THE UNIVERSITY OF CHICAGO

THE DEVELOPMENT OF IN VIVO PHAGE-ASSISTED DIRECTED EVOLUTION PLATFORMS TO MODULATE PROTEIN-PROTEIN INTERACTIONS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTORATE OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY
VICTORIA COCHRAN XIE

CHICAGO, ILLINOIS
AUGUST 2023

I dedicate this dissertation to all the ambitious young women who did, do, and will find themselves in STEM. May we continue to break the mold and redefine what it looks like to be a scientist.

"The universe constantly and obediently answers to our conceptions; whether we travel fast or slow, the track is laid for us. Let us spend our lives in conceiving then."

-Henry David Thoreau, *Walden*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would first like to thank my advisor, Bryan Dickinson, for his support and mentorship throughout my PhD. Years ago, a graduate student mentor gave me the following advice: when choosing a PhD advisor, pick someone who you want to become, scientifically, professionally, and personally. Even in my sixth (and final) year of my PhD, I am still happy that I chose to become more like Bryan. I continue to be inspired by his creativity and enthusiasm, and I am truly grateful for his encouragement, approachability, and care.

I would next like to thank all the Dickinson Lab members I have met along the way. Each of you have contributed to making the lab a great place to do science, whether through inspiring me with your hard work and dedication or by providing encouragement and laughter in day-to-day life. I especially want to thank those I have directly collaborated with. In addition to being such a steady, grounding, and skilled project partner, Dr. Jinyue Pu is largely responsible for building my synthetic biology foundation. I thank Dr. Matthew Styles for his great assistance in review writing and am grateful and impressed by his vast knowledge and understanding of his field. I also thank Sandrine Legault for bravely taking on my final project alongside me and for bringing her excellent attention to detail and kindness with her. I am also very grateful to Dr. Somayeh Ahmadiantehrani for her assistance with figures and day-to-day lab logistics, and for the positivity she brings.

I would also like to thank my external collaborators, Dr. Brian Metzger and Prof. Joe Thornton, for venturing into a challenging yet incredibly fruitful project with me, Jin, and Bryan. I thank Prof. Weixin Tang and Prof. Joe Piccirilli for graciously agreeing to be on my committee and for your scientific advice and encouragement. I am also grateful to the many others in the Chemistry Department and beyond who have helped me these past 6 years.

None of this would have been possible without my personal support system. I could not name all the ways my husband Lijia has aided in my PhD journey, scientifically, men-

tally, emotionally, and physically. He is my home. Mocha, my loyal mini Aussie, has also been an incredible source of comfort and joy. Thank you to my mom and sister for their constant companionship and encouragement, even from afar. There are too many others to name, so I will end with truly thanking all of my family, friends, and mentors for their love and for enriching my life.

ABSTRACT

Though biomolecular interactions regulate virtually all cell processes, controlling biomolecular interactions, both for the purposes of scientific inquiry and for correcting aberrant interactions, remains a difficult challenge. In the first chapter of my dissertation, I highlight the ways that current directed evolution platforms have been able to generate tools to control interactions among the biomolecules in the central dogma and suggest current challenges and future opportunities. As others have noted in their deeming of protein-protein interactions (PPIs) as "undruggable," PPIs have proven particularly challenging to modulate, not only due to their complex and varied biophysical properties, but also because hitting off-target interactions routinely poses an issue. In chapter 2, I describe a platform I developed that aims to address such specificity issues by enabling the directed evolution of specific protein binders. We use the technology, termed PPI specificity phage-assisted continuous evolution (PACE), to evolve varied binding profiles of extant and ancestral BCL-2 family proteins, which enabled further insight regarding the roles of chance and contingency in the evolution of this protein family. I went on to adapt the PPI specificity PACE technology to directly select for inhibition of a PPI rather than protein binding alone, a property that does not always confer inhibition. This work is detailed in the creation of PANCS-PPI*i* in chapter 3, where I establish platform parameters with 3 distinct model PPIs and go on to both improve PPI inhibition of an existing inhibitor and also identify a de novo PPI inhibitor. I conclude in chapter 4 by summarizing and contextualizing my work and looking to future opportunities in the field.

LIST OF PUBLICATIONS BASED ON WORK IN THIS THESIS†

1. **Xie, V.C.***, Pu J.*, Metzger B.P.H.*, Thornton J.W., and Dickinson, B.C. Contingency and chance erase necessity in the experimental evolution of ancestral proteins. *eLife*, **2021**, **10**:e67336.

2. **Xie, V.C.**, Styles M.J., and Dickinson, B.C. Methods for the directed evolution of biomolecular interactions. *Trends Biochem. Sci.*, **2022**, *47*:5, 403-416.

3. **Xie, V.C.***, Legault S.*, Styles M.J., and Dickinson, B.C. An *in vivo* phage-assisted directed evolution platform to directly identify protein-protein interaction inhibitors. *Manuscript in preparation*

†. The following chapters of this dissertation contain sections and figures adopted from the listed publications with modifications. Chapter 1: publication 2; Chapter 2: publication 1; Chapter 3: publication 3.

⋆. Denotes equal contribution

CHAPTER 1

METHODS FOR THE DIRECTED EVOLUTION OF BIOMOLECULAR INTERACTIONS

Noncovalent interactions between biomolecules such as proteins and nucleic acids coordinate all cellular processes through changes in proximity. Tools that perturb these interactions are and will continue to be highly valuable for basic and translational scientific endeavors. By taking cues from natural systems, such as the adaptive immune system, we can design directed evolution platforms that can generate proteins that bind to biomolecules of interest. In recent years, the platforms used to direct the evolution of biomolecular binders have greatly expanded the range of types of interactions one can evolve. Herein, we review recent advances in methods to evolve protein-protein, protein-RNA, and protein-DNA interactions.

1.1 The importance of manipulating biomolecular interactions

Noncovalent interactions between biomolecules DNA, RNA, proteins, lipids, and sugars underlie all biophysical processes in the cell. Biological signaling is largely driven by proximity between biomolecules[1,2] ; therefore, whether biomolecules are near one another, interacting, or critically, not interacting, is central to the organization and function of the cell. Moreover, aberrant interactions between biomolecules are often the drivers of disease and can be targeted with inhibitors for therapeutic development. For example, the interactions between BCL-2 family proteins and their binders can be blocked by a proteinprotein interaction (PPI) inhibitor to treat cancer[3,4] . Biomolecular interactions can also be reprogrammed for beneficial purposes, exemplified by the recent explosion of chimeric antigen receptor (CAR-T) cell therapies[5] and proteolysis targeting chimeras (PROTACs)[6] , which engineer cells to respond to novel antigens and redirect protein

degradation pathways to target proteins, respectively. As such, methods to understand, reprogram, and create de novo biomolecular interactions are increasingly important to understanding molecular biology and creating biotechnologies.

In this chapter, I detail the use of directed evolution as a method to modulate biomolecular interactions. In particular, I highlight continuous evolution, and most prominently phage-assisted continuous evolution (PACE), as a powerful technology for evolving proteins to interact with proteins and nucleic acids. I also describe additional methods that can engineer proteins to interact with DNA, RNA, and other proteins, as well as experimental campaigns to create multipartner, or higher-order, interactions.

1.2 Directed evolution as a technique to evolve biomolecular interactions

Advances in structural methods, such as cryo-electron microscopy (cryoEM), paired with advances in machine learning-based computational approaches such as AlphaFold2 and RoseTTAFold, have led to a dramatic increase in our ability to study and predict the structures of biomolecules[7–10] . By contrast, understanding whether and how a given set of biomolecules interact[11] or reprogramming their interaction though defined mutations, remains challenging. However, one technology for the creation of PPIs, evolved by nature, has proven wildly successful as an engineering tool: the immune system. The mammalian immune system is capable of rapidly creating antibodies that bind to a target antigen, often another protein, as part of the bodys defense system. The basis for the immune systems capacity to solve these complex biophysical puzzles is its use of evolution, essentially, selecting for antibodies that bind to a given epitope. The rapid diversification, selection, and amplification of antibodies allows for the immune system to identify high-affinity interactions[12,13] . This process can and has been harnessed to create antibodies

for a given epitope of interest, revolutionizing basic science and medicine in turn[14,15] . However, this natural evolutionary process cannot be used to evolve biomolecules other than antibodies. To fill this technological gap, researchers must engineer experimental evolution approaches in the laboratory.

Early work toward harnessing evolution in the laboratory focused on selection methods for enzymes and PPIs. In recognition of this foundational work, the 2018 Nobel Prize in Chemistry was awarded for the directed evolution of enzymes to Francis Arnold and for the phage display of peptides and antibodies to George Smith and Sir Gregory Winter. Arnolds work, which focuses on engineering enzymes by library generation and screening to catalyze new chemical reactions, illuminated how screening individual mutants can be used not only to endow biomolecules with improved or even novel functions[16–18] , but also to uncover fundamental knowledge about how the biomolecules function and evolve[19,20] . The use of evolution in biocatalysts has been extensively reviewed elsewhere[21–23] . Phage display, invented by Smith, Winter and colleagues, is analogous to the selection of antibodies by the immune system and uses enrichment of phage-encoded protein libraries as a method for identifying novel binders or ligands[24,25] . Phage display was subsequently expanded to related technologies, such as yeast display[26] , mRNA display[27] , and ribosome display[28] . These early approaches toward harnessing evolution in the laboratory highlight the power of technologies that can properly and specifically focus evolution on a desired outcome, which is referred to as directed evolution[29–32] .

Experimental evolution approaches can generally be categorized as screens, which involve performing individual assays on each variant, often using robotics, or selections, where a fitness advantage is used to enrich variants with desired activities (**Figure 1.1**). Formally, evolution is the process of repeated rounds of diversification and selection. By adding a round of mutagenesis or further diversification to a screen and performing a second round of screening, as is often done in the context of enzyme engineering, such a

process would then be classified as directed evolution. Selections entail testing a library of variants simultaneously, where some experimental process is used to enrich variants with a given fitness level. Selection methods can be further subdivided into isolation-based methods, where target variants are physically separated from the population [e.g., display technologies or FACS (fluorescence-activated cell sorting)], or growth-based selection approaches, where organismal or viral fitness and/or growth are directly tied to the fitness of the target molecule. Likewise, selections can be run either with or without re-diversification/mutagenesis to identify fit variants within the starting library. Whether by screening or selection, several rounds of diversification and identification of active variants allows for the directed evolution of a biomolecule toward a new function.

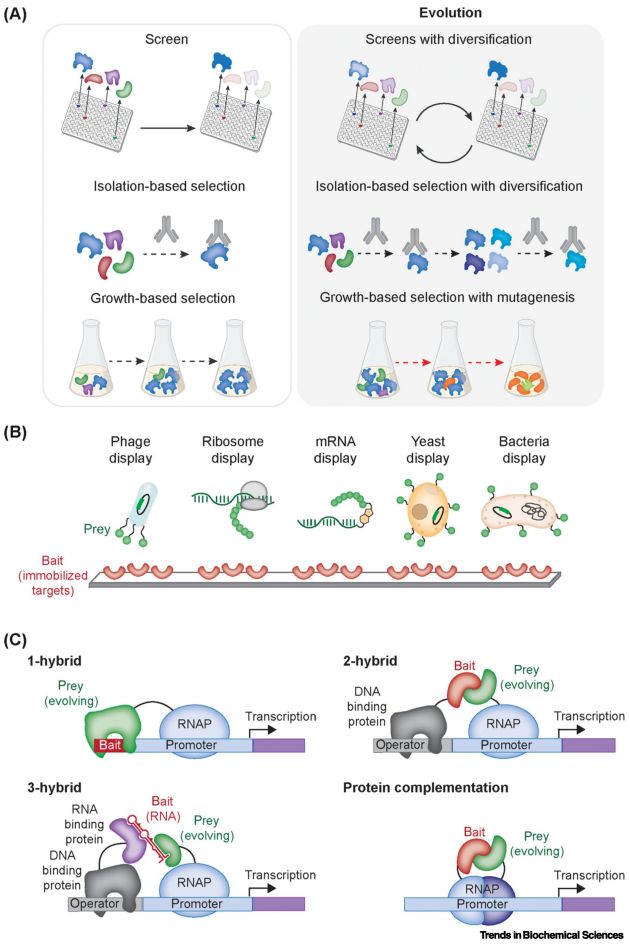


Figure 1.1 Overview of screening and selection methods used for the directed evolution of biomolecular interactions.
(A) Methods to identify mutants with a specific function include (left) screening individual variants to characterize function and selecting to enrich variants with the desired function either by isolation or growth. Experimental evolution (right) can be carried out by repeating rounds of diversification and identification of active variants either with screens or selections. (B) Types of surface display methods. In general, the bait target is immobilized to a surface, and prey (binding entities) are displayed in various manners, for example, on the surface of phage, yeast, bacteria, etc., and flowed over the immobilized targets. (C) Biosensors enable function to be linked to fitness when surface immobilization is not used. Transcription-based biosensors such as the one-, two-, and three-hybrid constructs induce the localization of RNAP to the promoter region of a reporter gene (i.e., fluorescent protein or luciferase) for screens and isolation-based selections or a gene required for host fitness (i.e., antibiotic resistance cassettes or gIII in PACE) for growth-based selections. Additionally, protein complementation can be used in a similar manner for isolation-based selections (e.g., split fluorescent protein or luciferase) and growth-based selection (split RNAP). Abbreviations: PACE, phage-assisted continuous evolution; RNAP, RNA polymerase.

Selections offer several benefits over screens: assaying larger libraries, having tunable fitness thresholds, in some cases featuring negative selections that allow for assessing multiple characteristics of a biomolecule at once, and in general, fewer labor- and instrument-intensive processes. However, the construction of the selection platform is critical to successful directed evolution campaigns. The primary challenges in designing a selection platform are (i) linking the genotype of the evolving biomolecule to a function of interest and (ii) linking the function of interest to the fitness of the genotype. For example, phage are well-suited hosts for the creation of a selection platform to evolve peptide and protein binders: phage genomes are naturally tied to the peptides/proteins they encode (linking genotype to the biomolecule); phage coat proteins can be fused to the evolving peptide/protein to facilitate phage binding to a target protein via the displayed peptide/protein (linking the biomolecule to function); and finally, bound phage can be isolated from nonbinding phage (linking function to fitness). While powerful, phage display is an example of noncontinuous evolution, requiring researcher intervention to replicate and possibly mutagenize the phage-encoded biomolecules for additional rounds of selection; this constraint limits these evolutions to only a few rounds of selection.

Continually generating diversity while enriching fit variants via growth-based selections is referred to as continuous evolution. Due to the challenges of linking the fitness of a host to a desired function of a biomolecule, continuous evolution approaches were, until recently, limited to selections that directly link fitness to function, such as selecting for antibiotic resistance[33,34] , and *in vitro* systems, such as self-replicating RNA ligases[35] . However, the past decade has brought about an explosion of new continuous evolution technologies to solve this challenge[36–45] For instance, PACE links phage propagation (host fitness) in Escherichia coli carrying a mutagenesis plasmid to the phenotype (function) of a gene within the phage (**Figure 1.2A**)[38,46] . The link between phage fitness and target activity is established by the inducible expression of pIII, a required phage protein,

which is provided by the host *E. coli* cells. PACE, once developed for a desired function of interest (that is, once a robust link between target activity and pIII expression is engineered), can enable hundreds of rounds of selection in days with minimal researcher intervention. For additional information on current continuous *in vivo* evolution methods, see **Box 1**. As mature display-based technologies continue to find new applications and novel continuous evolution systems continue to develop, an expansive experimental evolution toolkit for probing and engineering biomolecular interactions is emerging. This review sets out to highlight recent advances in approaches that use evolution to probe the interactions between biomolecules, organized by technologies for engineering interactions with proteins, interactions with RNA, interactions with DNA, and molecules that influence the interactions of pairs of biomolecules (higher-order interactions). In each case, these directed evolution approaches are leading to basic insights into how biomolecular interactions have evolved and to novel biotechnologies and therapeutic approaches.

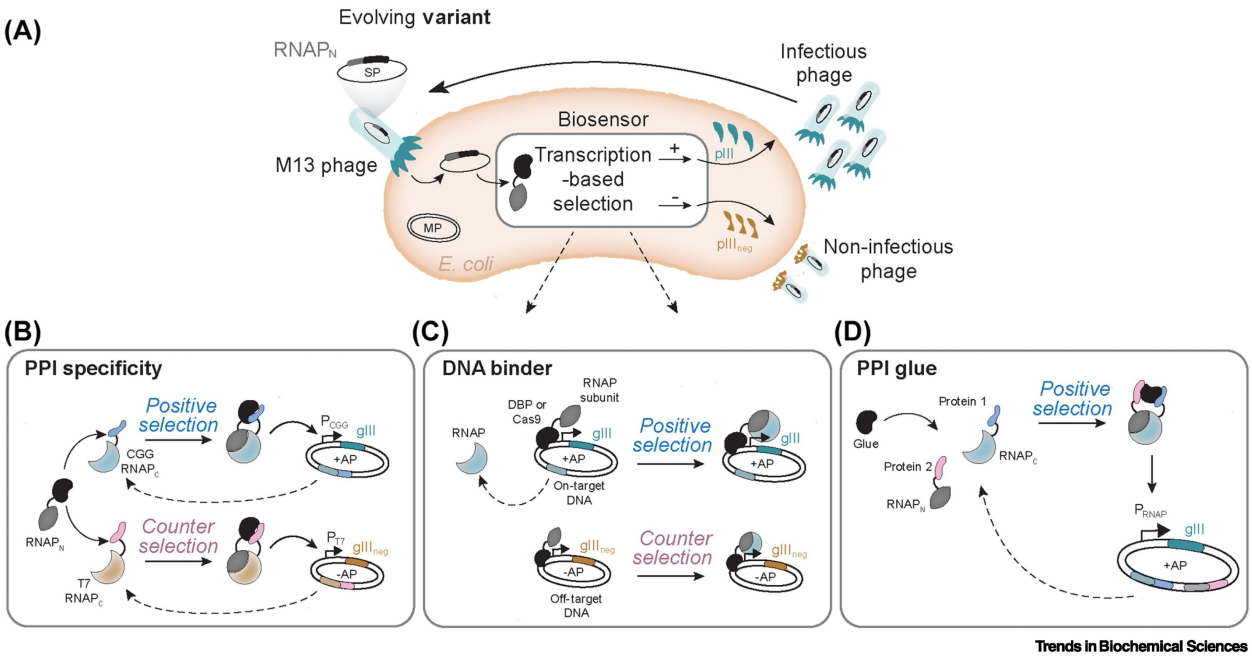


Figure 1.2 (A) Phage-assisted continuous evolution (PACE) biosensors for the evolution of biomolecular interactions, including (B) specific proteinprotein interactions (PPIs), (C) DNA binders, and (D) PPI glues.

(A) General schematic for how PACE works. Bacteriophage carry a plasmid that encodes an evolving protein of interest. Phage infect host Escherichia coli cells that contain plasmids that encode a transcription-based biosensor that drives pIII to produce infectious phage for positive selection and a dominant negative form of pIII (pIII$_{neg}$) to create non-infectious phage for negative selection. (B) Phage carry an evolving protein fused to the N-terminus of proximity-dependent split RNAP. Host *E. coli* cells express two proteins fused to two orthogonal C-terminal T7 RNAP fragment variants with different promoter specificities. If the evolving protein variant interacts with the target protein of interest, this reconstitutes an RNAP that binds to the CGG promoter, which triggers pIII production and phage replication. If the variant binds the counterselection protein, it reconstitutes an RNAP that binds the T7 promoter and leads to the production of a dominant negative phage protein, pIII$_{neg}$, which lowers phage fitness. (C) Proteins that bind DNA, including transcription factors and Cas9 effectors, can be evolved by fusing them to the subunit of an RNAP and encoding the fusion in phage. Positive selection is driven by the protein binding to a specific sequence upstream of the RNAP promoter to drive pIII expression, and negative selection can be driven via nonspecific binding triggering pIII$_{neg}$ production. (D) A genetically encoded bifunctional molecule is expressed by phage and can drive pIII production and phage propagation if it binds to two partners fused to split halves of the split RNAP biosensor to reconstitute active RNAP. Abbreviations: RNAP, RNA polymerase; DBP, DNA-binding protein.

1.3 Engineering interactions between proteins

Along with phage display, many other surface display technologies have been developed that can evolve peptides and proteins to bind protein targets. These include mRNA[47] , ribosome[3] , yeast[48,49] , and bacteria display[50] (**Figure 1.1B**). Display technologies have proven particularly fruitful for the evolution of antibodies, as covered in a recent review[51] , and each display technology has its advantages and drawbacks. For instance, while phage display can attain larger library sizes, bacteria and yeast display can accommodate larger proteins. Additionally, certain protein targets are more compatible with yeast display, such as those that are insoluble when expressed by bacteria or those containing post-translational modifications specific to eukaryotic hosts. For more details on display technologies, see a recent review by Park[52] .

Although display methods can work well for engineering extracellular interactions, as in discovering ligands for G-protein coupled receptors (GPCRs)[53] and plasma proteins[54] , one drawback is that they do not evolve proteins to function in an intracellular context. Biomolecular interactions can depend on a variety of cellular factors, from metabolite concentrations to localization; thus, evolution in a more native biological context is advantageous. For this reason, *in vivo* evolution systems have gained popularity in recent years, and those that select for binding generally use biosensors that adhere to an n-hybrid or protein complementation assay approach (**Figure 1.1C**). In these methods, a protein of interest (bait) and the protein under selection pressure (prey) are each fused to additional proteins, and binding of the prey to the bait protein results in bringing these components into close proximity which assemble to form some sort of selection output. As illustrated in **Figure 1.1C**, n-hybrid systems typically involve the localization of an RNA polymerase (RNAP) or other transcription inducer to a reporter gene such as GFP. Similarly, protein complementation approaches can utilize split fluorescent proteins or other optical reporters for assays by screening or isolation via FACS for selections, or a DNA-

binding protein/transcription factor pair or split RNAP to produce a protein that allows for survival in a growth-based selection. Biosensors have been widely used for analyzing and screening PPIs, which has been highlighted in previous reviews[55,56] . For example, two hybrid-based systems have been used to extensively map potential interactions between the proteome[57] , and recent split luciferase reporter systems enabled the rapid screening of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) antibodies[58] and endosomal disruption stimuli[59] . Protein complementation technologies have also been employed in directed evolution campaigns to yield useful binders for applied purposes and advanced study of evolution itself, biochemistry, and structural biology[60] .

Both display and complementation technologies have been linked to viral replication to create powerful *in vivo* experimental evolution platforms that can generate protein binders (**Table 1.1**). The PACE technology incorporates protein complementation when used to evolve protein binders. For example, by fusing an insect receptor protein to a DNA-binding domain and Bacillus thuringiensis -endotoxin (Bt toxin) to an RNAP subunit, the Liu laboratory used PACE to evolve Bt toxin to bind the insect receptor and overcome resistance[61] . A recently published paper from the laboratory also now enables the evolution of binders that contain disulfide bonds[62] . Additionally, our laboratory developed a protein binder PACE system based on complementation of split RNAP[63] . We performed deep mutational scanning of the Ras/Raf interaction, by generating a library of Raf variants and enriching for Ras binding without mutagenesis a technique we dubbed phage-assisted continuous selection (PACS)[64] . Technologies are also emerging for experimental evolution in eukaryotic cells. For instance, autonomous hypermutation yeast surface display, deemed AHEAD, combines yeast display with OrthoRep to facilitate continuous evolution of protein binders in yeast[48] . It has been used to evolve camelid single-domain antibodies, or nanobodies, that bind to targets such as the SARS-CoV-2 S glycoprotein. Moreover, efforts at directed evolution in mammalian cells (reviewed by

Hendel and Shoulders) are aiming to use adenovirus or RNA virus variants as a vector, analogous to the role of phage in PACE[32] . While the field still faces challenges, these technologies do have potential for evolving a variety of activities, including protein binders. Indeed, in its premier paper, viral evolution of genetically actuating sequences, or VEGAS, was used to evolve nanobodies that bind to GPCRs[41] .

All the aforementioned approaches measure, select for, or evolve a desired PPI. However, preventing interactions with an undesired protein is often just as critical as interacting with a desired protein, both in terms of understanding the emergence of molecular recognition through evolution and for developing selective biotechnologies. Counterselections or secondary screens can be deployed, but they are then decoupled from activity evolved in the primary evolution/screen. For example, in a recent work which will be described in detail in Chapter 2, we developed a new PACE-based system for evolving selective PPIs[65] (**Figure 1.2B**). To accomplish this, we employed two separate RNAP-based protein complementation systems using our groups proximity-dependent split RNAP biosensor technology[63,66] . In this system, the phage-encoded protein of interest is simultaneously and continuously evolving to interact with a target protein and to not interact with a non-target protein. We used this platform to explore the roles of chance and contingency in the evolution of binding specificity of the BCL-2 family proteins. However, the core platform can, in principle, be used to engineer novel specificity into PPIs of interest. Another recent example of PACE with negative selection was used to reprogram the binding specificity of proteases[67] , a class of proteins that both interact with and cleave specific proteins based on sequence motifs. Though improving the enzymatic activity of proteases is often a focus, engineering specificity in their interactions with their protein targets has proven difficult to achieve until now.

Table 1.1 Examples of continuous evolution methods to evolve binders to nucleic acids and proteins.

Binding partner	Method	Application	Evolution environment	Refs
Protein	PACE	Evolve a binder of an insect receptor to prevent antibiotic resistance	*E. coli*	[62]
Protein	PACE	Evolve specific BCL-2 family protein binders to study the evolution of PPI specificity	*E. coli*	[66]
Protein	PACE	Evolve a disulfide bond-containing protein binder to Her2	*E. coli*	[63]
Protein	OrthoRep	Evolve SARS-CoV-2 S glycoprotein nanobodies	Yeast	[49]
Protein	Adenoviral PACE/VEGAS	Evolve GPCR nanobodies	Mammalian cells	[40,41]
DNA	PACE	Evolve RNAPs with different promoter specificities	*E. coli*	[83,84]
DNA	PACE	Evolve DNA-binding proteins to bind various sequence motifs	*E. coli*	[8587]
DNA	PACE	Evolve dCas9 variants with broadened PAM compatibility	*E. coli*	[99,100]
Higher-order interactions	PACE	Evolve bifunctional binder to a zipper peptide and ULK1	*E. coli*	[103]

1.4 Engineering interactions with RNA

In its simplest form, engineering interactions between nucleic acids such as RNA and DNA can be very straightforward. For instance, binding to a single-stranded DNA (ssDNA) sequence can be achieved via a complementary single-stranded DNA or RNA molecule through easily programmable WatsonCrickFranklin base-pair interactions. However, engineering proteinnucleic acid interactions can be quite challenging, as there is no easily

decipherable code to program proteins to interact with a specific nucleic acid. In nature, proteins have evolved to recognize specific RNA motifs through interactions with specific base sequences, the chemical modification states of bases, and through interactions with RNA structures. These RNA binding proteins are involved in regulating RNA turnover, translation, localization, splicing, and post-transcriptional modifications[68] . Directed evolution techniques have been applied to alter the binding specificity of proteins with each type of proteinRNA interaction using both *in vitro* binding methods, such as phage display, and *in vivo* three-hybrid biosensors (**Figure 1.1C**). These studies have been used to engineer novel RNA specific binding interactions, as well as to study the evolution of RNA binding proteins.

Pumilio and FBF homology (PUF) proteins can recognize single-stranded RNA (ssRNA) through sequence-specific interactions. Typically, PUF proteins are composed of eight 36-amino acid repeat Pumillo homology (PUM) domains flanked by N- and C-terminal regions; these domains form a crescent shape with eight ssRNA nucleotides binding to the concave face of the protein. Each PUM domain recognizes a specific RNA base via three conserved side chains [a tbipartite recognition motif (TRM)], and thus, the specific PUM domains and their order dictate the sequence specificity of the PUF protein[69] . Site-directed mutagenesis of the TRM followed by screening variants via electrophoretic mobility shift assays (EMSAs) has facilitated the interconversion of PUM domains that bind to adenine, uracil, and guanine[70] ; however, naturally occurring cytosine PUM domains have not been discovered. Filipovska et al. used site-directed random mutagenesis paired with a yeast three-hybrid growth-based screening assay to identify PUM domain mutants capable of binding to cytosine and thus, created a universal code for RNA recognition by PUF proteins[71] . These sequence-specific PUF proteins have been used in a wide variety of applications, such as the development of a sequence-specific RNA endonuclease[72] . Selection-based techniques have been used to determine the RNA

sequence specificity for other PUF proteins[73] and to study the evolution of homologs that recognize different lengths of RNA sequences (8–10 nt)[74] .

RNA can form well-defined structures that proteins in nature have evolved to recognize, and directed evolution can be used to evolve proteins that bind specific RNA structures. For example, phage display has been used to evolve antibody fragments (Fabs) that bind to the internal ribosome entry site (IRES) of hepatitis C virus (HCV)[75] and competitively inhibit binding of ribosomal proteins to the HCV IRES. Yeast display was utilized to reprogram the human protein U1A to bind a structured element from the HIV viral RNA genome[76] , and this protein–RNA interaction serves as a key building block for our group's CIRTS platform for engineering RNA regulatory proteins[77,78] . A weakly active dCas13a variant was diversified using random mutagenesis and then selected by FACS to improve the ability of dCas13a to target and repress translation of mRNA targets[79] . Lastly, with similar goals as the counter selections employed to evolve specificity into PPIs, novel selection strategies, such as library-versus-library selection, have been used to evolve orthogonal RNA–RNA binding protein pairs[80] .

In addition to evolving proteins that bind to mRNA, directed evolution has also been used to evolve tRNA–aminoacyl-tRNA synthetase (aaRS) pairs[81] . In this study, the authors first computationally identified potential orthogonal tRNAs and tested for orthogonality against natural aaRS proteins in *E. coli* and for function with their cognate aaRS in an *E. coli* host. However, when the tRNAs that passed these screens were recoded for an amber suppression codon, they no longer functioned with their cognate aaRS. Thus, directed evolution was performed, generating libraries of aaRS variants followed by screening for fluorescence, which occurs if an aaRS interacts with the modified tRNA to enable translation though a GFP stop codon. This yielded additional orthogonal tRNA–aaRS pairs that can be used to further assist genetic code expansion efforts.

1.5 Engineering interactions with DNA

Categories of proteins that naturally interact with DNA include polymerases, nucleases, and transcription factors. In the initial report on PACE, the system was shown to be capable of evolving several proteinDNA interactions, including a recombinase and T7 RNAP[38] . In each example, the key property needed to drive pIII production, and thus phage replication, is binding of an evolving protein to DNA. Since these initial studies, PACE has been used to evolve RNAPs with orthogonal promoter specificity[82,83] and sequence-specific DNA-binding proteins called TALENs (transcription activator-like effector nucleases)[84] . PACE has also recently been used in combination with rational design to engineer a DNA E-box motif binder based on the Myc/Max transcription factors[85,86] (**Figure 2C**). In this paper, the PACE platform used a one-hybrid approach to evolve a DNA binder, which proved superior to previous yeast and bacterial one-hybrid noncontinuous evolution campaigns in that increasing the selection pressure over time was used to combat false positives.

CRISPR-Cas proteins have found widespread use in the development of DNA editing technologies, yet they tend to have substantial off-target activity and require a PAM recognition motif at the target DNA[87–90] . While RNA guide optimization and the crystallization of Cas9 have enabled rational design efforts to improve specificity[91–95] , several directed evolution campaigns have also been carried out to improve the utility of these systems. The more specific Sniper-Cas9 variant was generated by creating Cas9 libraries via error-prone PCR and XL1-red competent cells and screening them in *E. coli* cells for cell survival, which depended on them targeting an exact guide match and not targeting a close mismatch that was encoded in the genomic DNA[96] . Screening for on- and off-target activity among error-prone PCR-generated Cas9 libraries was also used to create evoCas9 with improved specificity[97] . In this report, screens were done in yeast where targeting an exact guide match led to cell survival and targeting a mismatch led to white

colonies, whereas no off-target activity created red colonies. PACE has also been used to evolve Cas9 with increased specificity for a specific PAM and for variants with broader PAM compatibility using a similar one-hybrid approach as was employed for the DNA box motif directed evolution[46,98] (**Figure 1.2C**).

Though generating technologies that regulate DNA has been a large focus in the field of evolving DNA interactions, progress has also been made in understanding the evolution of natural biological interactions with DNA. For example, multireplicate PACE evolutions of T7 RNAP to achieve novel promoter specificities led to insights into how path dependence impacts evolutionary trajectories[82,83] . In another example, ancestral sequence reconstruction combined with deep mutational scanning revealed how an ancient transcription factor evolved to achieve novel DNA specificities[99] . This study found that many alternative protein sequences conferred the given functions, further highlighting the role of contingency through permissive mutations that emerged in history.

1.6 Engineering higher-order interactions

Thus far in this chapter, I have focused on evolution technologies that engineer two-partner interactions. However, evolution is also able to solve more complex problems, such as engineering multipartner interactions[100] . One recent example from our laboratory introduced an experimental evolution system to engineer proteinprotein interaction glues molecules that bind to two different target proteins to bring them in proximity to each other[101] . The evolution strategy, termed re-PPI-G, mimics our previous PACE designs: two proteins are fused to the N- and C-terminal halves of the split RNAP biosensor (**Figure 1.2D**). These fragments are both expressed in the *E. coli* host cell, and thus, do not undergo evolution. Rather, another glue fragment is encoded in phage, which evolves to bring the two proteins together, promoting RNAP reassembly and subsequent pIII production and phage replication. After optimization, we tested the system by evolv-

ing a zipper peptide fused to ULK1 to better interact with both the partner zipper peptide and ULK1s partner, GABARAP. This technology holds promise as a means to rewire proteinprotein interaction networks, just as current small-molecule PROTAC technologies do[6] .

Coevolving biomolecular interactions has been an exciting field in recent years as well. Coevolution has primarily been achieved through mutating each partner individually followed by screening to assess changes in activity, though selection platforms are emerging as well[102] , and advances in deep mutational scanning technologies have enabled further progress. To date, coevolution technologies have advanced biochemical knowledge of PPIs and allowed the generation of orthogonal binding partners and signaling pathways[103,104] . However, though real-time whole organism evolution experiments by nature enable continuous coevolution[105,106] , directed experimental continuous coevolution has yet to be achieved.

Additional higher-order interactions include protein packaging, which occurs when viral capsids encapsulate their RNA or DNA genomes. Directed evolution was used to create a nonviral protein cage based on Aquifex aeolicus lumazine synthase (AaLS) building blocks that could package a tagged HIV protease[107] . To do this, error-prone PCR was used to create a library of AaLS building blocks, which were then screened in *E. coli* for their ability to encapsulate HIV protease, which is otherwise toxic to the cell. The cage was subsequently evolved using similar directed evolution campaigns to package its own mRNA[108] and more efficiently protect the enclosed RNA from nucleases, resulting in a structure that mimics those found in natural viruses[109] . Computational approaches have also been used to generate de novo nonviral icosahedral capsid scaffolds based on viral structures to address the basic scientific question of what is necessary for capsid formation[110,111] . Computationally derived scaffolds were then mutagenized by Kunkel mutagenesis and screened in *E. coli* for the ability to encapsulate and protect their own

RNA from challenges such as heat and other environments, where survivors could then be harvested and sequenced to link genotype to successful phenotype. The evolution campaigns found capsids that achieved RNA packaging activity *in cellulo* and in mouse models[111] , and an additional deep mutational scanning library provided insight into biochemically important characteristics of nucleocapsids, for example, hydrophobic cores and positively charged capsid interiors. This series of events spectacularly highlights a trend present in this review: the synergy between computational approaches followed by experimental evolution to generate robust biomolecular interactions. Just as molecular docking has proved a powerful technique in small-molecule inhibitor development, so too could computational methods provide potential starting points for evolving new interactions.

Although this chapter has focused on evolving proteins to interact with components of the central dogma (**Figure 1.3**), it is also worth noting a few related advances. Biomolecules can be decorated with various modifications that go beyond nucleotides and natural amino acids. Examples where directed evolution has been used to generate proteins that recognize such modifications include proteins that can bind proteins with glycan and sulfotyrosine additions[112,113] and the reprogramming of RNA reverse transcriptase to interact with specific nucleic acid methylation sites[114] . Moreover, systematic evolution of ligands by exponential enrichment (SELEX) technologies create DNA and RNA aptamers capable of binding proteins and small molecules and are often used for detection purposes[114] . Just as recent advances have highlighted additional roles that nucleic acids can play other than the traditionally ascribed role of information storage and transfer, progress has also been made in engineering nucleic acidnucleic acid interactions that go beyond that of simple base pairing. Efforts to expand the genetic code have required experimental evolutions of tRNAribosome interactions to enable quadruplet codons[115,116] . While others have focused on engineering orthogonal tRNAtRNA synthetase pairs, self-aminoacylating

tRNA ribozymes, dubbed Flexizymes, were evolved, which allow for genetic code expansion[117] , and evolution of the ribosome itself is being done to allow further expansion of what types of synthetic proteins can be made[115,118,119] . Furthermore, DNA enzymes, which are not known to exist in nature, have been evolved for a variety of purposes over the past 2030 years, as reviewed elsewhere[120] .

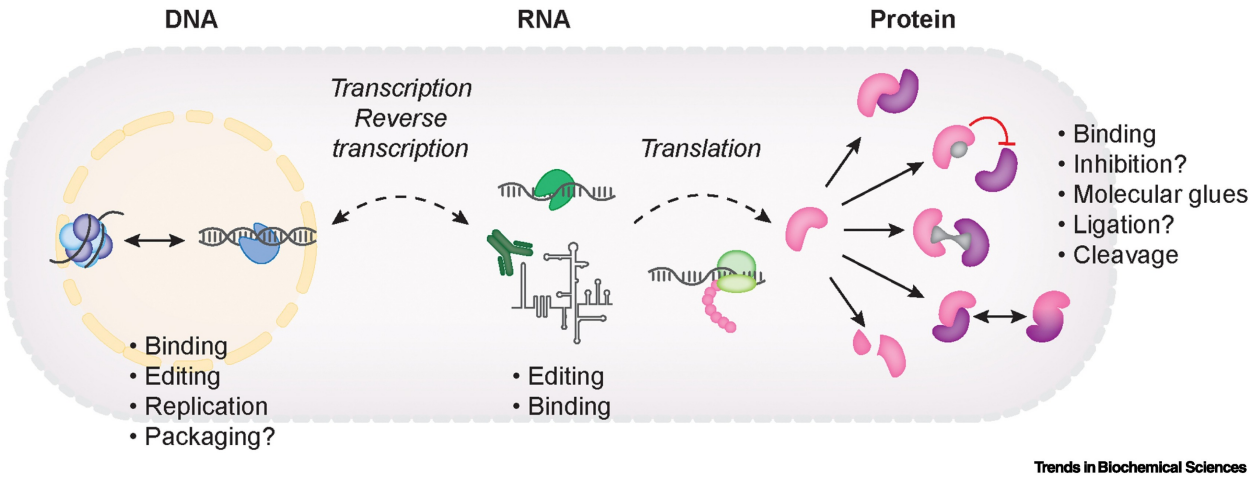


Figure 1.3 Directed evolution has been employed, or could in principle be employed, to evolve proteins that interact with DNA, RNA, and other proteins to facilitate biomolecular interactions in the above areas.

1.7 Concluding remarks

As our understanding of the diverse interactomes of the biomolecules of life continues to expand, so too has our ability to evolve biomolecular interactions. Experimental evolution can approximate replaying the tape of life and thus give insight into the molecular basis of evolving interactions, and it can also generate new or improved interactions for use in studying biology and creating novel biotechnologies. Despite significant successes and advances over the past 30 years, the full potential of evolution has still not been harnessed as a design approach for evolving biomolecular interactions. The development of biosensor technologies has spurred advances in evolving different interaction types and promises to continue to do so (see **Box 2** for outstanding questions). Novelty is

certainly important for progress, yet efficiency and selectivity are also paramount. Easily deployable methodologies, robust continuous evolution systems, advanced automation, and computational approaches for improved library design will ensure that experimental evolution becomes more accessible and successful in the coming years.

1.8 Supplementary notes

1.8.1 Eukaryotic continuous directed evolution

Like PACE, eukaryotic continuous directed evolution efforts aim to evolve biomolecules by linking a desired activity to survival and, critically, to focus the evolution on a desired gene. One approach that has been developed is to engineer a native retrotransposon to elevate mutation rates of its corresponding cargo[39] . Additionally, OrthoRep uses an orthogonal error-prone DNA polymerase that specifically drives replications and mutations of a gene expressed in a plasmid in yeast[45] . A similar platform is seen in the recently realized compatibility of EvolvR for use in bacteria and yeast; here, dCas9 is linked to an error-prone DNA polymerase such that targeted mutagenesis is possible[121] . As in PACE, the challenge then becomes linking activity to fitness, which can be done though employing various biosensors to evolve activities such as catalysis and binding. Though the replication rate of yeast is lower than that of phage, meaning the evolutionary process in principle takes longer than PACE to achieve the same number of rounds of mutagenesis and selection a noteworthy advantage of yeast continuous evolution is that activity can be evolved in a eukaryotic cellular context, which is arguably more suited to evolving biomolecules that can function in humans than the bacterial environment. Advances such as combining OrthoRep with automated continuous culture technologies[122] and with yeast surface display[48] showcase the great potential of using yeast as a conduit for directed evolution.

