactions were incubated at 37 °C for 45 min, followed by heating to 80 °C and slow cooling to 22 °C at 0.1 °C/s in a temperature-controlled block. The hybridized constructs were directly used for heat-shock transformation of chemically competent NEB Turbo *E. coli* cells according to the manufacturer's instructions. Transformants were selected on 1.8% agar-2xYT plates supplemented with the appropriate antibiotic(s).

For SP cloning, the hybridized constructs were purified using EconoSpin purification columns (Epoch Life Sciences), eluted using 25 µL 10% glycerol, and transformed into electrocompetent S2060 cells carrying the phage-responsive AP

pJC175e, which produces functional pIII in response to phage infection (this strain is henceforth referred to as S2208). Following recovery for 3-4 h at 37 °C using 2xYT (United States Biological) media, the culture was centrifuged and the supernatant was purified using a 0.22 μm PVDF Ultrafree centrifugal filter (Millipore). The supernatant was diluted serially in 100-fold increments and used in plaque assays using log-phase S2208 cells. Following overnight at 37 °C, single plaques were picked into 2xYT media and grown for 12-18 h in a 37 °C shaker at 230 rpm. The supernatant was purified again to yield clonal phage stocks. In all cases, cloned plasmids and phages were prepared using the TempliPhi 500 Amplification Kit (GE Life Sciences) according to the manufacturers protocol and verified by Sanger sequencing.

Plaque assays. S1030[14] or S2060[16] cells transformed with the AP of interest were grown in 2xYT (United States Biological) liquid media supplemented with the appropriate antibiotics to an OD_{600} of 0.6-0.9. The phage supernatant was diluted serially in three, 100-fold increments to yield four total samples (undiluted, 10^2-fold, 10^4-fold, and 10^6-fold diluted) to be used for infections. For each sample, 150 μL of cells were added to 10 μL of phage that had been filtered using a 0.22 μm PVDF Ultrafree centrifugal filter (Millipore). Within 1-2 min of infection, 1 mL of warm (~55°C) top agar (7 g/L bacteriological agar in 2xYT) was added to the phage/cell mixture, mixed by pipetting up and down once, and plated onto quartered plates that had been previously poured with 2 mL of bottom agar (18 g/L bacteriological agar in 2xYT) in each quadrant. The plates were then grown overnight at 37 °C before plaques could be observed.

Phage-Assisted Continuous Evolution (PACE). Host cell cultures, lagoons, media, and the PACE apparatus were as previously described[14]. Recombined selection phage harboring gene III (rSP) will poison a PACE experiment by outcompeting the evolving SP. We have noted that the likelihood of rSP occurrence in an SP stock increases with extended standing culture growth during the initial SP stock preparation. To reduce the likelihood of rSP formation, all SPs are repurified prior to any continuous evolution experiments. Briefly, SPs were plaqued on S2208 cells. A single plaque was picked into 2 mL 2xYT (United States Biological) supplemented with the appropriate antibiotics and grown until the culture reached mid log-phase (OD_{600} 0.5-0.8). The culture was centrifuged using a tabletop centrifuge for 2 min at 10,000 *G*, followed by supernatant filtration using a 0.22 μm PVDF Ultrafree centrifugal filter (Millipore). This short growth period routinely yields titers of 10^6-10^8 pfu/mL and was found to minimize the occurrence of rSP during PACE experiments.

To prepare the PACE strain, the AP and MP were co-transformed into electrocompetent S1030 cells (see above) and recovered using Davis rich media[14] (DRM) to ensure MP repression. Transformations were plated on 1.8% agar-2xYT containing 50 μg/mL carbenicillin, 40 μg/mL chloramphenicol, 10 μg/mL fluconazole, 10 μg/mL amphotericin B, 100 mM glucose (United States Biological) and grown for 12-18 h in a 37 °C incubator. Following overnight growth, four single colonies were picked and resuspended in DRM, then serially diluted and plated on 1.8% agar-2xYT containing 50 μg/mL carbenicillin, 40 μg/mL chloramphenicol, 10 μg/mL fluconazole, 10 μg/mL amphotericin B, and either 100 mM glucose or 100 mM arabinose (Gold Biotechnology) and grown for 12-18 h in a 37 °C incubator. Concomitant with this plating step, the

dilution series was used to inoculate liquid cultures in DRM supplemented with 50 μg/mL carbenicillin, 40 μg/mL chloramphenicol, 10 μg/mL tetracycline, 50 μg/mL streptomycin, 10 μg/mL fluconazole, 10 μg/mL amphotericin B and grown for 12-18 h in a 37 °C shaker at 230 rpm. Following confirmation of arabinose sensitivity using the plate assay, cultures of the serially diluted colonies still in log-phase growth were used to seed a 25-mL starter culture for the PACE chemostat.