Continuous directed evolution in mammalian cells is also a fast-growing yet challenging area of research. Like PACE, mammalian cell continuous evolution approaches seek to use viruses as a conduit for directed evolution, in one method by a double-stranded DNA adenovirus[40] and in another by a single-stranded RNA Sindbis alphavirus, the latter of which is known as VEGAS[41] . For a recent perspective on this field, please see ref[32] .

1.8.2 Outstanding questions

How does nature compare to laboratory evolution, and how can the lessons of natural evolution inform how to better deploy evolutionary principles in the laboratory?

Can we employ multiple negative and/or positive selections simultaneously to evolve more than one characteristic in a molecule (e.g., allostery)?

What methods can be used to evolve biomolecular interactions not yet amenable to directed evolution, such as interactions with lipids, in vivo continuous coevolution, and proteinprotein interaction inhibitors? I address the latter in Chapter 3 with my recent advent of PANCS-PPI*i*.

Can we evolve biomolecular interactions radically different than those found in nature (e.g., synthetic translational machinery for sequence defined polymers and protein materials, control of phase transition/protein condensates)?

Current continuous evolution techniques generally require some small level of activity how can one generate truly novel function without pre-existing function (e.g., with robust de novo libraries) using continuous evolution techniques?

As a continuous evolution technology, PACE has been expanded to evolve a variety of interactions can we adapt similar selection strategies for yeast and mammalian continuous technologies?.

CHAPTER 2

CONTINGENCY AND CHANCE ERASE NECESSITY IN THE EXPERIMENTAL EVOLUTION OF ANCESTRAL PROTEINS

The roles of chance, contingency, and necessity in evolution are unresolved because they have never been assessed in a single system or on timescales relevant to historical evolution. My colleagues and I combined ancestral protein reconstruction and a new continuous evolution technology to mutate and select proteins in the B-cell lymphoma-2 (BCL-2) family to acquire proteinprotein interaction specificities that occurred during animal evolution. By replicating evolutionary trajectories from multiple ancestral proteins, we found that contingency generated over long historical timescales steadily erased necessity and overwhelmed chance as the primary cause of acquired sequence variation; trajectories launched from phylogenetically distant proteins yielded virtually no common mutations, even under strong and identical selection pressures. Chance arose because many sets of mutations could alter specificity at any timepoint; contingency arose because historical substitutions changed these sets. Our results suggest that patterns of variation in BCL-2 sequences and likely other proteins, too are idiosyncratic products of a particular and unpredictable course of historical events.

2.1 Introduction

The extent to which biological diversity is the necessary result of optimization by natural selection or the unpredictable product of random events and historical contingency is one of evolutionary biologys most fundamental and unresolved questions[123–126] . The answer would have strong implications not only for our understanding of evolutionary processes but also for how we should analyze the particular forms of variation that exist today. For example, if diversity primarily reflects a predictable process of adaptation to

distinct environments, then a central goal of biology would be to explain how the characteristics of living things help to execute particular functions and improve fitness[127] . By contrast, if diversity reflects chance sampling from a set of similarly fit possibilities, then the variation itself is of little interest because it does not affect biological properties or shape future evolutionary outcomes; the goal of biology would be to identify the invariant characteristics of natural systems and explain how they contribute to function[128–131] . Finally, if diversity reflects contingency a strong dependence of future outcomes on initial conditions or subsequent events, also known as path-dependence then the outcomes of evolution would be predictable only given complete knowledge of the constraints and opportunities specific to each set of conditions[106,132–134] ; the goal of biology would then be to characterize these constraints and opportunities, their mechanistic causes, and the historical events that shaped them.

Many studies have provided insight into the ways that chance, contingency, and necessity can affect the evolution of molecular sequences and functions, but the relative importance of these factors during evolutionary history remains unresolved because they have never been measured in the same system, and their effects over long evolutionary time scales have not been characterized. For example, experiments on ancestral proteins have shown that particular historical mutations have different effects when introduced into different ancestral backgrounds suggesting contingency but they do not reveal the extent to which context-dependence actually influenced evolutionary outcomes; further, these historical trajectories happened only once, so they cannot elucidate the effect of contingency relative to chance[135–144] . Experimental evolution studies could, in principle, characterize both chance and contingency if they had sufficient replication from multiple starting points, but to date no study has done so; furthermore, no study has imposed selection on historical proteins to acquire functions that changed during history, so their relevance to historical evolution is not clear[83,145–156] . Studies of phenotypic convergence in

nature suggest some degree of repeatability at the genetic level (reviewed in refs[157–160] , but these studies rarely involve replicate lineages from the same starting genotypes, and evolutionary conditions are seldom identical; as a result, similarities and differences among lineages cannot be attributed to chance, contingency, or necessity. Furthermore, these studies have typically involved closely related species or populations and therefore do not measure the effects of chance and contingency that might be generated during long-term evolution.

The ideal experiment to determine the relative roles of chance, contingency, and necessity in historical evolution would be to travel back in time, re-launch evolution multiple times from each of various starting points that existed during history, and allow these trajectories to play out under historical environmental conditions[123] . By comparing outcomes among replicates launched from the same starting point, we could estimate the effects of chance; by comparing those from different starting points, we could quantify the effects of contingency that was generated along historical evolutionary paths (**Figure 2.1**). Necessity would be apparent if the same outcome recurred in every replicate, irrespective of the point from which evolutionary trajectories were launched and changes that occurred subsequently: in that case, evolution would be both deterministic (free of chance) and insensitive to initial and intervening conditions (noncontingent). Although time travel is currently impossible, we can approximate this ideal design by reconstructing ancestral proteins as they existed in the deep past[161] and using them to launch replicated evolutionary trajectories in the laboratory under selection to acquire the same molecular functions that evolved during history.

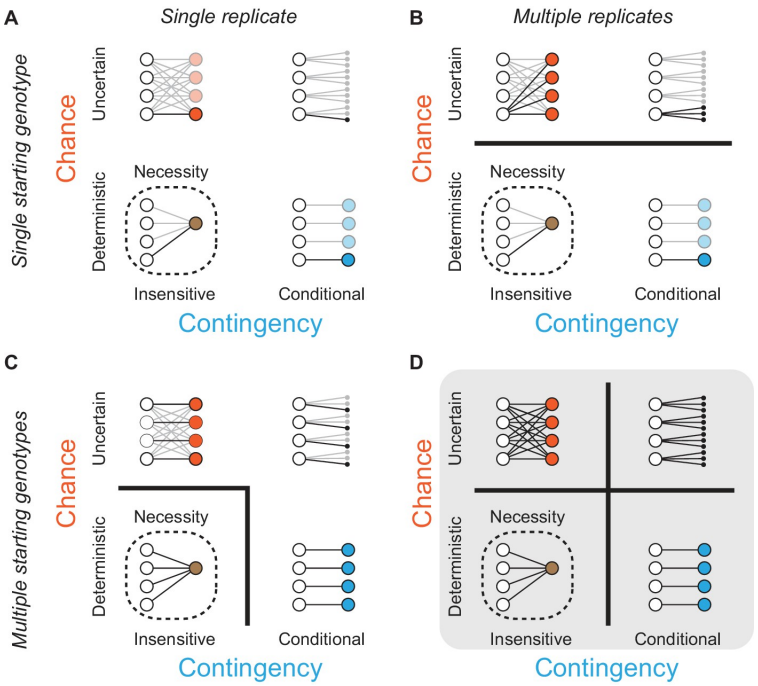


Figure 2.1 Assessing the effects of chance and contingency during evolution. Each panel (A-D) shows the capacity of one experimental design to detect chance and contingency; the quadrants within each panel show evolutionary scenarios with varying degrees of chance and contingency. Chance (y-axis within each panel) is defined as random occurrence of events from a probability distribution in which multiple events have probability > 0 given some defined starting point; in the absence of chance, evolution is deterministic because a single outcome always occurs from any starting genotype. Contingency (x-axis within each panel) is defined as differences in this probability distribution given different starting or subsequent conditions; in the absence of contingency, outcomes are insensitive to these conditions, and all starting points lead to the same outcome or set of outcomes. Lines connect starting genotypes (white circles) to evolutionary outcomes. Quadrants show evolution under the influence of chance (orange), contingency (blue), or both (black); outcomes are necessary (brown, with dotted line) when neither chance nor contingency is important. Potential trajectories that are not observed because of deficiencies in experimental design are shown with reduced opacity. Thick black lines between quadrants in (AD) separate evolutionary scenarios that can be distinguished from each other given each design. (A) Assessing one evolutionary replicate from one starting point provides no information about the extent to which chance, contingency, or necessity shape the outcome. (B) Assessing multiple replicates from one starting point can detect chance but provides no information about contingency. (C) Assessing one replicate each from multiple starting points can detect necessity or its absence, but cannot not distinguish between chance and contingency. (D) Studying multiple replicates from multiple starting genotypes allows chance, contingency, and necessity to be distinguished.

Here we implement this strategy using the B-cell lymphoma-2 (BCL-2) protein family as a model system and the specificity of proteinprotein interactions (PPIs) as the target of selection. BCL-2 family proteins are involved in the regulation of apoptosis[162–165] through PPIs with coregulators[166–169] . Although there are many dimensions to BCL-2 family proteins cellular effects, different binding specificities for coregulator proteins are a critical determinant of their particular biological functions. Among BCL-2 family members, the myeloid cell leukemia sequence 1 protein (MCL-1) class strongly binds both the BID and NOXA coregulators, whereas the BCL-2 class (a subset of the larger BCL-2 protein family) strongly binds BID but not NOXA (**Figure 2.2A**)[170] . The two classes share an ancient evolutionary origin: both are found throughout the Metazoa[171,172] and are structurally similar, using the same cleft to interact with their coregulators (**Figure 2.2B, Figure 2.3**), despite having only 20% sequence identity.

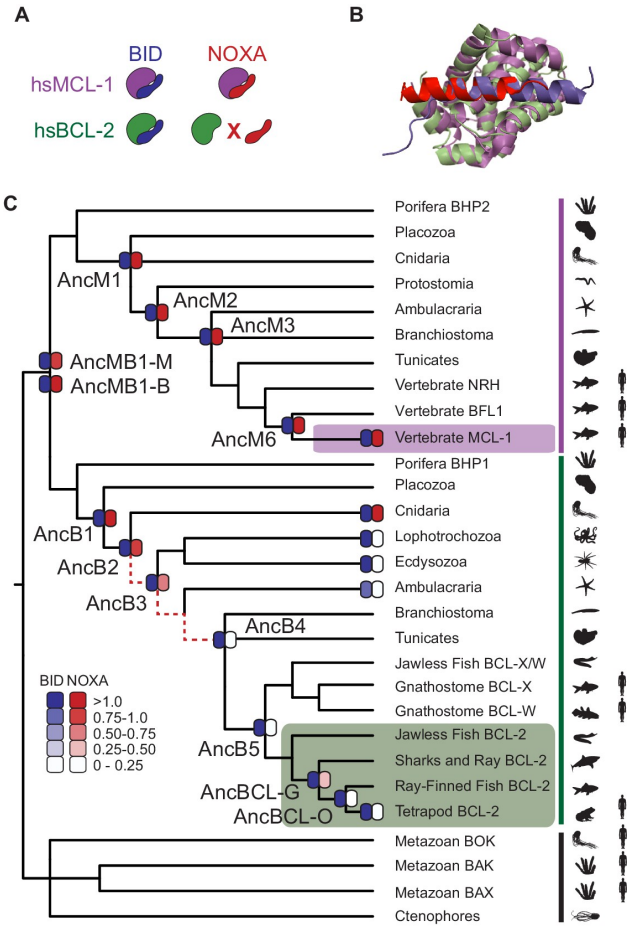


Figure 2.2 BID specificity was acquired during vertebrate BCL-2 evolution.
(A) Protein binding specificities of extant BCL-2 family members. Human MCL-1 (hsMCL-1, purple) strongly binds BID (blue) and NOXA (red), while human BCL-2 (hsBCL-2, green) strongly binds BID but not NOXA. (B) Crystal structures of MCL-1 (purple) bound to NOXA (red, PDB 2nla), and BCL-xL (green, a closely related paralog of BCL-2) bound to BID (blue, PDB 4qve). (C) Reduced maximum likelihood phylogeny of BCL-2 family proteins. Purple bar, MCL-1 class; green bar, BCL-2 class. The phylogeny was rooted using as outgroups the paralogs BOX, BAK, and BAX (black bar). Heatmaps indicate BID (blue) and NOXA (red) binding measured using the luciferase assay. Each shaded box shows the normalized mean of three biological replicates. Red dotted lines, interval during which NOXA binding was lost, yielding BID specificity in the BCL-2 proteins of vertebrates (green box). Purple box, vertebrate MCL-1. Silhouettes, representative species in each terminal group. AncMB1-M and -B are alternative reconstructions using different approaches to alignment ambiguity (see Materials and methods). For complete phylogeny, see **Figure 2.5**.

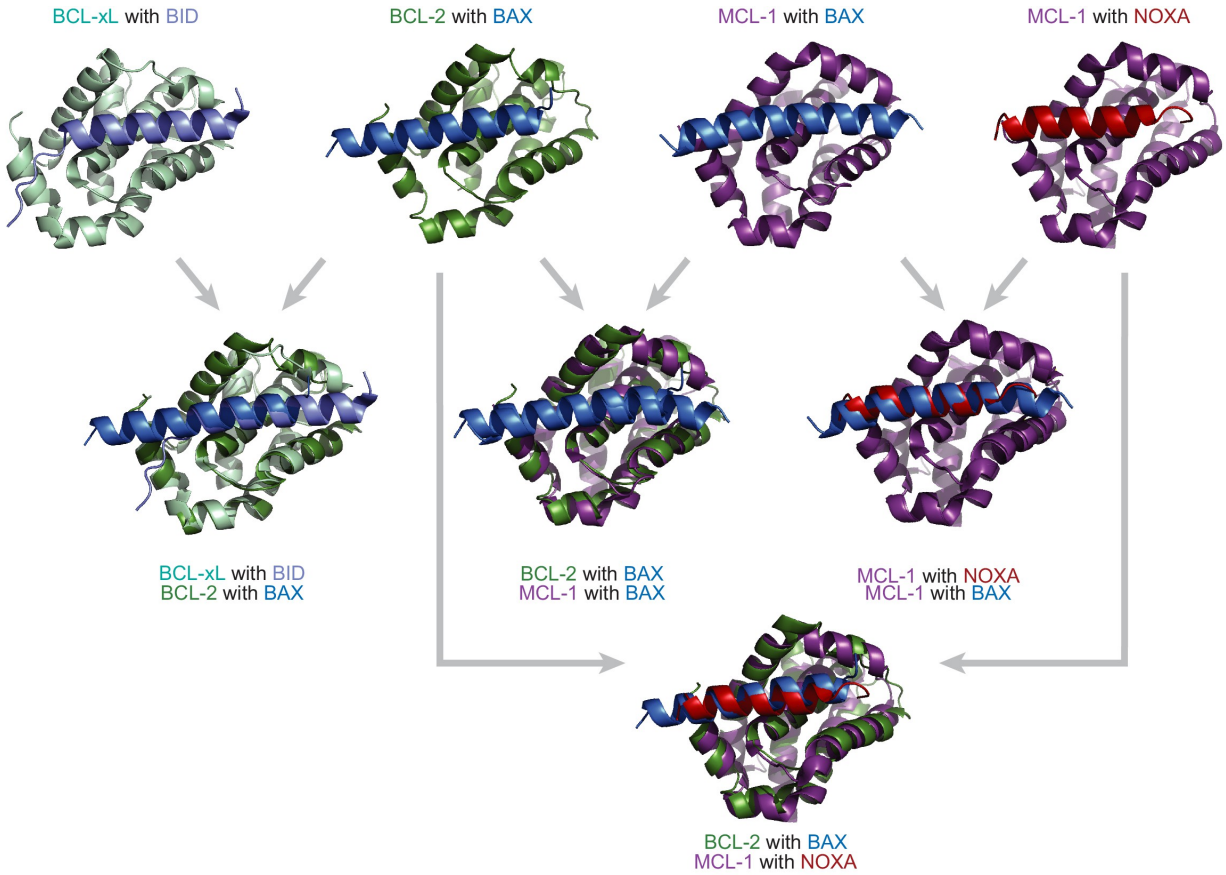


Figure 2.3 BCL-2 family proteins are structurally similar but have different binding profiles.
Crystal structures and overlays of BCL-xL (a vertebrate paralog of BCL-2, light green) bound to BID (light blue; PDB: 4qve); BCL-2 (green) bound to BAX (a protein with a BID-like binding profile, blue; PDB: 2xa0); MCL-1 (purple) bound to BAX (blue; PDB: 3pk1); and MCL-1 bound to NOXA (red; PDB: 2nla). The BCL-2 family proteins bind the coregulator proteins at the same interface.

To drive the evolution of new PPI specificities, we developed a new high-throughput phage-assisted continuous evolution (PACE) system[38] that can simultaneously select for and against particular PPIs[63,173] . We applied this technique to a series of reconstructed ancestral BCL-2 family members, repeatedly evolving each starting genotype to acquire PPI specificities found among extant family members. By comparing sequence outcomes among PACE replicates from the same starting point, we quantified the role of chance in the evolution of historically relevant molecular functions under strong and identical selec-

tion pressures; by comparing outcomes of PACE initiated from different starting points, we quantified the effect of contingency generated by the sequence changes that accumulated during these proteins histories. This design also allowed us to characterize how these factors have changed over phylogenetic time and dissect the underlying genetic basis by which they emerged.

2.2 Results

2.2.1 BID specificity is derived from an ancestor that bound both BID and NOXA

We first characterized the historical evolution of PPI specificity in the BCL-2 family using ancestral protein reconstruction (**Figure 2.4**). We inferred the maximum likelihood phylogeny of the family, which recovered the expected sister relationship between the metazoan BCL-2 and MCL-1 classes (**Figure 2.2C, Figure 2.5**). We then reconstructed the most recent common ancestor (AncMB1) of the two classes a gene duplication that occurred before the last common ancestor (LCA) of all animals and 11 other ancestral proteins that existed along the lineages leading from AncMB1 to human BCL-2 (hsBCL-2) and to human MCL-1 (hsMCL-1) (**Supplementary file 1**).

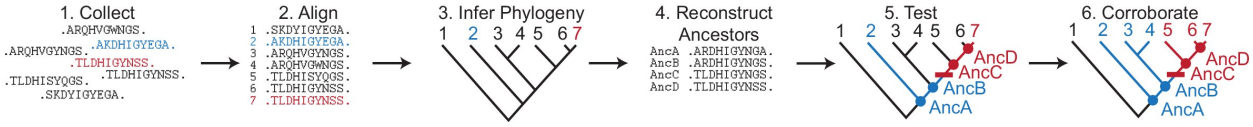


Figure 2.4 Ancestral sequence reconstruction procedure in schematic form. (1) Sequences are collected, including those of proteins conferring different functions (red v. blue)and others with unknown functions (black). These sequences may be orthologs from various species, paralogs related by gene duplication events, or both. (2) Sequences are aligned. (3) A phylogeny is inferred. (4) Using the inferred phylogeny, the aligned sequences, and a model of sequence evolution, the most likely state at each ancestral node is determined. (5) Ancestral sequences are synthesized and tested for function. (6) Functional differences among successive ancestral proteins indicate functional changes during evolutionary history (red bar).

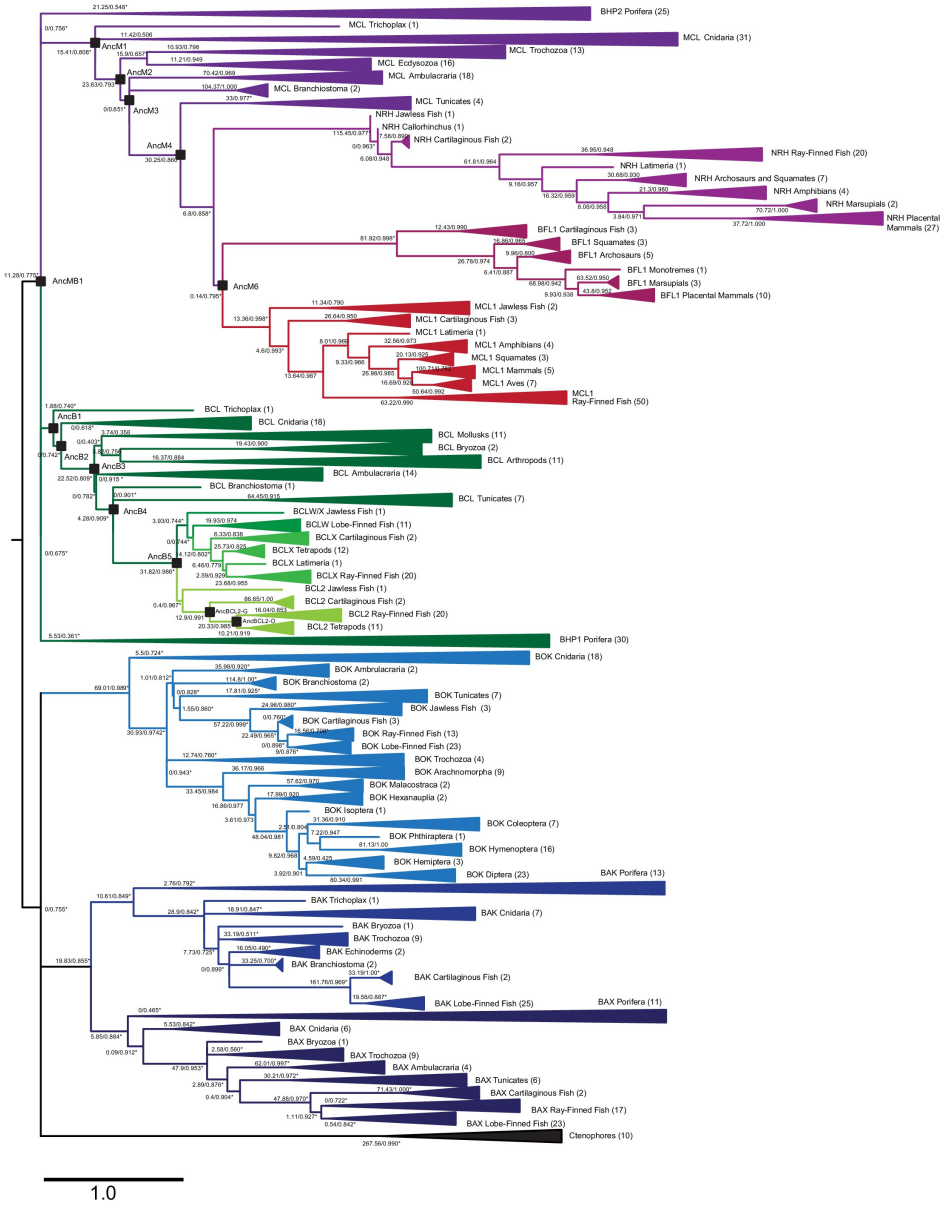


Figure 2.5 Maximum likelihood phylogeny of BCL-2 family proteins.
Light green, vertebrate BCL-2; light-medium and dark-medium green, vertebrate BCLX and BCLW, respectively; dark green, non-vertebrate sequences most closely related to vertebrate BCL-2; red, vertebrate MCL-1; maroon, vertebrate BFL1; light purple, vertebrate NRH; dark purple, non-vertebrate sequences most closely related to vertebrate MCL-1; dark blue, BAX; medium blue, BAK; light blue, BOK; black, ctenophore sequences. Parentheses, number of sequences in each clade. Black squares, ancestral sequences reconstructed and tested. Node labels, approximate likelihood ratio statistics and transfer bootstrap values. Asterisks, nodes constrained to be congruent with known taxonomic relationships.

We synthesized genes coding for these proteins and experimentally assayed their ability to bind BID and NOXA using a proximity-dependent split RNA polymerase (RNAP) luciferase assay (**Figure 2.6**).[63] AncMB1 bound both BID and NOXA, as did all ancestral proteins in the MCL-1 clade and hsMCL-1 (**Figure 2.2C, Supplementary file 1**). Ancestral proteins in the BCL-2 clade that existed before the LCA of deuterostomes also bound both BID and NOXA, whereas BCL-2 ancestors within the deuterostomes bound only BID, just as hsBCL-2 does. This reconstruction of history was robust to uncertainty in the ancestral sequences: experiments on AltAll proteins at each ancestral node which combine all plausible alternative amino acid states (posterior probability > 0.2) in a single worst-case alternative reconstruction also showed that BID specificity arose within the BCL-2 clade (**Figure 2.6, Supplementary file 2**).

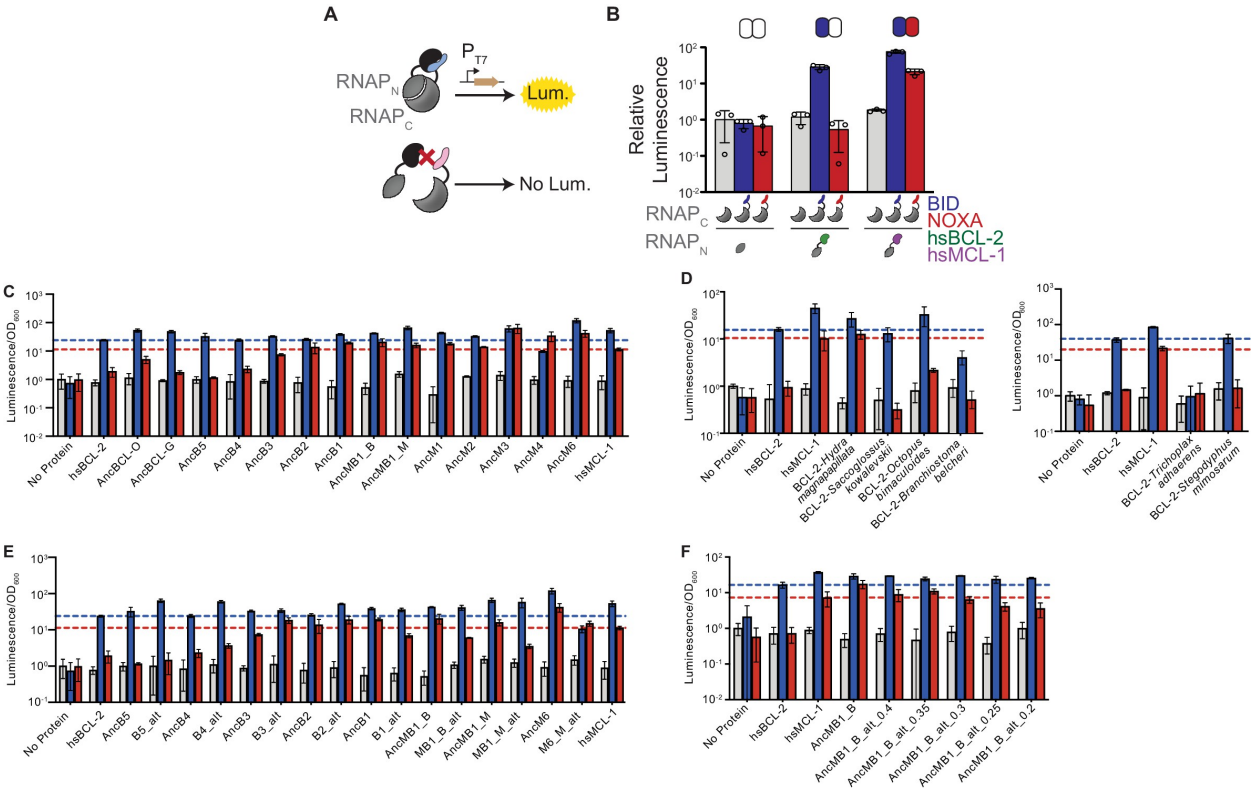


Figure 2.6 Binding of BID and NOXA to extant and ancestral proteins.
(A) Schematic of the luciferase reporter assay to assess PPIs. If a BCL-2 family protein (black) binds a coregulator protein (blue), the split T7 RNAP biosensor (gray) assembles and drives luciferase expression. If a coregulator (pink) is not bound, no luciferase is expressed. (B) Interactions of human BCL-2 and MCL-1 with BID (blue bars) and NOXA (red) in the luciferase assay, compared to no-coregulator control (gray). Activity is scaled relative to no-coregulator control with no-BCL-2 protein. Columns and error bars, mean $\pm$ SD of three biological replicates (circles). Shaded boxes above show the same data in heatmap form: BID activity is normalized relative to hsBCL-2 with BID; NOXA activity is normalized to hsMCL-1 with NOXA. (C) Interactions of ancestral reconstructed proteins with BID (blue) and NOXA (red) in the luciferase assay, compared to no-coregulator control (gray). Activity is scaled relative to no-coregulator control with no-BCL-2 family protein. Columns and error bars, mean $\pm$ SD of three biological replicates. hsBCL-2 with BID (dashed blue line). hsMCL-1 with NOXA (dashed red line). (D) Same as (C), but for extant species Hydra magnapapillata (Cnidaria), Octopus bimaculoides (Lophotrochozoa), Saccoglossus kowalevskii (Hemichordata), Branchiostoma belcheri (Cephalochordata), Trichoplax adhaerens (Placozoa), and Stegodyphus mimosarum (Ecdysozoa). (E) Same as (C), but contains alternative reconstructions (Alt) for each ancestral protein, which combine all plausible alternative amino acid states (PP > 0.2) in a single worst-case alternative reconstruction. (F) Same as (C), but contains multiple alternative reconstructions for AncMB1_B. In each case, all plausible alternative amino acid states with PP greater than the listed value are included in a single worst-case alternative reconstruction.

To further test this inferred history, we characterized the coregulator specificity of extant BCL-2 class proteins from taxonomic groups in particularly informative phylogenetic positions. Those from Cnidaria were activated by both BID and NOXA, whereas those from protostomes and invertebrate deuterostomes were BID-specific (**Figure 2.2C, Figure 2.6, Supplementary file 1**). These results corroborate the inferences made from ancestral proteins, indicating that BID specificity evolved when the ancestral ability to bind NOXA was lost between AncB2 (in the ancestral eumetazoan) and AncB4 (in the ancestral deuterostome).

2.2.2 A directed continuous evolution system for rapid changes in PPI specificity

To rapidly evolve BCL-2 family proteins to acquire the same PPI specificities that existed during the familys history, we developed a new PACE system (**Figure 2.7A-B, Figure 2.8**)[38] . Previous PACE systems have evolved binding to new protein partners using a bacterial 2-hybrid approach[61] , but evolving PPI specificity requires simultaneous selection for a desired PPI and against an undesired PPI. For this purpose, we used two orthogonal proximity-dependent split RNAPs that recognize different promoters in the same cell and if reconstituted by a PPI activate transcription of positive and negative selectable markers. Specifically, the N-terminal fragment of RNAP was fused to the BCL-2 protein of interest and encoded in the phage genome, and two C-terminal RNAP fragments (RNAPc), each fused to a different BCL-2 coregulator, were encoded on host cell plasmids. One RNAPc is fused to the selected-for coregulator and drives expression of an essential viral gene (gIII) when reconstituted by binding to the BCL-2 protein; the other RNAPc, fused to the counter-selected coregulator, drives expression of a dominant-negative version of gIII (Pu et al., 2017a). Phage containing BCL-2 variants that bind the positive selection protein but not the counterselection protein produce infectious phage.

After optimizing this system, we used activity-dependent plaque assays and phage growth assays to confirm that it imposes strong selection for the PPI specificity profiles of extant hsBCL-2 and hsMCL1 (**Figure 2.7D**).

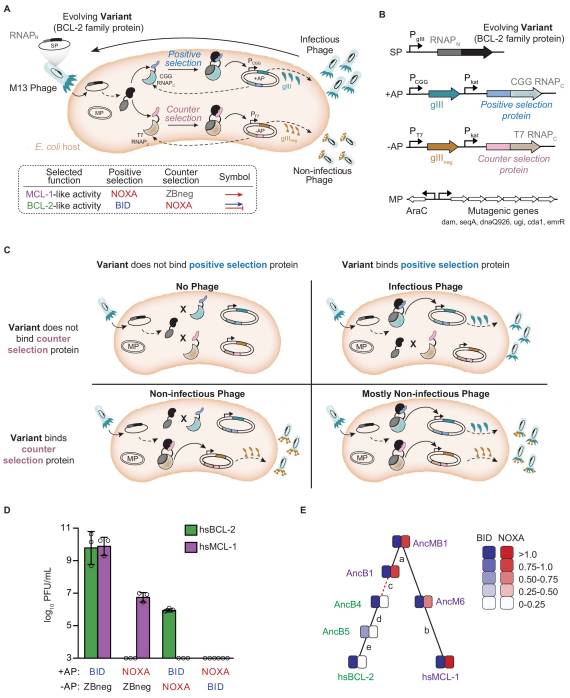


Figure 2.7 Continuous directed evolution of specificity in modern and ancestral BCL-2 family proteins.

(A) Top: Components of the PACE system for evolving PPI specificity. The protein targeted for altered specificity (black) is fused to the N-terminus of RNAP ($RNAP_N$, dark gray) and placed into the phage genome (SP, selection plasmid). Host cells carry accessory plasmids (+AP and AP) that contain the C-terminus of RNAP (RNAPc) fused to peptides for which specificity is desired (blue, positive selection protein; pink, counterselection protein). Binding of the target protein to either the selection protein or counterselection protein reconstitutes a functional RNAP. Binding of RNAP to the corresponding promoter results in the expression of either gIII (teal) or $gIII_{neg}$ (gold). gIII is necessary to produce infectious phage. $gIII_{neg}$ is a dominant-negative version of gIII which results in the production of non-infectious phage. An arabinose-inducible mutagenesis plasmid in the system (MP) increases the mutation rate of the evolving protein. Bottom: PACE schemes for evolving PPI specificities. (B) Plasmid maps of the SP, APs, and MP. (C) Selection for protein variants with the desired specificity. Infection by a phage carrying a protein variant that (Upper left) binds neither the positive selection nor the counterselection protein results in production of little to no progeny phage, (Upper right) binds only the positive selection protein results in expression of gIII and production of infectious phage, (Lower left) binds only the counterselection protein results in expression of $gIII_{neg}$ and production of non-infectious phage, (Lower right) binds the positive selection and counterselection proteins results in expression of both gIII and $gIII_{neg}$, leading to production of primarily non-infectious phage. (D) Phage growth assays to assess selection and counterselection. Detection limit 10^3 PFU/mL. Bars show mean $\pm$ SD of three replicates (circles). (E) Phylogenetic relations of starting genotypes used in PACE.

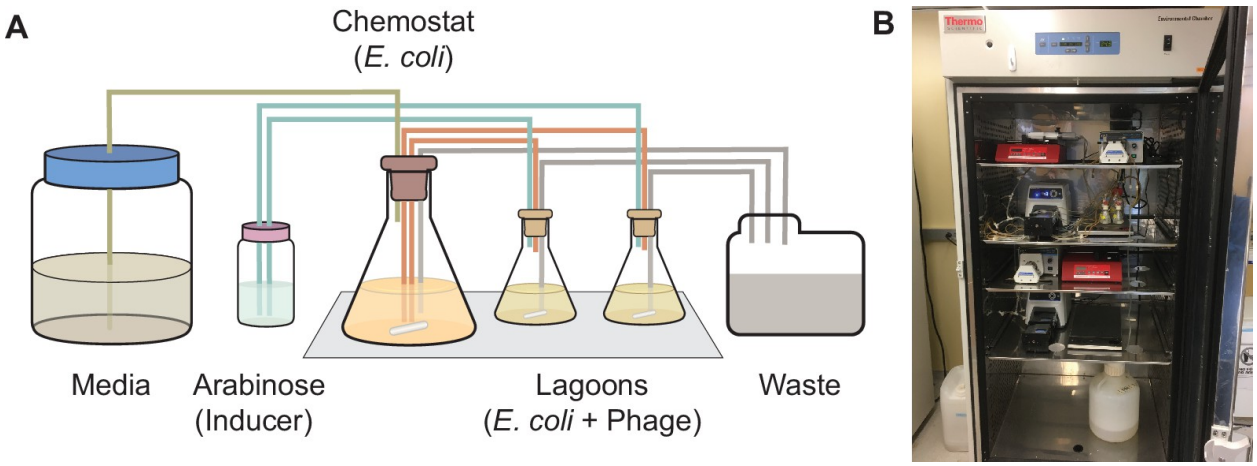


Figure 2.8 Using PACE to evolve target PPI specificity of BCL-2 family proteins. (A) Schematic of a PACE experiment. Davis Rich carboy media flows into the chemostat, which contains *E. coli* with the positive selection (+AP), counterselection (AP), and mutagenesis plasmids (MP). The cells then flow into the lagoons, which contain phage with the evolving BCL-2 family protein. Arabinose is pumped into the lagoons to induce the mutagenesis plasmid in the *E. coli*. Both chemostats and lagoons are connected to the waste to maintain proper volume, cell density, and flow rate. (B) Picture of representative PACE experiment from this work.

The simplicity of this platform allowed us to drive extant and reconstructed ancestral proteins to recapitulate or reverse the historical evolution of the BCL-2 familys PPI specificity in multiple replicates in just days, without severe experimental bottlenecks. Three proteins that bound both BID and NOXA hsMCL-1, AncM6, and AncB1 were selected to acquire the derived BCL-2 phenotype, retaining BID binding and losing NOXA binding. Conversely, hsBCL-2, AncB5, and AncB4 were evolved to gain NOXA binding, reverting to the ancestral phenotype (**Figure 2.7C,E, Figure 2.9**). For each starting genotype, we performed four replicate experimental evolution trajectories (Supplementary file 3). Each experiment was run for 4 days, corresponding to approximately 100 rounds of viral replication[38] . All trajectories yielded the target PPI specificity, which we confirmed by experimental analysis of randomly isolated phage clones using activity-dependent plaque assays and *in vivo* and *in vitro binding* assays (**Figure 2.9, Figure 2.10A,B, Figure 2.11**). As in prior PACE experiments, variation in the selected phenotype was observed among

individual phage isolates within the final populations[83] , presumably because of large populations, high mutation rates, and/or inadequate time for fixation.

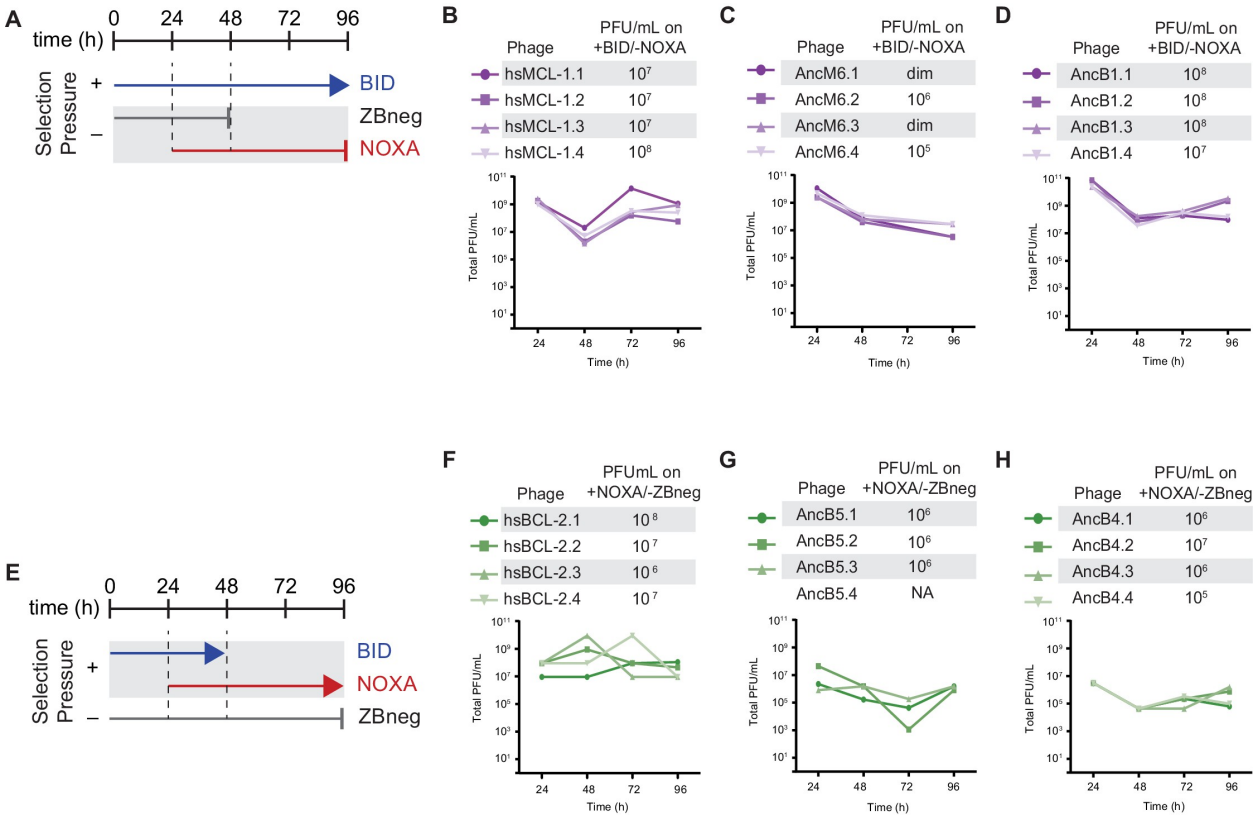


Figure 2.9 Selection schemes and phage titers for changes in PPI specificity. (A) Timeline of PACE experiments when hsMCL-1, AncM6, and AncB1 were evolved to lose NOXA binding. ZBneg is a control zipper peptide. (B) Phage titers (PFU/mL) over time (bottom) and activity-dependent phage titers at the end of the PACE experiments (top) when hsMCL-1 was evolved to lose NOXA binding. Activity-dependent plaque assays used plasmids 28-46 and Jin 487. (C) Same as (E) for AncM6. dim means plaques were visible but weak, and therefore not quantifiable. (D) Same as (E) for AncB1. (E) Timeline of PACE experiments when hsBCL-2, AncB5, and AncB4 were evolved to gain NOXA binding. (F) Phage titers (PFU/mL) over time (bottom) and activity-dependent phage titers at the end of the PACE experiments (top) when hsBCL-2 was evolved to gain NOXA binding. Activity-dependent plaque assays used plasmids 2848 and 2939. (G) Same as (H) for AncB5. (H) Same as (H) for AncB4.

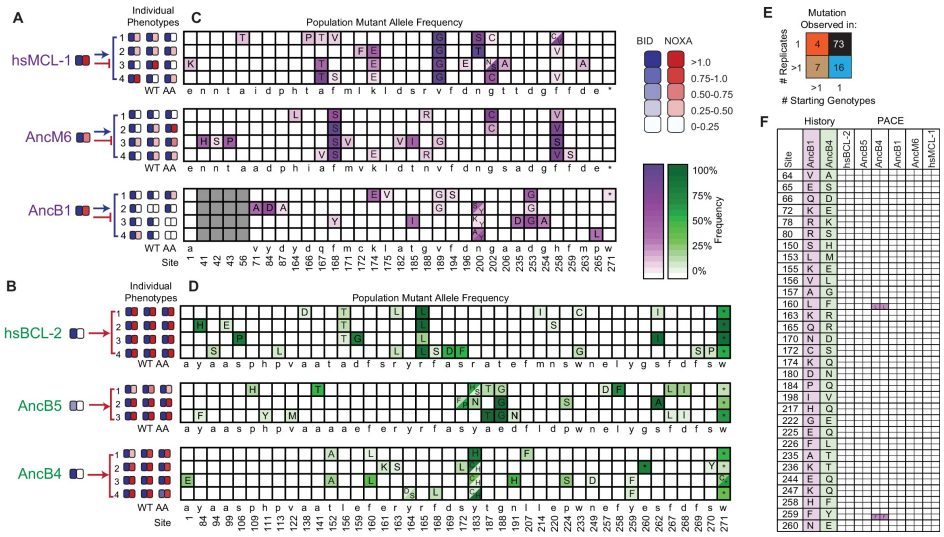
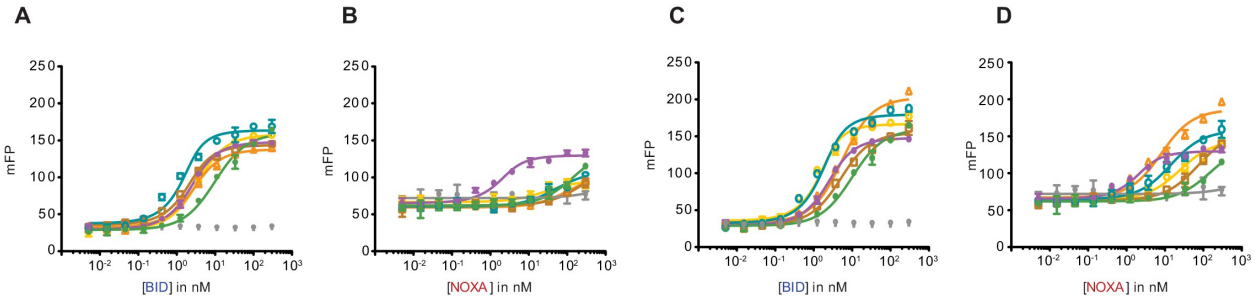


Figure 2.10 Chance and contingency shape evolutionary outcomes.
(A) Phenotypic outcome of PACE experiments when proteins with MCL-1-like specificity were selected to maintain BID and lose NOXA binding. For each starting genotype, the BID (blue) and NOXA (red) binding activity of the starting genotype and three phage variants isolated from each evolved replicate (number) are shown as heatmaps. (B) Phenotypic outcome of PACE experiments when proteins with BCL-2-like specificity were selected to gain NOXA binding. (C) Frequency of acquired states in PACE experiments when proteins with MCL-1-like specificity were selected to maintain BID and lose NOXA binding. Rows, outcomes of each replicate trajectory. Columns, sites that acquired one or more non-wild-type amino acids (letters in cells) at frequency >5%; color saturation shows the frequency of the acquired state. Site numbers and wild-type amino acid (WT AA) states are listed. Gray, sites that do not exist in AncB1. (D) Frequency of acquired states when BCL-2-like proteins were selected to gain NOXA binding. (E) Repeatability of acquired states across replicates. The 100 non-WT states acquired in all experiments were categorized as occurring in 1 or >1 replicate trajectory from 1 or >1 unique starting genotype, with the number in each category shown. The vast majority of states evolved in just one replicate from one starting point (black). (F) Historical substitutions that contributed to the change in PPI specificity rarely occur or revert during PACE. Rows, substitutions that historically occurred between AncB1 and AncB4, the ancestral proteins that flank the loss of NOXA on the phylogeny. For each substitution, columns show whether the historical ancestral or derived state was acquired in PACE trajectories from each ancestral starting point. Purple and green boxes, PACE acquisition of ancestral or derived state, respectively, in each replicate. White boxes, neither state acquired.



Variant	K_d (nM) BID	NOXA
GST only	>1000	>1000
hsBCL-2	87.6	948
hsMCL-1	14.7	9.79
L1-1	5.38	489
L3-1	25.3	949
L3-3	9.29	>1000
L4-3	18.7	>1000

Variant	K_d (nM) BID	NOXA
GST only	>1000	>1000
hsBCL-2	87.6	948
hsMCL-1	14.7	9.79
B1-3	7.21	137
B2-1	43.3	340
B3-1	33.4	53.4
B4-2	5.16	186

Figure 2.11 Fluorescence polarization of PACE-evolved variants.
(A) BID fluorescence polarization for hsMCL-1 variants evolved to lose NOXA binding. Bars are the mean of three replicates; error bars, SD. mFP, normalized measured fluorescent polarization. Kd estimates are shown below in the table. (B) Same as (A), but for NOXA binding. (C) BID fluorescence polarization for hsBCL-2 variants evolved to gain NOXA binding. (D) Same as (C), but for NOXA binding.

2.2.3 Chance and contingency erase necessity in the evolution of PPI specificity

We used deep sequencing to compare the sequence outcomes of evolution across trajectories initiated from the same and different starting points (**Figure 2.12**). Necessity was almost entirely absent. Across all trajectories, 100 mutant amino acid states at 75 different sites evolved to frequency > 5% in at least one replicate (**Figure 2.10C,D, Figure 2.13, Supplementary file 4**). Of these acquired states, 73 appeared in only a single trajectory, and only four arose in more than one replicate from multiple starting points (**Figure 2.10E, Figure 2.14**). When selection was imposed for binding to both BID and NOXA, no states were predictably acquired in all trajectories from all starting points. The only mutation universally acquired under any selection regime was a nonsense mutation

at codon 271, which was acquired in all trajectories selected for BID specificity, but experimental analysis of this mutation shows that it has no detectable effect on coregulator binding (**Figure 2.15**).

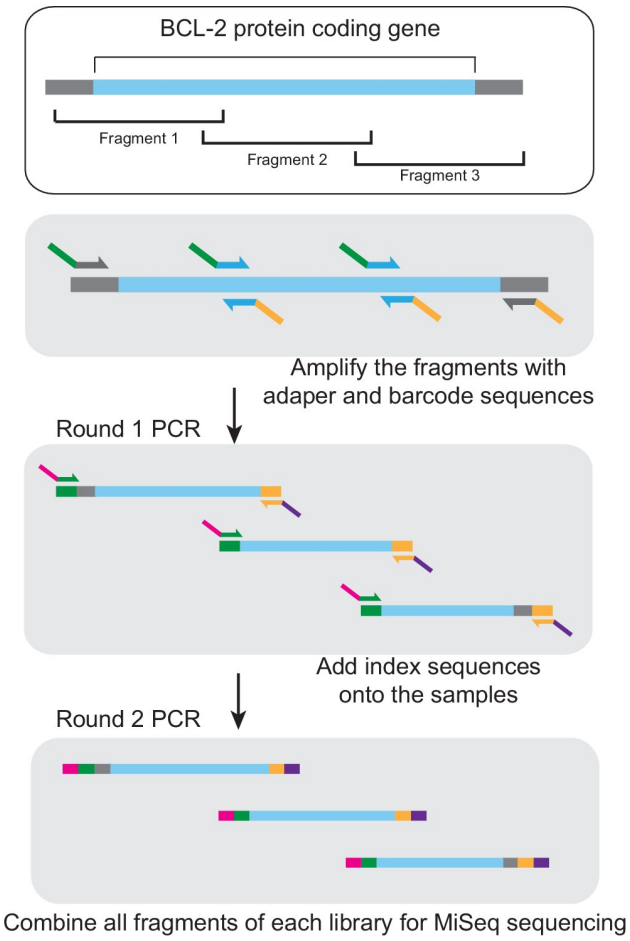


Figure 2.12 MiSeq library preparation.
After isolation of phage DNA, the coding region of the evolving BCL-2 family protein was amplified in three overlapping fragments, each of which was smaller than 300 bp. The DNA fragments were then amplified using sequence-specific primers. MiSeq adapters were added in a second PCR step. These fragment libraries were combined and used for MiSeq sequencing. Blue, target gene coding region. Gray, adjacent vector sequence. Green, forward adapter and barcode sequence. Orange, reverse adapter and barcode sequence. Magenta, index one sequence. Purple, index two sequence.

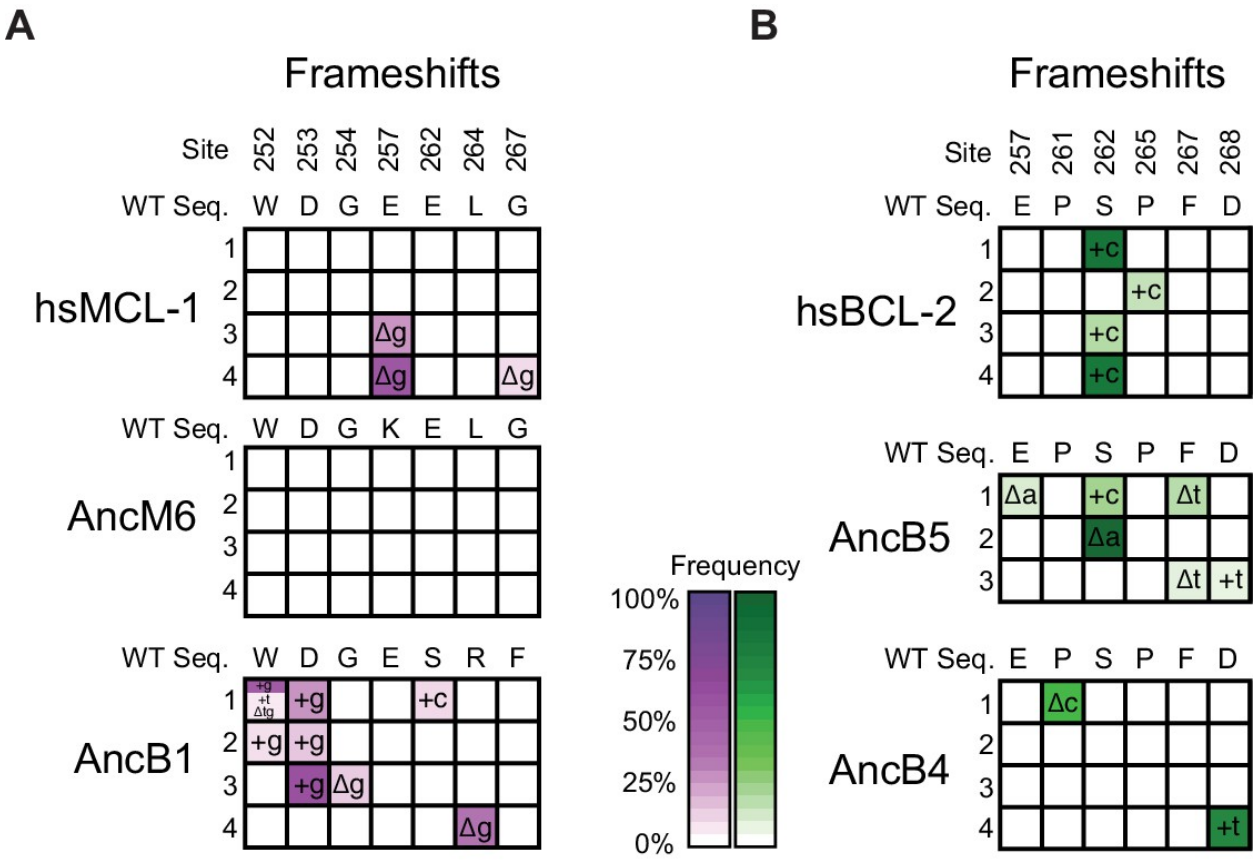


Figure 2.13 Frequency of insertions and deletions during PACE.
(A) Allele frequency of frameshifts in replicate PACE experiments started from hsMCL-1, AncM6, and AncB1 evolved to lose NOXA binding. Site numbers and wild-type (WT) amino acid states are listed above each sequence. Each row represents an independent replicate population. Non-wild-type insertions and deletions that reached >5% in frequency are shown, with frequency proportional to color saturation. Split cells show populations with multiple non-WT states > 5%. Plus (+) indicates an addition of a nucleotide. Delta (△) indicates a deletion of a nucleotide. (B) Same as (A), but for replicate PACE experiments of hsBCL-2, AncB5, and AncB4 evolved to gain NOXA binding.

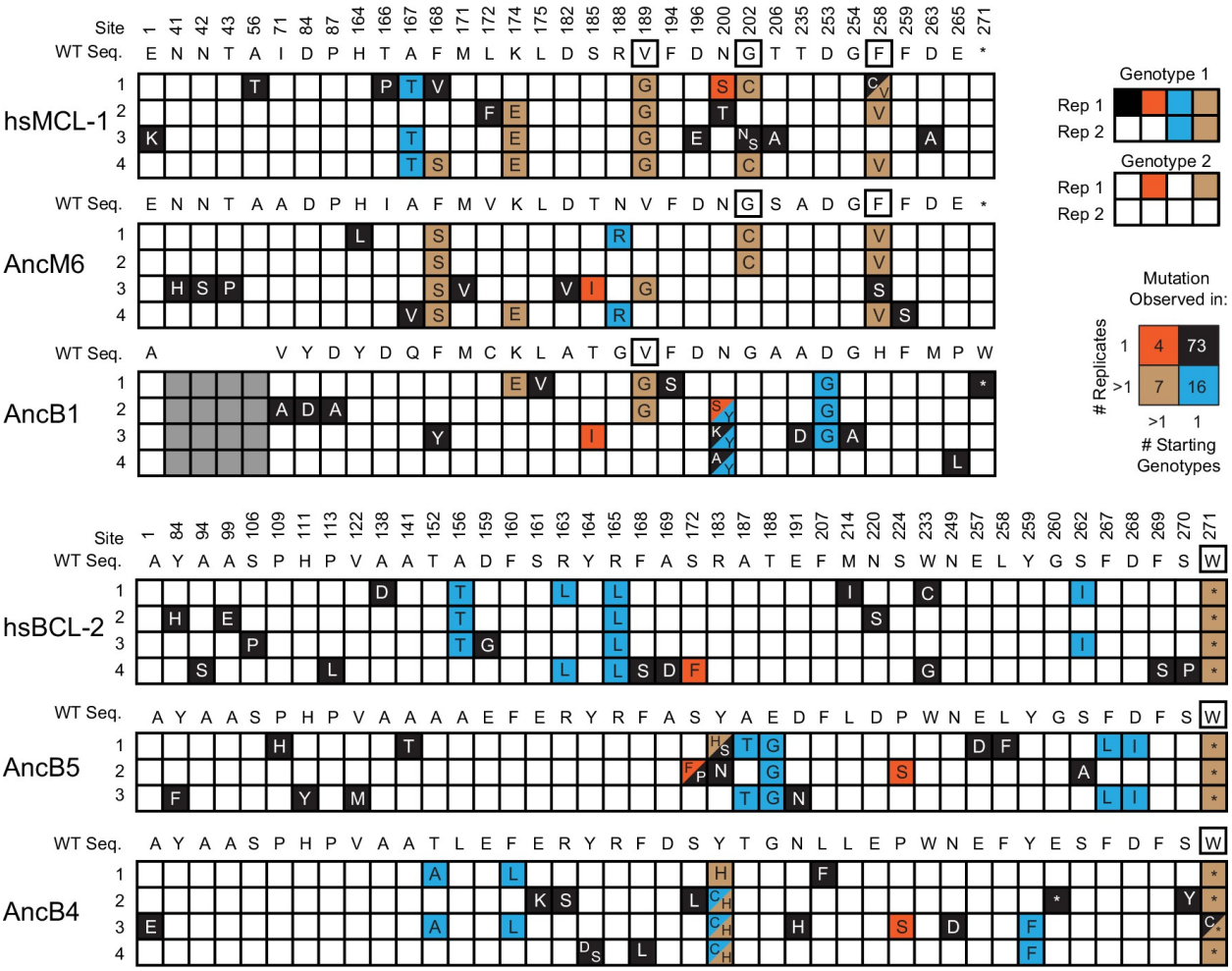


Figure 2.14 Categories of the 100 non-WT states observed for each non-WT state. Black box with white letters, mutant states observed in only one replicate. Teal, mutant states observed in multiple replicates from the same starting genotype. Orange, mutant states observed in a single replicate from multiple different starting genotypes. Brown, mutant states observed in multiple replicates from the same starting genotype and in at least one other replicate from a different starting genotype. Black box outline, mutant states observed in multiple replicates from the same starting genotype and from multiple replicates from a different starting genotype. Gray boxes are sites that do not exist in a particular sequence.

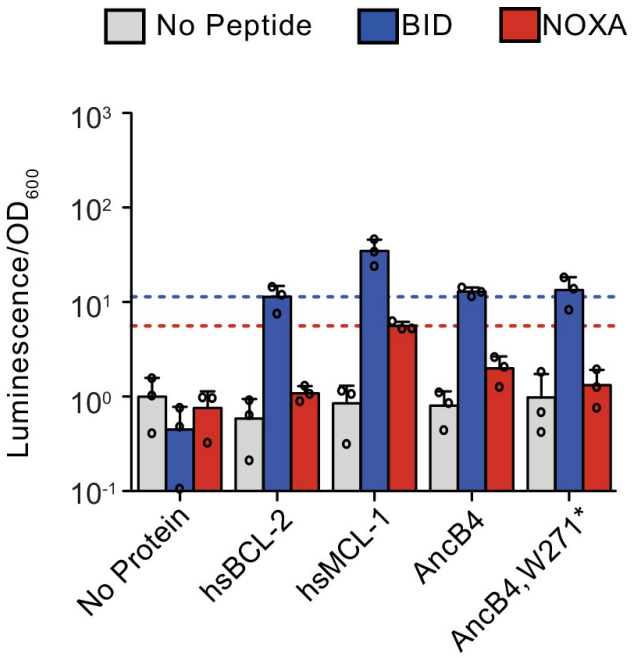


Figure 2.15 Effect of w271* mutation on BID and NOXA binding.
Activity is scaled relative to the control experiment with no- BCL-2 family protein and no-coregulator peptide. Bars show the mean ± SD of three biological replicates (circles). Gray bar, no-coregulator peptide. Blue bar, BID. Red bar, NOXA. Blue dotted lines mark the average signal of hsBCL-2 with BID, and red dotted lines mark the average signal of hsMCL-1 with NOXA.

Both chance and contingency contributed to this pervasive unpredictability. Pairs of trajectories launched from the same starting point differed, on average, at 78% of their acquired states, indicating a strong role for chance. Pairs that were launched from different starting points (but selected for the same PPI specificity) differed at an average of 92% of acquired states, indicating an additional role for contingency.

These starting points are separated by different amounts of evolutionary divergence, so to understand the extent of contingency over the timescale of metazoan evolution, we compared trajectories launched from AncB1 to those launched from hsMCL-1 (the two most distant genotypes that were selected for BID specificity). Of 34 states acquired in these experiments, only three occurred in at least one trajectory from both starting points. Of 40 states acquired in trajectories launched from AncB4 and hsBCL-2 (the two

most distant proteins that were selected to gain NOXA binding), only one occurred in any trajectories from both starting points. Together, contingency generated across long phylogenetic timescales and chance therefore make sequence evolution in the BCL-2 family almost entirely unpredictable.

These experiments indicate an almost complete lack of necessity in the evolution of PPI specificity in PACE. To gain insight into the extent of necessity in the historical evolution of BCL-2 PPI specificity, we asked whether substitutions that occurred during the phylogenetic interval when NOXA binding was lost (between AncB1 and AncB4) were either repeated or reversed during PACE trajectories to lose or regain NOXA binding from any starting point (**Figure 2.10F, Figure 2.16, Figure 2.17**). In PACE experiments to lose NOXA binding from proteins that initially bound both peptides, none of the acquired states recapitulated substitutions from the branch on which NOXA binding was historically lost. In PACE experiments to reacquire NOXA binding from proteins with BCL2-like specificity for BID, only two states reversed historical substitutions that occurred on that branch. Both of these reacquisitions occurred in PACE trajectories launched from AncB4, the immediate daughter node of this branch, suggesting that in other proteins, contingency accumulated over phylogenetic time restricted their accessibility. Furthermore, both of these states were acquired in only a subset of trajectories from AncB4, indicating a role for chance even from this starting point. Some substitutions that occurred during other historical intervals were recapitulated or reversed during PACE trajectories, indicating that these states are compatible with BCL-2 family protein functions, but these substitutions could not have contributed to historical changes in PPI specificity, which remained unchanged on these branches. Our experiments therefore suggest strong effects of chance and contingency in the historical evolution of BCL-2's derived PPI specificity.

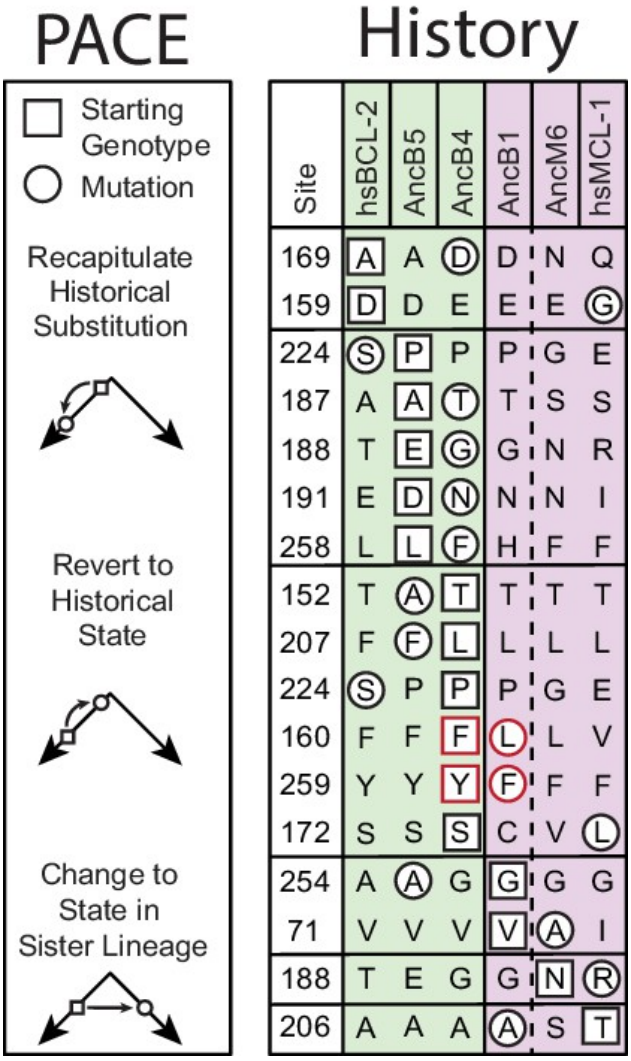


Figure 2.16 Historical distribution of PACE mutations.
Historical WT states for each starting genotype are listed. Green, hsBCL-2 like function. Purple, hsMCL-1 like function. Solid vertical line, historical interval in which function changed. Dashed vertical line, location of the phylogeny root. For each PACE mutation, the genotype on which it arose is in a square. The nearest historical state that the mutation matches is in a circle. PACE mutations can either recapitulate historical substitutions, revert to historical states, or switch to a state found in a sister lineage. PACE mutations that revert historical states that changed during the interval at the same time as the change in function or outlined in red.

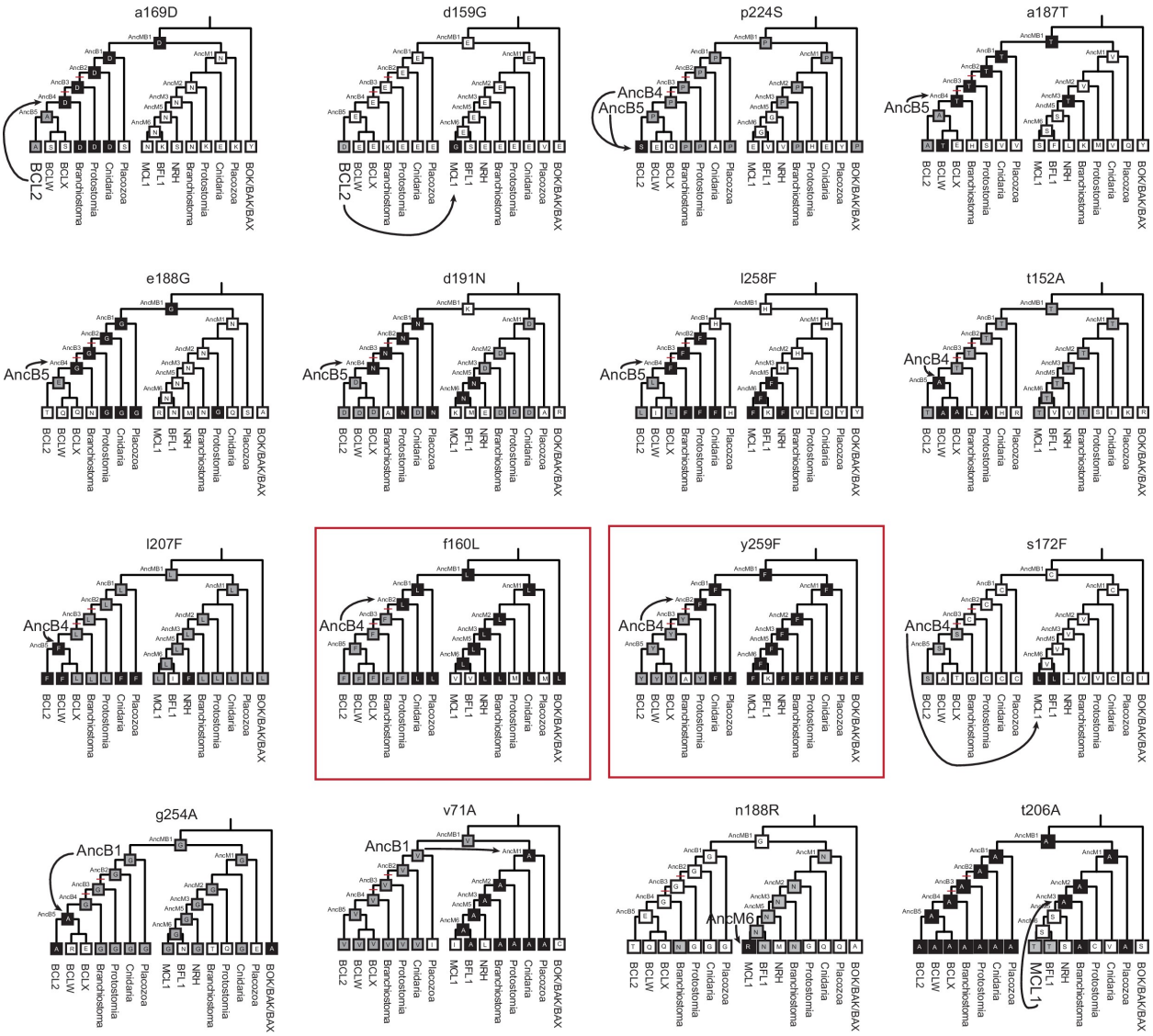


Figure 2.17 Phylogenetic recapitulation of PACE mutations.
Mutation state and position are given above each cladogram. Lowercase letters, WT state for PACE. Uppercase letters, mutant state for PACE. Each cladogram shows the estimated most likely state for each ancestral node and the majority state for each extant clade. Gray boxes; same WT state as the sequence in which the PACE mutation emerged. Black boxes; same WT state as the PACE mutation. Arrows point from the starting genotype for PACE (larger font) to the closest genotype with the PACE mutation. Red boxes show the two instances in which substitutions that occurred during the historical interval in which NOXA specificity was lost (red hash marks on phylogeny) also occurred during PACE.

2.2.4 Historical contingency is the major cause of sequence variation under selection for new functions

We next sought to directly quantify the relative effects of chance and historically generated contingency on sequence outcomes in our experiments. We analyzed the genetic variance defined as the probability that two variable sites, chosen at random, are different in state within and between trajectories from the same and different starting genotypes. To estimate the effects of chance, we compared the genetic variance between replicates initiated from the same starting genotype (V_g) to the within-replicate genetic variance (V_r). We found that V_g was on average 30% greater than V_r, indicating that chance causes evolution to produce divergent genetic outcomes between independent lineages even with strong selection for a change in function (**Figure 2.18A**). We quantified contingency by comparing the pooled genetic variance among replicates from different starting genotypes (V_t) to that among replicates from the same starting genotype (V_g). Contingencys effect was even larger than that of chance, increasing V_t by an average of 80% across all pairs of starting points compared to V_g when selecting for a new function. Together, chance and contingency had a multiplicative effect, increasing the genetic variance among trajectories from different starting genotypes (V_t) by an average of 2.4-fold compared to the genetic variance within trajectories (V_r). The effects of chance and contingency were not significantly different between PACE experiments in which protein interactions were gained and those in which they were lost (**Figure 2.19**).

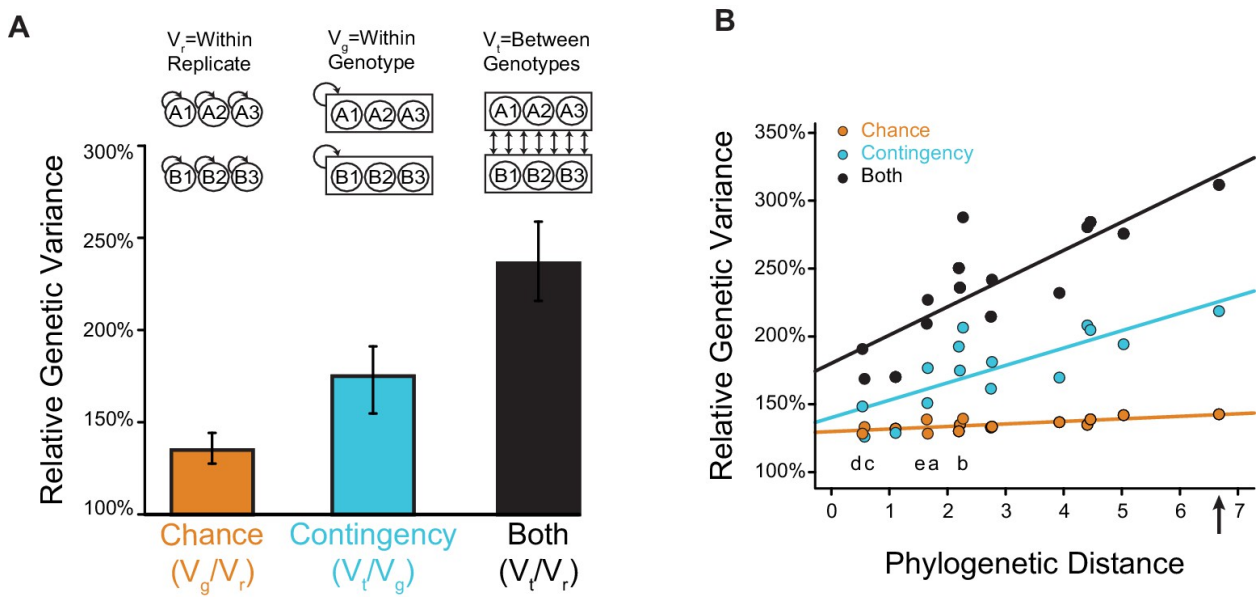


Figure 2.18 Effects of chance and contingency.
(A) Variation in evolutionary sequence outcomes caused by chance (orange), contingency (teal), and both (black). Inset: schematic for estimating the effects of chance and contingency. Chance was estimated as the average genetic variance among replicates from the same starting genotype (V_g) divided by the within-replicate genetic variance (V_r). Contingency was estimated as the average genetic variance among replicates from different starting genotypes (V_t) divided by the average genetic variance among replicates from the same starting genotype (V_g). Combined effects of chance and contingency were estimated as the average genetic variance among replicates from different starting genotypes (V_t) compared to the within-replicate genetic variance (V_r). Genetic variance is the probability that two randomly drawn alleles are different in state. Error bars, 95% confidence intervals on the mean by bootstrapping PACE replicates. (B) Change in the effects of chance and contingency over phylogenetic distance. Each point is for a pair of starting proteins used for PACE, comparing the phylogenetic distance (the total length of branches separating them, in substitutions per site) to the effects of chance (orange), contingency (teal), or both (black), when PACE outcomes are compared between them. Solid lines, best-fit linear regression. Letters indicate the phylogenetic branch indexed in Figure 2.7E. The combined effect of chance and contingency increased significantly with phylogenetic distance (slope=0.19, $p=2x10^5$), as did the effect of contingency alone (slope=0.11, p=0.007). The effect of chance alone did not depend on phylogenetic distance (slope=0.02, p=0.5). The combined effect of chance and contingency increased significantly faster than the effect of contingency alone (slope=0.08, p=0.04). Arrow, phylogenetic distance between extant hsMCL-1 and hsBCL-2 proteins, which share AncMB1 as their most recent common ancestor.

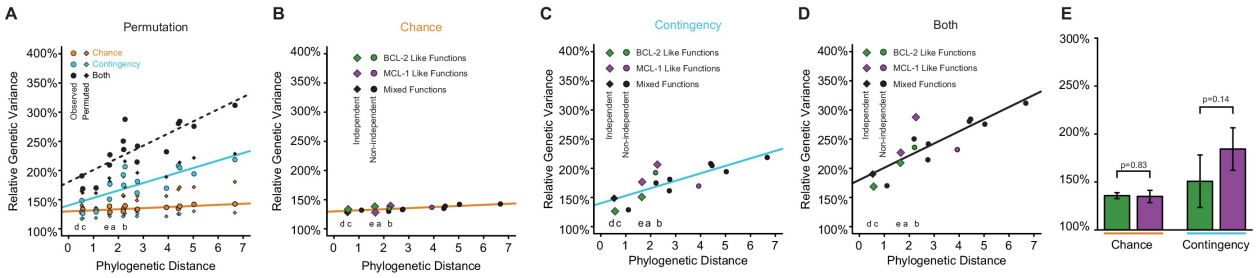


Figure 2.19 Change in chance and contingency over time.
(A) Relationship between phylogenetic distance between pairs of starting genotypes for experimental evolution (ancestral or extant proteins, as the total branch lengths separating them) and the effects of chance (orange), contingency (teal), or both (black) on the outcomes of evolution between them. Lines are best fits from linear models. Circles are observed values. Diamonds are averages of 1000 permutations of starting genotype labels. This shuffling of genotype labels results in more genetic variance among samples from the same starting genotype than the observed data, and less genetic variance between samples from different starting genotypes than the observed data. Letters indicate the specific branch from Figure 3E. (B) Change in chance over time. Green, both starting genotypes had BCL-2 like function. Purple, both starting genotypes had MCL-1 like function. Black, starting genotypes differed in function. Phylogenetically independent comparison are shown as diamonds. The effect of chance did not change with phylogenetic distance when restricting analysis to comparisons that are phylogenetically independent (slope=0.042, p=0.71) and genotypes selected for the same function (slope=0.029, p=0.82). (C) Change in contingency over time. Green, both starting genotypes had BCL-2 like function. Purple, both starting genotypes had MCL-1 like function. Black, starting genotypes differed in function. Phylogenetically independent comparison are shown as diamonds. The effect of contingency increased with phylogenetic distance and was marginally significant when restricting analysis to comparisons that are phylogenetically independent (slope=0.31, p=0.07), and genotypes selected for the same function (slope=0.42, p=0.05). (D) Change in the combined effect of chance and contingency over time. Green, both starting genotypes had BCL-2 like function. Purple, both starting genotypes had MCL-1 like function. Black, starting genotypes differed in function. Phylogenetically independent comparison are shown as diamonds. The combined effect of chance and contingency increased with phylogenetic distance when restricting analysis to comparisons that are phylogenetically independent (slope=0.50, p=0.009) and genotypes selected for the same function (slope=0.63, p=0.01). (E) Effects of chance and contingency do not depend on the selection regime. Each column shows the portion of genetic variance among trajectories that was caused by chance or contingency, relative to the within-population variance (see **Figure 2.18A**). Green, trajectories in which BCL-2 like starting genotypes were selected to gain NOXA binding. Purple, trajectories in which MCL-1 like starting genotypes were selected to lose NOXA binding but maintain BID binding (purple). Error bars, 95% confidence intervals on the mean. p-values estimated by t-test.

The preceding analyses do not account for phylogenetic structure or the extent of divergence between starting points. We therefore assessed how chance and contingency changed with phylogenetic distance using linear regression (**Figure 2.18B, Figure 2.19**). We found that the effect of contingency on genetic variance increased significantly with phylogenetic divergence among starting points. The effect of chance did not increase with divergence, but the combined effect of contingency and chance increased even more rapidly than contingency alone because the total impact on genetic variance of these two factors is multiplicative by definition.

We next compared the impact of contingency to that of chance as phylogenetic divergence increases. On the timescale of metazoan evolution, contingencys effect (an increase in genetic variance by about 100%) was three times greater than that of chance when evolution was launched from extant starting points whose LCA was AncMB1, near the base of Metazoa (**Figure 2.18B**). The combined effect of chance and contingency on this timescale was a 3.2-fold increase in variance among single trajectories launched from these starting points. Even across the shortest phylogenetic intervals we studied, contingencys effect was larger than that of chance, although to a smaller extent. Taken together, these data indicate that contingency, magnified by chance, steadily increases the unpredictability of evolutionary outcomes as protein sequences diverge across history.

2.2.5 Contingency is caused by epistasis between historical substitutions and specificity-changing mutations

Contingency is expected to arise in our experiments if historical substitutions (which separate ancestral starting points) interact epistatically with mutations that occur during PACE, causing the mutations that can confer selected PPI specificities to differ among starting points. To experimentally test this hypothesis and characterize underlying epistatic interactions, we first identified sets of candidate causal mutations that arose repeatedly

during PACE replicates from each starting genotype. We then verified their causal effect on specificity by introducing only these mutations into the protein that served as the starting point for the PACE experiment in which they were observed and measuring their effects on BID and NOXA binding. We found that all sets were sufficient to confer the selected-for specificity in their native background (**Figure 2.20A,B**).