Once the starter culture had reached log-phase density (OD_{600} 0.5-0.8), the 25-mL culture was added directly to 175 mL of fresh DRM in the chemostat. The chemostat culture was maintained at 200 mL and grown at a dilution rate of 1.5-1.6 vol/hr as previously described[14]. Lagoons flowing from the chemostats were maintained at 40 mL, and diluted as described for each experiment. Lagoons were supplemented with 25 mM arabinose to induce the MP for 8-16 h prior to infection with packaged SP. Samples were taken at the indicated time points, centrifuged at 10,000 *G* for 2 min, then filtered with a 0.2 μm filter and stored overnight at 4°C. Phage aliquots were titered by plaque assay on S2208 cells (total phage titer) and S1030 or S2060 cells (rSP titer) for all time points.

Mutagenesis during PACE. The basal mutation rate of replicating filamentous phage in *E. coli* (7.2×10^{-7} substitutions/bp/generation) is sufficient to generate all possible single but not double mutants of a given gene in a 40-mL lagoon following one generation of phage replication. For the 2,139 bp *rpoZ-Cry1Ac* target, a basal mutation rate of 7.2×10^{-7} substitutions/bp/generation applied to 2×10^{10} copies of the gene (a single generation) in a 40-mL lagoon yields ~3.1×10^{7} base substitutions, easily enough to cover all 6,417 single point mutants but not all double mutants. Arabinose induction of our first generation

mutagenesis plasmid (MP), MP1, increased the phage mutation rate by ~100-fold, yielding $7.2x10^{-5}$ substitutions/bp/generation, yielding ~$3.1x10^{9}$ substitutions spread over $2x10^{10}$ copies of the gene after a single generation. This enhanced mutation rate is sufficient to cover all possible single mutants ($6.4x10^{3}$ possibilities) and double mutants ($4.1x10^{7}$ possibilities), but no triple mutants ($2.6x10^{11}$ possibilities) after a single phage generation. Our recent efforts to enhance mutagenesis in PACE yielded the improved MP6 system[2], which increases the phage mutation rate by an additional 100-fold as compared to MP1, resulting in $7.2x10^{-3}$ substitutions/bp/generation, yielding ~$3.1x10^{11}$ substitutions spread over $2x10^{10}$ copies of the gene after a single generation. This elevated mutation rate is sufficient to cover all possible single mutants ($6.4x10^{3}$ possibilities), double mutants ($4.1x10^{7}$ possibilities), and many triple mutants ($2.6x10^{11}$ possibilities) after a single phage generation.

Luciferase assays. Complementary plasmids (CPs) were co-transformed with an AP of interest into electrocompetent S1030[14] or S2060[16] cells and plated onto 1.8% agar-2xYT plates with 50 μg/mL carbenicillin and 100 μg/mL spectinomycin. After overnight growth at 37 °C, single colonies were each picked into 2 mL DRM supplemented with 50 μg/mL carbenicillin, 100 μg/mL spectinomycin, 10 μg/mL tetracycline, 50 μg/mL streptomycin, 10 μg/mL fluconazole, 10 μg/mL amphotericin B and grown for 12-18 h in a 37 °C shaker at 230 rpm. Following overnight growth, cultures were diluted 1000-fold in a 96-well deep well plate containing 500 μL DRM with 50 μg/mL carbenicillin, 100 μg/mL spectinomycin and the indicated arabinose, IPTG or anhydrotetracycline (ATc) concentration to induce protein expression from either the AP or CP. Constitutive APs

and CPs were used where no inducer concentration is given. After growth with shaking at 37 °C for 4-5 hours, 150 μL of each culture was transferred to a 96-well black wall, clear bottom plate (Costar), and the OD_{600} and luminescence for each well was measured on an Infinite M1000 Pro microplate reader (Tecan). The OD_{600} of a well containing only media was subtracted from all sample wells to obtain a corrected OD_{600} value for each well. The raw luminescence value for each well was then divided by that well's corrected OD_{600} value to obtain the luminescence value normalized to cell density. Each variant was assayed in at least biological triplicate, and the error bars shown reflect the standard deviations of the independent measurements.