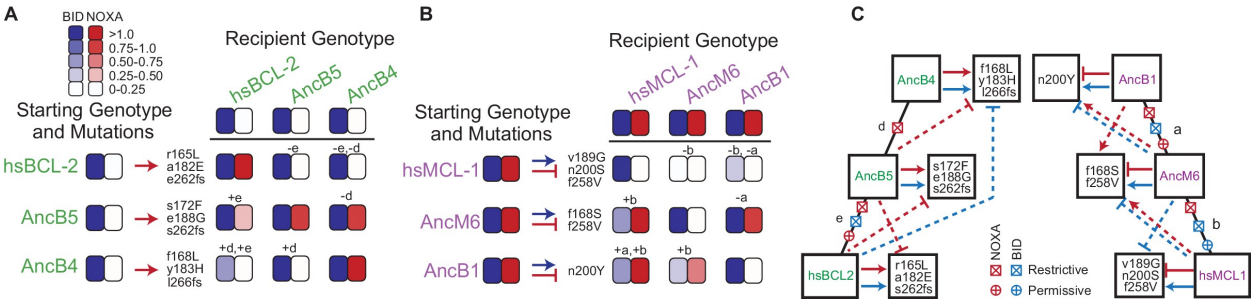


Figure 2.20 Sources of contingency.
(A) Epistatic incompatibility of PACE mutations in other historical proteins. Effects on activity are shown when amino acid states acquired in PACE under selection to acquire NOXA binding (red arrows) are introduced into ancestral and extant proteins. The listed mutations that occurred during PACE launched from each starting point (rows) were introduced as a group into the protein listed for each column. Observed BID (blue) and NOXA (red) activity in the luciferase assay for each mutant protein are shown as heatmaps (normalized mean of three biological replicates). Letters indicate the phylogenetic branch in Figure 3E that connects the PACE starting genotype to the recipient genotype. Plus and minus signs indicate whether mutations were introduced into a descendant or more ancestral sequence, respectively. (B) Effects on activity when amino acids acquired in PACE under selection to lose NOXA binding and acquire BID binding are introduced into different ancestral and extant proteins, represented as in (A). (C) Epistatic interactions between historical substitutions and PACE mutations. Restrictive historical substitutions (X) cause mutations that alter PPI specificity in an ancestor to abolish either BID (blue) or NOXA (red) activity when introduced into later historical proteins. Permissive substitutions (+) cause PACE mutations that alter PPI specificity in a descendent to abolish either BID or NOXA activity in an ancestor. Arrow, gain or maintenance of binding. Blunt bar, loss of binding. Mutations that confer selected functions in PACE are shown in the boxes at the end of solid arrows or bars. Solid lines, functional changes under PACE selection. Dashed lines, functional effects different from those selected for when PACE-derived mutations are placed on a different genetic background.

We then introduced these mutations into the other starting proteins that had been subject to the same selection regime and performed the same assay (**Figure 2.20A,B**).

Eleven of 12 such swaps failed to confer the PPI specificity on other proteins that they conferred in their native backgrounds. These swaps compromised binding of BID, failed to confer the selected-for gain or loss of NOXA binding, or both. The only case in which the mutations that conferred the target phenotype during directed evolution had the same effect in another background was the swap into AncB4 of mutations that evolved in AncB5 the most similar genotypes of all pairs of starting points in the analysis. Contingency therefore arose because historical substitutions that occurred during the intervals between ancestral proteins made specificity-changing mutations either deleterious or functionally inconsequential when introduced into genetic backgrounds that existed before or after those in which the mutations occurred.

To characterize the timing and effect of these epistatic substitutions during historical evolution, we mapped the observed incompatibilities onto the phylogeny (**Figure 2.20C**). We inferred that restrictive substitutions evolved on a branch if mutations that arose during directed evolution of an ancestral protein compromised coregulator binding when swapped into descendants of that branch. Conversely, we inferred that permissive substitutions evolved on a branch if mutations that arose during directed evolution compromised coregulator binding when swapped into more ancient ancestral proteins.

We found that both permissive and restrictive epistatic substitutions occurred on almost every branch of the phylogeny and affected both BID and NOXA binding. The only exception was the branch from AncB4 to AncB5, on which only restrictive substitutions affecting NOXA binding occurred. This is the branch immediately after NOXA function changed during history; it is also the shortest of all branches examined and the one with the smallest effect of contingency on genetic variance (**Figure 2.18B**). Even across this branch, however, the PACE mutations that restore the ancestral PPI specificity in AncB4 can no longer do so in AncB5. These results indicate that the paths through sequence space leading to historical PPI specificities changed repeatedly during the BCL-2 familys

history, even during intervals when the proteins PPI binding profiles did not evolve.

2.2.6 Chance is caused by degeneracy in sequence-function relationships

For chance to strongly influence the outcomes of adaptive evolution, multiple paths to a selected phenotype must be accessible with similar probabilities of being taken. This situation could arise if several different mutations (or sets of mutations) can confer a new function or if mutations that have no effect on function accompany function-changing mutations by chance. To distinguish between these possibilities, we measured the functional effect of different sets of mutations that arose in replicates when hsMCL-1 was evolved to lose NOXA binding (**Figure 2.21A, Figure 2.22**). One mutation (v189G) was found at high frequency in all four replicates, but it was always accompanied by other mutations, which varied among trajectories. We found that v189G was a major contributor to the loss of NOXA binding, but it had this effect only in the presence of the other mutations, which did not decrease NOXA binding on their own. Mutation v189G therefore required permissive mutations to occur during directed evolution, and there were multiple sets of mutations with the potential to exert that effect; precisely which permissive mutations occurred in any replicate was a matter of chance. All permissive mutations were located near the NOXA binding cleft, suggesting a common mechanistic basis (**Figure 2.21B**).

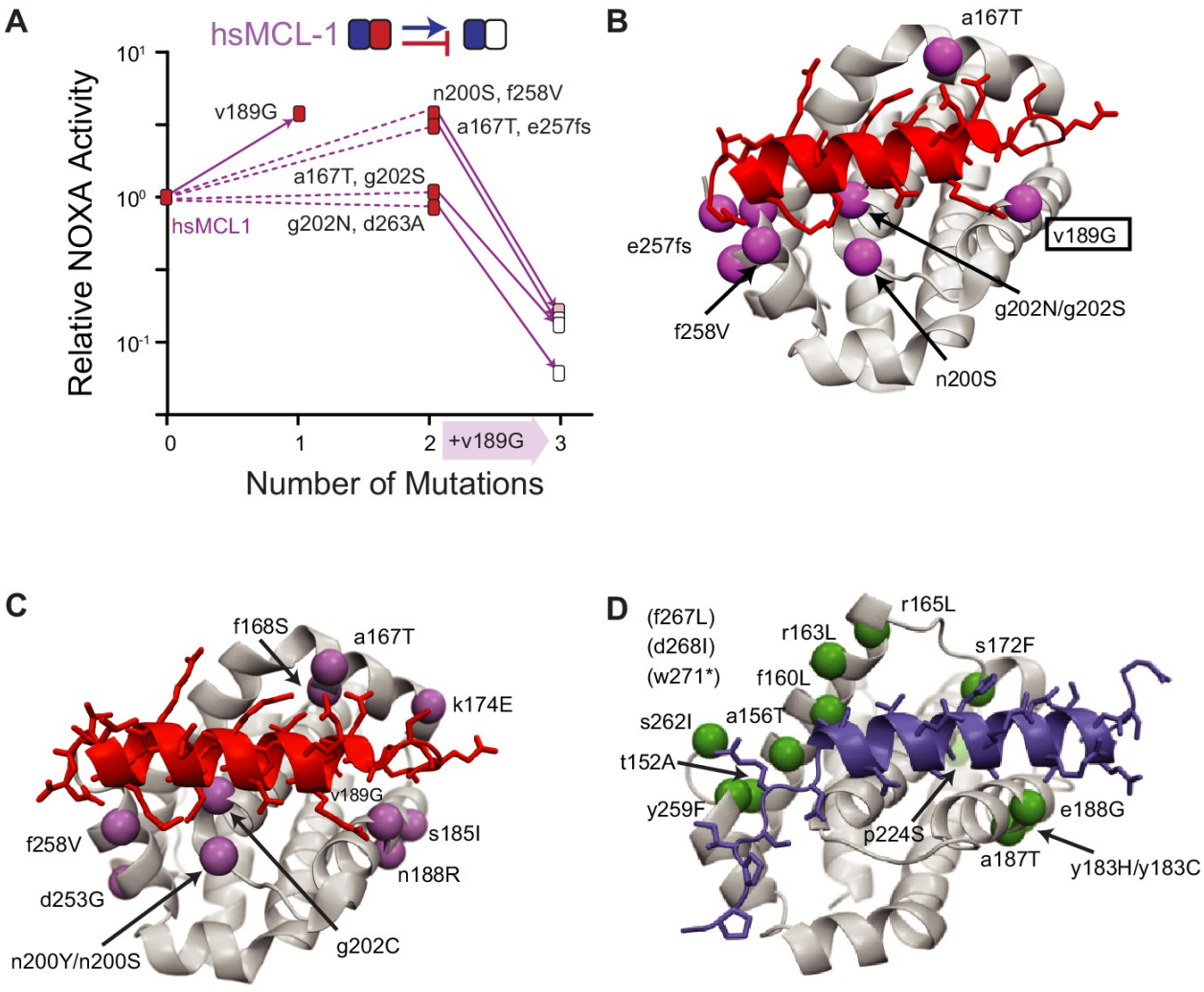


Figure 2.21 Sources of chance.
(A) Dissecting the effects of sets of mutations (white boxes) that caused hsMCL-1 to lose NOXA binding during four PACE trajectories. Filled boxes show the effect of introducing a subset of mutations into hsMCL-1 (normalized mean relative from three biological replicates). Solid lines show the effect of introducing v189G, which was found in all four sets. Dotted lines, effects of the other mutations in each set. (B) Structural location of mutations in (A). Alpha-carbon atom of mutated residues are shown as purple spheres on the structure of MCL-1 (light gray) bound to NOXA (red, PDB 2nla). (C) Location of repeated mutations when hsMCL-1, AncM6, and AncB1 were selected to lose NOXA binding (purple spheres), represented on the structure of MCL-1 (gray) bound to NOXA (red, PDB 2nla). (D) Location of repeated mutations when hsBCL-2, AncB5, and AncB4 were selected to gain NOXA binding (green spheres), on the structure of hsBCL-xL (gray) bound to BID (blue, PDB 4qve).

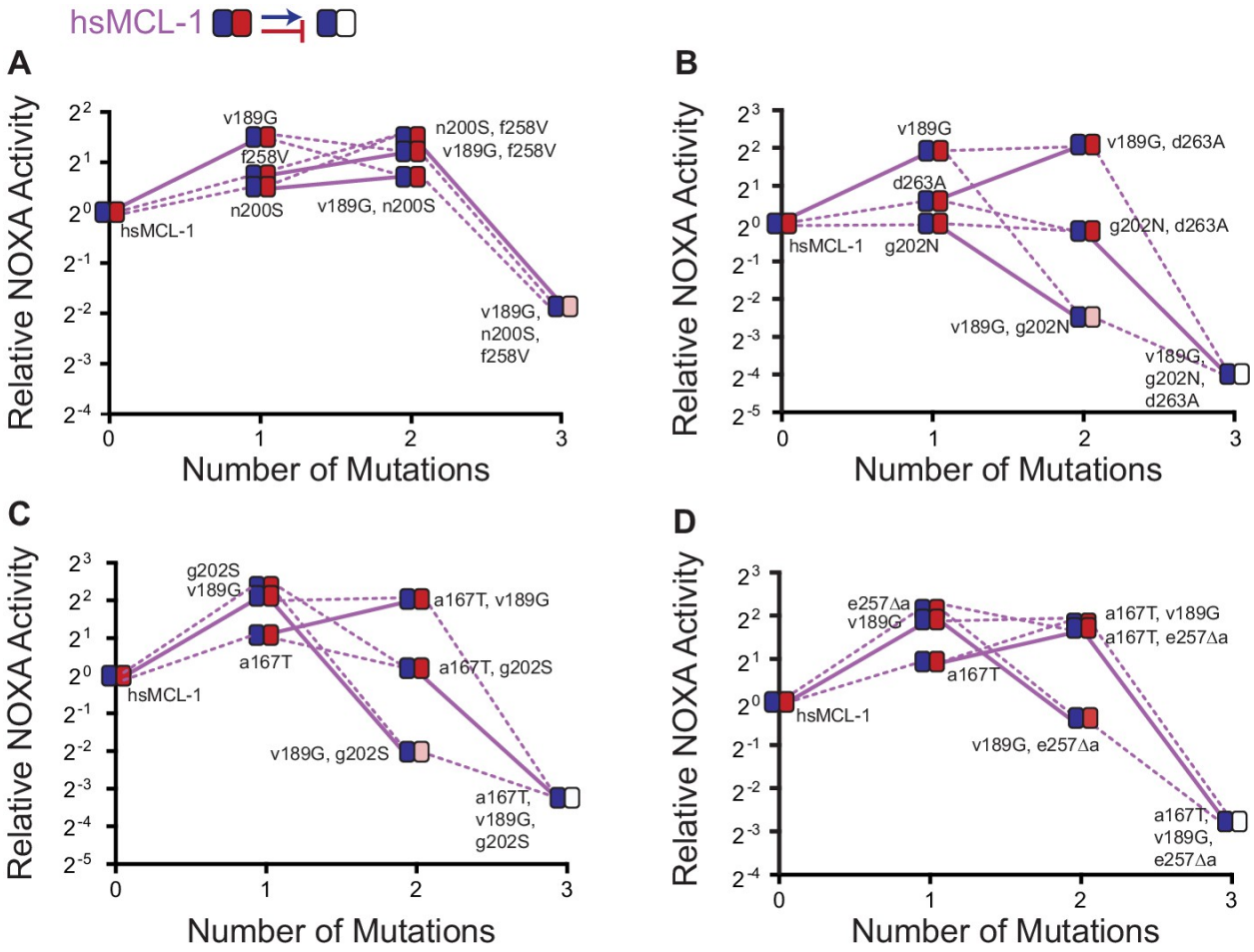


Figure 2.22 Effects on NOXA binding of hsMCL-1 PACE-derived mutations. Each panel shows NOXA binding (y-axis) for a unique variant as additional mutations are added (x-axis). Values are the mean of three biological replicates. Heatmaps show the effects of each mutation on BID (blue) and NOXA (red) activity, and each shaded box represents the normalized mean of three biological replicates. Lines connect genotypes that differ by a single mutation. Solid lines show the effects of the v189G mutation. Dashed lines show the effects of all other mutations. Mutations come from variants L1-1 (A), L3-1 (B), L3-3 (C), and L4-3 (D).

Other starting genotypes showed a similar pattern of multiple sets of mutations capable of conferring the selected function (**Figure 2.23**). In addition, when mapped onto the protein structure, all sites that were mutated in more than one replicate either directly contacted the bound peptide or were on secondary structural elements that did so (**Figure 2.21CD**), suggesting a limited number of structural mechanisms by which PPIs

can be altered. Taken together, these results indicate that chance arose because from each starting genotype, there were multiple mutational paths to the selected specificity; partial determinism arose because the number of accessible routes was limited by the structure-function relationships required for peptide binding in this family of proteins.

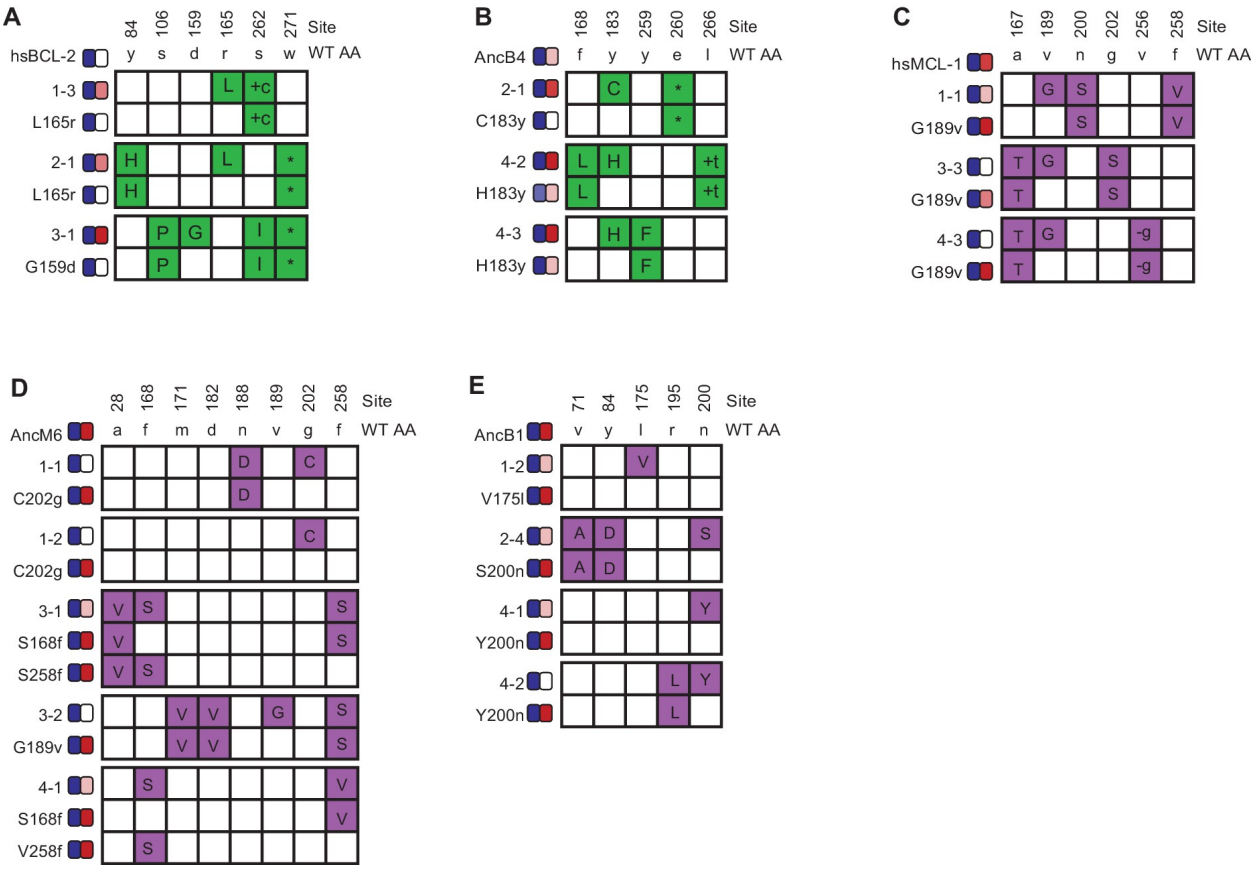


Figure 2.23 Phenotypic effects of reverting frequent PACE-derived mutations. Individual variants were isolated from PACE experiments that selected for the gain of NOXA binding in hsBCL-2 (A) and AncB4 (B) and the loss of NOXA binding in hsMCL-1 (C), AncM6 (D), and AncB1 (E). For each variant, non-WT states are colored. Sites and WT amino state are indicated at top. Heatmaps on the left show binding to BID and NOXA in the luciferase assay for each variant and their corresponding mutant without the key mutation. Each shaded box represents the normalized mean of three biological replicates.

2.2.7 Partial determinism is attributable to a limited number of function-changing mutations

We next analyzed the genetic basis for the limited degree of determinism that we observed in our experiments. Specifically, we sought to distinguish whether, from a given BID-specific starting point, only a few genotypes can confer NOXA binding while retaining BID binding or, alternatively, whether there are many such genotypes, but under strong selection a few are favored over others.

We performed PACE experiments in which we selected hsBCL-2 to retain its BID binding, without selection for or against NOXA binding; we then screened for variants that fortuitously gained NOXA binding using an activity-dependent plaque assay (**Figure 2.24A,B**). All four replicate populations produced clones that neutrally gained NOXA binding at a frequency of 0.1% to 1% lower than when NOXA binding was directly selected for but five orders of magnitude higher than when NOXA binding was selected against (**Figure 2.24A, Figure 2.25**). From each replicate, we then sequenced three NOXA-binding clones and found that all but one of them contained mutation r165L (**Figure 2.24B**), which also occurred at high frequency when the same protein was selected to gain NOXA binding (**Figure 2.26**). We introduced r165L into hsBCL-2 and found that it conferred significant NOXA binding with little effect on BID binding (**Figure 2.27**). Several other mutations appeared repeatedly in clones that fortuitously acquired NOXA binding, and these mutations were also acquired under selection for NOXA binding (**Figure 2.24B, Figure 2.27**). A similar pattern of common mutations was observed in AncB4 and AncB5 clones that fortuitously or selectively evolved NOXA binding (**Figure 2.28**). These observations indicate that the partial determinism we observed arises because from these starting points only a few mutations have the potential to confer NOXA binding while retaining BID binding.

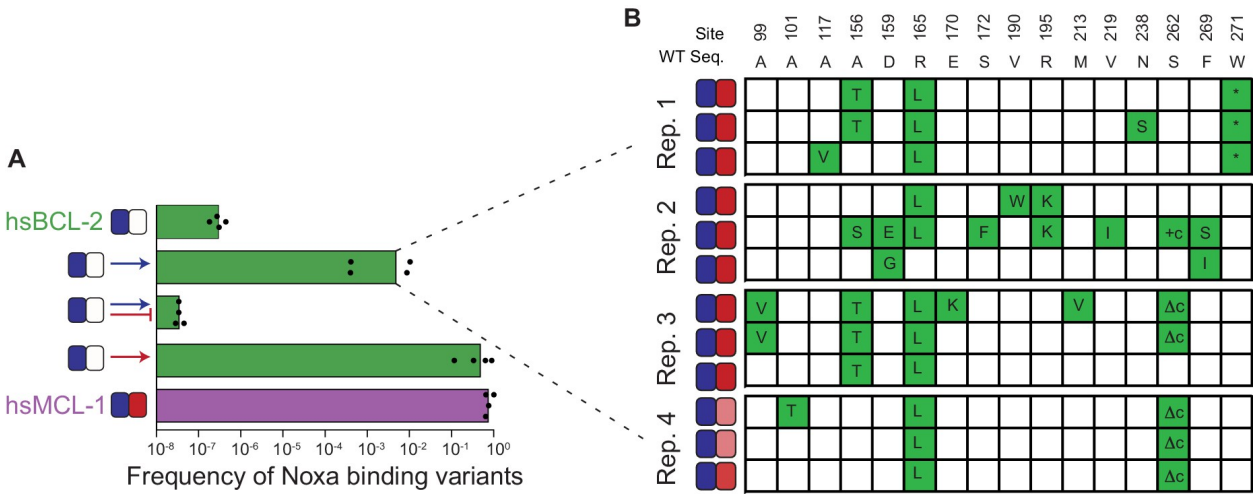


Figure 2.24 Sources of determinism.
(A) Evolution of NOXA-binding phage under various selection regimes. Frequency was calculated as the ratio of plaque forming units (PFU) per milliliter on *E. coli* cells that require NOXA binding to the PFU on cells that require BID binding to form plaques. Wild-type hsBCL-2 (green) and hsMCL-1 (purple) are shown as controls. Arrow, positive selection for function. Bar, counterselection against function. Blue, BID. Red, NOXA. Bars are the mean of four trajectories for each condition (points). (B) Phenotypes and genotypes of hsBCL-2 variants that evolved NOXA binding under selection to maintain only BID binding. Sites and WT amino state are indicated at top. For each variant, non-WT states acquired are shown in green. Heatmaps show binding to BID and NOXA in the luciferase assay for each variant (normalized mean of three biological replicates).

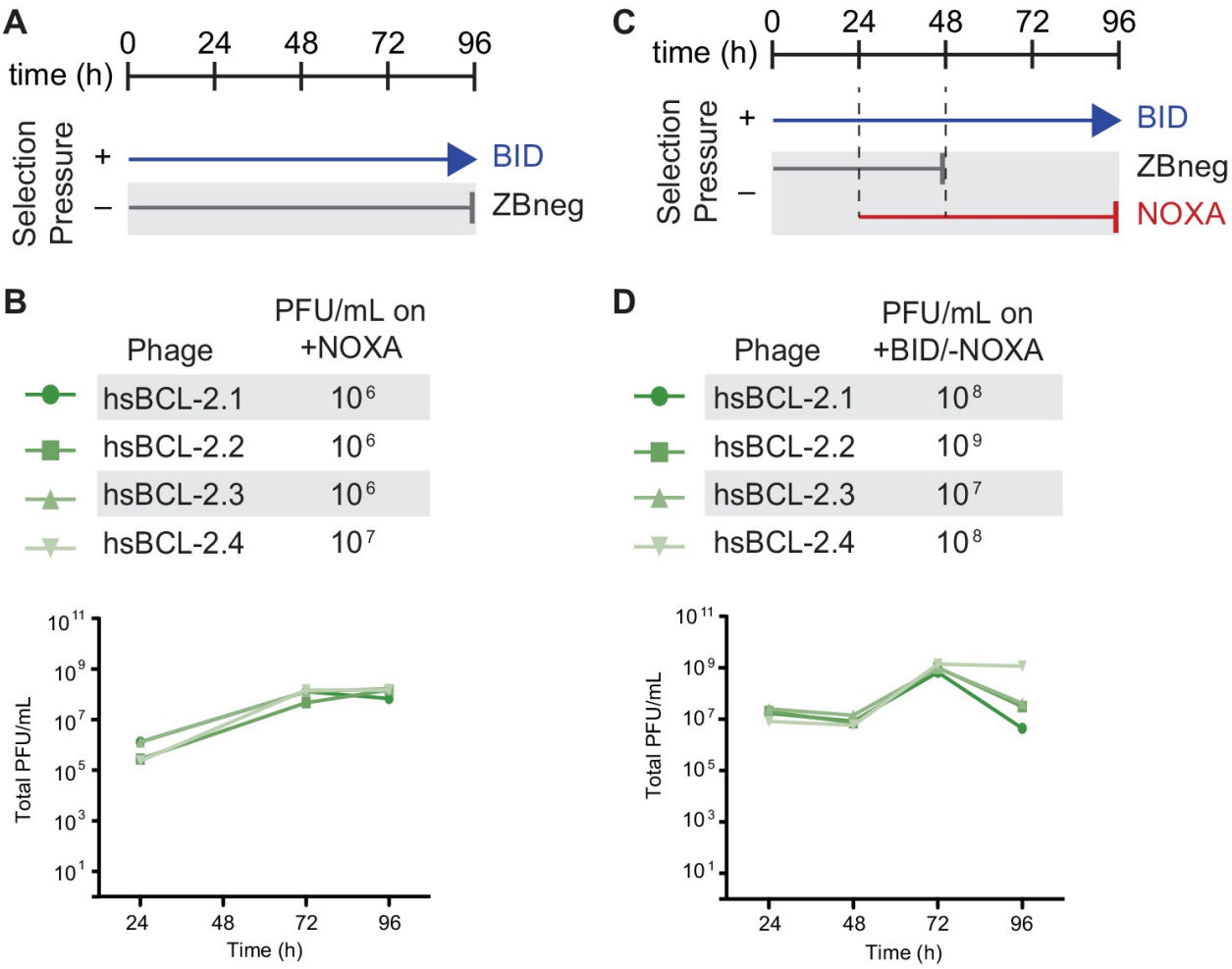


Figure 2.25 Selection schemes and phage titers for fortuitous NOXA binding of hsBCL2.

(A) Timeline of PACE experiments when hsBCL-2 was evolved with positive selection to maintain only BID binding. Selection conditions shown as arrows and blunt bars: arrow, selection for binding to BID (blue); blunt bar, selection against binding to ZBneg (gray). (B) Phage titers (PFU/mL) over time (bottom) and activity-dependent phage titers on NOXA at the end of the PACE experiment (top) when hsBCL-2 was evolved to maintain BID binding. Activity-dependent plaque assays used plasmid 2848. (C) Timeline of PACE experiments when hsBCL-2 was evolved with positive selection to maintain BID binding and negative selection against NOXA binding. Selection conditions shown as arrows and blunt bars: arrow, selection for binding to BID (blue); gray blunt bar, selection against binding to Zbneg; red blunt bar, selection against binding to NOXA. (D) Same as (B), but for hsBCL-2 evolved to bind BID and not NOXA. Activity-dependent plaque assays used plasmids 28-48 and Jin 487.

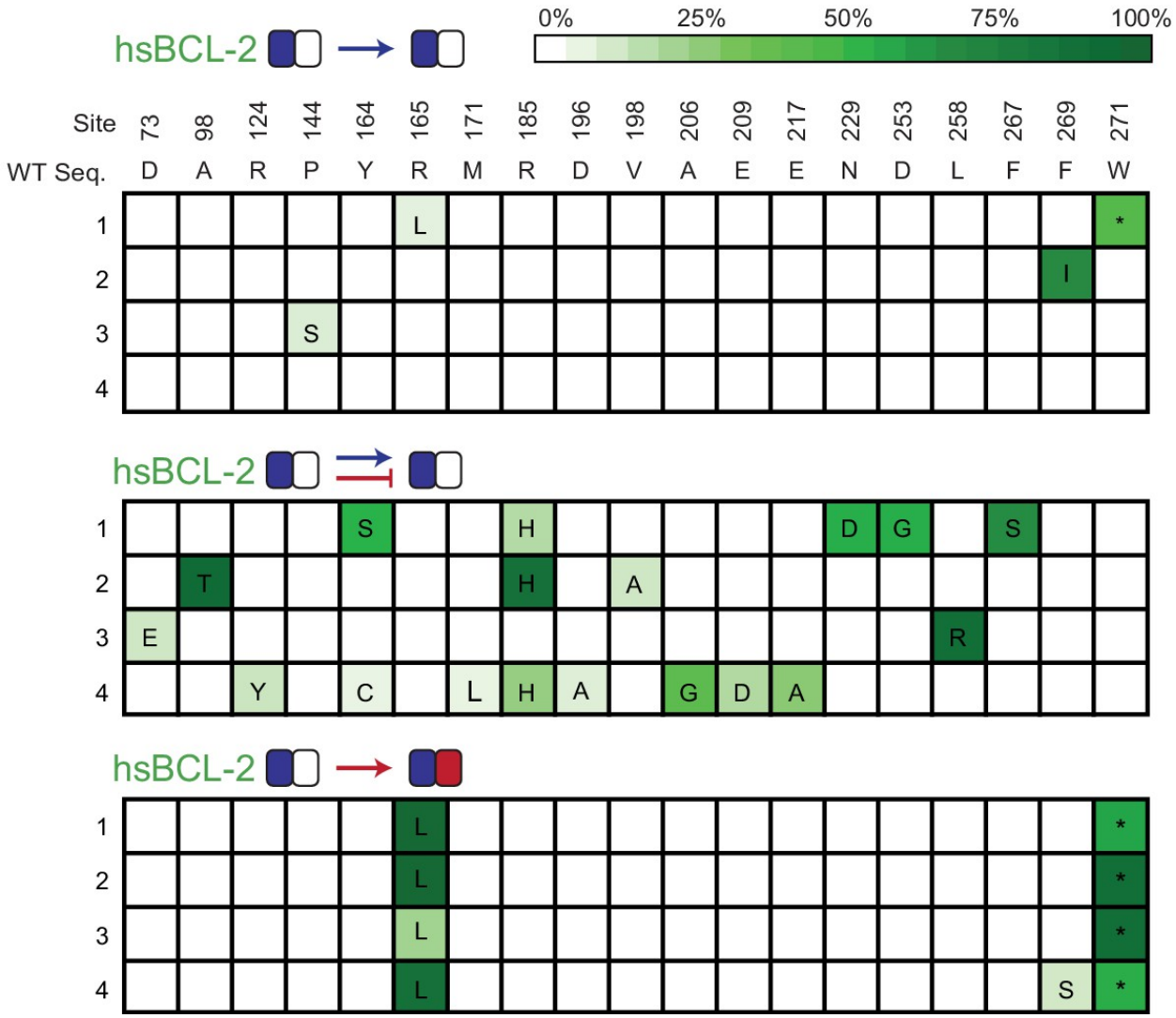


Figure 2.26 Allele frequency of non-wt states during PACE.
Allele frequency of non-wild-type states when hsBCL-2 was evolved to maintain BID binding (top) or when hsBCL-2 was evolved to simultaneously maintain BID binding and lose NOXA binding (middle). For comparison, the same sites are also shown for when hsBCL-2 was evolved to gain NOXA binding (bottom). Site numbers and wild-type (WT) amino acid states are listed above each sequence. Each row represents an independent replicate population. Non-wild-type amino acids that reached > 5% in frequency are shown, with frequency proportional to color saturation.

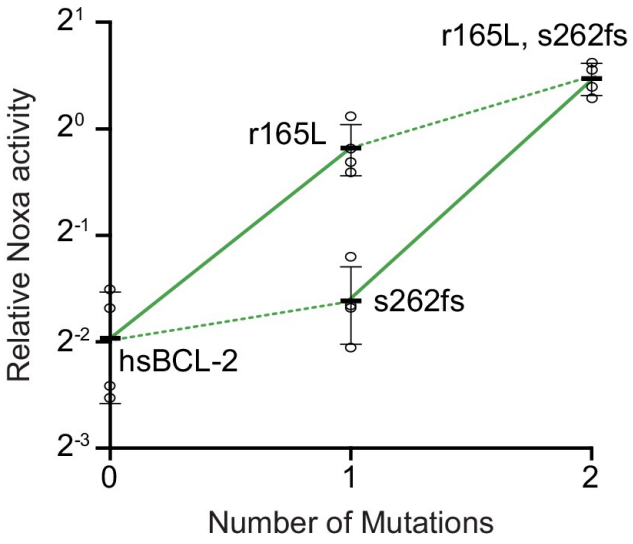


Figure 2.27 Effect on NOXA binding of the key r165L mutation.
Bars are the mean ± SD of three biological replicates (circles). Solid lines show the effects of the r165L mutation while dotted lines show the effect of a frameshift (fs) at site 262.

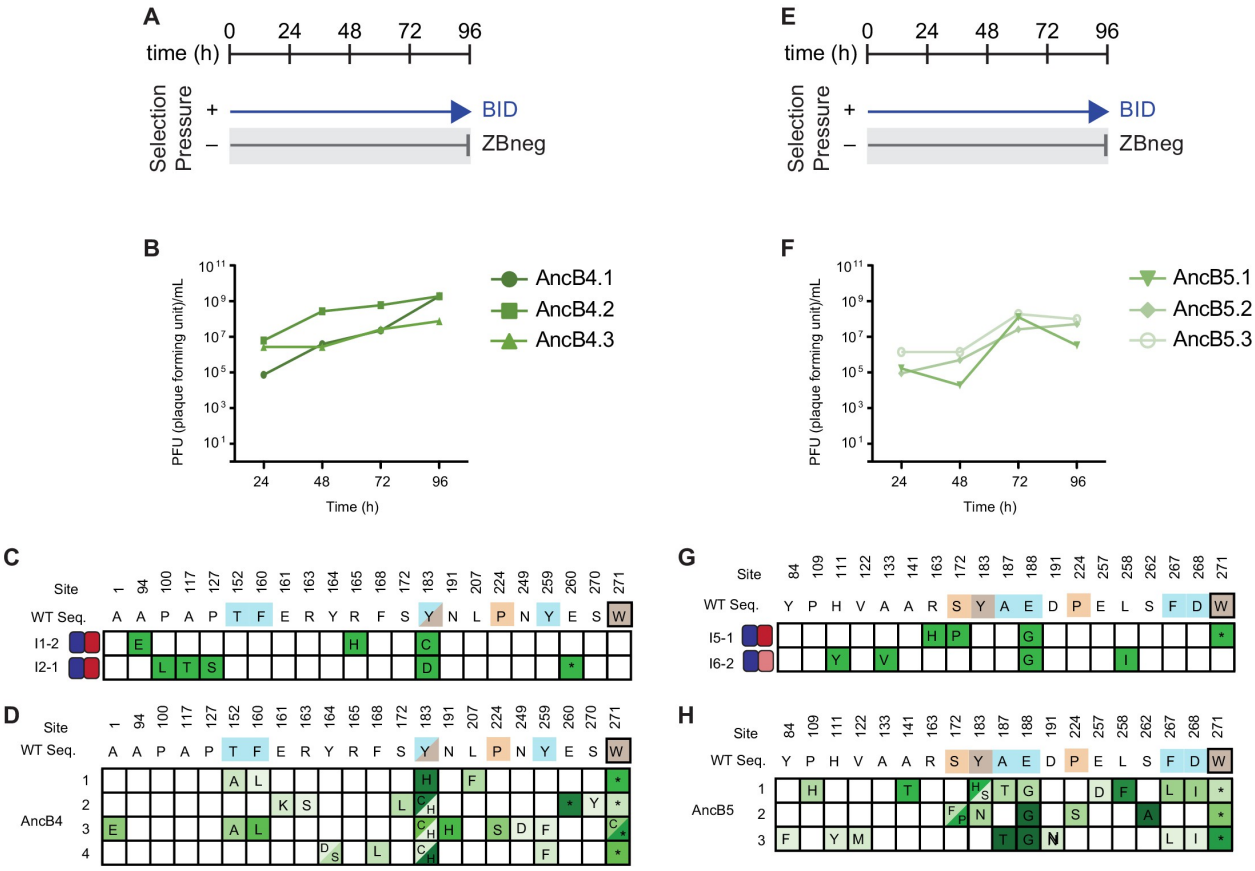


Figure 2.28 Selection and phage titers for fortuitous NOXA binding of AncB4 and AncB5.

(A) Timeline of PACE experiments when AncB4 was evolved with positive selection to maintain only BID binding. Selection conditions shown as arrows and blunt bars: arrow, selection for binding to BID (blue); blunt bar, selection against binding to ZBneg (gray). (B) Phage titers (PFU/mL) over time when AncB4 was evolved to maintain BID binding. (C) Phenotypes and genotypes of individual AncB4 variants that were isolated from PACE when selecting for BID binding and screened for the gain of NOXA binding. Site numbers and wild-type (WT) amino acid states are indicated at the top. Heatmaps on the left show binding to BID (blue) and NOXA (red) in the luciferase assay for each variant, and each shaded box represents the normalized mean of three biological replicates. (D) Non-wild-type amino acid states that reached >5% in frequency are shown for PACE when AncB4 was evolved to gain NOXA binding, for comparison with (C). Frequency is proportional to color saturation. Split cells show populations with multiple non-WT states > 5%. Each row represents an independent replicate lagoon. Color of WT state indicate if the mutation was seen among multiple replicates of the same starting genotype (teal), a single replicate from multiple starting genotypes (orange), or in multiple replicates and multiple starting genotypes (brown). Black box outline indicates mutant states observed in multiple replicates from the same starting genotype and from multiple replicates from a different starting genotype. (E) Same as (A) but for AncB5. (F) Same as (B) but for AncB5. (G) Same as (C) but for AncB5. (H) Same as (D) but for AncB5.

2.2.8 Contingency can affect accessibility of new functions

Although we found that chance and contingency strongly influenced sequence outcomes in our experiments, all trajectories acquired the historically relevant PPI specificities that were selected for, indicating strong necessity at the level of protein function. This was true whether evolution began from more promiscuous starting points that bound both BID and NOXA or from more specific proteins that bound only BID.

To further probe the evolutionary accessibility of new functions, we used PACE to select for a PPI specificity that never arose during historical evolution binding of NOXA but not BID. We found that trajectories launched from hsMCL-1 (which binds both coregulators) readily evolved the selected phenotype, but two PACE-evolved variants of hsBCL-2, which had acquired the same PPI profile as hsMCL-1, went extinct under the same selection conditions (**Figure 2.29, Figure 2.30**). The inability of the derived hsBCL-2 genotypes to acquire NOXA specificity was not attributable to a general lack of functional evolvability by these proteins because they successfully evolved in a separate PACE experiment to lose their NOXA binding but retain BID binding (**Figure 2.31**). These results establish that contingency can influence the accessibility of new functions and that the sequence by which a specific functional phenotype is encoded can play important roles in subsequent phenotypic evolution.

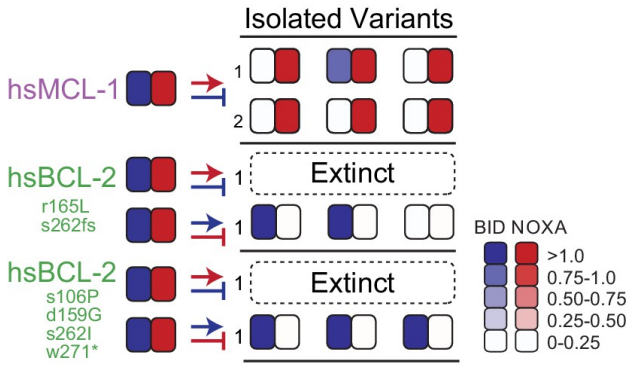


Figure 2.29 Contingency affects the evolution of novel specificity.
Starting genotypes that can bind both BID and NOXA (left) were selected to lose only BID or NOXA binding. Heatmaps show binding to BID and NOXA in the luciferase assay for each starting genotype (on the left) and for three individual variants picked at the end of one or more PACE trajectories (index numbers). Each box displays the normalized mean of three biological replicates for one variant. Trajectories initiated from starting points produced by PACE (green) and then selected for a non-historical function (loss of BID binding) went extinct.