High-throughput sequencing and oligotype analysis. Raw reads are deposited in the NCBI Sequence Read Archive with accession number PRJNA293870, and all custom scripts used in analysis are available at http://github.com/MonsantoCo/BadranEtAl2015. Illumina reads obtained from each time point were mapped to the SP055-rpoZ-cMyc-Cry1Ac1-d123 reference sequence using bowtie v2.1.0[35], and the resulting SAM files were combined into a single BAM file using samtools v0.1.19[36]. This BAM file was used as input to freebayes v0.9.21-12-g92eb53a[37] to call SNPs, using the command "freebayes --use-best-n-alleles 1 --pooled-continuous --use-reference-allele --theta 500000000 --min-alternate-fraction 0.01 --ploidy 1 --region SP055-rpoZ-cMyc-Cry1Ac1-d123:2833-4971." The analysis is encapsulated in the custom script "ill.callsnps.sh." PacBio polymerase reads were demultiplexed with RS_Resequencing_Barcode.1 workflow provided by PacBio. Polymerase reads with quality score lower than 0.80 (defined by the PacBio scoring algorithm) or shorter than 50bp were filtered. High quality reads were processed

into subreads after sequencing primers and adaptors were removed. Circular consensus reads (or reads-of-inserts) were obtained by calling consensus of subreads generated from the same polymerase reads. These circular consensus reads were mapped to the SP055-rpoZ-cMyc-Cry1Ac1-d123 reference sequence using BLASR v1.3.1.142244[38], and the alignment was exported as an aligned FASTA sequence using the custom script "SAMtoAFA.py." The aligned FASTA was used as input to the oligotyping platform[27], manually specifying entropy components as the positions at which the Illumina data defined informative SNPs. Only oligotypes that occur at >1% in at least one sample were retained. This methodology resulted in informative changes at 25 of the 27 specified components. Oligotypes with gaps at the specified components, likely due to indels in the PacBio sequencing or alignment, were reassigned to other oligotypes with nucleotides in those positions only when it could be done unambiguously, and discarded otherwise, resulting in a total fraction abundance < 1 in Fig. 5D. The resulting oligotype-percent abundance matrix was read into R and analyzed using the custom script "PedigreeAndMullerPlot.R." The pedigree was refined manually, assuming that single-mutant derivatives of previous oligotypes are due to *de novo* mutation, while double, triple, or greater mutations that can be explained by recombination of previously observed oligotypes was due to recombination, since the latter are highly unlikely to arise by multiple point mutation after the start of the PACE experiment.

High-throughput primary *Bt* toxin preparation and analysis. Wild-type *Cry1Ac* was cloned into the *Bt* expression vector pMON262346 using BspQ1 endonuclease restriction sites. Consensus PACE-evolved *Cry1Ac* variants were synthesized (Gen9) and cloned