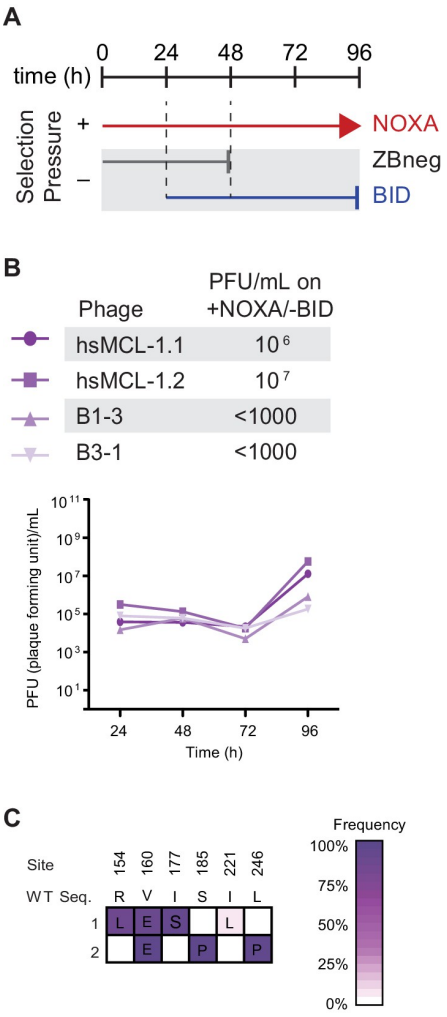


Figure 2.30 Selection scheme and phage titers for the gain of NOXA specificity. (A) Timeline of PACE experiments where hsMCL-1 and two previously evolved NOXA-binding hsBCL-2 variants were evolved to maintain NOXA binding and lose BID binding. Selection conditions: arrow, selection for binding NOXA (red); blunt bar, selection against binding a specific peptide (BID [blue] or ZBneg [gray]). (B) Phage titers (PFU/mL) over time (bottom) and activity-dependent phage titers at the end of the PACE experiment (top) where hsMCL-1 and NOXA-binding hsBCL-2 variants were evolved for binding NOXA and against BID. Activity-dependent plaque assays used plasmids 28-48 and Jin 518. Limit of detection = 10^3 PFU/mL. (C) Allele frequency of non-wild-type states after hsMCL-1 was evolved to maintain NOXA binding and lose BID binding. Site numbers and wild-type (WT) amino acid states are listed above each sequence. Each row represents an independent replicate lagoon. Non-wild-type amino acid frameshifts that reached >5% in frequency are shown, with frequency proportional to color saturation.

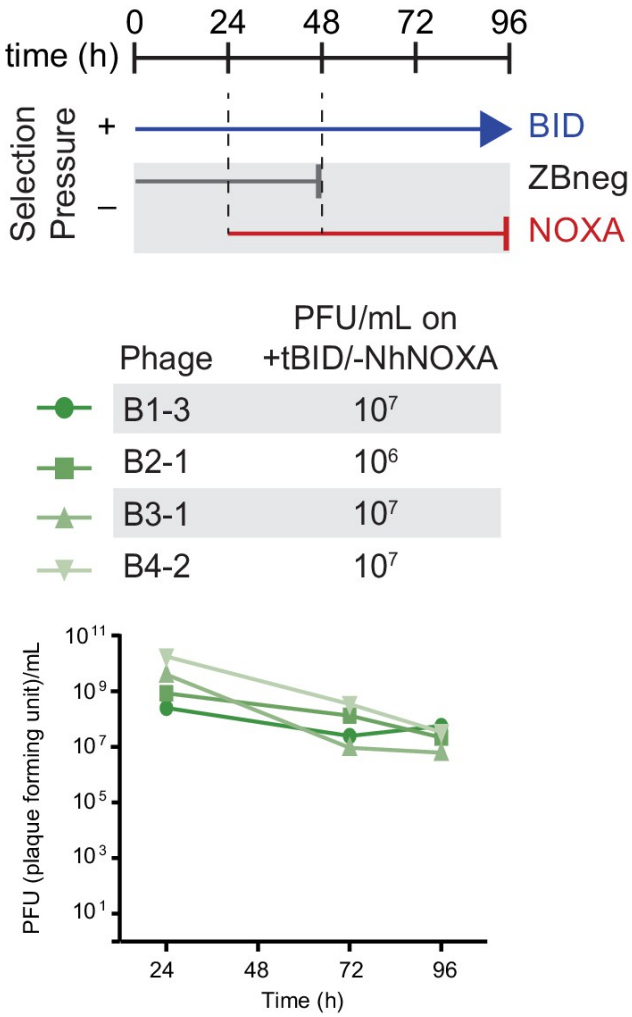


Figure 2.31 Selection scheme and phage titers for the regain of BID specificity. Phage titers (PFU/mL) over time (bottom) and activity-dependent phage titers at the end of the PACE experiment (top) where NOXA-binding hsBCL-2 variants were evolved to lose NOXA binding. Activity-dependent plaque assays used plasmids 2846 and Jin 487.

2.3 Discussion

The two major paradigms of 20th-century evolutionary biology-the adaptationist program[127] and the neutral theory of molecular evolution[174] - focus on either necessity or chance, respectively, as the primary mode of causation that produces natural variation in molecular sequences. Neither of these schools of thought admits much influence from

contingency or history. From an adaptationist perspective, variation is caused by natural selection, which generates optimal forms under different environmental conditions. Differences in protein sequence or other properties are interpreted as the result of adaptive changes that improved a molecules ability to perform its function in the species particular environment[175–178] . For neutralists, variation reflects the influence of chance in choosing among biologically equivalent possibilities, and conservation reflects purifying selection, both of which are viewed as largely unchanging across sequences in an alignment. For example, conserved portions of molecular sequences are interpreted as essential to structure and function, whereas differences in sequence alignments reflect a lack of constraint[179–181] . In neither worldview, does the particular state of a system strongly reflect its past or shape its evolutionary future. Recent work has shown that contingency might affect the sequence outcomes of evolution[135,136,141,143,146,182–187] , echoing themes raised in paleontology[123,124] and developmental biology[188,189] . Despite these recent findings, the dominance of the adaptationist and neutralist worldviews, and the continuing rhetorical battle between them[190,191] , has obscured the possibility that contingency might join selection, drift, and mutation as a primary factor shaping the outcomes of evolution.

We found that contingency generated by sequence change over phylogenetic timescales plays a profound role in BCL-2 family protein sequence evolution under laboratory selection for new functions. The mutations that rose to high frequency during experimental evolution were almost completely different among evolutionary trajectories initiated from historical starting points separated by long phylogenetic distances. We observed a strong role for chance (because trajectories launched from the same starting point evolved extensive differences from each other) and an even greater effect of contingency (because pools of trajectories launched from different starting genotypes evolved even greater differences). When combined, chance and contingency erased virtually all traces of neces-

sity between individual trajectories initiated from distantly related starting points. With the exception of a single truncation mutation that does not affect the selected-for function, the only predictable sequence states were those that remained unchanged from the starting point in all trajectories, presumably because they are unconditionally necessary for both PPIs tested and were therefore conserved by purifying selection.

Contingency and chance are distinct but interacting modes of causality; our experiments allowed us to disentangle their individual effects and interactions. By calculating genetic variance among replicates from the same starting point and among pooled replicates from different starting points, we quantified the effect of chance and contingency, respectively. The total effect of chance and contingency together, that is, genetic variance among replicates from different starting points, is by definition the product of the separate effects of chance and contingency. This quantitative relationship reflects the intrinsic interaction between chance and contingency in evolutionary processes[133,192] . At any point in history, numerous sets of mutations were accessible, and chance determined which ones occurred. These chance events then determined the steps that could be taken during future intervals, because of contingency. Without chance, contingency - dependence of the accessibility of future trajectories on the proteins state - would never be realized or observed: all phylogenetic lineages launched from a common ancestor would always lead to the same intermediate steps and thus the same ultimate outcomes. Conversely, without contingency, chance events would have no impact on the accessibility of other mutations because every path that was ever open would remain forever so, irrespective of the random events that happen to take place. The outcomes of evolution from a common ancestral starting point are therefore unpredictable when intermediate steps shape future possibilities (contingency), and those intermediate steps cannot be predicted because multiple possibilities are accessible at any point in history (chance).

Our experimental design approximates but does not quite achieve the ideal design of

multireplicate evolution from ancestral starting points under historical conditions, because the conditions we imposed during PACE differ in several ways from those that pertained during historical evolution. Many factors that give rise to chance, contingency, and necessity are likely to be similar between history and our experiments. For example, factors related to a proteins sequence-structure-function relations, such as the number of mutations that can produce a particular function and the nature of epistasis among them, play a key role in chance and contingency and are shared between PACE and history. Other aspects of our design may underestimate the effects of chance and contingency during history. For example, the population genetic parameters in our experimental conditions favor determinism because they involve very large population sizes, strong selection pressures, and high mutation rates, all directed at a single gene. If population sizes during historical BCL-2 family evolution involved smaller populations, weaker selection, lower mutation rates, and a larger genetic target size for adaptation, as seems likely, then chance would have played an even larger role during history than in our experiments. In addition, we used human BID and NOXA as fixed binding partners, but during real evolution these proteins would have varied in sequence as well, introducing opportunities for chance and contingency to further affect the sequence outcomes of BCL-2 evolution.

Some differences between our design and the biological setting of historical BCL-2 family evolution could have overestimated chances historical role. We selected for PPI interactions with two particular peptides, leaving out many potential cellular binding partners. PACE takes place in the cytosol of *E. coli* cells, but BCL-2 evolution occurred in animal cells, and natural BCL-2 proteins are partially membrane-bound. These additional dimensions of BCL-2 biological function could have imposed additional selective constraints on the evolution of BCL-2 family proteins historically, reducing the number of functionally equivalent genotypes available to chance. We used peptide fragments from coregulator proteins rather than full-length BID and NOXA; however, the peptide-binding

cleft is cytosolic, and recent work indicates that relative affinity of BCL-2 family proteins is similar between peptides and full-length coactivators, although absolute affinity is typically higher in the latter case[164] . Whether these differences quantitatively affect chance and contingency in PACE versus historical evolution is unknown. Finally, because our experimental design imposed selection for new PPI specificities, it does not reveal the effects of chance and contingency under different selective regimes, such as purifying selection to maintain an existing function, which may or may not be similar.

We studied a particular protein family as a model, but we expect that qualitatively similar results may apply to many other proteins. Epistasis is a common feature of protein structure and function, so the accumulating effect of contingency across phylogenetic time in the BCL-2 family will probably be a general feature of protein evolution, although its rate and extent are likely to vary among protein families and timescales[138,187,193,194] . The influence of chance depends upon the existence of multiple mutational sets that can confer a new function; this kind of degeneracy is likely to pertain in many cases: greater determinism is expected for functions with very narrow sequencestructurefunction constraints, such as catalysis[150,151,160,195–197] , than those for which sequence requirements are less strict, such as substrate binding[99,146,156,198] . Consistent with this prediction, when experimental evolution regimes have imposed diffuse selection pressures on whole organisms, making loci across the entire genome potential sources of adaptive mutations, virtually no repeatability has been observed among replicates[155,199] .

The method that we developed for rapid evolution of PPI specificity has several advantages that can be extended to other protein families. First, by using PACE, many replicates can be evolved in parallel across scores or hundreds of generations in just days, with minimal need for intervention by the experimentalist[38] . Second, our split RNAP design for acquiring new PPIs has fewer components than previous methods for this purpose, such as two-hybrid designs; this makes it considerably easier to tune and

optimize and therefore to extend to other protein systems. Third, unlike approaches that attempt to evolve specific PPIs by alternating selection and counterselection through time, our platform simultaneously imposes selection and counterselection within the same cell, thus selecting for specificity directly. By combining these elements in a single system, our platform should allow rapid multireplicate evolution of new cytosolic PPI specificities in a variety of protein families.

Our results have implications for efforts to engineer proteins with desired properties. We found no evidence that ancestral proteins were more or less evolvable than extant proteins: the selected-for phenotypes readily evolved from both extant and ancestral proteins with the same starting binding capabilities. Moreover, chances effect was virtually constant across 1 billion years of evolution, indicating that the number of accessible mutations in the deep past that could confer a selected-for function was apparently no greater than it is now. Nevertheless, the strong effect of contingency that we observed on sequence evolution and its partial role in the acquisition of new functions per se suggests that efforts to produce proteins with new functions by design or directed evolution will be most effective and will lead to more diverse sets of sequence outcomes, if they use multiple different protein sequences as starting points, ideally separated by long intervals of sequence evolution. Ancestral proteins can be useful for this purpose simply because they provide routes to functions that were inaccessible from extant protein, even if those routes are not fundamentally different in number or kind.

Finally, our work has implications for understanding the processes of protein evolution and the significance of natural sequence variation. Our observations suggest that sequencestructurefunction associations apparent in sequence alignments are to a significant degree the result of contingent constraints that were transiently imposed or removed by chance events during history[99,137,138,143] . Evolutionary explanations of sequence diversity and conservation must therefore explicitly consider the historical trajectories by

which sequences evolved, in contrast to the largely history-free approaches of the dominant schools of thought in molecular evolution. Our findings suggest that present-day BCL-2 family proteins and potentially many others, as well are largely physical anecdotes of their particular unpredictable histories: their sequences reflect the interaction of accumulated chance events during descent from common ancestors with necessity imposed by physics, chemistry, and natural selection. Apparent design principles in the pattern of variability and conservation in extant proteins reflect not how things must be to perform their functions, or even how they can best do so. Rather, todays proteins reflect the legacy of opportunities and limitations that they just happen to have inherited.

2.4 Materials and methods

Table 2.1 Key resources table.

Reagent type or resource	Designation	Source or reference	Identifiers
E. coli strain	S1030	ref[200]	
E. coli strain	1059	ref[200]	
E. coli strain	NEB 10-beta	NEB	Cat#C3019I
E. coli strain	BCL21 (DE3)	NEB	Cat#C2530H
Peptide, recombinant protein	BID	GenScript	This study- human BID peptide used for fluorescence polarization
Peptide, recombinant protein	NOXA	Genscript	This study- human NOXA peptide used for fluorescence polarization
Commercial assay or kit	DNA clean and concentrator kit	Zymo	Cat#D4013
Commercial assay or kit	MiSeq Reagent Kit v3	Illumina	Cat#MS-1023003
Chemical compound, drug	Q5 DNA Polymerase	NEB	Cat#M0491
Chemical compound, drug	Phusion DNA polymerase	ThermoFisher Scientific	Cat#F518L
Chemical compound, drug	IPTG	bioWORLD	Cat#21530057
Chemical compound, drug	His60 Ni Superflow Resin	Takara	Cat#635660
Software, algorithm	Geneious	Geneious	10.1.3
Software, algorithm	R	CRAN	3.5.1
Software, algorithm	RStudio	RStudio	1.1.456
Software, algorithm	PROT Test	ref[201]	3.4.2
Software, algorithm	RAXML-ng	ref[202]	0.6.0

2.4.1 Phylogenetics

Amino acid sequences of the human BCL-2, BCLW, BCL-xL, MCL-1, NRH, BFL1, BAK, BAX, and BOK paralogs were used as starting points for identifying BCL-2 family members in other species. For each paralog, tblastn and protein BLAST on NCBI BLAST were used to identify orthologous sequences between January and March of 2018[203] . Sequences for each paralog were aligned using MAFFT (G-INS-I) with the allowshift option and unalignlevel set at 0.1. For each paralog, phylogenetic structure was determined using fasttree 2.1.11 within Geneious 10.1.3. Missing clades based on known species relationships were then identified, and specific tblastn searches were used within Afrotheria (taxid:311790), Marsupials (taxid:9263), Monotremes (taxid:9255), Squamata (taxid:8509), Archosauria (taxid:8492), Testudinata (taxid:8459), Amphibia (taxid:8292), Chondrichthyes (taxid:7777), Actinopterygii (taxid:7898), Dipnomorpha (taxid:7878), Actinistia (taxid:118072), Agnatha (taxid:1476529), Cephalochordata (taxid:7735), and Tunicata (taxid:7712) as needed. Additional sequences were added by downloading genome and transcriptome data for tuatara[204] , sharks and rays[205] , gar[206] , 2018), ray-finned fish[207] , lamprey[208] , hagfish[209] , Ciona savignyi[206] , tunicates[210] , echinoderms[211] , porifera[212] , and ctenophores[213] . In each case, local BLAST databases were created in Geneious and searched using tblastn. Finally, we used BCL-2DB to add missing groups as needed[214] .

After collection of sequences, each paralog was realigned using MAFFT (G-INS-I) with the allowshift option and unalignlevel set at 0.1. Based on known species relationships, lineage-specific insertions were removed and gaps manually edited. Only a single sequence was kept among pairs of sequences differing by a single amino acid and sequences with more than 25% of missing sites were removed. For difficult to align sequences, sequences were modeled on the structures of human BCL-2 family members using SWISS-Model to identify likely locations of gaps[215] . Finally, paralogs were profile

aligned to each other, and paralog-specific insertions were identified.

In total, 151 amino acid sites from 745 taxa were used to infer the phylogenetic relationships among BCL-2 family paralogs. PROT Test 3.4.2 was used to identify the best-fit model among JTT, LG, and WAG, with combinations of observed amino acid frequencies (+F), gamma distributed rate categories (+G), and an invariant category (+I)[201] . From this, JTT + G + F had the highest likelihood and lowest Aikake Information Criterion score. RAXML-ng 0.6.0 was then used to identify the maximum likelihood tree using JTT+G12+F0 (12 gamma rate categories with maximum likelihood estimated amino acid frequencies)[202] . Finally, we enforced monophyly within each paralog for the following groups: lobe-finned fish (n = 9), ray-finned fish (n = 9), jawless fish (n = 5), cartilaginous fish (n = 8), tunicates (n = 4), branchiostoma (n = 4), chordates (n = 5), ambulacraria (n = 5, hemichordata +echinodermata), deuterostomia (n = 5), protostomia (n = 5), cnidaria (n = 5), and porifera (n = 4) (values in parenthesis are number of identified paralogs in each group) and used RAXML-ng with JTT + G12 + F0 to identify the best tree given these constraints (Supplementary Data Phylogenetic.Data.zip).

Overall, we recovered three clades: a pro-apoptotic clade; a clade containing the BCL-2, BCLW, and BCLX vertebrate paralogs and BCL non-vertebrate sequences; and a clade containing the MCL-1, BFL1, and NRH vertebrate paralogs and MCL non-vertebrate sequences. We used the pro-apoptotic clade as the outgroup to the two anti-apoptotic clades. Within the BCL-2 clade, the majority of vertebrates contained all three copies. However, the exact relationship among the paralogs was unclear; only two copies were identified within jawless fish and their phylogenetic placement had weak support. Non-vertebrate clades tended to have good support and only a single copy. However, support for these groups following established species relationships was often limited. The MCL-1 clade contained the fastest evolving paralogs of the BCL-2 family. As with the BCL-2-like clade, only two copies were found within the jawless fish and the exact sister relationships

among paralogs was unclear. Non-vertebrates contained only a single copy, but as with the BCL-2-like clade, support for relationships following established species relationships was often weak.

The BCL-2-like and MCL-1-like paralogs formed a clade with the BHP1 and BHP2 sequences from porifera. The sister relationships among these four clades were unresolved. In addition, we recovered a sister relationship between the BAK and BAX paralogs. While both paralogs contained copies from porifera, these clades evolved quickly and had relatively low support, and they may be artifactual. We identified only a single clade of ctenophores. Finally, the placement of BOK was unresolved; BOK may be sister to the BAK/BAX clade or an outgroup to all clades and the most ancient copy of the BCL-2 family.

2.4.2 Ancestral reconstruction

Posterior probabilities of each amino acid at each site were inferred using Lazarus[216] to run codeml within PAML. We used the same model and alignment as used to infer the phylogeny. We used the branch lengths and topology of the constrained maximum likelihood phylogeny found by raxml-ng.

We first reconstructed the LCAs of all BCL-2 and MCL-1 like sequences, AncMB1-M, using the maximum likelihood state for each alignable site. We then reconstructed a series of ancestors from AncMB1 to modern human MCL-1. These included AncM1 (LCA of MCL-1-related sequences), AncM2 (LCA of MCL-1- related deuterostomes and protostomes), AncM3 (LCA of MCL-1-related deuterostomes), AncM4 (LCA of MCL-1-related urochordates and chordates), AncM5 (LCA of MCL-1, BFL1, and NRH like copies in vertebrates), AncM6 (LCA of MCL-1 and BFL1 like copies), AncMCL-1 (LCA of MCL-1 like copies), AncMCL-1-G (LCA of MCL-1 like Gnathostomes), AncMCL-1-O (LCA of MCL-1 like Osteichthyes), and AncMCL-1-T (LCA of MCL-1 like Tetrapods), AncMCL-1-

A (LCA of MCL-1 like Amniotes), and AncMCL-1-M (LCA of MCL-1 like Mammals). In each case, the sequence of each ancestor used the maximum likelihood state at each site, with gaps inserted based on parsimony. We used the modern sequences of human MCL-1 to fill in portions of the sequence that showed poor alignment and could not be reconstructed, including both the N and C terms, as well as the loop between the first and second alpha helices. Average posterior probabilities for ancestors in the MCL-1 clade ranged from 0.73 (AncM6) to 0.98 (AncMCL-1-M) with an average of 0.83 (sd 0.08) (Supplementary file 2).

For the BCL-2 like clade, we also reconstructed AncMB1, this time using human BCL-2 sequence to fill in the N and C terms and the loop between the first and second alpha helices (AncMB1-B). We then reconstructed sequences from AncMB1 to modern human BCL-2. These included AncB1 (LCA of BCL-2-related sequences), AncB2 (LCA of BCL-2-related Bilaterian and Cnidaria), AncB3 (LCA of BCL-2-related deuterostomes and protostomes), AncB4 (LCA of BCL-2 deuterostomes), AncB5 (LCA of BCL-2, BCLW, and BCLX like copies in vertebrates), AncBCL-2 (LCA of BCL-2 like copies), AncBCL-2-G (LCA of BCL-2 like gnathostomes), AncBCL-2-O (LCA of BCL-2 like osteichthyes), and AncBCL-2-T (LCA of BCL-2 like tetrapods), using human BCL-2 sequences for the N and C terms and the loop between the first and second alpha helices. Average posterior probabilities for ancestors in the BCL-2 clade ranged from 0.87 (AncB1) to 0.95 (AncBCL-2-T) with an average of 0.9 (sd 0.04).

2.4.3 Test of robustness of ancestral inference

To determine the robustness of our conclusions on the phenotype of ancestral sequences, we synthesized and cloned alternative reconstructions for key ancestors. In each case, sequences contained the most likely alternative state with posterior probability > 0.2 for all such sites where such a state existed. Alternative reconstructions con-

tained an average of 24 alternative states and represent a conservative test of function (min: 4, max: 44, Supplementary file 2). In our luciferase assay, all but two alternative reconstructions retained similar BID and NOXA binding as the maximum likelihood ancestral sequences. The first alternative reconstruction that differed from the maximum likelihood reconstruction was AltAncB3, which bound both BID and NOXA, while the ML for AncB3 bound BID, but NOXA only weakly. As a result, the exact branch upon which NOXA binding was lost historically is not resolved by this data.

The second alternative reconstruction that differed from the ML reconstruction was AltAncMB1-B, which had weaker NOXA binding than the ML reconstruction. To further test the robustness of AncMB1-B to alternative reconstructions, we synthesized and tested additional reconstructions that included only alternative amino acids with posterior probabilities greater than 0.4 (n = 3), 0.35 (n = 7), 0.3 (n = 13), and 0.25 (n = 18), and compared these to AncMB1-B and the 0.2 AltAncMB1-B (n = 21) (values in parentheses are number of states that differ from the ML state). We found that the 0.4, 0.35, and 0.3 alternative reconstructions bound both BID and NOXA, while the 0.25 and 0.2 alternative reconstructions had diminished NOXA binding.

Finally, we synthesized and tested modern sequences from key groups to determine the robustness of our inference on the timing of NOXA binding loss. These included BCL-2-related sequences from groups that diverged prior to the predicted loss of NOXA binding (Trichoplax adhaerens and Hydra magnapapillata), sequences from groups that diverged around the time of predicted NOXA binding loss (Octopus bimaculoides and Stegodyphus mimosarum), or sequences from groups predicted to have diverged after NOXA binding lost (Saccoglossus kowalevskii and Branchiostoma belcheri). In each case, we used human BCL-2 sequence to replace extant N and C terms and the loop between the first and second alpha helices. The T. adhaerens and B. belcheri sequences were non-functional in our luciferase assays, binding neither BID nor NOXA. However, recent work has compre-

hensively characterized binding in BCL-2 family members within T. adhaerens, finding that the BCL copy can bind both BID and NOXA as predicted[217] . H. magnapapillata bound both BID and NOXA in our assay and the remaining sequences bound only BID, suggesting a loss of NOXA binding prior to the divergence of protostomes and deuterostomes in the BCL-2 related clade, consistent with the conclusion drawn using reconstructed proteins.

2.4.4 Escherichia coli strains

E. coli 10-beta cells were used for cloning and were cultured in 2xYT media. *E. coli* BL21 (BE3) cells were used for protein expression and were cultured in Luria-Bertain (LB) broth. *E. coli* S1030 cells cultured in LB broth were used for activity-dependent plaque assays, phage growth assays, and luciferase assays. S1030 cells cultured in Davis Rich media were used for PACE experiments[200] . *E. coli* 1059 cells were used for cloning phage and assessing phage titers and were cultured in 2xYT media.

2.4.5 Cloning and general methods

Plasmids were constructed by using Q5 DNA Polymerase (NEB) to amplify fragments that were then ligated via Gibson Assembly. Primers were obtained from IDT, and all plasmids were sequenced at the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility. Vectors and gene sequences used in this study are listed in Supplementary file 5, with links to fully annotated vector maps on Benchling. Key vectors are deposited at Addgene, and all vectors are available upon request. The following working concentrations of antibiotics were used: 50 μg/mL carbenicillin, 50 μg/mL spectinomycin, 40 μg/mL kanamycin, and 33 μg/mL chloramphenicol. Protein structures and alignments were generated using the program PyMOL[218] .

2.4.6 Luciferase assays

Cloned expression vectors contained the following: (1) a previously evolved, isopropyl -D-1-thiogalactopyranoside (IPTG)-inducible N-terminal half of T7 RNAP[?] fused to a BCL-2 family protein; (2) the C-terminal half of T7 RNAP fused to a peptide from a BH3-only protein; and (3) T7 promoter-driven luciferase reporter. Chemically competent S1030 *E. coli* cells[200] were prepared by culturing to an OD600 of 0.3, washing twice with a calcium chloride-HEPES solution (60 mM CaCl2, 10 mM HEPES pH 7.0, 15% glycerol), and then resuspending in the same solution. Vectors were transformed into chemically competent S1030 cells via heat shock at 42°C for 45 s, followed by 1 hr recovery in 2xYT media, and then plated on agar with the appropriate antibiotics (carbenicillin, spectinomycin, and chloramphenicol) to incubate overnight at 37°C. Individual colonies (three to four biological replicates per condition) were picked and cultured in 1 mL of LB media containing the appropriate antibiotics overnight at 37°C in a shaker. The next morning, 50 μL of each culture was diluted into 450 μL of fresh LB media containing the appropriate antibiotics, as well as 1 μM of IPTG. The cells were incubated in a shaker at 37°C, and OD600 and luminescence measurements were recorded between 2.5 and 4.5 hr after the start of the incubation. Measurements were taken on a Synergy Neo2 Microplate Reader (BioTek) by transferring 150 ţL of the daytime cultures into Corning black, clear-bottom 96-well plates. Data were analyzed in Microsoft Excel and plotted in GraphPad Prism, as previously reported[63] .

2.4.7 Protein purification

Protein expression hsBCL-2, hsMCL-1, and evolved variants were constructed as N-terminal 6xHis-GST tagged proteins. The recombinant proteins were expressed in BL21 *E. coli* (NEB) and purified following standard Ni-NTA resin purification protocols (ThermoFisher Scientific)[114] . Briefly, BL21 *E. coli* containing an N-terminal 6xHis-GST tagged

BCL-2 family protein were cultured in 5 mL LB with carbenicillin overnight. The following day, the culture was added to 0.5 L of LB with carbenicillin, incubated at 37°C until it reached an OD600 of 0.6, induced with IPTG (final concentration: 200 ţM), and cultured overnight at 16°C. The cell pellet was harvested by centrifugation followed by resuspension in 30 mL of lysis buffer (50 mM Tris 1 M NaCl, 20% glycerol, 10 mM TCEP, pH 7.5) supplemented by protease inhibitors (200 nM Aprotinin, 10 ţM Bestatin, 20 ţM E-64, 100 ţM Leupeptin, 1 mM AEBSF, 20 ţM Pepstatin A). Cells were lysed via sonication and were then centrifuged at 12,000 g for 40 min at 4°C. Solubilized proteins, located in the supernatant, were incubated with His60 Ni Superflow Resin (Takara) for 1 hr at 4°C, and the protein was eluted using a gradient of imidazole in lysis buffer (50250 mM). Fractions with the protein, as determined by SDS-PAGE, were concentrated in Ulta-50 Centrifugal Filter Units (Amicon, EMD Millipore). Proteins were purified via a desalting column with storage buffer (50 mM TrisHCl [pH 7.5], 300 mM NaCl, 10% glycerol, 1 mM DTT) and further concentrated. The concentration of the purified BCL-2 family proteins was determined by BCA assay (ThermoFisher Scientific), and they were flash-frozen in liquid nitrogen and stored at 80°C.

2.4.8 Fluorescence polarization binding assays

Fluorescent polarization (FP) was used to measure the affinity of BCL-2 family proteins with peptide fragments of the BH3-only proteins in accordance with previously described methods[219] . hsBCL-2, hsMCL-1, and evolved variants were purified as described above. The fluorescent NOXA and BID peptides (95+% purity) were synthesized by GenScript and were N-terminally labeled with 5-FAM-Ahx and C-terminally modified by amidation. These peptides were dissolved and stored in DMSO. Corning black, clear-bottom 384-well plates were used to measure FP, and three replicates were prepared for each data point. Each well contained the following 100 ţL reaction: 20 nM BH3-only pro-

tein, 0.05 nM to 3 ţM of BCL-2 family protein (1/3 serial dilutions), 20 mM Tris (pH 7.5), 100 mM NaCl, 1 mM EDTA, and 0.05% pluronic F-68. FP values (in milli-polarization units; mFP) of each sample were read by a Synergy Neo2 Microplate Reader (BioTek) with the FP 108 filter (485/530) at room temperature 5-15 min after mixing all the components. Data were analyzed in GraphPad Prism 8, using the following customized fitting equation, to calculate K_d[114] :

$$= B + C(D + K_d + x - \sqrt{(D + K_d + x)^2 - 4Dx})$$

where y is normalized measured FP, x is the concentration of BCL-2 protein, D is the concentration of the BH3-only protein, B and C are parameters related to the FP value of free and bound BH3-only protein, and Kd is the dissociation constant.

2.4.9 Phage-assisted continuous evolution

PACE was used to evolve hsBCL-2, hsMCL-1, and ancestral proteins in accord with previously reported technical methods[38,63,173,200] using a new vector system. Briefly, combinations of accessory plasmids and the MP6 mutagenesis plasmid[36] were transformed into S1030 *E. coli*., plated on agar containing the appropriate antibiotics (carbenicillin, kanamycin, and chloramphenicol) and 10 mM glucose, and incubated overnight at 37°C. Colonies were grown overnight in 5 mL of LB containing the appropriate antibiotics and 20 mM glucose. Davis Rich media was prepared in 510 L carboys and autoclaved, and the PACE flasks and corresponding pump tubing were autoclaved as well. The following day, PACE was set up in a 37°C environmental chamber (Forma 3960 environmental chamber, ThermoFisher Scientific). For each replicate, an overnight culture was added to 150 mL of Davis Rich carboy media in chemostats and grown for 23 hr until reaching an OD600 of approximately 0.6. Lagoons containing 20 μL of phage from saturated phage

stocks (10^{8-9} phage) were then connected to the chemostat. Magnetic stir bars were used to agitate chemostats and lagoons. The chemostat cultures were flowed into the lagoons at a rate of approximately 20 mL/h. Waste output flow rates were adjusted to maintain a constant volume of 20 mL in the lagoons, 150 mL in the chemostat, and an OD600 close to 0.6 in the chemostat. A 10% w/v arabinose solution was pumped into the lagoons at a rate of 1 mL/h. If the experiment included a mixing step (two separate chemostats flowed together into one lagoon for a mixed selection pressure), a chemostat was prepared the next day (as described above) and connected to the lagoons. During this step, lagoon volumes were increased to 40 mL, and the arabinose inflow rate was increased to 2 mL/h. After disconnecting the first chemostat the next day, the lagoon volumes and arabinose inflow were both lowered to 20 mL and 1 mL/h, respectively. During the experiment, samples were collected from the lagoons every 24 hr and centrifuged at 13,000 rpm for 3 min to collect the phage-containing supernatant, as well as the cell pellet for DNA extraction. PACE experiments are listed in Supplementary file 3. A single replicate of AncB5 was removed because of contamination. No statistical method was used to determine the number of replicates as only four independent replicate experiments could be performed simultaneously. During PACE, the media volume of each lagoon turned over once per hour for 4 days, or 100 times. For a phage population to survive this amount of dilution, a similar number of generations must have occurred between the starting phage and the phage in the lagoon at the end of the experiment[38] . This is expected to be a conservative estimate; as a more fit phage rises in frequency in the population, it will undergo a greater number of generations than less-fit phage in the population. The mutagenesis plasmid MP6 induces a mutation rate of approximately $6x10^{6}$ per bp per generation. The BCL-2 family proteins used in the PACE experiments were 230 amino acids long, indicating that a mutation occurred on average every 250 phage replications. Phage population sizes ranged from 10_{5} per mL to 10^{10} per mL over the course of a PACE experiment,

indicating a rate of 40040,000,000 new mutations every generation. Conservative estimates thus suggest that a during each individual replicate, phage populations sampled at least 40,000 mutations, and upwards of 4 Œ 109 mutations. While not all mutations were equally likely each generation because MP6 enriches for transitions (i.e. GA, AG, CT, and TC), the high number of mutations sampled suggests that the vast majority of possible single point mutations (approximately 230x3x4 = 2760 potential mutations) were sampled over the course of each experiment, with higher population sizes generating all potential single point mutations each generation.

2.4.10 Plaque assays

Plaque assays were performed on 1059 *E. coli* cells[84,200] , which supply gene III (gIII) to phage in an activity-independent manner, to measure phage titers. Additionally, activity-dependent plaque assays were done on S1030 *E. coli* containing the desired accessory plasmids to determine the number of phage encoding a BCL-2 family protein with a given peptide-binding profile. All cells were grown to an OD600 of approximately 0.6 during the day. Four serial dilutions were done in Eppendorf tubes by serially pipetting 1 L of phage into 50 ţL of cells to yield the following dilutions: 1/50, 1/2500, 1/125,000, and 1/6,250,000. 650 ţL of top agar (0.7% agar with LB media) was added to each tube, which was then immediately spread onto a quad plate containing bottom agar (1.5% agar with LB media). Plates were incubated overnight at 37°C. Plaques were counted the following day, and plaque forming units (PFU) per mL was calculated using the following equation:

$$PFU = 1000 \times A \times 50^{\text{4-B}}$$

where A is the number of plaques in a given quadrant, and B is the quadrant number where the phage were counted, in which one is the least dilute quadrant and four is the

most dilute quadrant.

2.4.11 Phage growth assays

Phage growth assays were performed by adding the following to a culture tube and shaking at 37°C for 6 hr: 1 mL of LB with the appropriate antibiotics (carbenicillin and kanamycin), 10 ţL of saturated S1030 *E. coli* containing the accessory plasmids of interest, and 1000 phage. Phage were then isolated by centrifugation at 13,000 rpm for 3 min, and PFU was determined by plaque assays using 1059 *E. coli* and the plaque assay protocol described above.

2.4.12 High-throughput sequencing library construction

PACE samples were collected from each lagoon every 24 hr. The lagoon samples were centrifuged at 13,000 rpm for 3 min on a bench top centrifuge to separate supernatant and cell pellet. The phage-containing supernatants were stored at 4°C prior to the creation of sequencing libraries. To prepare Illumina sequencing libraries, each phage sample was cultured overnight with 1059 *E. coli* cells, followed by phage DNA purification (Qiagen plasmid purification reagent buffer), P1 (catalog number 19051), P2 (catalog number 19052), N3 (catalog number 19064), PE (catalog number 19065), and spin column for DNA (EconoSpin, catalog number 1920250). The resulting DNA concentration was 50 ng/μL. Freshly generated DNA samples were then used as template for PCR amplification. For each library sample, we amplified three overlapping fragments of the BCL-2 family protein, which are 218241 bp in length (Figure 2.12). Each primer also included 69 Ns to introduce length variation (Supplementary file 4). In total, 12 PCRs were used for each library. Phusion DNA polymerases and buffers (ThermoFisher Scientific, catalog number F518L) were used in the first PCR round to amplify all three fragments for

all library sequencing. The 25 μL reaction contained: 0.5 μL of 50 mM MgCl2, 0.75 μL of 10 mM dNTP, 0.75 μL Phusion DNA polymerase, 20 ng library DNA, and 0.5 μL of 10 μM primer (each). The PCR were run on a C1000 Touch Thermal Cycler (Bio-Rad), with the following parameters: 98°C for 1 min, followed by 16 cycles of 98°C for 12 s, 58°C for 15 s and 72°C for 45 s, and finally 72°C for 5 min. PCR were purified using the ZYMO DNA clean and concentrator kit (catalog number D4013) and 96 well filter plate (EconoSpin, catalog number 2020001). The DNA products were dissolved in 30 μL ddH2O. All 12 reactions for each library were combined, and 1 ţL was used as the template for a second PCR round. PCR components and thermocycler parameters were the same as above, except that the annealing temperature was 56°C, and only 15 rounds of amplification were conducted. The primer and sample combinations are listed in Supplementary file 4. PCRs were then purified following the same procedure as previous step. Equal volumes of all 72 library samples were combined and concentration was measured using a Qubit 4 Fluorometer. The total DNA sample was 2.68 ng/μL (equivalent to 10 nM, according to the average length of PCR fragments). DNA samples were diluted to 4 nM from step 4 following the Illumina MiSeq System Denature and Dilute Libraries Guide and then diluted to 12 pM for high-throughput sequencing. The final sample contained 100 ţL of 20 pM PhiX spike-in plus 500 ţL of the 12 pM library sample. Sequencing was performed on the Illumina MiSeq System using MiSeq Reagent Kit v3 (600-cycle) with paired-end reads according to the manufacturers instructions.

2.4.13 Processing of Illumina data

Illumina sequencing yielded 22 million reads, 13 million of which could be matched to a specific sample (Supplementary file 4). One replicate for AncB5 was found to be contaminated and removed from further analysis. To process the remaining data, we first used Trim Galore with default settings to trim reads based on quality (https://www.bioin-

formatics.babraham.ac.uk/projects/trim_galore/). Then, we used BBMerge, a script in BBTools (https://jgi.doe.gov/data-and-tools/bbtools/), to merge paired-end reads. Next, we used Clumpify to remove repeated barcode sequences. We then used Seal to identify and bin reads by sample and fragment. Finally, we used BBDuk to remove any primer or adapter sequence present. Scripts and reference sequences are available on Github (Thornton, 2021).

2.4.14 Illumina sequencing analysis

Reads were binned by experiment and then aligned to the appropriate WT sequence using Geneious (low sensitivity, five iterations, gaps allowed). Sequences were then processed in R to remove sequences containing Ns or that were not full length. Insertions found in less than 1% of the population and sites that extended outside of the coding region were removed from all sequences. Remaining gaps were standardized among replicates and within an experiment. Finally, allele frequencies were calculated for each site and amino acid, as well as remaining insertions and deletions.

2.4.15 Quantifying the effects of chance and contingency on the outcomes of evolution

See ref[65]

2.4.16 Data availability

The high throughput sequencing data of evolved BCL-2 family protein variants were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) databases. They can be accessed via BioProject: PRJNA647218. The processed sequencing data are available on Dryad (https://doi.org/10.5061/dryad.866t1g1ns).

The coding scripts and reference sequences for processing the data are available on Github (https://github.com/JoeThorntonLab/BCL2.ChanceAndContingency).