into the *Bt* expression vector pMON262346 using Hot Fusion[33]. Reversion mutants of consensus Cry1Ac PACE variants were generated via PCR with Phusion High-Fidelity DNA polymerase (New England Biolabs) and mutant primers followed by Hot Fusion into the *Bt* expression vector pMON262346. The resulting plasmids were transformed into the protease-deficient *Bt* strain EG10650[39] for protein expression. Cells were grown from single colonies in 96-well plates (Thermo Scientific, AB-0932) overnight in 400 µL Brain Heart Infusion Glycerol (BHIG) media (VWR) supplemented with 5 µg/mL chloramphenicol. Overnight cultures were used to prepare glycerol stocks (15% glycerol final concentration) and stored at -80 °C for future protein expression. Following overnight growth, 10 µL of each culture was used to inoculate 1 mL of complete *C2* medium[40] containing 5 µg/mL chloramphenicol in 96-well plates. The plates were incubated at 26 °C with vigorous shaking at 550 rpm in a Multitron shaking incubator (Infors HT) for 72 hr. The cells were harvested by centrifugation at 3,200 *G* for 15 min at 4 °C. The supernatant was decanted and a single 3.5 mm glass bead was added to each well of the plate. The pellet was then resuspended in 1 ml of TX wash buffer composed of 10 mM Tris-HCl, pH 7.5, 0.005% Triton X-100 supplemented with 25 units/mL Benzonase® (EMD Millipore), and 2 mM $MgCl_2$, incubated at room temperature for 30-60 min (with vigorous vortexing every 10 min), then centrifuged at 3,200 *G* for 15 min at 4 °C. The resulting pellet was resuspended and centrifuged under identical conditions two additional times.

The washed spore/crystal pellet from each 1-mL culture was solubilized in the 96-well plate using 300 µL of solubilization buffer composed of 50 mM CAPS, pH 11, and 10 mM DTT, then incubated while shaking at room temperature for 1 h. The insoluble

debris was pelleted by centrifugation at 3,200 *G* for 15 min at 4 °C, and 200 µL of the supernatant were transferred to a sterile U-bottom 96-well plate. To each well, 10 µL of 0.2 mg/mL trypsin in 1 M Tris-HCl, pH 7.5 was added. The mixture was incubated at 37°C for 2 h while shaking at 150 rpm, followed by quenching using 2 µL 0.1 M PMSF. The solution was filtered using a Millipore multiscreen plate with a 0.22 µm membrane. Protein stability was assessed by SDS-PAGE and quantified using spot densitometry. Proteins purified using this protocol were tested in downstream insect cell assays.

Secondary *Bt* toxin purification and analysis. *Bt* glycerol stocks described above were used for large-scale protein expression and purification. A 2-mL starter culture of BHIG medium supplemented with 5 µg/mL chloramphenicol was inoculated from the glycerol stocks and grown overnight at 280 rpm in a 28°C shaker. The following day, the saturated culture was transferred into 500 mL complete C2 medium containing 5 µg/mL chloramphenicol in a 2 L baffled flask and grown for an additional 72 h at 26°C while shaking at 280 rpm. Sporulation and crystal formation in the culture was verified by optical microscopy of a 2-µL aliquot of the saturated *Bt* culture. Upon confirmation of crystals, the partially lysed sporulated cells were harvested by centrifugation at 10,000 *G* for 12 min at 4°C. The pellet was then resuspended in 100 mL TX wash buffer composed of 10 mM Tris-HCl, pH 7.5, and 0.005% Triton X-100 supplemented with 0.1 mM PMSF, 25 units/mL Benzonase® (Sigma-Aldrich), and 2 mM $MgCl_2$, incubated at room temperature for 30-60 min (with vigorous vortexing every 10 min), then centrifuged at 3,200 *G* for 15 min at 4°C. The resulting pellet was resuspended and centrifuged under identical conditions two additional times.

The washed spore/crystal pellet was solubilized in 120 mL 50 mM CAPS, pH 11, 10 mM DTT at room temperature for 1 h while shaking at 130 rpm. The solubilized protein was separated from the insoluble debris by centrifugation at 35,000 *G* for 20 min at 4 °C. The supernatant was transferred to a fresh flask, and then supplemented with 10 mL 0.2 mg/mL Trypsin in 1 M Tris-HCl at pH 7.5. The mixture was incubated at 30 °C for 2-6 h with shaking at 150 rpm and trypsinization was monitored by SDS-PAGE. Once the trypsin digestion reaction was complete, the mixture was centrifuged at 3,200 *G* for 15 min at 4 °C. The clear supernatant was removed and mixed with PMSF to 1 mM final concentration. The sample was loaded on a 5-10 mL Q-Sepharose (GE Healthcare) anion exchange column at a flow-rate of 4 mL/min and the trypsin resistant core of the toxin was eluted in 25 mM sodium carbonate, pH 9 supplemented with 200-400 mM NaCl. Fractions containing the toxin tryptic cores were pooled, concentrated (Millipore Amicon Ultra-15 centrifugal filter Units, Fisher) and loaded on a Hiload Superdex 200 gel filtration column using an ÄKTA chromatography system (GE Healthcare, United Kingdom). The column was pre-equilibrated and run with 25 mM sodium carbonate at pH 10.5 supplemented with 1 mM β-mercaptoethanol. Only the monomer peak of the toxin fractions was collected in each case and concentrated to 1-3 mg/mL. The final protein concentration was quantified by spot densitometry. The quality of the trypsinized toxin was assessed using the peptide mass fingerprinting (PMF) method that was based on in-gel digestion of proteins by trypsin and mass spectrometry (MS) analysis of the resulted peptides.