The following data sets were generated Xie VCPu JMetzger BPHThornton JWDickinson BC (2020) NCBI Bioproject ID PRJNA647218. Experimental evolution of BCL2 family ancestral proteins. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA647218 Xie VCPu JMetzger BPHThornton JWDickinson BC (2020) Dryad Digital Repository BCL2-Chance and Contingency. https://doi.org/10.5061/dryad.866t1g1ns

2.5 Supplementary files

See ref[65]

CHAPTER 3

AN *IN VIVO* PHAGE-ASSISTED DIRECTED EVOLUTION PLATFORM TO DIRECTLY IDENTIFY PROTEIN-PROTEIN INTERACTION INHIBITORS

Though the mis-regulation of protein-protein interactions (PPIs) has been linked to many diseases, the scientific community lacks tools that specifically perturb a given PPI, and the process of identifying or creating such a molecule often takes years. We set out to address this technological gap by developing a rapid system for the selection of PPI inhibitors (PPIis). We designed and optimized a phage-assisted continuous evolution (PACE)-based platform that directly selects for genetically encoded PPI inhibitors through utilizing a split RNA polymerase (RNAP) biosensor. Upon establishing our systems compatibility with a set of clinically relevant PPIs, we optimized the platform for selecting binders of the KRas-Raf, P53-MDM2, and MYC-MAX interactions and performed a deep-mutational scan of a Raf-based inhibitor for the KRas-Raf interaction We then used the platform to discover a de novo affibody inhibitor of the P53/MDM2 interaction, which we found bound to MDM2 and inhibited the interaction in a mammalian cell-based assay as well. Based on this model study, we believe this platform can be used with a variety of targets to rapidly generate PPI inhibitors for use in biological research and as starting points for drug development.

3.1 Introduction

An estimated 130,000 human protein-protein interactions (PPIs)[220] regulate virtually every cellular process, including replication[221–223] , translation[224,225] , and signal transduction[226–229] . Advances in unbiased proteome-wide PPI mapping methods such as 2-hybrid screens[230–232] , proximity labeling technologies[233–236] , and advanced imag-

ing[237–239] have enabled an unparalleled look at which proteins interact with one another in a cell. Dysregulation of specific PPIs drive pathology in humans[240–244] , and therefore represent therapeutic targets for disease intervention[245,246] . As such, molecules that specifically disrupt target PPIs are critical for both assigning functional significance to mapped PPIs as well as for creating next-generation therapeutics. However, discovery and development of PPI inhibitors is challenging, due to the observed inherent difficulty in discovering selective inhibitors of protein complexes, thereby rendering many PPIs to be considered undruggable targets[247–250] .

The most common approach for PPI inhibitor discovery entails finding a competitive binding partner to one of the proteins involved in the interaction, usually through binding at the PPI interface. This can be done via high-throughput screening[251–253] , rational design[254] , computational modeling[255] , and binding-based directed evolution platforms[256–258] . However, the underlying issue for all these approaches is that binding does not necessarily confer inhibition; the function selected for is not the desired end function. For instance, these techniques often discount the possibility of finding alternative modes of inhibition, such as allostery. Phenotypic screening provides a mechanistically-agnostic alternative to identify PPI inhibitors[259] , but lacks generalizability and remains relatively low-throughput, and, once again, the functional readout is not the same as the desired property of PPI inhibition. An optimal strategy to discover inhibitors would be both high-throughput and directly select for inhibition of a preformed PPI in a mechanistically-agnostic manner.

By linking PPI inhibition to *in vivo* selection systems, advances in directed evolution platforms can enable high-throughput, direct, and mechanistically-agnostic identification of PPI inhibitors. For example, a bacterial reverse 2-hybrid system was used to link inhibition of a PPI to cell survival, which enabled discovery of a compound that inhibits the HIV p6 and human TSG101 interaction from a 106-member lanthipeptide library screen[260] .

Additionally, a yeast reverse 2-hybrid was used to develop an assay to detect inhibitors of the p53-MDM2 PPI[261] . However, despite being first described over 27 years ago[262] , relatively few successes of *in vivo* screening campaigns using reverse 2-hybrid systems have been reported[263–265] .

While powerful, existing *in vivo* reverse 2-hybrid systems often suffer from false positive (non-specific or off-target inhibitors) and false negatives (incomplete library sampling), and are often therefore technically limited to screening libraries of 10^6 molecules. On the contrary, continuous *in vivo* evolution technologies, which link viral replication to defined functions of interest of molecules encoded in the viral genome, have led to advanced gene editing platforms[98] , reprogrammed proteases[67] , engineered tRNAs[266] , and improved and diversified enzymes[267,268] . Recently, we established the first *in vivo* continuous evolution platform that can select for specific PPIs[65] using our proximity-dependent split RNAP biosensor technology[63,269] as an alternative to a traditional 2-hybrid. The critical advantages of our split RNAP biosensor approach are: 1) a broad dynamic range, both in terms of RNA output and linear sensitivity to an array of PPI affinities[64] ; 2) detection of multiple interactions simultaneously in cells[270] , allowing for careful focusing of the selection pressure on a single target interaction; and 3) extremely facile engineering, analogous to a FRET or split luciferase sensor[101,271] . Therefore, we sought to develop a new *in vivo* PPI inhibitor selection system using the principles of Phage-Assisted Continuous Evolution (PACE)[38,200] that leverages our split RNAP biosensors to enable comprehensive screening of large phage-encoded libraries of peptides or proteins for novel PPI inhibitors.

In this work, we develop Phage-Assisted Non-Continuous Selection for PPI inhibitors (PANCS-PPI*i*), which directly links the life cycle of an M13 bacteriophage to the selective disruption of a target PPI. After optimizing PANCS-PPI*i* for the detection of inhibitors of three PPIs related to human health (KRas-Raf, P53-MDM2, and MYC-MAX), we show-

case in mock selections that the system can reproducibly and efficiently enrich one active inhibitor from a pool of 10^9 inactive inhibitors, encoded in phage, in a few days, simply by serially diluting the phage populations on engineered selection *E. coli* cells. We then use PANCS-PPI*i* to perform a deep mutational scan (DMS) of a cell-penetrating peptide fused to Raf as a competitive inhibitor of the KRas-Raf complex, which revealed novel mutations that modulate this interaction. Finally, we performed a de novo selection of an affibody protein library against the P53-MDM2 interaction, which yielded a novel inhibitor. Together, this work showcases the capacity of PANCS-PPI*i* for robust, rapid, and mechanistically agnostic discovery of novel PPI inhibitors from $10^{8\text{-}9}$-member phage libraries.

3.2 Results

3.2.1 Design of PANCS-PPIi

PACE-based platforms control the fitness of M13 bacteriophage by removal of the necessary pIII coat protein from the phage genome, rendering the phage unable to replicate on *E. coli* cells. Upon infection of engineered *E. coli* cells, the needed pIII is supplied to the phage through an inducible promoter, such that the phage propagate at a rate that corresponds to the ability of a phage-encoded evolving proteins ability to activate that promoter (**Figure 3.1**). For additional control, a dominant negative form of pIII, pIIIneg, can also be produced from a separate inducible system, which lowers the phage replication rate[200] , thereby allowing for both positive (on-target) and negative (off-target) selections. Focusing *in vivo* selection pressures in a system like PACE to inhibit of single target PPI and avoid cheaters (i.e. variants which can survive the selection pressure but do not have the desired activity) requires a system that can monitor multiple PPIs at once. We recently developed a PACE system that uses our split RNAP-based biosensors to evolve proteins to selectively bind to one target over another, by using the capacity of the split RNAP

technology to measure multiple PPIs simultaneously[65] . We reasoned that we could build from this split RNAP dual PPI biosensor selection to devise a new selection scheme that focuses the selection pressure on specific inhibition of a single target PPI.

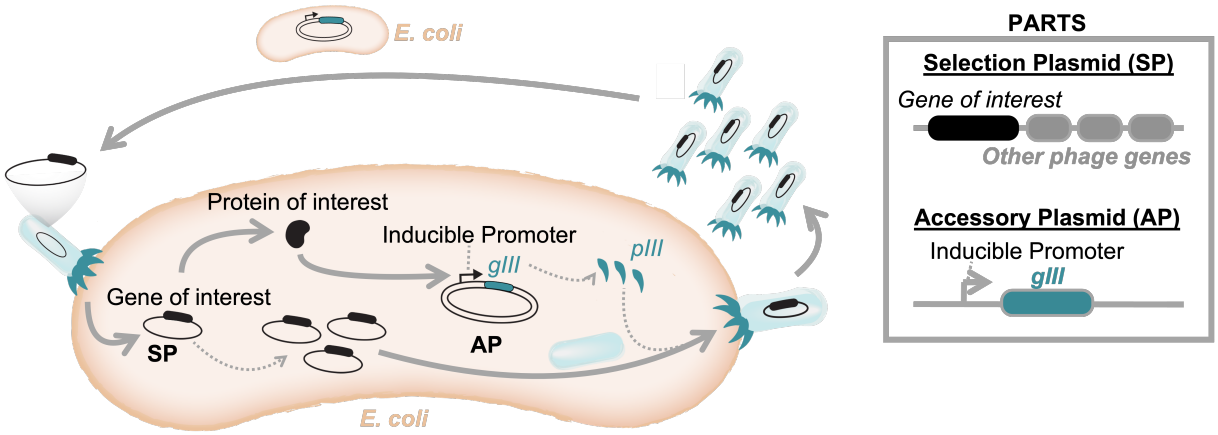


Figure 3.1 In vivo phage-assisted directed evolution platforms, in general.
In vivo phage-assisted directed evolution platforms function by encoding a gene of interest (black) into the phage genome (SP) in place of the essential gIII. gIII (teal) is instead encoded in an accessory plasmid (AP) in the E. coli and placed under control of an inducible promoter. The researcher must engineer a biosensor to link desired protein activity with the induction of gIII. In this case, phage with proteins with the desired activity will be able to propagate on these engineered E. coli cells.

After attempting several strategies (see **Supplementary Note**), we devised PANCS-PPI*i*, which links the production of an inhibitor for a target PPI (PPI partner 1 and PPI partner 2) to viral replication (**Figure 3.2A,B**). In PANCS-PPI*i*, one accessory plasmid (AP1) encodes for: 1) constitutively expressed zipper peptide 2 (ZP2) fused to zipper peptide A (ZA) and PPI partner 1, 2) constitutively expressed zipper peptide B (ZB) fused to the C-terminal half of the split CGG RNAP ($RNAP_{C(CGG)}$), which transcribes from the CGG promoter (P_{CGG}) when activated, and 3) P_{CGG}-driven gIII, which encodes for the pIII protein. A second accessory plasmid (AP2) encodes for: 1) constitutively expressed PPI partner 2 fused to the C-terminal half of the split T7 RNAP ($RNAP_{C(T7)}$), which transcribes

from the T7 promoter (P_{T7}) when activated, and 2) P_{T7}-driven gIIIneg, which encodes for the pIIIneg protein. Prior to phage infection, ZA and ZB form a PPI, and PPI partner 1 and PPI partner 2 form a PPI, but neither gIII or $gIII_{neg}$ are transcribed, because the RNAP N-terminal split fragment is not present in the system. We then engineered M13 bacteriophage that express: 1) the evolved RNAP N-terminal fragment ($RNAP_N$) fused to zipper peptide 1 (ZP1), which binds ZP2, and 2) a potential inhibitor molecule of the target PPI. Upon phage infection, ZP1 binds to the ZP2 complexes, triggering both gIII and $gIII_{neg}$ production, thus inhibiting phage replication. However, if the inhibitor encoded by the phage disrupts the target PPI, $gIII_{neg}$ production is blocked, allowing that phage variant to replicate. Critically, the only difference between the two trimolecular complexes is the target PPI versus ZA-ZB, so inhibitors of the RNAP biosensor or ZA-ZB will prevent phage growth by not allowing the production of pIII; therefore, the selection pressure is entirely focused on disruption of the target PPI. With the conceptual framework of PANCS-PPI^i established, we next sought to optimize the system and assess its performance characteristics as a selection platform for PPI inhibitors.

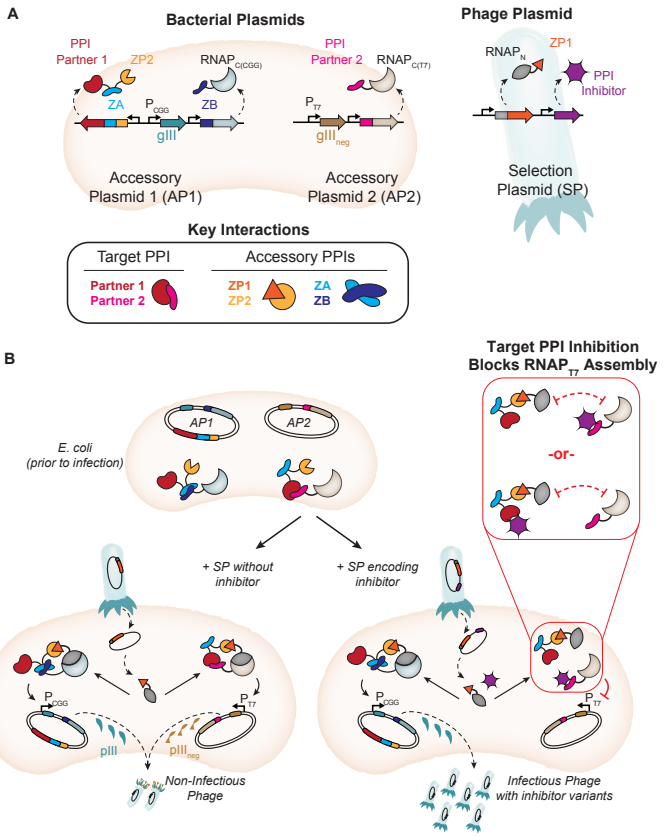


Figure 3.2 Overview of PANCS-PPI*i*.
(A) Schematic of the genetically-encoded components of PANCS-PPI*i*. (Left) Two accessory plasmids (APs) are encoded in S1030 *E. coli* cells. AP1 encodes 1) zipper peptide 2 (ZP2, light orange) fused to zipper peptide A (ZA, light blue) and PPI partner 1 (red); 2) zipper peptide B (ZB, dark blue) fused to the C-terminal half of the RNAP that binds to the CGG promoter ($RNAP_{C(CGG)}$, light gray); and 3) Gene III (gIII, teal) under the control of the CGG promoter (P_{CGG}). AP2 encodes 1) PPI partner 2 (pink) fused to the C-terminal half of the RNAP that binds to the T7 promoter ($RNAP_{C(T7)}$, light brown) and 2) a dominant negative form of gene III ($gIII_{neg}$, brown) under the control of the T7 promoter (P_{CT7}). (Right) The phage genome includes a portion that expresses 1) the N-terminal half of the RNAP ($RNAP_N$, dark gray) fused to zipper peptide 1 (ZP1, dark orange) and 2) a genetically-encoded inhibitor (purple). (Bottom) The key interactions that take place in the system include the binding of 1) PPI partner 1 to PPI partner 2; 2) ZP1 to ZP2; and 3) ZA to ZB. (B) (Top) Prior to phage infection, the components of AP1 and AP2 are expressed such that a mixture of two complexes are formed. Upon phage infection, both $RNAP_N$-ZP2 and inhibitor are expressed. (Left) If the inhibitor does not inhibit any of the key interactions, RNAPN will be brought into proximity with $RNAP_{C(CGG)}$ and $RNAP_{C(T7)}$, and both gIII and $gIII_{neg}$ will be expressed, leading to the production of non-infectious phage. (Right) If the inhibitor only inhibits the binding of PPI partner 1 and PPI partner 2, $RNAP_N$ will only be brought into proximity with $RNAP_{C(CGG)}$ such that only gIII will be expressed, which will lead to the propagation of phage with this encoded inhibitor.

3.2.2 Split RNA polymerase biosensors can detect PPI inhibitors

We selected three well-studied, clinically important, yet biophysically-distinct PPIs around which we developed the PANCS-PPI*i* platform: KRas-Raf, P53-MDM2, and MYC-MAX (**Figure 3.3A**). We first assessed whether each PPI could be detected by the split RNAP system in *E. coli* using a luciferase reporter assay (**Figure 3.3B**)[101] . We found enhanced luminescence signal, as triggered by PPI-dependent transcription, of each known PPI pair, but not for pairs that do not interact (**Figure 3.3C,D,E**). We also tested and confirmed the binding of Raf to two highly-prevalent yet challenging oncogenic mutants of KRas: G12D and G12V (**Figure 3.4**)[272,273] . These data confirm that the split RNAP biosensors can detect a range of disease-relevant PPIs in *E. coli* and control gene expression outputs based on each target PPI.

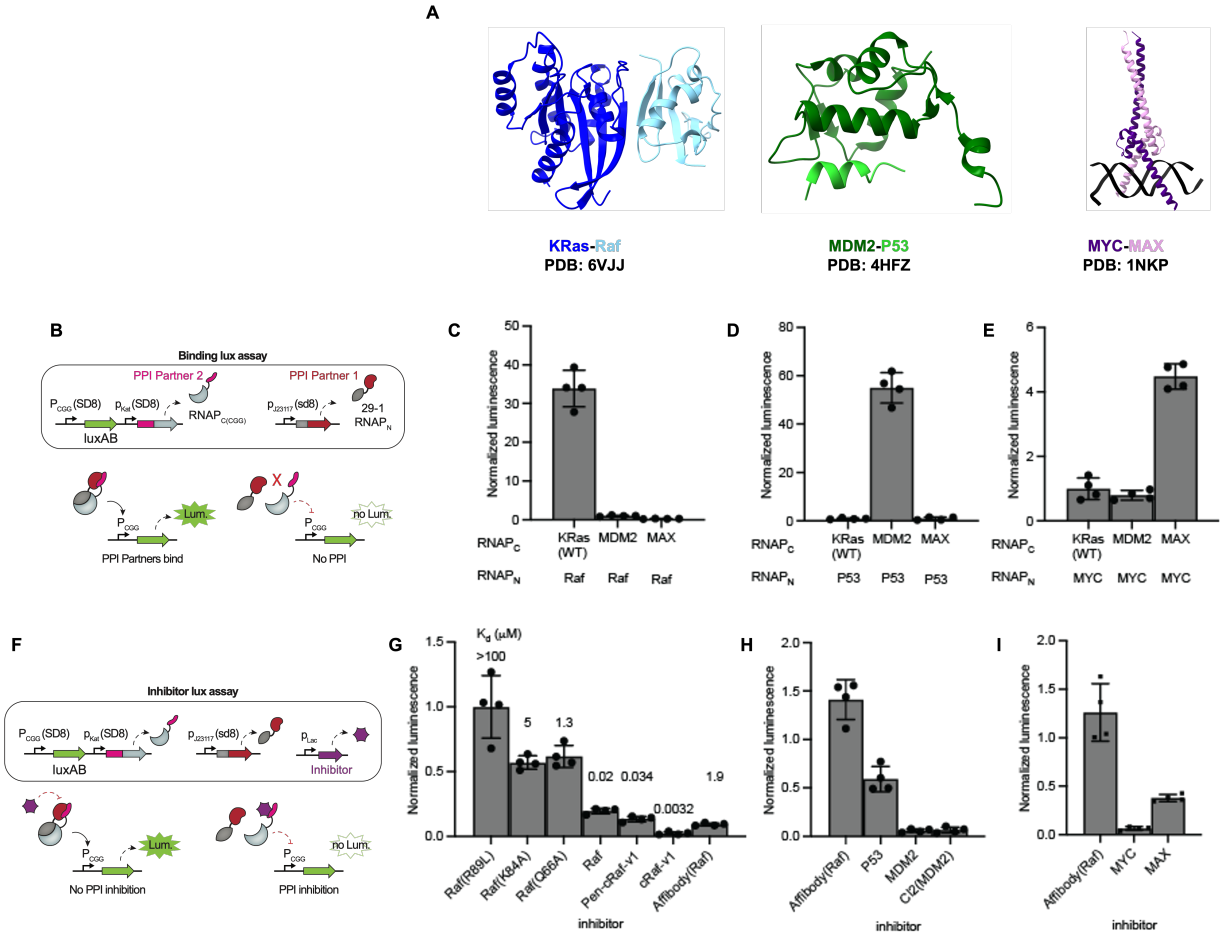


Figure 3.3 *In vivo E. coli* luciferase assays to assess PPIs and PPI inhibition. (A) Crystal structures of the 3 model PPIs used in this study. (B) Schematic of the genetically-encoded components of the binding luciferase (lux) assay. One plasmid encodes PPI partner 1 fused to $RNAP_N$. Another plasmid encodes 1) PPI partner 2 fused to $RNAP_{C(CGG)}$ and 2) a P_{CGG}-driven luciferase component (luxAB). If PPI partner 1 binds to PPI partner 2, then the $RNAP_{CGG}$ will be reconstituted and drive luminescence. If PPI partner 1 does not bind to PPI partner 2, the $RNAP_{CGG}$ will not be reconstituted; thus, no luminescence will result. (C,D,E) The *E. coli* luciferase binding assay when testing binding across the 3 model PPIs. (F) Schematic of the genetically-encoded components of the inhibitor lux assay. Both of the plasmids from (B) are present as well as a vector with a genetically-encoded inhibitor under the control of an IPTG-inducible promoter (P_{Lac}). If the inhibitor does not prevent the binding of PPI partner 1 to PPI partner 2, then the RNAPCGG will be reconstituted and drive luminescence. If the inhibitor does prevent PPI partner 1 binding to PPI partner 2, the RNAPCGG will not be reconstituted; thus, no luminescence will result. (G,H,I) The *E. coli* luciferase inhibitor assay when testing known genetically-encoded inhibitors across the 3 model PPIs. Bars show mean $\pm$ SD of four replicates (circles).

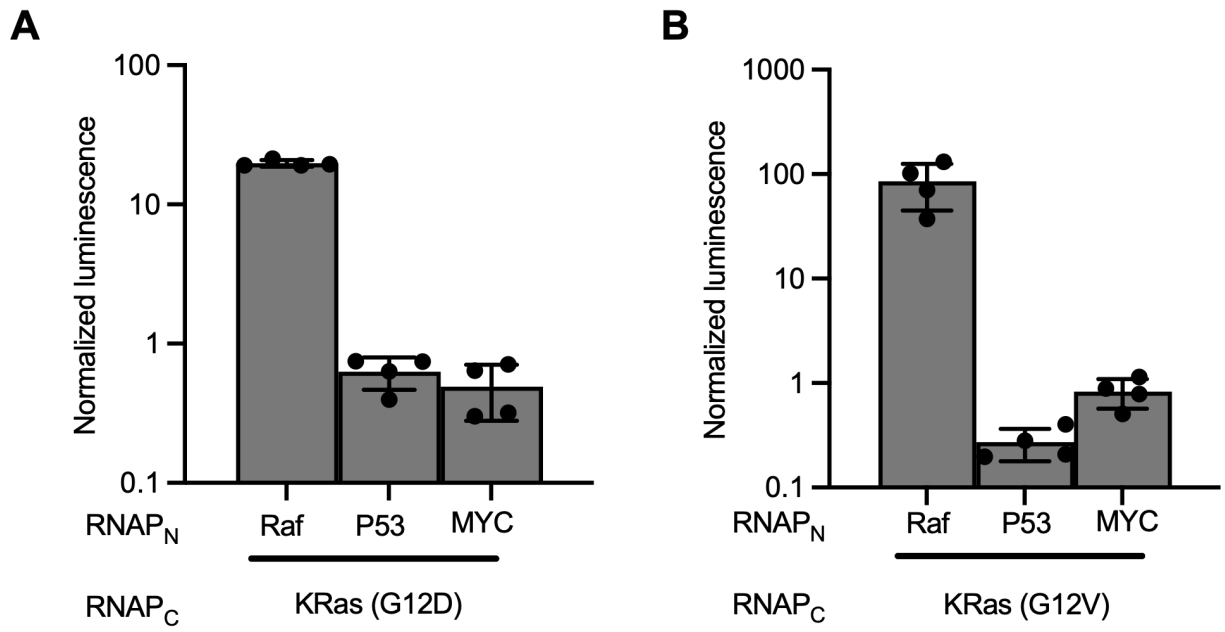


Figure 3.4 *In vivo E. coli* luciferase binding assays for KRas mutants.
The *E. coli* luciferase binding assay when testing binding of (A) KRas(G12D) and (B) KRas(G12V) for Raf, P53, and MDM2. Bars show mean ± SD of four replicates (circles).

Next, we aimed to assess whether the split RNAP PPI detection system could detect PPI inhibitors. We modified the luciferase assay to detect for PPI inhibition by adding an additional plasmid that encodes for an IPTG-driven expressed protein that may or may not inhibit the target PPI. In this system, if the additional protein successfully inhibits the target PPI, this will prevent the split RNAP from reconstituting, and therefore decrease the production of luciferase (**Figure 3.3F**). We used the KRas-Raf PPI as the first model system and found that expression of Raf(R89L), which does not bind KRas and is therefore not a competitive inhibitor, had no impact on the activity in the reporter, which therefore served as a negative control (**Figure 3.5**). However, we found that expression of Raf(K84A) or Raf(Q66A), which are weak KRas binders and therefore likely weak competitive inhibitors, knocked down the signal by 50%, while expression of Raf, a 20 nM KRas binder, knocked down the signal by 70% (Figure 3.3G). We then tested engineered Raf-based inhibitors, including cRaf-v1 (a 3.2 nM binder) and Pen-cRaf-v1 (a 34

nM binder)[274] , which showed even more inhibition. Finally, we found a Raf affibody a 1.9 μM binder)[275] also knocked down the signal. We performed analogous assays with the KRas(G12D)-Raf and KRas(G12V)-Raf PPIs as well. (**Figure 3.6**). Taken together, these data show that the split RNAP reporter can read out KRas-Raf PPI inhibition in changes in gene expression across a range of inhibitor strengths.

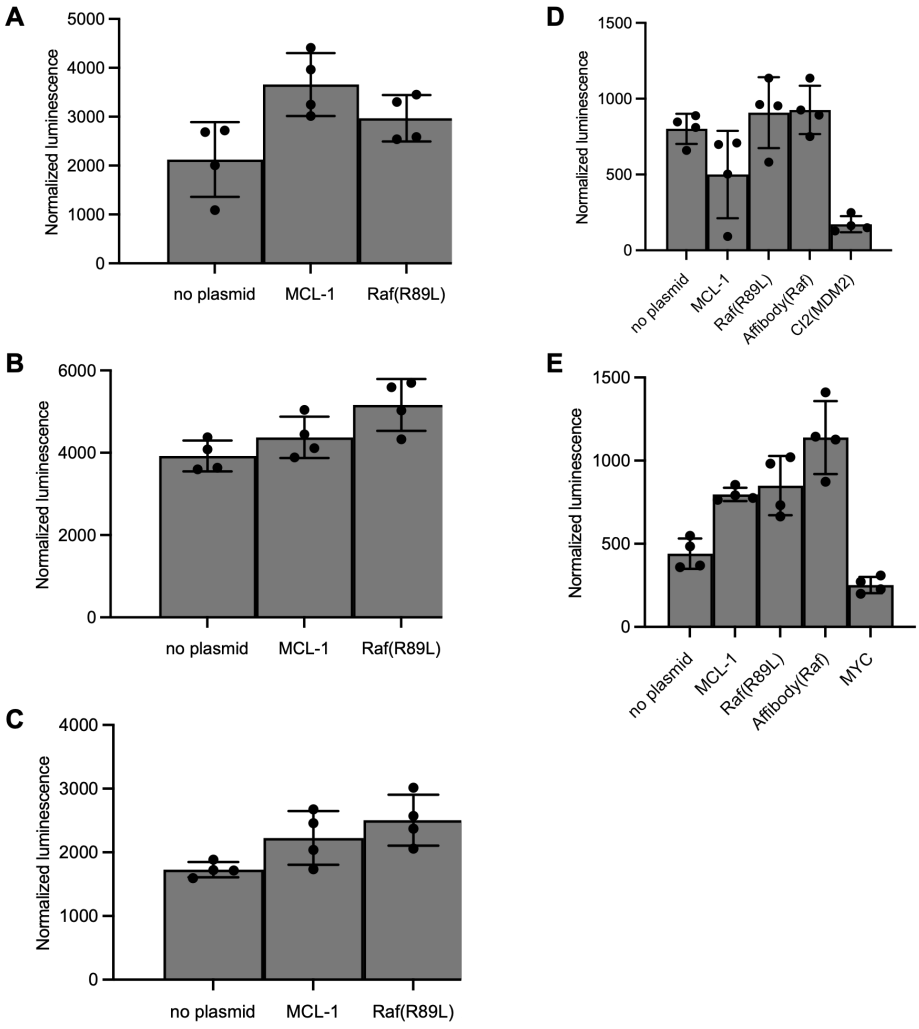


Figure 3.5 *In vivo E. coli* luciferase inhibitor assays to assess negative controls. The *E. coli* luciferase inhibitor assays when testing a selection of negative controls for the inhibition of (A) KRas-Raf, (B) KRas(G12D)-Raf, (C) KRas(G12V)-Raf, (D) P53-MDM2, and (E) MYC-MAX. Bars show mean $\pm$ SD of four replicates (circles).

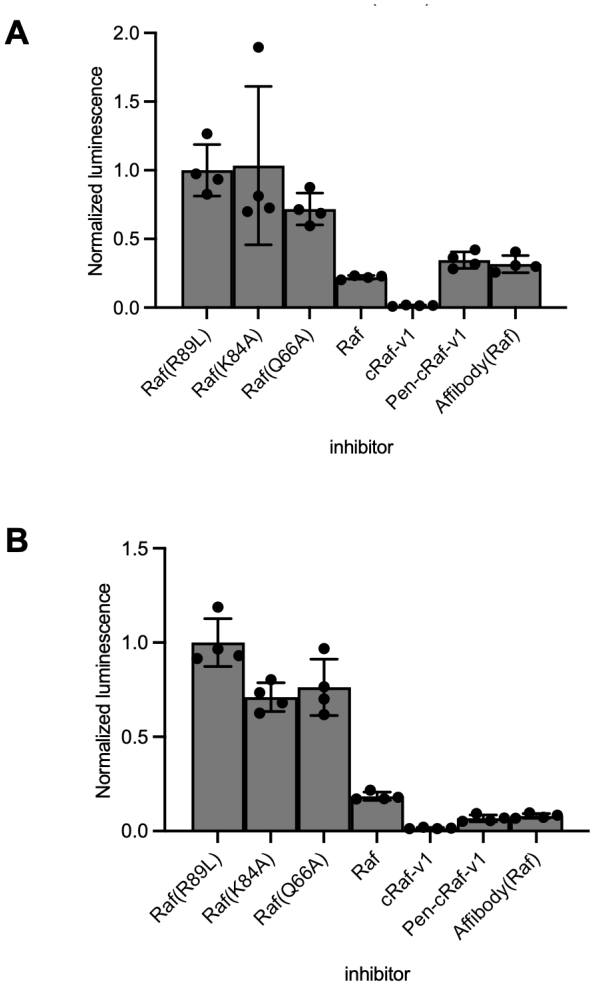


Figure 3.6 *In vivo E. coli* luciferase inhibitor assays for KRas mutants.
The *E. coli* luciferase inhibitor assay when testing known inhibitors of (A) KRas(G12D)-Raf and (B) KRas(G12V)-Raf. Bars show mean ± SD of four replicates (circles).

To confirm the generality of this result, we also tested whether P53-MDM2 inhibitors can be detected. Again, compared to expression of a Raf affibody as a negative control, expression of either P53 or MDM2, or an engineered chymotrypsin inhibitor 2 protein (CI2) binder of MDM2[276] , inhibits the signal (**Fig. 3.3H**). Likewise, we found that expression of either MYC or MAX as a competitive inhibitor in the MYC-MAX PPI detection system also inhibited the signal compared to the Raf affibody negative control (**Figure 3.3I**). Taken together, these data indicate the split RNAP sensor can detect the inhibition of range of disease-relevant PPIs in *E. coli* and encode that inhibition information in changes in gene

expression. We next sought to deploy the biosensors in the PANCS-PPI*i* system.

3.2.3 Mock selections to validate PANCS-PPIi

We first tested the assembly of our trimolecular complexes in our PANCS-PPI*i* system via luciferase assay and found it functioned as designed (**Figure 3.7**). We then cloned and optimized the components of PANCS-PPI*i* using competitive inhibitors of KRas-Raf, P53-MDM2, and MYC-MAX as models to validate the system. Broadly speaking, we would expect that phage encoding an active PPI inhibitor for a target PPI of interest should have a faster replication rate than phage without an inhibitor, and only replicate on cells containing the AP for the on-target PPI for that inhibitor. To quantify phage replication rates, we performed overnight phage growth assays (**Figure 3.8A**). In this assay 1000 phage are added to 1 mL LB cultures along with 10 μL of saturated S1030 cells containing PANCS-PPI*i* APs ($10^{6\text{-}7}$ cells) and are incubated overnight at 37°C, and then the phage population is measured by plaque assay. Using this assay, we varied the strengths of the ribosomal binding site (RBS) of each component of the system to alter the concentrations of the fused proteins and the output of gIII and gIII_{neg} to optimize phage propagation in the presence of each known inhibitor, using the range of inhibitor strengths to determine the optimal conditions (**Figure 3.9,10,11,12,13,14**). Under the optimized conditions, positive control phage, which replicate by bypassing the selection system, propagate from 1000 input phage to 10^{11} phage overnight (10^{7}-fold growth), while empty phage, which turn on the selection but do not encode an inhibitor, do not replicate at all (1000 input and 1000 are present after overnight incubation). Critically, phage that encode strong inhibitors for each of the three target PPIs restore phage replication to 10^{10} phage during the overnight growth assay, representing a 10,000,000-fold growth fitness advantage based on the presence of the PPI inhibitor (**Figure 3.8B**). Critically, the substantial growth fitness advantage is only observed when an inhibitor matches the target PPI of the APs,

indicating the selection pressure is focused on inhibition of the target PPI.

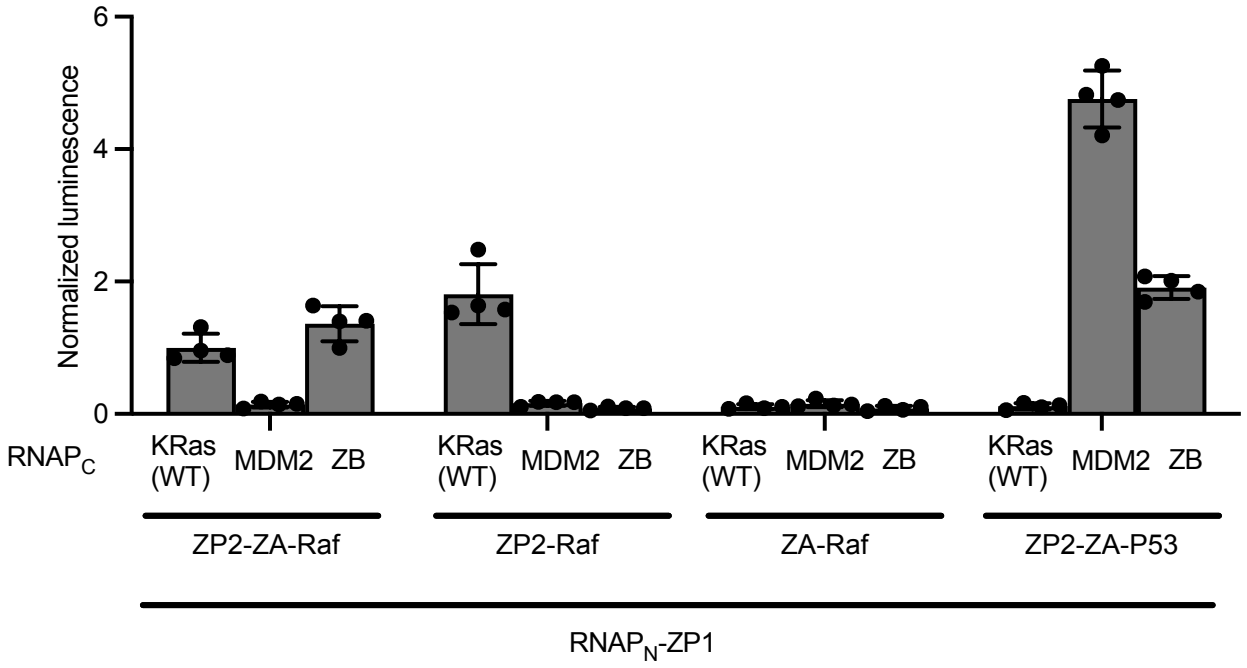


Figure 3.7 *In vivo E. coli* luciferase binding assays to assess trimolecular complex formation.

The *E. coli* luciferase binding assay when testing for trimolecular complex formation among the following component: 1) $RNAP_N$-ZP1; 2) KRas, MDM2, or ZB-$RNAP_{C(CGG)}$; and 3) ZP2-ZA-Raf, ZP2-Raf, ZA-Raf, and ZP-ZA-P53. Bars show mean ± SD of four replicates (circles).

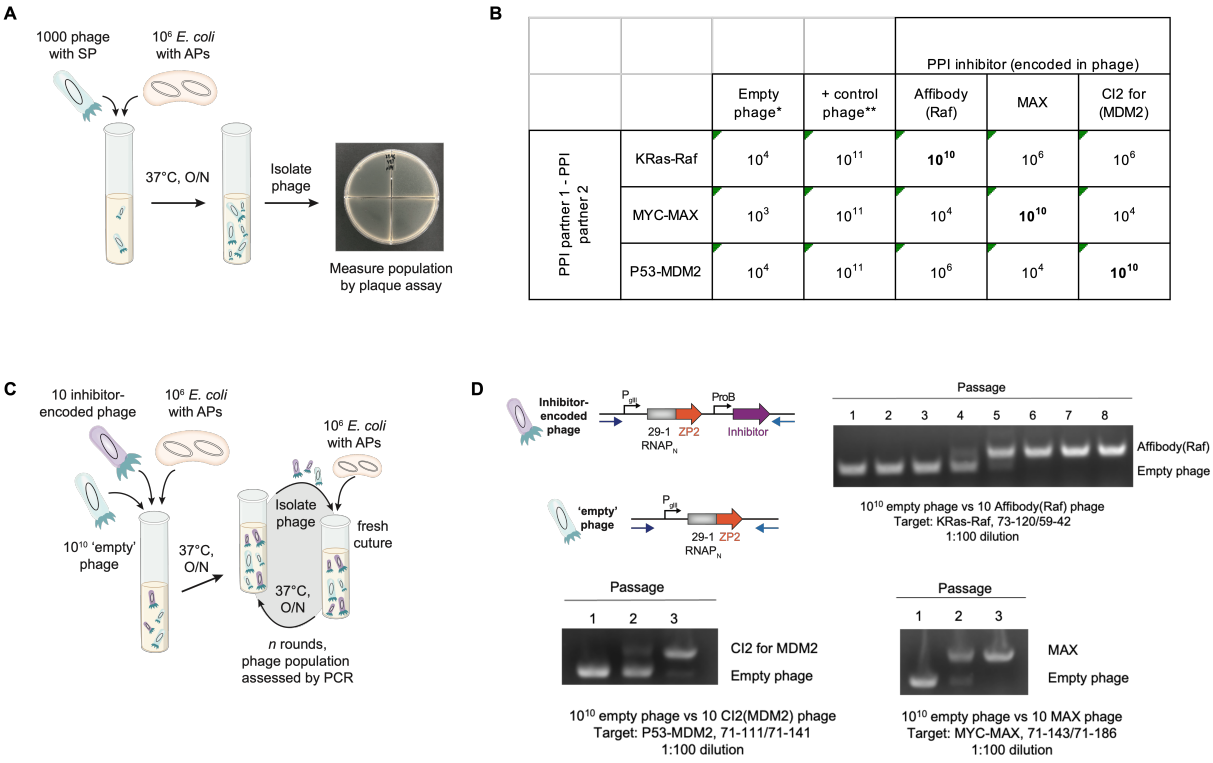
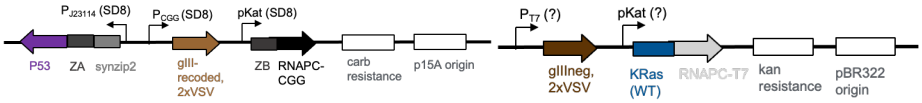
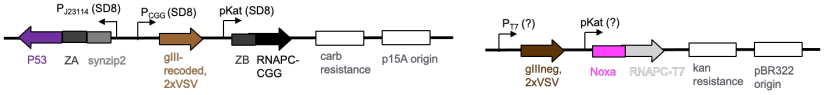


Figure 3.8 Validation of the PANCS-PPI*i* platform.
(A) Schematic demonstrating the process of a phage growth assay. 1000 phage with SP and 10^6 E. coli with APs are added to 1 mL LB, shaken overnight at 37°C, followed by phage isolation and population measurement by plaque assay. (B) Phage growth assays with 2 control phage and 3 known inhibitor-encoding phage grown on the plasmids encoding the 3 model PPIs (same as reported in (D)). Number reported is phage forming units per mL (PFU/mL) of one replicate. (C) Schematic demonstrating the process of a mock PANCS. 10 inhibitor-encoded phage, 10^{10} empty phage, and 10^6 *E. coli* with APs are added to 1 mL LB and shaken overnight at 37°C. The following day, phage are isolated and a fraction of that population is added to a fresh LB with 10^6 *E. coli* with APs and shaken overnight at 37°C. The process is repeated and analyzed by PCR, as shown in (D). (D) (Top left) Relevant plasmid maps of the inhibitor-encoded phage and empty phage. Black arrow indicates forward primer binding site. Blue arrow indicates reverse primer binding site. The PCR product of the inhibitor-encoded phage is longer than that of the empty phage. Gels with PCR products of passages from the mock PANCS of Raf-binding Affibody-encoded phage on Raf-KRas APs of (Top right), phage with a chymotrypsin inhibitor 2 (CI2)-based binder of MDM2 on P53-MDM2 APs (Bottom left), and MAX-encoded phage on MYC-MAX APs. *Empty phage do not encode an inhibitor. ** +control phage replicate by bypassing the selection conditions.

	Name	phage: 53-78, +control phage	phage: 73-119, empty phage
+AP	71-123	1.00E+11	2.00E+03
	71-125	1.00E+03	3.00E+03



+AP 71-111		phage (name, inhibitor): 53-78, +control phage	73-119, empty phage	71-114, MDM2	73-121, Pen-Raf
-AP (name, RBS of pT7, RBS of pKat)	59-43, sd8, sd8	1.00E+11	1.00E+10	1.00E+10	1.00E+10
	59-42, SD8, sd8	1.00E+11	1.00E+10	1.00E+10	1.00E+10
	66-82 SD8, SD8	1.00E+11	1.00E+07	1.00E+09	1.00E+10



+AP 73-120		phage (name, inhibitor): 53-78, +control phage	73-119, empty phage	71-114, MDM2	73-121, Pen-Raf
-AP (name, RBS of pT7, RBS of pKat)	23-49, sd8, sd8	1.00E+11	1.00E+08	5.00E+09	1.00E+11
	Jin 487, SD8, sd8	1.00E+11	1.00E+08	1.00E+08	1.00E+11
	28-67 SD8, SD8	1.00E+11	1.00E+07	1.00E+08	1.00E+11

Figure 3.9 A set of control phage growth assays to test PANCS-PPI*i*.
Plasmid maps are shown above corresponding phage growth assay data. Number reported is PFU/mL of one replicate.

		+AP 73-120							
		phage (name, inhibitor)							
		53-78, +control phage	73-119, empty phage	73-121, Pen-Raf	73-122, Pen-cRaf-v1	73-123, Pen-Raf (Q66A)	73-118, Raf (R89L)	71-157, Monobody (KRas)	71-158, Affibody (Raf)
-AP (name, RBS of PT7, RBS of pKat)	59-55, sd5, sd2	1.00E+11	1.00E+10	1.00E+10	1.00E+10	1.00E+10	1.00E+10	1.00E+11	1.00E+11
	66-95, sd5, sd5	1.00E+11	1.00E+08	1.00E+10	1.00E+10	1.00E+10	1.00E+08	1.00E+11	1.00E+11
	66-94, sd8, sd5	1.00E+11	5.00E+07	1.00E+10	5.00E+08	1.00E+10	1.00E+08	1.00E+11	1.00E+11
	59-43, sd8, sd8	1.00E+11	1.00E+05	1.00E+10	1.00E+09	1.00E+07	1.00E+05	5.00E+06	1.00E+09
	59-42, SD8, sd8	1.00E+11	5.00E+04	1.00E+10	1.00E+09	5.00E+05	5.00E+04	1.00E+06	1.00E+10
	66-82, SD8, SD8	1.00E+11	3.00E+03	2.50E+06	7.50E+06	6.00E+03	<10^3	4.00E+03	9.50E+04

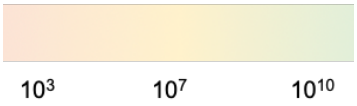


Figure 3.10 Phage growth assays on Raf-KRas APs.
Number reported is PFU/mL of one replicate.

+AP 73-120									
		phage (name, inhibitor)							
		53-78, +control phage	73-119, empty phage	73-121, Pen-Raf	73-122, Pen-cRaf-v1	73-123, Pen-Raf (Q66A)	73-118, Raf (R89L)	71-131, Raf (Q66A)	71-158, Affibody (Raf)
-AP (name, RBS of PT7, RBS of pKat)	20-79, sd8, sd5	1.00E+11	1.00E+08	1.00E+10	1.00E+10	1.00E+10	1.00E+10	1.00E+10	1.00E+10
	73-124, sd8, sd8	1.00E+11	5.00E+04	1.00E+08	1.00E+08	5.00E+07	1.00E+08	1.00E+08	1.00E+08
	32-10, SD8, sd8	1.00E+11	3.00E+03	1.00E+07	7.50E+07	4.00E+04	5.00E+04	1.00E+04	1.00E+10
	32-11, SD8, SD8	1.00E+11	<10^3	1.30E+04	4.40E+04	1.00E+03	1.00E+03	<10^3	4.00E+04

Figure 3.11 Phage growth assays on Raf-KRas(G12D) APs.
Number reported is PFU/mL of one replicate.

		+AP 73-120							
		phage (name, inhibitor)							
		53-78, +control phage	73-119, empty phage	73-121, Pen-Raf	73-122, Pen-cRaf-v1	73-123, Pen-Raf (Q66A)	73-118, Raf (R89L)	71-131, Raf (Q66A)	71-158, Affibody (Raf)
-AP (name, RBS of PT7, RBS of pKat)	72-25, sd8, sd5	1.00E+10	2.05E+06	1.00E+08	1.00E+07	5.00E+08	1.00E+08	2.50E+08	1.00E+10
	72-26, sd8, sd8	1.00E+11	8.00E+04	1.00E+08	1.00E+08	1.00E+09	1.00E+09	5.00E+08	1.00E+10
	72-27, SD8, sd8	1.00E+11	1.70E+04	5.00E+06	1.75E+06	6.00E+05	3.80E+04	4.40E+04	2.50E+08
	72-28, SD8, SD8	1.00E+10	2.00E+03	1.00E+06	1.00E+06	1.00E+04	3.60E+04	1.00E+03	1.00E+06

Figure 3.12 Phage growth assays on Raf-KRas(G12V) APs.
Number reported is PFU/mL of one replicate.

+AP 71-111						
		phage (name, inhibitor)				
		53-78, +control phage	73-119, empty phage	71-114, MDM2	71-115, P53	71-156, CI2 (MDM2)
-AP (name, RBS of PT7, RBS of pKat)	71-134, sd5, sd2	1.00E+11	1.00E+10	1.00E+11	1.00E+10	1.00E+10
	71-135, sd8, sd2	1.00E+11	1.00E+09	5.00E+09	1.00E+11	1.00E+11
	71-136, SD8, sd2	1.00E+11	5.00E+09	1.00E+11	1.00E+11	1.00E+11
	71-140, sd5, sd5	1.00E+11	1.00E+05	5.00E+06	1.00E+07	1.00E+10
	71-141, sd8, sd5	1.00E+11	1.00E+04	1.00E+05	1.00E+07	1.00E+10
	66-169, sd8, sd8	1.00E+11	<10^3	7.00E+03	<10^3	<10^3
	66-168, SD8, sd8	1.00E+11	<10^3	3.00E+03	<10^3	5.00E+04
	66-167, SD8, SD8	5.00E+10	<10^3	<10^3	1.00E+03	1.00E+05

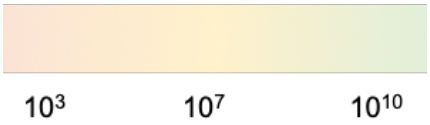
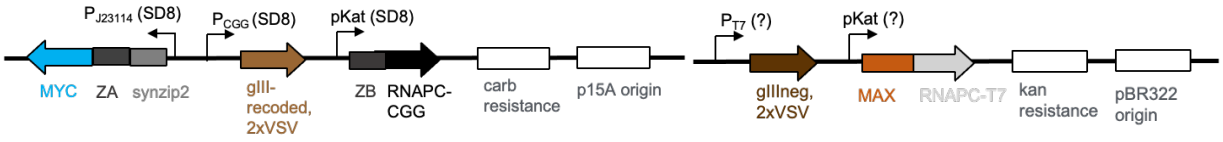


Figure 3.13 Phage growth assays on P53-MDM2 APs.
Number reported is PFU/mL of one replicate.



+AP 71-143							
		phage (name, inhibitor)					
		53-78, +control phage	73-119, empty phage	71-149, MYC	71-150, MYC_DBD	71-151, MAX	71-155, Telobody (MYC)
-AP (name, RBS of PT7, RBS of pKat)	71-192, sd8, sd8	1.00E+11	8.00E+04	3.75E+06	1.00E+10	1.00E+11	1.00E+10
	71-189, SD8, sd8	1.00E+11	5.00E+04	1.75E+06	1.00E+10	1.00E+11	1.00E+08
	71-186, SD8, SD8	1.00E+11	3.00E+03	4.00E+04	1.00E+10	1.00E+10	1.80E+04

Figure 3.14 Phage growth assays on MYC-MAX APs.
Number reported is PFU/mL of one replicate.

Given the dramatic growth rate differences of the PPI inhibitor-encoding phage in the PANCS-PPI*i* system, we postulated that simple serial dilutions would allow enrichment of phage encoding PPI inhibitors from pools of phage without inhibitors. Given that typical phage libraries contain 10^9 variants based on cloning approaches[277] , we aimed to optimize the system to enrich 1 active variant in a sea of 10^9 inactive variants to ensure we comprehensively cover typical libraries. To perform mock selections, we mixed 10^{10} negative control phage with 10 phage encoding a target inhibitor, and then grew those mixed phage populations up on *E. coli* the optimized vectors for PANCS-PPI*i*. After 16 hours of growth, we then reseeded fresh tubes of engineered *E. coli* cells with phage in a serial dilution, allowed them to grow up again for 16 hours, and repeated this serial dilution process up to 8 times over 3-5 days (**Figure 3.8C**). We monitored the phage population distribution at each step by PCR using primers that land up and downstream of the site in the phage genome where the inhibitor is encoded or not, which yields a larger size PCR

product for phage encoding an inhibitor. The initial input library only shows inactive phage by PCR, as the 10 active phage are below the limit of detection. However, after just 3-8 rounds of serial dilutions, the inactive phage populations de-enrich for each target PPI, and the active, PPI-encoding, phage appear (**Figure 3.8D, Figure 3.15**), representing at least a 10^{14}-fold relative enrichment of PPI inhibitor-encoding phage over non-inhibitor control phage in just a few days (starting with 1 active/10^9 inactive, ending with at least $10^9/10^5$). Given the strong performance of PANCS-PPI*i* in these mock selections, we next aimed to challenge the system with PPI inhibitor libraries encoded in phage.

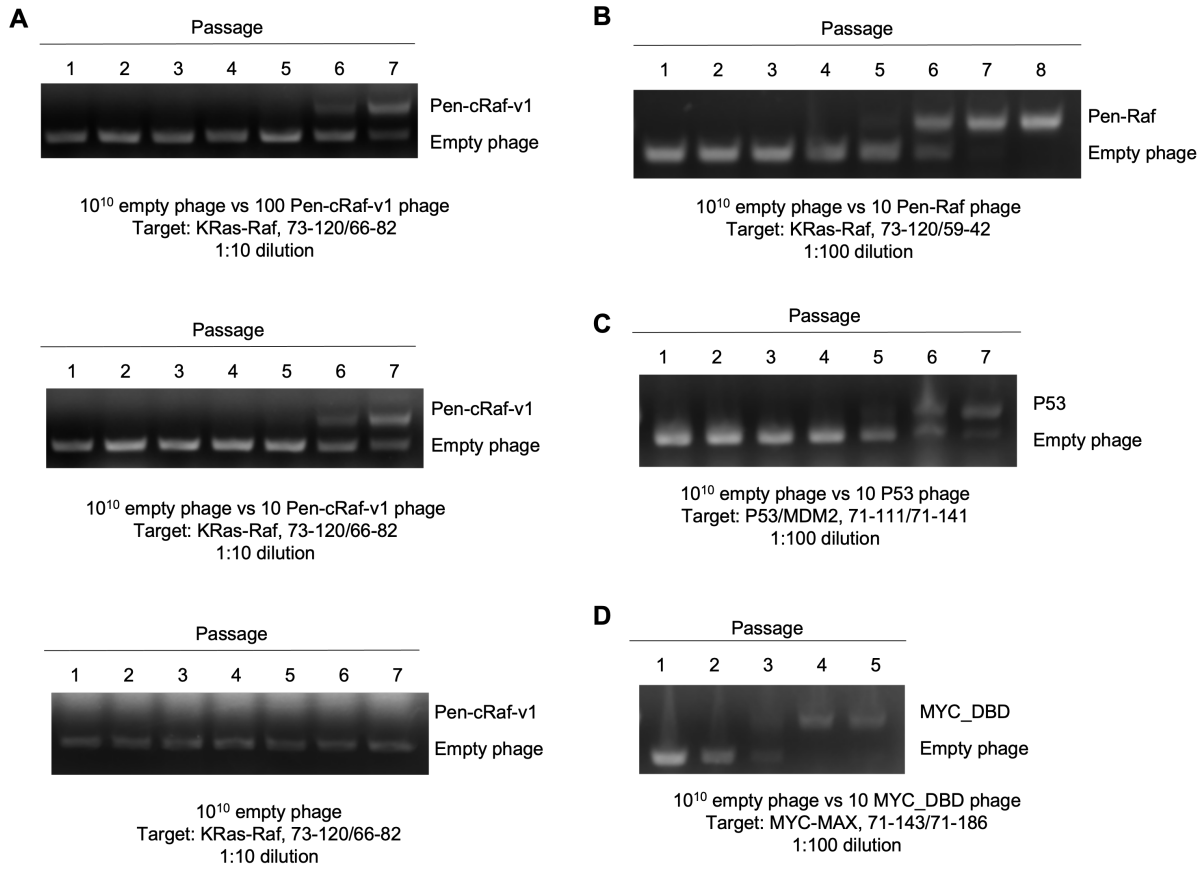


Figure 3.15 Additional mock PANCS-PPI*i* experiments.
Gels with PCR products of passages from the mock PANCS of (A) Pen-cRaf-v1-encoding phage on one set of Raf-KRas APs starting with 100 Pen-cRaf-v1 phage (Top), 10 Pen-cRaf-v1 phage (Middle), or 0 Pen-cRaf-v1 phage (Bottom); (B) Pen-Raf-encoding phage on another set of Raf-KRas APs, (C) P53 phage on P53-MDM2 APs, and (D) MYC phage on MYC-MAX APs.

3.2.4 Deep mutational scan of Raf-based inhibitors of KRas-Raf

We next tested whether PANCS-PPI*i* could be used to perform a deep mutational scan of a competitive PPI inhibitor, targeting Raf-based inhibitors of the KRas-Raf interaction as a model. While cell-penetrating peptides (Pen) can be used to enhance the uptake of protein-based KRas-Raf inhibitors, appending a Pen onto Raf, or engineered variants of Raf, decreases its affinity for KRas and its ability to inhibit the complex[274]. We hypothesized that mutations near the interface could enhance KRas-Raf inhibition by modulating KRas binding in the context of the cell-penetrating peptide. We therefore selected 4 positions at the PPI interface to randomize on Pen-Raf (**Figure 3.16A**, PDB: 6VJJ), using NNK degenerate codons. Because this library theoretically consists of 194,481 unique protein variants and our transformation efficiency was $10^{6\text{-}8}$, we are confident that we could sample essentially every variant to perform a comprehensive mutational scan of these sites. Furthermore, we subcloned out 10 random variants and found that each which were different mutants and all were worse inhibitors than the parent Pen-Raf (**Figure 3.16B**), which is expected based on the library size.

With the Pen-Raf phage library in hand, we performed 8 passages of the library on the KRas-Raf PANCS-PPI*i* system with 2 different AP combination strengths. At the first round of each selection, we also started with mixture of 10^9 phage from the library and 109 empty phage. This empty phage spike in enables us to track the de-enrichment of phage that do not encode a functional inhibitor by PCR, so that we can track the progress of the selections in terms of de-enriching inactive variants. After 4-6 rounds of selection, both conditions resulted in complete de-enrichment of the negative control phage (**Figure 3.16C,D**), indicating the selections were successful and complete. Phage population numbers and activity-dependent plaque assays further suggested that the ending populations largely contained functional inhibitors (**Figure 3.17**). We then cloned out 10 random variants from each library at the end of the selection and assessed their ability

to inhibit KRas-Raf using the PPI inhibitor lux assay (**Figure 3.16E,F**). Although each variant had a different genotype, 100% of the variants were active inhibitors, and excitingly, all were equivalent or better inhibitors than the parent Pen-Raf. We are in the process of further characterizing the activity of these variants by *in vitro* binding assays and assessing the population throughout the selections via high-throughput sequencing to identify advantageous and disadvantageous residues. Taken together, this DMS experiment showed that PANCS-PPI*i* can enrich activate variants from phage-encoded inhibitor libraries and can engineer improved inhibition properties of protein variants, all while providing high-throughput mutational data.

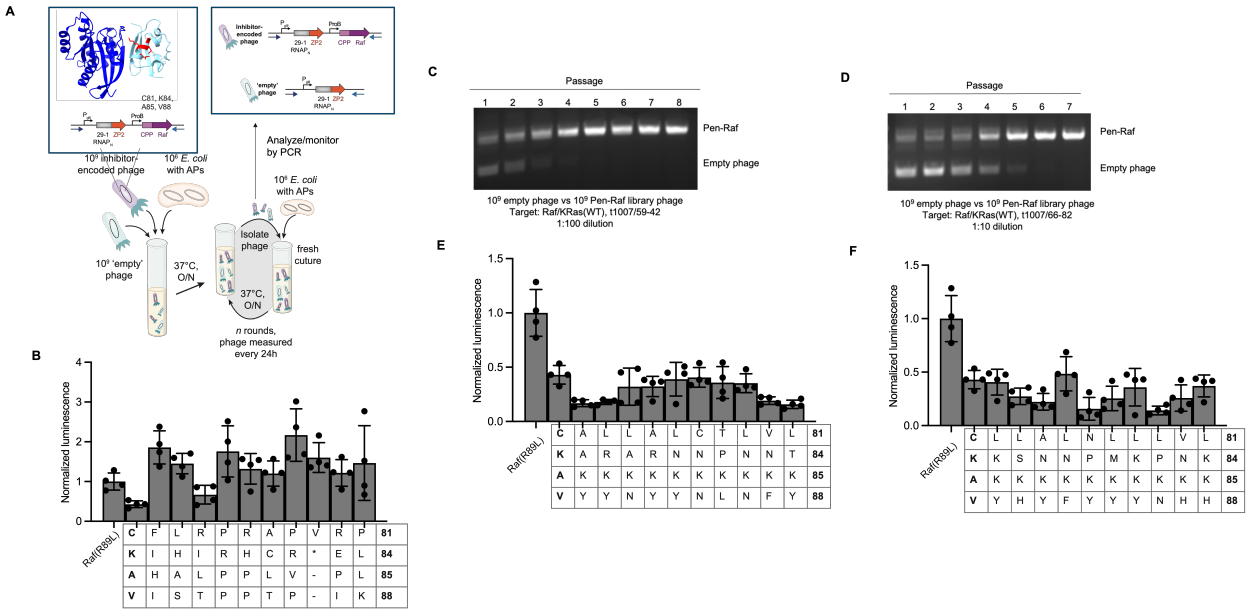


Figure 3.16 A deep mutational scan of Pen-Raf as an inhibitor of KRas-Raf using PANCS-PPI*i*

(A) Schematic demonstrating the process of PANCS-PPI*i* with the Pen-Raf library. Four residues, located at the KRas binding interface in the KRas-Raf crystal structure (PDB: 6VJJ), were randomized using NNK degenerate codons. 10^9 Pen-Raf library phage, 10^9 empty phage, and 10^6 *E. coli* with APs were added to 1 mL LB and shaken overnight at 37°C. The following day, phage were isolated and a fraction of that population was added to a fresh LB with 10^6 *E. coli* with APs and shaken overnight at 37°C. The process was repeated for a total of 8 passages and analyzed by PCR, as shown in (C) and (D). (B) The *E. coli* luciferase inhibitor assay when testing 10 random variants from the starting Pen-Raf library. (C) Gel with PCR products of passages from the Pen-Raf PANCS-PPIi experiments with condition 1. (D) Gel with PCR products of passages from the Pen-Raf PANCS-PPIi experiments with condition 2. (E) The *E. coli* luciferase inhibitor assay when testing 10 random variants from passage 8 from condition 1. (F) The *E. coli* luciferase inhibitor assay when testing 10 random variants from passage 8 from condition 2. For *E. coli* luciferase inhibitor assay graphs, the residues corresponding to each randomized site are shown below the bar for the corresponding variant. The bolded left-most column is the wildtype Pen-Raf, and the bolded right-most column corresponds to the site number in full-length Raf. Bars show mean ± SD of four replicates (circles).

		+AP 73-120						
		phage (name, inhibitor)						
		73-119, empty phage	73-121, Pen-Raf	73-122, Pen-cRaf-v1	73-124, Pen-Raf (Q66A)	Pen-Raf library, P0	Pen-Raf library, P8 from 59-42	Pen-Raf library, P8 from 66-82
-AP (name, RBS of PT7, RBS of pKat)	59-42, SD8, sd8	$<10^3$	$2x10^9$	$1.7x10^{10}$	$<10^3$	$3.2x10^9$	$2*10^{10}$	$2x10^{10}$
	66-82, SD8, SD8	$<10^3$	$<10^3$	$<10^3$	$<10^3$	$<10^3$	$5*10^9$	$1x10^9$

Figure 3.17 Activity-dependent plaque assays of populations from the Pen-Raf PANCS-PPI*i* experiments.
Number reported is PFU/mL of one replicate.

3.2.5 De novo selection of a P53-MDM2 PPI inhibitor

Rather than starting from a library of an active PPI inhibitor and targeting improvement, we next assessed whether PANCS-PPI*i* could discover PPI inhibitors from random libraries, de novo. For the inhibitor library, we decided to use an affibody scaffold[278] , due to its small size and robust folding capability. After cloning an affibody scaffold into the inhibitor position on the SP, we cloned an affibody library in the phage by randomizing 17 positions on the protein scaffold, chosen based on previous work for protein binder discovery (**Figure 3.18A**)[279] . Theoretically, this library consists of 10^{14} variants, but based on cloning estimates, the final phage library contained 10^8 individual variants.

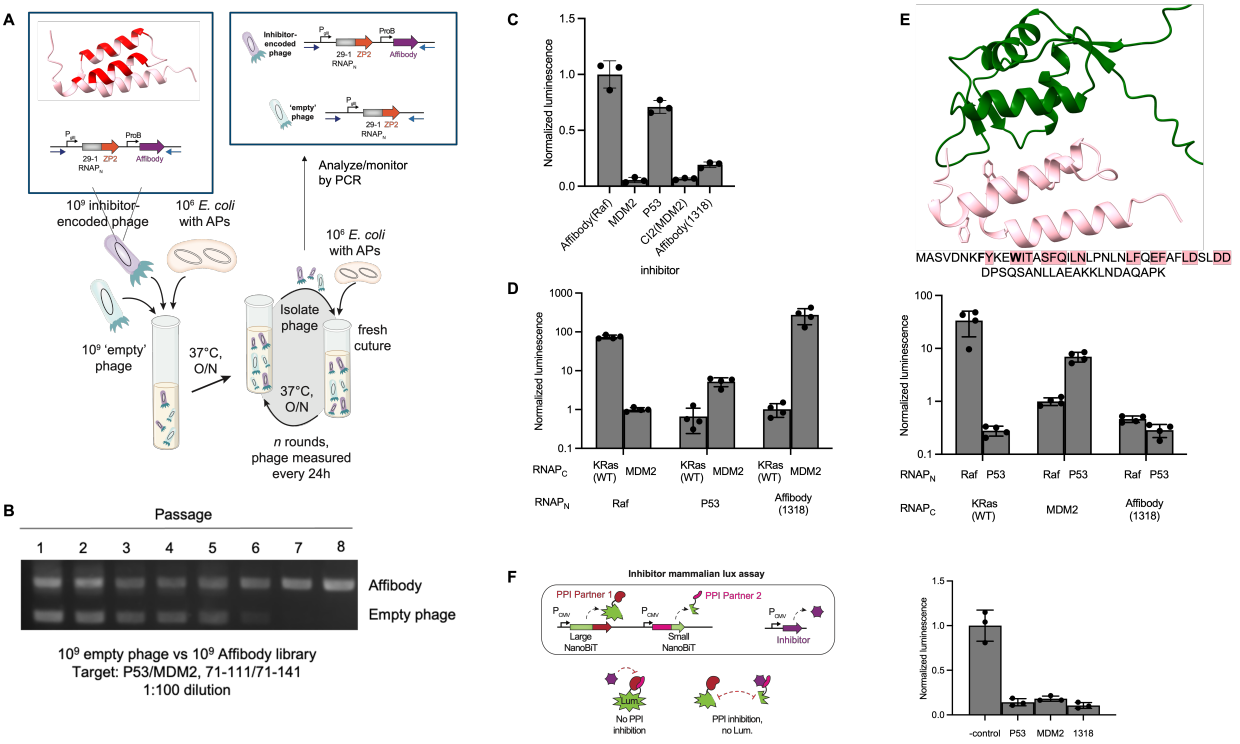


Figure 3.18 PANCS-PPI*i* identifies a de novo affibody inhibitor of the P53-MDM2 interaction.

(A) Schematic demonstrating the process of PANCS-PPI*i* with the affibody library. Residues in two alpha-helices along one face of the affibody scaffold were randomized. 10^9 affibody library phage, 10^9 empty phage, and 10^6 *E. coli* with APs were added to 1 mL LB and shaken overnight at 37°C. The following day, phage were isolated and a fraction of that population was added to a fresh LB with 10^6 *E. coli* with APs and shaken overnight at 37°C. The process was repeated for a total of 8 passages and analyzed by PCR, as shown in (B). (C) The *E. coli* luciferase inhibitor assay when testing the dominant variant (1318) that emerged from PANCS-PPI*i*. (D) The *E. coli luciferase* binding assays when testing the dominant variant (1318) that emerged from PANCS-PPI*i*. (E) An Alphafold2-generated prediction of the 1318 affibody (pink) binding to MDM2 (dark green). The affibody sequence is shown below, with the randomized sites highlighted in pink and the FXXX motif in bold. (F) (Left) Schematic of the genetically-encoded components of the inhibitor mammalian luciferase (lux) assay. One plasmid encodes PPI partner 1 fused to the large NanoBiT fragment as well as PPI partner 2 fused to the small NanoBiT fragment. Another plasmid contains a genetically-encoded inhibitor. If PPI partner 1 binds to PPI partner 2, then the luciferase protein will be reconstituted and produce luminescence. If the inhibitor prevents the binding of PPI partner 1 and PPI partner 2, the luciferase protein will not be reconstituted; thus, no luminescence will result. (Right) The mammalian luciferase inhibitor assay when testing the dominant variant (1318) that emerged from PANCS-PPI*i*.

With the affibody inhibitor library in hand, we subjected it to PANCS-PPI*i* for P53-MDM2 inhibitors. As we did for the DMS, we initiated each selection with the library alone as well as a 1:1 ratio of library (10^9) and empty (10^9) phage, so that we could use PCR to determine if/when the negative control phage (and other non-functional library phage) were effectively de-enriched. After 8 passages, we found the negative control phage were non-detectable by PCR, while the affibody populations remained strong (**Figure 3.18B**). We used activity-dependent plaque assays to further confirm that the populations at passage 8 contained active inhibitors (**Table 3.1**).

Table 3.1 Activity-dependent plaque assays of populations from the affibody PANCS-PPIi experiment. Number reported is PFU/mL of one replicate.

	Affibody library P0	Affibody library P8
Population	8.75×10^9	2.8×10^7
Population with activity on target PPI APs (71-111/71-141)	3.5×10^5	1.8×10^7

We next cloned six random affibody variants from the phage at the end of the selection into the PPI inhibitor assay vector, which revealed the libraries had largely converged on a single genotype, "1318." We tested whether 1318 is a P53-MDM2 inhibitor using the *E. coli* PPI inhibitor, which confirmed 1318 is a P53-MDM2 inhibitor (**Figure 3.18C**). To decipher the mechanism of action of 1318, we then tested whether it binds to P53 or MDM2 directly using the luciferase binding assay, which revealed strong MDM2 binding, but no P53 binding (**Figure 3.18D**). Indeed, AlphaFold2 predicted a high-confidence structure of the interaction between 1318 and MDM2, showing key interaction of a FXXXW motif on 1318 with MDM2 (**Figure 3.18E, Figure 3.19**)[280] , a well-known recognition motif for this protein[281] . Furthermore, 1318 showed inhibition of the P53-MDM2 interaction in HEK293T cells in a split NanoLuc reporter assay (**Figure 3.18F**), which demonstrates its biological activity outside of *E. coli* cells and in a more native mammalian environment. Taken together, these results confirm that PANCS-PPI*i* can discover PPI inhibitors de

novo from random libraries.

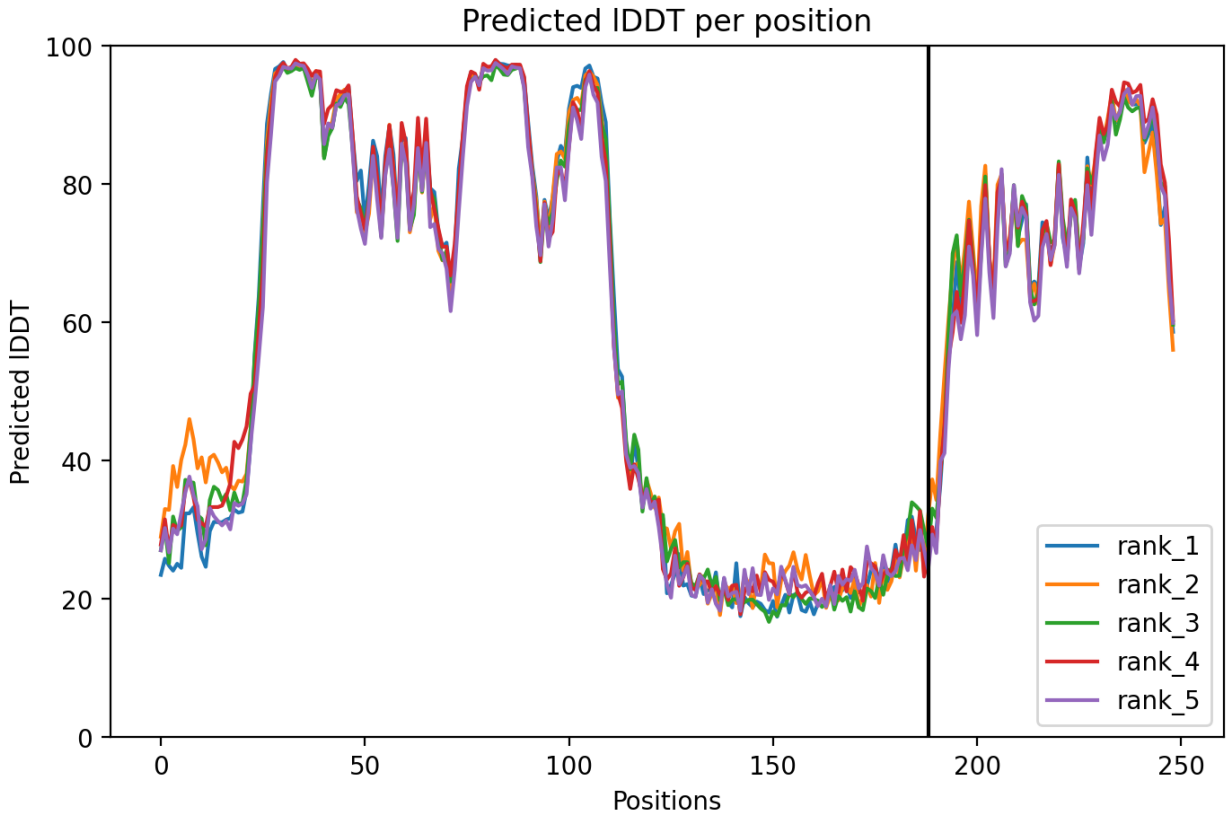


Figure 3.19 Predicted local distance difference test (lDDT) per position of Alphafold2-generated model of affibody 1318 binding to MDM2.
Left corresponds to MDM2, and right corresponds to 1318 affibody.

3.3 Discussion

Here, we developed PANCS-PPI*i*, an *in vivo* phage-based directed evolution platform that selects directly for molecules that disrupt PPIs. We showcased the ability of PANCS-PPI*i* to search through large libraries of 10^8 or more genetically encoded molecules and identify PPI inhibitors, in just a few days. Moreover, the selection process simply involves serial dilutions on *E. coli* cells, with each round of selection taking less than 5 minutes worth of researcher work and without any specialized equipment. Finally, we showed PANCS-PPI*i* is robust, and can reproducibly enrich active variants and reproducibly drive

extinction events of active variants, avoiding the pitfalls of cheaters and background that often plaque traditional selection-based methods. We found PANCS-PPI*i* can not only the enrich 1 active inhibitor in a population of 1 billion non-functional inhibitors after just a few rounds of selection, but the de-enrichment of the non-functional inhibitors was complete. This extinction of inactivate variants is a key strength of the method, as ascertaining whether a selection worked or not is as simple as determining whether the phage survived. PANCS-PPI*i* therefore provides a high-throughput method to directly select for PPI inhibitors.

In this work, we demonstrated the compatibility of PANCS-PPI*i* in terms of PPI targets across three distinct, clinically important PPIs. The platform should be broadly applicable for any PPI that functions in *E. coli*. For each new compatible target, we outlined a straightforward method to tune the selection pressures of the system by adjusting gene expression via altering the RBSs of the promoters on the APs, using simple phage replication assays to assess the selection pressures. For the targets presented here, we demonstrated a range of tunability such that inhibitors of varying strengths could be selected for, depending on the properties of the the library and target PPI. In terms what can be screened for, PANCS-PPI*i* allows any genetically encoded biomolecule to act as an inhibitor. Though in this work we used protein scaffolds, a variety of scaffolds could be used, including small molecules such as cyclic peptides or other natural product-like molecular scaffolds[282] .

The selection itself is technically easy to carry out. No proteins need to be purified; the researcher only needs to clone the APs and SPs needed, and the selection simply consists of passaging phage, which can be done in a medium-throughput manner with replicates and conditions via the use of deep-well plates and multichannel pipettes. With the use of the empty phage spike-in, the selections are also easily and quickly monitored via PCR. Depending on the target and conditions, one could isolate functional inhibitors in

as few as 3 passages. Since passages can be as short as 6 hours, this means one could go from a library to a confident inhibitor hit in less than 24 hours, as we demonstrated in the case of the P53-MDM2 and MYC-MAX interactions (**Figure 3.8D**).

An advantage of this PANCS-PPI*i* over display technologies is the system directly selects for PPI disruption, rather than simply binding to one of the two PPI partners. While in this work the inhibitors discovered functioned as competitive binders, this mechanistically agnostic selection should allow for the discovery if inhibitors with novel mechanisms of inhibition, such as through interactions with allosteric sites. Another advantage of PANCS-PPI*i* is that the inhibitor is expressed without a tag, which could impede or enhance the activity and is not feasible for some molecular scaffolds, and that the inhibitor must function in *E. coli*, amongst a sea of other potential off-targets. In this work, we used a simple zipper peptide pair (ZA-ZB) as the counter-selection off-target PPI, which is critical for focusing the selection pressure on the target PPI and avoiding things like RNAP inhibitors or other cheaters. However, the counter-selection, off-target PPI could in principle be any desired off-target PPI of interest (**Figure 3.20**), allowing for the screening of mutant- or isoform-selective inhibitors, as we demonstrated recently in a related PPI selectivity technology[65] . In conclusion, PANCS-PPI*i* is a simple to execute and robust method to directly select for inhibitors of PPIs in a mechanistically-agnostic manner, which will open up new opportunities for the discovery of inhibitors for use as biological tools or as leads in drug discovery campaigns.

+AP 72-29					
		phage (name, inhibitor)			
		73-119, empty phage	71-115, P53	73-122, Pen-cRaf-v1	71-158, Affibody (Raf)
-AP (name, RBS of PT7, RBS of pKat)	No -AP	3.52E+09	6.00E+10	1.00E+08	1.00E+09
	20-76, sd5, sd2	3.00E+05	2.50E+06	<10^3	1.70E+04
	20-78, sd5, sd5	3.00E+04	4.00E+04	1.00E+03	1.30E+04
	20-79, sd8, sd5	1.00E+04	1.10E+03	<10^3	1.60E+04
	T43, sd8, sd8	1.00E+03	<10^3	<10^3	1.60E+04
	32-10, SD8, sd8	1.00E+03	<10^3	<10^3	2.40E+04
	32-11, SD8, SD8	<10^3	<10^3	<10^3	6.00E+04

+AP 72-29					
		phage (name, inhibitor)			
		73-119, empty phage	71-115, P53	73-122, Pen-cRaf-v1	71-158, Affibody (Raf)
-AP (name, RBS of PT7, RBS of pKat)	No -AP	3.52E+09	6.00E+10	1.00E+08	1.00E+09
	72-23, sd5, sd2	5.00E+05	2.50E+06	<10^3	2.60E+04
	72-24, sd5, sd5	5.00E+05	1.50E+06	<10^3	1.90E+04
	72-25, sd8, sd5	6.00E+04	3.50E+05	<10^3	3.00E+04
	72-26, sd8, sd8	4.00E+04	<10^3	<10^3	1.70E+04
	72-27, SD8, sd8	2.00E+04	1.00E+03	<10^3	7.00E+03
	72-28, SD8, SD8	5.00E+03	<10^3	<10^3	1.00E+04

Figure 3.20 Phage growth assays on AP combinations that select for inhibitors of KRas(G12D)-Raf but not KRas-Raf (top) and of KRas(G12V)-Raf but not KRas-Raf (bottom).
Number reported is PFU/mL of one replicate.

3.4 Methods

3.4.1 General methods

Protein structures were generated using UCSF ChimeraX[283] . *E. coli* 10-beta cells (NEB) were used for cloning and were cultured in 2xYT media. *E. coli* BL21 (BE3) cells (NEB) were used for protein expression and were cultured in Luria-Bertain (LB) broth. *E. coli* S1030 cells[200] were used for activity-dependent plaque assays, phage growth assays, luciferase assays, and selections and were cultured in LB broth. *E. coli* 1059 cells[200] were used for cloning phage and plaque assays and were cultured in 2xYT media for cloning and LB for plaque assays. The following working concentrations of antibiotics were used: 50 μg/mL carbenicillin, 40 μg/mL kanamycin, and 33 μg/mL chloramphenicol. Human embryonic kidney (HEK) cell line 293T (female, ATCC) was maintained at 37°C with 5% carbon dioxide in DMEM (L-glutamine, high glucose, sodium pyruvate, phenol red, Corning) with 10% fetal bovine serum (FBS, Gemini Benchmark) and 1x penicillin/streptomycin (GIBCO/Life Technologies). Cells were passaged at a ratio of 1:10 to 1:20 every 2-3 days when at approximately 90-100% confluency by washing with Dulbecco's phosphate-buffered saline (PBS) and treating with Trypsin-EDTA 0.25% (GIBCO) to lift.

3.4.2 Cloning

Plasmids were constructed by amplifying fragments using Q5 DNA Polymerase (NEB) and ligating the fragments via Gibson Assembly. Primers were obtained from IDT, and all plasmids were sequenced at the University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility. The plasmids used in this study are available upon request.

3.4.3 Luciferase assays

For the *E. coli* luciferase binding assays, one vector contained the following the previously-evolved N-terminal half of RNAP (Zinkus-Boltz et al., 2019) fused to one protein, and another vector encoded both the C-terminal half of CGG RNAP fused to one protein and the CGG promoter-driven luciferase reporter. The *E. coli* luciferase inhibitor assays contained both vectors above as well as a vector containing an isopropyl -D-1-thiogalactopyranoside (IPTG)-inducible inhibitor, and the trimolecular complex binding assay likewise contained both vectors above as well as a vector containing an isopropyl -D-1-thiogalactopyranoside (IPTG)-inducible entity with various protein fusions. Chemically competent S1030 *E. coli* cells were prepared by culturing to an OD600 of 0.3, washing twice with a calcium chloride/HEPES solution (60 mM CaCl2, 10 mM HEPES pH 7.0, 15% glycerol), resuspending in the same solution, and flash-freezing in liquid nitrogen and storing at -80°C. Vectors were transformed into chemically competent S1030 cells via heat shock at 42°C for 45 s, followed by 1 hr recovery in 3Œ volume of 2xYT media, and then plated on agar with the appropriate antibiotics and left to incubate overnight at 37°C. Individual colonies (three to four biological replicates per condition) were picked and cultured in 1 mL of LB media containing the appropriate antibiotics overnight at 37°C in a shaker. The next morning, 50 μL of each culture was diluted into 450 μL of fresh LB media containing the appropriate antibiotics. For cells containing the IPTG-inducible inhibitor plasmid, each culture also contained 1 mM of IPTG, and for cells containing the IPTG-inducible trimolecular complex fragment, each culture contained 0.1 mM of IPTG. The cells were incubated in a shaker at 37°C, and OD600 and luminescence measurements were recorded between 3 and 4 hours after the start of the incubation. Measurements were taken on a Synergy Neo2 Microplate Reader (BioTek) by transferring 125 μL of the daytime cultures into Corning black, clear-bottom 96-well plates. Data were analyzed in Microsoft Excel and plotted in GraphPad Prism, as previously reported[63] .