***T. ni* receptor fragment expression and purification.** Custom expression vectors

pMON251427 and IS0008 (same as pMON251427 but with wild-type TnCAD) were used to express TnTBR3 and TnCAD fragments in *Escherichia coli.* Both vectors contain an N-terminal MBP-TVMV protease cleavage site tag[41] and a C-terminal 6x histidine tag flanking the receptor fragment of interest, with the ORF driven by the T7 promoter. Expression vectors were transformed into commercial BL21 (λDE3) competent cells (Life Technologies) that had been previously transformed with TVMV protease expression vector (pMON101695; encodes constitutive TVMV protease from a pACYC184 (New England Biolabs) backbone). A single colony was inoculated in 2 mL of LB media supplemented with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol, and grown at 37°C for 4 h to generate a starter culture, which was used to prepare glycerol stocks and stored at -80°C for the future protein expression. A second starter culture was inoculated using the BL21 (λDE3) strain glycerol stocks in 2 mL of LB media supplemented with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol and grown in a 25°C shaker (280 rpm) for 15 h. The culture was transferred into 500 mL of Terrific Broth medium (24g/L yeast extract, 12 g/L tryptone, and 5 g/L glucose) supplemented with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol, and grown at 37 °C for 4 h at 280 rpm, then transferred to 15 °C and grown for and additional 48 h after supplementation with IPTG to a final concentration of 0.1 mM.

The cells were harvested by centrifugation at 10,000 *G* for 12 min at 4 °C. The bacterial cell pellet was resuspended in affinity buffer A (25 mM Tris-HCl at pH 8.0, 0.5 M NaCl, 15 mM imidazole, and 0.2 mM $CaCl_2$) containing 125 units/mL of Benzonase (EMD Millipore), 10,000 units/mL of Chicken egg white lysozyme (Sigma Aldrich) and 1x BugBuster (Novagen). The cell slurry was incubated at room temperature for 15 min,

followed by sonication using Cell Disruptor W-0375 (Heat Systems-Ultrasonics) at 45% Duty Cycle (output No.5) for 30 seconds with 60 second rests for a total of three cycles. The cell lysate was centrifuged at 35,000 *G* for 20 min at 4ºC. The supernatant was loaded onto a 5-mL Ni-NTA column that had been pre-equilibrated using affinity buffer A. After extensive washing with affinity buffer A, the receptor fragment was eluted with the affinity buffer B (25 mM Tris-HCl at pH 8.0, 0.1 M NaCl, 250 mM imidazole, 0.2 mM $CaCl_2$). Fractions containing the receptor fragment were pooled, concentrated and loaded on a Hiload Superdex 200 gel filtration column using an ÄKTA chromatography system (GE Healthcare, United Kingdom). The column was pre-equilibrated and run with 25 mM Tris-HCl at pH 8.0, 0.1 M NaCl, 0.2 mM $CaCl_2$. Dimer and monomer peaks of the *T. ni* TBR3 and CAD fractions were collected separately and concentrated to 1-2 mg/mL. Only TnTBR3 and TnCAD monomers were used for Cry1Ac1 binding studies.

Fluorescence thermal shift (FTS) assays. All assays were performed using a BioRad CFX96 real-time PCR thermal cycler, enabling thermal manipulations and dye fluorescence detection. The fluorescence sensitive dye SYPRO orange (Life Technologies, S6650) was used at a 5x concentration in all assays. The temperature was increased by 0.5 °C each cycle over a temperature range of 25−90 °C. Assay reactions were performed in 96-well white PCR plates (Bio-Rad, No. HSP9631), and heat-sealed (Thermo Scientific, No. ALPS3000) to reduce volume loss through evaporation. The data was analyzed using the CFX manager software.

Protein–protein interaction affinity measurement. The Octet^{Qk} (ForteBio) and the Dip

and Read™ Ni-NTA (NTA) biosensors were used to measure the affinity of Cry1Ac and its variants to immobilized 6xHis-TnCAD or TnTBR3 receptor fragments in 25 mM Tris-HCl at pH 8.5, 0.1 M NaCl, 0.1 mg/ml BSA, 0.05% Tween 20 according to the manufacturer's instructions. Octet Data Acquisition 7.1.0.100 software was used for data acquisition, and ForteBio Data Analysis 7 software was used for data analysis. At least four readings at different Cry1Ac1 concentrations (2–100 nM) were used for each receptor fragment-Bt toxin interaction and a global fit was used to calculate binding affinities.

Insect cell assays. Sf9 cells (Life Technologies) were plated in Sf-900™ III SFM (Life Technologies) at a density of 50,000 cells/well in a 96-well optical bottom black plate (Nunc, Thermo Scientific). The cells were incubated at 27 °C overnight to allow for adherence to the plate. Following overnight incubation, the medium was aspirated from the cells and 100 μL of p3 or p4 generation (third or fourth generation of baculovirus amplification in Sf9 cells following initial transfection with plasmid) recombinant baculovirus encoding each receptor diluted in SFM was added to each well. The plates were kept in a humidified environment to prevent evaporation and incubated at 27 °C for 48 h. Receptor expression was confirmed by western blotting. Toxins were diluted to the same protein concentration in 25 mM sodium carbonate at pH 11, supplemented with 1 mM β-mercaptoethanol, followed by an additional 10-fold dilution in unsupplemented Grace's Media with 2 μM SYTOX Green Nucleic Acid Stain (Life Technologies, S7020). The media was removed from the wells without disturbing the attached cells, and the diluted toxins or buffer controls were added to respective wells. The fluorescence was

measured on a CLARIOstar microplate reader (BMG Labtech) after incubation for 4 h. The fluorescence intensity of control cells expressing β-glucuronidase (GUS) was subtracted from wells expressing the variable receptor fragments with or without toxins. Replicates were averaged and signal was plotted for each toxin condition.

Primary insect diet bioassays. Insect diet bioassays using the evolved consensus Cry1Ac variants were performed as previously described[42]. Briefly, 200 mL of artificial diet in 96-well plates were overlaid with 20 mL aliquots of toxin Bt spore/crystal or Bt crystal suspension, dried, after which wells were infested with neonate insect eggs suspended in 0.2% agar, dried again, sealed with Mylar sheets, and incubated at 20 °C, 60% RH, in complete darkness for 5 days. The plates were scored on day 5 for larval mortality and growth stunting. Each assay was carried out in three independent biological replicates with eight insects per replicate.

Secondary insect diet bioassays. An inbred Bt-susceptible laboratory strain of *T. ni*, (designated the Cornell strain)[43], and a Cry1Ac-resistant strain nearly isogenic to the Cornell strain, GLEN-Cry1Ac-BCS[44], were maintained on a wheat germ-based artificial diet at 27 °C with 50% humidity and a photoperiod of 16 h light and 8 h dark[43]. Diet surface overlay bioassays were conducted to determine the insecticidal activity of the toxins in the susceptible and Cry1Ac-resistant *T. ni*, as previously described[43]. Briefly, 200 μL of toxin solution was spread on the surface of 5 mL of artificial diet in 30-mL plastic rearing cups (diet surface area was ~7 cm^2), and 10 neonatal larvae were placed into each rearing cup after the toxin solution had dried. For each bioassay, 7-8

concentrations of the toxin were used and each treatment included five replicates (50 larvae total per concentration). Larval growth inhibition (neonates that did not reach 2nd instar after 4 days) and mortality were recorded after 4 days of feeding. The observed larval growth inhibition and mortality were corrected using Abbott's formula[45]. Both IC_{50} and LC_{50} values and their 95% confidence intervals were calculated by probit analysis using the computer program POLO (LeOra Softare, 1997).

4.8 – References

1 Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503, (2011).

2 Badran, A. H. & Liu, D. R. Development of Potent In Vivo Mutagenesis Plasmids with Broad Mutational Spectra. *Nat Communications* **6**, 8425, (2015).

3 Prado, J. R. *et al.* Genetically engineered crops: from idea to product. *Annual review of plant biology* **65**, 769-790, (2014).

4 Pardo-Lopez, L., Soberon, M. & Bravo, A. Bacillus thuringiensis insecticidal three-domain Cry toxins: mode of action, insect resistance and consequences for crop protection. *FEMS microbiology reviews* **37**, 3-22, (2013).

5 James, C. *Global Status of Commercialized Biotech/GM Crops: 2014. ISAAA Brief No, 49. ISAAA: Ithaca, NY*. (2014).

6 Tabashnik, B. E., Brevault, T. & Carriere, Y. Insect resistance to Bt crops: lessons from the first billion acres. *Nature biotechnology* **31**, 510-521, (2013).

7 Adang, M. J., Crickmore, N. & Jurat-Fuentes, J. L. Diversity of Bacillus thuringiensis Crystal Toxins and Mechanism of Action. *Adv Insect Physiol* **47**, 39-87, (2014).

8 Gomez, I. *et al.* Role of receptor interaction in the mode of action of insecticidal Cry and Cyt toxins produced by Bacillus thuringiensis. *Peptides* **28**, 169-173, (2007).

9 Gahan, L. J., Gould, F. & Heckel, D. G. Identification of a gene associated with Bt resistance in Heliothis virescens. *Science* **293**, 857-860, (2001).

10 Morin, S. *et al.* Three cadherin alleles associated with resistance to Bacillus thuringiensis in pink bollworm. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5004-5009, (2003).

11 Yang, Y., Chen, H., Wu, Y., Yang, Y. & Wu, S. Mutated cadherin alleles from a field population of Helicoverpa armigera confer resistance to Bacillus thuringiensis toxin Cry1Ac. *Applied and environmental microbiology* **73**, 6939-

6944, (2007).

12 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. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9007-9012, (2013).

13 Leconte, A. M. *et al.* A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* **52**, 1490-1499, (2013).

14 Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature chemical biology* **10**, 216-222, (2014).

15 Dickinson, B. C., Packer, M. S., Badran, A. H. & Liu, D. R. A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nature communications* **5**, 5352, (2014).

16 Hubbard, B. P. *et al.* Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat Methods*, (2015).

17 Dove, S. L. & Hochschild, A. Conversion of the omega subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes & development* **12**, 745-754, (1998).

18 Wojcik, J. *et al.* A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nature structural & molecular biology* **17**, 519-527, (2010).

19 Dorsch, J. A. *et al.* Cry1A toxins of Bacillus thuringiensis bind specifically to a region adjacent to the membrane-proximal extracellular domain of BT-R(1) in Manduca sexta: involvement of a cadherin in the entomopathogenicity of Bacillus thuringiensis. *Insect biochemistry and molecular biology* **32**, 1025-1036, (2002).

20 Xie, R. *et al.* Single amino acid mutations in the cadherin receptor from Heliothis virescens affect its toxin binding ability to Cry1A toxins. *The Journal of biological chemistry* **280**, 8416-8425, (2005).

21 Baxter, S. W. *et al.* Parallel evolution of Bacillus thuringiensis toxin resistance in

lepidoptera. *Genetics* **189**, 675-679, (2011).

22 Tiewsiri, K. & Wang, P. Differential alteration of two aminopeptidases N associated with resistance to Bacillus thuringiensis toxin Cry1Ac in cabbage looper. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14037-14042, (2011).

23 Zhang, X., Tiewsiri, K., Kain, W., Huang, L. & Wang, P. Resistance of Trichoplusia ni to Bacillus thuringiensis toxin Cry1Ac is independent of alteration of the cadherin-like receptor for Cry toxins. *PloS one* **7**, e35991, (2012).

24 Hua, G., Jurat-Fuentes, J. L. & Adang, M. J. Bt-R1a extracellular cadherin repeat 12 mediates Bacillus thuringiensis Cry1Ab binding and cytotoxicity. *The Journal of biological chemistry* **279**, 28051-28056, (2004).

25 Nagamatsu, Y., Koike, T., Sasaki, K., Yoshimoto, A. & Furukawa, Y. The cadherin-like protein is essential to specificity determination and cytotoxic action of the Bacillus thuringiensis insecticidal CryIAa toxin. *FEBS letters* **460**, 385-390, (1999).

26 Peng, D., Xu, X., Ye, W., Yu, Z. & Sun, M. Helicoverpa armigera cadherin fragment enhances Cry1Ac insecticidal activity by facilitating toxin-oligomer formation. *Applied microbiology and biotechnology* **85**, 1033-1040, (2010).

27 Eren, A. M. *et al.* Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in ecology and evolution / British Ecological Society* **4**, (2013).

28 Chougule, N. P. *et al.* Retargeting of the Bacillus thuringiensis toxin Cyt2Aa against hemipteran insect pests. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8465-8470, (2013).

29 Fujii, Y. *et al.* Affinity maturation of Cry1Aa toxin to the Bombyx mori cadherin-like receptor by directed evolution. *Molecular biotechnology* **54**, 888-899, (2013).

30 Song, X., Kain, W., Cassidy, D. & Wang, P. Resistance to Bacillus thuringiensis Toxin Cry2Ab in Trichoplusia ni Is Conferred by a Novel Genetic Mechanism. *Applied and environmental microbiology* **81**, 5184-5195, (2015).

31 Carriere, Y., Crickmore, N. & Tabashnik, B. E. Optimizing pyramided transgenic Bt crops for sustainable pest management. *Nature biotechnology* **33**, 161-168, (2015).

32 Tabashnik, B. E. *et al.* Efficacy of genetically modified Bt toxins against insects with different genetic mechanisms of resistance. *Nature biotechnology* **29**, 1128-1131, (2011).

33 Fu, C., Donovan, W. P., Shikapwashya-Hasser, O., Ye, X. & Cole, R. H. Hot Fusion: an efficient method to clone multiple DNA fragments as well as inverted repeats without ligase. *PloS one* **9**, e115318, (2014).

34 Lund, A. M. *et al.* A versatile system for USER cloning-based assembly of expression vectors for mammalian cell engineering. *PloS one* **9**, e96693, (2014).

35 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-U354, (2012).

36 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, (2009).

37 Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *ArXiv e-prints* **1207**, 3907 (2012). <http://adsabs.harvard.edu/abs/2012arXiv1207.3907G%3E.

38 Chaisson, M. J. & Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC bioinformatics* **13**, 238, (2012).

39 Tan, Y. & Donovan, W. P. Deletion of aprA and nprA genes for alkaline protease A and neutral protease A from bacillus thuringiensis: effect on insecticidal crystal proteins. *Journal of biotechnology* **84**, 67-72, (2001).

40 Donovan, W. P. *et al.* Amino acid sequence and entomocidal activity of the P2 crystal protein. An insect toxin from Bacillus thuringiensis var. kurstaki. *The Journal of biological chemistry* **263**, 561-567, (1988).

41 Nallamsetty, S. *et al.* Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein expression and*

purification **38**, 108-115, (2004).

42 Baum, J. A. *et al.* Cotton plants expressing a hemipteran-active Bacillus thuringiensis crystal protein impact the development and survival of Lygus hesperus (Hemiptera: Miridae) nymphs. *Journal of economic entomology* **105**, 616-624, (2012).

43 Kain, W. C. *et al.* Inheritance of resistance to Bacillus thuringiensis Cry1Ac toxin in a greenhouse-derived strain of cabbage looper (Lepidoptera: Noctuidae). *Journal of economic entomology* **97**, 2073-2078, (2004).

44 Wang, P. *et al.* Mechanism of resistance to Bacillus thuringiensis toxin Cry1Ac in a greenhouse population of the cabbage looper, Trichoplusia ni. *Applied and environmental microbiology* **73**, 1199-1207, (2007).

45 Abbott, W. S. A method of computing the effectiveness of an insecticide. 1925. *Journal of the American Mosquito Control Association* **3**, 302-303, (1987).