3.4.4 Plaque assays

Plaque assays to assess total phage population were performed on 1059 *E. coli* cells, which supply gene III (gIII) to phage in an activity-independent manner. Additionally, activity-dependent plaque assays were done on S1030 *E. coli* containing the desired accessory plasmids (APs) to determine the number of phage in a population that could replicate on those APs and therefore had the activity needed to confer desired PPI inhibition. All cells were grown to an OD600 of approximately 0.6 during the day. Four serial dilutions were done in 1.2 mL 12-well tube strips (VWR) by serially pipetting 1 μL of phage into 50 μL of cells to yield the following dilutions: 1/50, 1/2500, 1/125,000, and 1/6,250,000. 650 μL of top agar (0.7% agar with LB media) was added to each tube, which was then immediately spread onto a quad plate containing already-solidified bottom agar (1.5% agar with LB media). Plates were incubated overnight at 37°C. Plaques were counted the following day, and plaque forming units (PFU) per mL was calculated using the following equation:

$$PFU = 1000 \times A \times 50^{\text{4-B}}$$

Where A is the number of plaques in a given quadrant, and B is the quadrant number where the phage were counted, in which one is the least dilute quadrant and four is the most dilute quadrant.

3.4.5 Phage growth assays

Phage growth assays were performed by adding the following to a culture tube and shaking overnight at 37°C: 1 mL of LB with the appropriate antibiotics, 10 μL of saturated S1030 *E. coli* containing the accessory plasmids of interest, and 1000 phage. Phage were then isolated by centrifugation at 16,000 rcf for 3 min, and PFU was determined by

plaque assays using 1059 *E. coli* and the plaque assay protocol described above.

3.4.6 Phage-Assisted Non-Continuous Selection (PANCS)

Phage-assisted non-continuous selection took place by doing successive phage growth assays, using phage dilutions for each passage. To begin, the following were added to each well of a 96 deep well plate (Fisherbrand), which was then shaken overnight at 37°C: 1 mL of LB with the appropriate antibiotics, 10 μL of saturated S1030 *E. coli* containing the accessory plasmids of interest, and 10^9 library-containing phage. An identical 96 deep well plate was likewise made and cultured that contained the above plus 10^9 empty phage in each well in order to monitor non-functional phage de-enrichment. (For the mock PANCS experiments, one plate was run with 10 phage with a functional inhibitor encoded and 10^{10} empty phage.) The next day, the phage were collected by centrifuging the deep well plates at 3000 g for 5 minutes and collecting the supernatant from each well. To being a new passage, the following were added to each well of a 96 deep well plate (Fisherbrand), which was then shaken for at least 6 hours at 37°: 1 mL of LB with the appropriate antibiotics, 10 μL of saturated S1030 *E. coli* containing the accessory plasmids of interest, and either a 1:10 (100 μL) or 1:100 dilution (10 μL) of phage from the previous passage. The selections were monitored by Phusion PCR (NEB, 50 uL reactions) using 1 μL of phage from the samples from the plate that contained the empty phage spike in control using the following primers: BR76 and JD1060, which result in a 735 bp fragment when used on just the empty phage and an at least 931 kb fragment when used on inhibitor-encoding phage (exact length depends on the inhibitor).

3.4.7 Protein purification

Proteins were cloned with a 6xHis-tag fused to their N-termini. The proteins were expressed in BL21 *E. coli* (NEB) and purified following standard Ni-NTA resin purification protocols (ThermoFisher Scientific)[114] . Briefly, BL21 *E. coli* containing the protein-encoded plasmid of interest were cultured in 5 mL LB with kanamycin overnight. The following day, the culture was added to 0.5 L of LB with kanamycin, incubated at 37°C until it reached an OD600 of 0.6, induced with IPTG (final concentration: 200 μM), and cultured overnight at 16°C. The cell pellet was harvested by centrifugation followed by resuspension in 30 mL of lysis buffer (50 mM Tris, 500 mM NaCl, pH 7.5) supplemented by protease inhibitors (200 nM Aprotinin, 10 μM Bestatin, 20 μM E-64, 100 μM Leupeptin, 1 mM AEBSF, 20 μM Pepstatin A). Cells were lysed via sonication (90% amplitude for 1 second on and 1 second off for 1 minute, followed by resting on ice for 2 minutes and then another cycle) total and were then centrifuged at 12,000 g for 40 min at 4°C. Solubilized proteins, located in the supernatant, were incubated with His60 Ni Superflow Resin (Takara) for 1 hr at 4°, and the protein was eluted using a gradient of imidazole in lysis buffer (50250 mM). Fractions with the protein, as determined by SDS-PAGE, were concentrated in Centrifugal Filter Units (Amicon, EMD Millipore). Proteins were further purified via a desalting column with storage buffer (50 mM Tris, 500 mM NaCl, 3 mM MgCl2, pH 7.5) and further concentrated. The concentrations of the purified proteins were determined by BCA assay (ThermoFisher Scientific), and they were flash-frozen in liquid nitrogen and stored at 20°.

3.4.8 Split NanoLuc mammalian cell assays

HEK293T cells were passaged into black 96-well clear bottom plates. Following 16 hours post-passaging, the cells reached 60% confluency, and each was transfected with 60 ng of the split reporter plasmid and 360 ng of the inhibitor expression plasmid in 20

uL Opti-MEM I Reduced Serum Medium (ThermoFisher Scientific) and 0.5 uL of Lipofectamine 2000. After 48 hours, media was replaced with 100 uL Opti-MEM I Reduced Serum Medium, and each well was treated and mixed with 25 uL of Nano-Glo Live Cell Reagent (Promega N2011) and luminescence measurements were immediately taken on a Synergy Neo2 Microplate Reader (BioTek). Data were analyzed in Microsoft Excel and plotted in GraphPad Prism.

3.5 Supplementary notes

Design 1: We first attempted to create the PANCS-PPI*i* platform by using the same SP and APs as the PACE platform we designed to evolve PPIs, but also adding the genetically encoded inhibitor into the SP (**Figure 3.21**). However, we found that even out of a theoretically monoclonal phage population of 10^9 phage, cheaters existed, purportedly through chance natural mutagenesis, that rendered the PPI partner 1 in the phage non-functional, thus bypassing the negative selection. Even when we had 2 copies of PPI partner 1 in the phage, there still existed cheater phage that removed the PPI partner 1 genes, thus bypassing the selection.

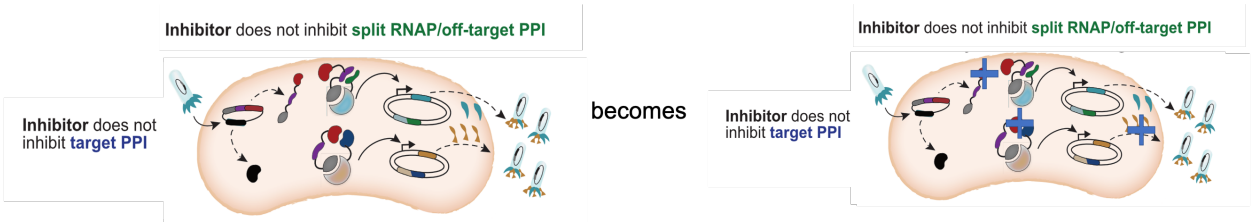


Figure 3.21 Schematic showing supplementary note, design 1.

Example 1: We added 10^9 of the following phage to S1030 *E. coli* cells with the following APs and performed 4 passages with 1:20000 dilution rate of phage (**Figure 3.22**) . The phage population over all passages was 1011 and phage variants at the end contained Raf variants in the phage that had the following inactivating mutations:

Q66taa(stop) and R89C.

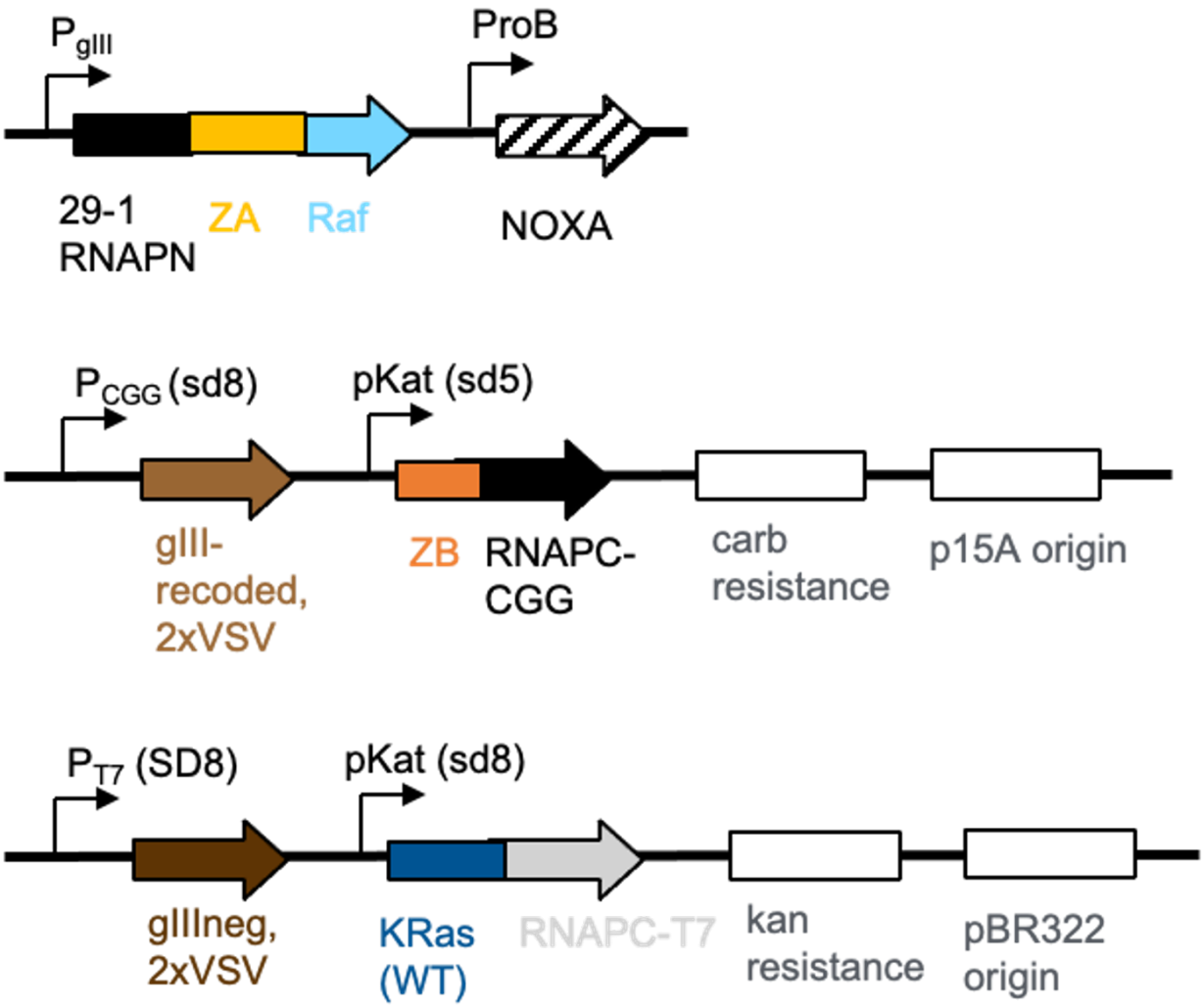


Figure 3.22 Plasmid maps corresponding to supplementary note, design 1, example 1.

Example 2: We added 10^9 of the following phage to S1030 *E. coli* cells with the following APs and performed 4 passages with 1:20000 dilution rate of phage (**Figure 3.23**) . The phage population over all passages was 10^{11} and phage variants at the end all contained the first Raf deleted and some also contained the Q66taa(stop) in the second Raf.

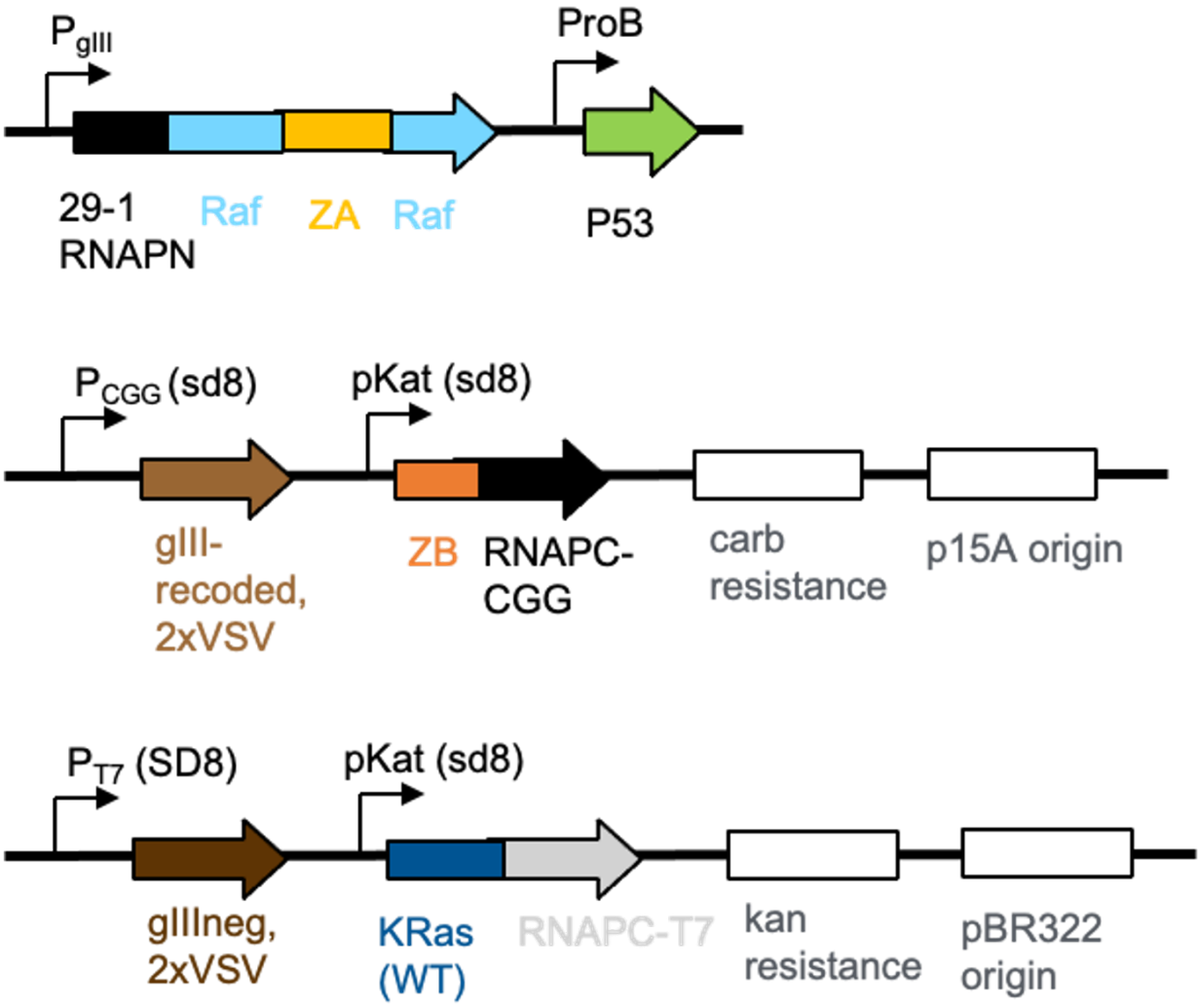


Figure 3.23 Plasmid maps corresponding to supplementary note, design 1, example 2.

Design 2: We then attempted to modify this platform by removing the $RNAP_N$-ZA-PPI partner 1 from the phage and encoding into an AP (**Figure 3.24**). However, this enabled the *E. coli* to produce gIII prior to phage infection, which is known to prevent phage infection and thus prohibit our selection to take place[284] . We experimentally saw this to be the case as well.

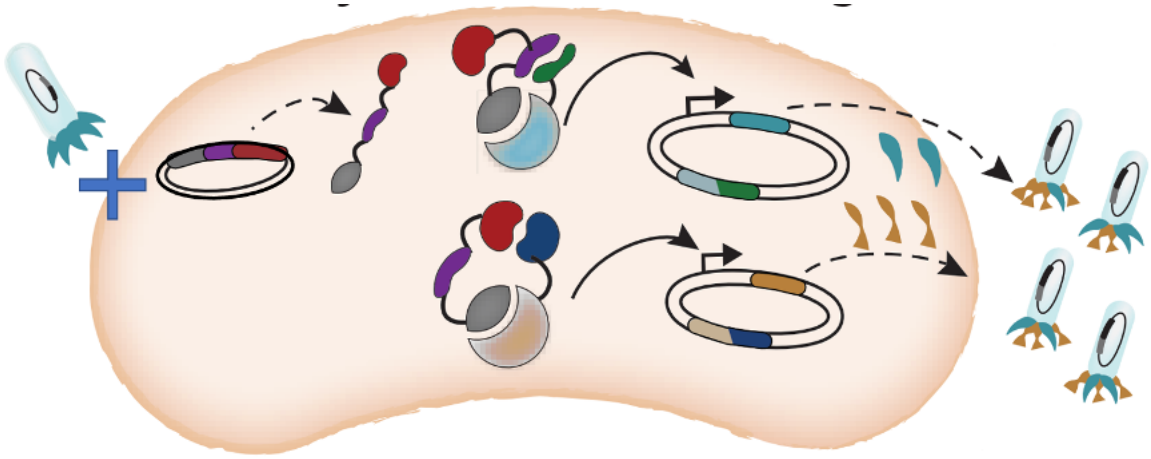
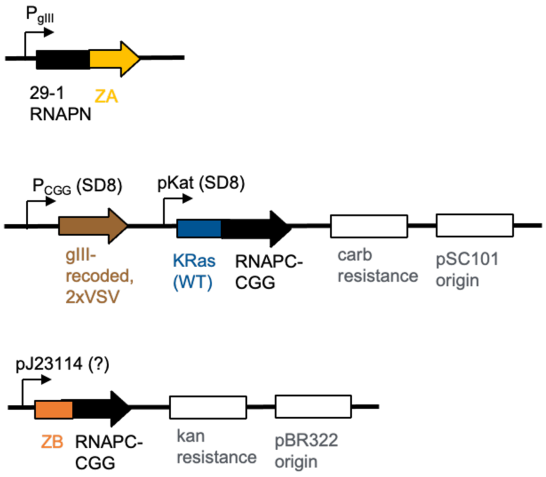
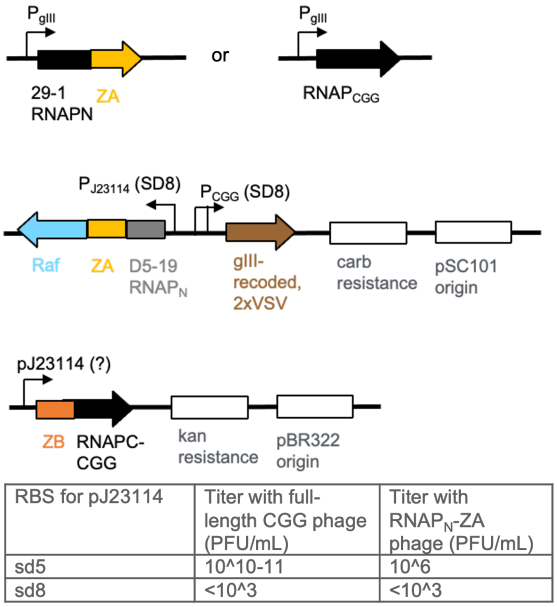


Figure 3.24 Schematic showing supplementary note, design 2.

Example: Overnight phage growth assays starting with 1000 phage resulted in the following (**Figure 3.25**):



RBS for pJ23114	Titer (PFU/mL)
sd5	10^5
sd8	10^11



RBS for pJ23114	Titer with full-length CGG phage (PFU/mL)	Titer with $RNAP_N$-ZA phage (PFU/mL)
sd5	10^10-11	10^6
sd8	<10^3	<10^3

Figure 3.25 Plasmid maps and phage growth assays corresponding to supplementary note, design 2.
Number reported is PFU/mL of one replicate.

CHAPTER 4

SUMMARY AND PERSPECTIVES

4.1 Summary

Just as with people and our relationships with one another, biomolecules and their interactions are complex, and interactions gone awry can lead to adverse consequences. Similarly, understanding and correcting aberrant interactions cannot be contained to a "one size fits all" approach. Specific tools are needed to modulate specific biomolecular interactions, and the number needed to control each interaction in the immense protein interactome alone is staggering. The task for finding even one tool for a protein-protein interaction (PPI) is so challenging, largely due to their biophysical properties, that PPIs were once deemed "undruggable." Though recent technological advances now render PPIs "challenging" rather than completely "undruggable," the term "challenging" magnifies when one accounts for how many interactions exist.

The work undertaken in this dissertation seeks to find tools to alleviate the challenges plaguing PPI modulation. One such challenge is the inability of current directed evolution platforms to directly identify specific binders. (Display platforms can do this indirectly through successive campaigns with different targets, yet this indirect method is liable to experimental bottlenecks along the way.) Therefore, the first project, described in Chapter 2, seeks to overcome this challenge and therefore demonstrates the advent of a new technological platform for finding specific protein binders. By adapting the phage-assisted continuous evolution (PACE) platform for use with a multi-dimensional split RNA polymerase (RNAP) biosensor, we were able to evolve proteins to bind one target but not another, even with biophysically-similar targets. The ability to evolve specificity represents the key advance provided in this platform, a novel activity distinct from evolving binding alone. We demonstrated the utility of the platform by conducting multi-replicate evolution

campaigns with multiple modern-day and ancestral protein starting points to alter their binding profiles. The advent of our technology combined with the power of ancestral sequence reconstruction and high-throughput sequencing to afford the unique ability to decouple the roles of chance and contingency in the evolution of the model protein family.

In the second project, described in Chapter 3, I set out to create another technology that endows activity beyond binding alone: here, in PPI inhibition. As has been mentioned, many platforms exist that select for protein binding (and now one also exists that selects for specific protein binding), and, in some cases, binding to one protein does confer inhibition of the interaction of that protein with its native binding partner. However, binding does not guarantee inhibition; a binder merely offers a chance of achieving inhibition via competitive binding or allostery.

The lack of a broadly-applicable platform for directly identifying inhibitors of PPIs motivated the creation of PANCS-PPI*i*. PANCS-PPI*i* takes inspiration from the PPI specificity PACE technology in Chapter 2, though getting the system to technically-function required some creative design (see **Chapter 3, Supplementary Note** as well as **Figure 3.2**). We demonstrated compatibility with 3 clinically-relevant, biophysically-distinct PPIs and robustness in the ability to not only enrich one functional inhibitor out of 1 billion non-functional inhibitors, but also in the ability to essentially eliminate the non-functional inhibitor population. We showed the ability of PANCS-PPI*i* to not only improve existing inhibitor activity, as seen in the deep mutational scan of a Raf-based inhibitor, but also to identify a novel inhibitor from a library de novo. Both Chapters 2 and 3, therefore, detail novel directed evolution technologies with the power to address critical unmet needs in PPI understanding and PPI modulator discovery.

4.2 Future Directions

In the PPI specificity PACE project described in Chapter 2, we also performed an evolution campaign and did successfully obtain a protein with a novel binding profile, not currently possessed by any protein in nature (nor any other synthetic protein, to our knowledge). Indeed, there is great potential for this technology to develop novel specific protein binders for a variety of targets. This potential has recently been further realized by the Dickinson lab and will hopefully be publicly available later in 2023 or 2024. The technology could also be pushed to its limits in finding isoform-specific binders with the ability to bind or not bind due to a single amino acid difference at a single protein site. Designing specificity around a single site has been notoriously difficult in the case of KRas isoforms, yet this technology would provide a unique approach in its ability to test for specific binding among billions of genetically-encoded molecules at a time.

While we claim that the PPI specificity PACE platform can evolve specific PPIs, we technically mean it can evolve one component (i.e. protein) of a specific PPI. Co-evolution has long been of fascination to evolutionary biologists, yet co-evolution is difficult to achieve in the lab, aside from tracking the evolution of whole viruses, cells, or organisms. Indeed, only a few instances of directed co-evolution are known[102,285] , and they typically are limited to only a few rounds of evolution. The PPI specificity PACE could prove amenable to facilitating the directed co-evolution of both proteins involved in a PPI, and, to the delight of evolutionary biologists who desire longer timeframes, could do so with hundreds of rounds of evolutions in just days. I have done some preliminary experiments to showcase the potential of this project, as shown in **Figure 4.1**.

Plasmid Name	Description
38-49	pBR322, P_{J23117} (sd8) $RNAP_N$-ZA
41-30	p15A, P_{CGG} (SD8), P_{Kat} (sd8) ZP2-$RNAP_{C(CGG)}$
37-67	p15A, P_{CGG} (SD8), P_{Kat} (sd8) KRas-$RNAP_{C(CGG)}$
41-31	cloDF13, P_{Lac} ZB-ZP1
42-34	cloDF13, P_{Lac} ZB-hsMCL-1, P_{Kat} (sd8) BID-ZP1

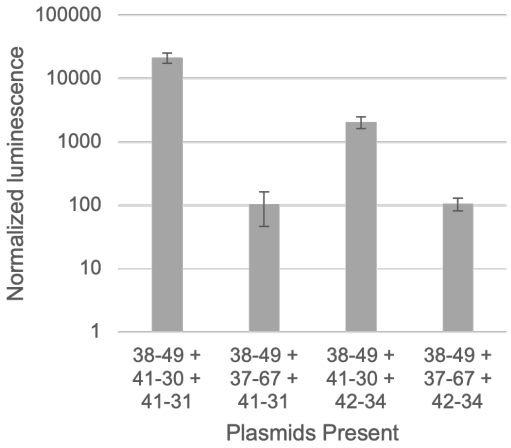


Figure 4.1 *E. coli* luciferase assay to test co-evolution of MCL-1 and BID feasibility. (Left) Table detailing the plasmids used in (right) E. coli luciferase assay. 1 μM of IPTG was used to induce the P_{Lac} plasmid. Bars show mean ś SD of four replicates.

The PANCS-PPI*i* platform established in Chapter 3 is also primed and ready for future novel inhibitor discovery. I am planning to continue making genetically-encoded libraries and running PANCS-PPI*i* experiments on a variety of targets during the rest of my time in the Dickinson lab, and my co-author will do the same even beyond my time. Due to its mechanistically-agnostic approach, we anticipate the ability of the platform to identify allosteric inhibitors, which would lead to further insight into this complex mechanism of inhibition. Because of its ease-of-use yet robustness, we anticipate that PANCS-PPI*i* can be used by any lab versed in molecular biology and cloning to generate inhibitors on-demand for use as biological tools or as starting points for drug development.

4.3 Perspectives

This thesis highlights two important applications of directed evolution platforms; directed evolution can be used as a tool for studying evolution and for the discovery and development of novel molecules. The latter is perhaps more commonly thought-of among

scientists: many pharmaceutical companies now have entire departments devoted to using directed evolution platforms like phage display for drug development purposes. That said, directed evolution can also offer insights into evolutionary processes, principles, and trajectories, as shown in Chapter 2. The expansion of directed evolution platforms should thus not only be of interest to those using it as a tool for molecular development, but also to evolutionary biologists looking for better experimental set-ups. The converse is true as well: scientists involved in directed evolution would do well to stay attuned to emerging developments from the evolutionary biology world to inform how to improve and diversity directed evolution development.

The two main projects of my dissertation have also allowed me to probe two divisions of directed evolution: continuous and non-continuous platforms. With rapid developments and expansions in both areas, the field too has taken note and attempted to assess the strengths and drawbacks of each[29,30,37] . In the context of my work, I view the distinctions as detailed in **Table 4.1**. In summary, I increasingly see the power of non-continuous directed evolution to identify de novo activity. It excels at the proverbial "finding a needle in a haystack" challenge. Perhaps because the continuous aspect of continuous directed evolution provides a harsher selection pressure, continuous evolution excels not at finding de novo activity, but at improving existing activity, perhaps the "sharpening the needle" challenge. To identify robust, novel molecules, it seems the optimal method is to first use non-continuous directed evolution to identify some activity, and then improve activity by continuous directed evolution. There was a series of papers from David Liu's lab where they followed this trajectory for developing base editors[286,287] , and, if things progress as they are, similar papers will likely come from the Dickinson lab as well.

Table 4.1 Continuous versus non-continuous directed evolution methods in this thesis.

	Continuous directed evolution	Non-continuous directed evolution
Rounds of evolution in 4 days	100s	~8
Researcher intervention	Minimal	Manually perform passages
Set-up	Elaborate	Easy
Overall strength	Improves existing activity	Finds 1 winner in a billion non-winners

Both main projects in my dissertation offer a few key advantages and disadvantages to be considered. Both rely on a biosensor, here a split RNAP biosensor, which poses an advantage in that it renders the system amenable to a variety of inputs and outputs and to a large range of tunability. Furthermore, both operate as a selection rather than a screen and function by linking activities to growth and depletion. On one hand, this allows for larger libraries (i.e. more possibilities) to be assessed, yet the linking of activity to growth currently limits the entities being evolved to only those that can be encoded genetically. Lastly, activity is assessed in a biological context, here in the *E. coli* cytoplasm, which offers an advantage in that the activity must take place amidst a myriad of other process that it could either be impacted by or impact itself. The *E. coli* cytoplasm does not accommodate all proteins equally- those that are membrane-bound or reliant on mammalian post-translational modifications for proper functioning are notable examples of incompatibility.

That said, technologies are continually being developed that overcome apparent incompatibilities: disulfide-bond containing proteins were thought to be incompatible with PACE, until periplasm PACE was developed[62] . Unnatural amino acid incorporation technologies[288] , engineered ribosomes[289] , and the like are continually expanding our toolbox and capabilities in *E. coli* as well. Furthermore, yeast and mammalian directed evo-

lution platforms, which are continually developing and improving, could one day prove amenable to the PPI specificity PACE and PANCS-PPI*i* platforms. Overall, I am hopeful that the platforms I developed and reported in this thesis will not only be used to aid in discovery efforts as they are, but will also continue to be reimagined into more robust technologies themselves.

REFERENCES

1. Miryala, S. K.; Anbarasu, A.; Ramaiah, S. Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* **2018**, *642*, 84–94.

2. Porras, P.; Barrera, E.; Bridge, A.; Del-Toro, N.; Cesareni, G. et al. Towards a unified open access dataset of molecular interactions. *Nature communications* **2020**, *11*, 6144.

3. Li, Q.; Cheng, L.; Shen, K.; Jin, H.; Li, H. et al. Efficacy and safety of Bcl-2 inhibitor venetoclax in hematological malignancy: a systematic review and meta-analysis of clinical trials. *Frontiers in pharmacology* **2019**, *10*, 697.

4. Kapoor, I.; Bodo, J.; Hill, B. T.; Hsi, E. D.; Almasan, A. Targeting BCL-2 in B-cell malignancies and overcoming therapeutic resistance. *Cell death & disease* **2020**, *11*, 941.

5. Sterner, R. C.; Sterner, R. M. CAR-T cell therapy: current limitations and potential strategies. *Blood cancer journal* **2021**, *11*, 69.

6. Gao, H.; Sun, X.; Rao, Y. PROTAC technology: opportunities and challenges. *ACS medicinal chemistry letters* **2020**, *11*, 237–240.

7. Pfab, J.; Phan, N. M.; Si, D. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2017525118.

8. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

9. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590–596.

10. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876.

11. Bryant, P.; Pozzati, G.; Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nature communications* **2022**, *13*, 1265.

12. Kaur, H.; Sain, N.; Mohanty, D.; Salunke, D. M. Deciphering evolution of immune recognition in antibodies. *BMC Structural Biology* **2018**, *18*, 1–15.

13. Schroeder Jr, H. W. The evolution and development of the antibody repertoire. *Frontiers in immunology* **2015**, *6*, 33.

14. Lu, R.-M.; Hwang, Y.-C.; Liu, I.-J.; Lee, C.-C.; Tsai, H.-Z. et al. Development of therapeutic antibodies for the treatment of diseases. *Journal of biomedical science* **2020**, *27*, 1–30.

15. Castelli, M. S.; McGonigle, P.; Hornby, P. J. The pharmacology and therapeutic applications of monoclonal antibodies. *Pharmacology research & perspectives* **2019**, *7*, e00535.

16. Chen, K.; Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences* **1993**, *90*, 5618–5622.

17. Yang, Y.; Arnold, F. H. Navigating the unnatural reaction space: directed evolution of heme proteins for selective carbene and nitrene transfer. *Accounts of Chemical Research* **2021**, *54*, 1209–1225.

18. Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H. et al. Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp 3 C–H functionalization. *Nature* **2019**, *565*, 67–72.

19. Spiller, B.; Gershenson, A.; Arnold, F. H.; Stevens, R. C. A structural view of evolutionary divergence. *Proceedings of the National Academy of Sciences* **1999**, *96*, 12305–12310.

20. Romero, P. A.; Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature reviews Molecular cell biology* **2009**, *10*, 866–876.

21. Chen, K.; Arnold, F. H. Engineering new catalytic activities in enzymes. *Nature Catalysis* **2020**, *3*, 203–213.

22. Almhjell, P. J.; Boville, C. E.; Arnold, F. H. *Chemical Society Reviews* **2018**, *47*, 8980–8997.

23. Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P.; Hayes, M. A. et al. Biocatalysis. *Nature Reviews Methods Primers* **2021**, *1*, 46.

24. Winter, G.; Griffiths, A. D.; Hawkins, R. E.; Hoogenboom, H. R. Making antibodies by phage display technology. *Annual review of immunology* **1994**, *12*, 433–455.

25. Smith, G. P.; Petrenko, V. A. Phage display. *Chemical reviews* **1997**, *97*, 391–410.

26. Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology* **1997**, *15*, 553–557.

27. Roberts, R. W.; Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences* **1997**, *94*, 12297–12302.

28. Hanes, J.; Plückthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences* **1997**, *94*, 4937–4942.

29. Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nature Reviews Genetics* **2015**, *16*, 379–394.

30. Wang, Y.; Xue, P.; Cao, M.; Yu, T.; Lane, S. T. et al. Directed evolution: methodologies and applications. *Chemical reviews* **2021**, *121*, 12384–12444.

31. Jang, S.; Kim, M.; Hwang, J.; Jung, G. Y. Tools and systems for evolutionary engineering of biomolecules and microorganisms. *Journal of Industrial Microbiology and Biotechnology* **2019**, *46*, 1313–1326.

32. Hendel, S. J.; Shoulders, M. D. Directed evolution in mammalian cells. *Nature methods* **2021**, *18*, 346–357.

33. Lee, H. H.; Molla, M. N.; Cantor, C. R.; Collins, J. J. Bacterial charity work leads to population-wide resistance. *Nature* **2010**, *467*, 82–85.

34. Toprak, E.; Veres, A.; Michel, J.-B.; Chait, R.; Hartl, D. L. et al. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature genetics* **2012**, *44*, 101–105.

35. Wright, M. C.; Joyce, G. F. Continuous in vitro evolution of catalytic function. *Science* **1997**, *276*, 614–617.

36. Badran, A. H.; Liu, D. R. In vivo continuous directed evolution. *Current opinion in chemical biology* **2015**, *24*, 1–10.

37. Morrison, M. S.; Podracky, C. J.; Liu, D. R. The developing toolkit of continuous directed evolution. *Nature chemical biology* **2020**, *16*, 610–619.

38. Esvelt, K. M.; Carlson, J. C.; Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **2011**, *472*, 499–503.

39. Crook, N.; Abatemarco, J.; Sun, J.; Wagner, J. M.; Schmitz, A. et al. In vivo continuous evolution of genes and pathways in yeast. *Nature communications* **2016**, *7*, 13051.

40. Berman, C. M.; Papa III, L. J.; Hendel, S. J.; Moore, C. L.; Suen, P. H. et al. An adaptable platform for directed evolution in human cells. *Journal of the American Chemical Society* **2018**, *140*, 18093–18103.

41. English, J. G.; Olsen, R. H.; Lansu, K.; Patel, M.; White, K. et al. VEGAS as a platform for facile directed evolution in mammalian cells. *Cell* **2019**, *178*, 748–761.

42. 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* **2013**, *6*, 1–11.

43. Wong, B. G.; Mancuso, C. P.; Kiriakov, S.; Bashor, C. J.; Khalil, A. S. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nature biotechnology* **2018**, *36*, 614–623.

44. Mancuso, C. P.; Lee, H.; Abreu, C. I.; Gore, J.; Khalil, A. S. Environmental fluctuations reshape an unexpected diversity-disturbance relationship in a microbial community. *Elife* **2021**, *10*, e67175.

45. Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K.; Javanpour, A. A.; Liu, C. C. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* **2018**, *175*, 1946–1957.

46. Miller, S. M.; Wang, T.; Liu, D. R. Phage-assisted continuous and non-continuous evolution. *Nature protocols* **2020**, *15*, 4101–4127.

47. Kamalinia, G.; Grindel, B. J.; Takahashi, T. T.; Millward, S. W.; Roberts, R. W. Directing evolution of novel ligands by mRNA display. *Chemical Society Reviews* **2021**, *50*, 9055–9103.

48. Wellner, A.; McMahon, C.; Gilman, M. S.; Clements, J. R.; Clark, S. et al. Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nature chemical biology* **2021**, *17*, 1057–1064.

49. Linciano, S.; Pluda, S.; Bacchin, A.; Angelini, A. Molecular evolution of peptides by yeast surface display technology. *Medchemcomm* **2019**, *10*, 1569–1580.

50. Navaratna, T.; Atangcho, L.; Mahajan, M.; Subramanian, V.; Case, M. et al. Directed evolution using stabilized bacterial peptide display. *Journal of the American Chemical Society* **2019**, *142*, 1882–1894.

51. Ministro, J.; Manuel, A. M.; Goncalves, J. Therapeutic antibody engineering and selection strategies. *Current Applications of Pharmaceutical Biotechnology* **2020**, 55–86.

52. Park, M. Surface display technology for biosensor applications: a review. 2020.

53. Manglik, A.; Kobilka, B. K.; Steyaert, J. Nanobodies to study G protein–coupled receptor structure and function. *Annual review of pharmacology and toxicology* **2017**, *57*, 19–37.

54. McMahon, C.; Baier, A. S.; Pascolutti, R.; Wegrecki, M.; Zheng, S. et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nature structural & molecular biology* **2018**, *25*, 289–296.

55. Li, P.; Wang, L.; Di, L.-j. Applications of protein fragment complementation assays for analyzing biomolecular interactions and biochemical networks in living cells. *Journal of proteome research* **2019**, *18*, 2987–2998.

56. Blaszczak, E.; Lazarewicz, N.; Sudevan, A.; Wysocki, R.; Rabut, G. Protein-fragment complementation assays for large-scale analysis of protein–protein interactions. *Biochemical Society Transactions* **2021**, *49*, 1337–1348.

57. Luck, K.; Kim, D.-K.; Lambourne, L.; Spirohn, K.; Begg, B. E. et al. A reference map of the human binary protein interactome. *Nature* **2020**, *580*, 402–408.

58. Elledge, S. K.; Zhou, X. X.; Byrnes, J. R.; Martinko, A. J.; Lui, I. et al. Engineering luminescent biosensors for point-of-care SARS-CoV-2 antibody detection. *Nature biotechnology* **2021**, *39*, 928–935.

59. Kilchrist, K. V.; Tierney, J. W.; Duvall, C. L. Genetically encoded split-luciferase biosensors to measure endosome disruption rapidly in live cells. *ACS sensors* **2020**, *5*, 1929–1936.

60. Jones, K. A.; Zinkus-Boltz, J.; Dickinson, B. C. Recent advances in developing and applying biosensors for synthetic biology. *Nano Futures* **2019**, *3*, 042002.

61. Badran, A. H.; Guzov, V. M.; Huai, Q.; Kemp, M. M.; Vishwanath, P. et al. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature* **2016**, *533*, 58–63.

62. Morrison, M. S.; Wang, T.; Raguram, A.; Hemez, C.; Liu, D. R. Disulfide-compatible phage-assisted continuous evolution in the periplasmic space. *Nature communications* **2021**, *12*, 5959.

63. Pu, J.; Zinkus-Boltz, J.; Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nature chemical biology* **2017**, *13*, 432–438.

64. Zinkus-Boltz, J.; DeValk, C.; Dickinson, B. C. A phage-assisted continuous selection approach for deep mutational scanning of protein–protein interactions. *ACS Chemical Biology* **2019**, *14*, 2757–2767.

65. Xie, V. C.; Pu, J.; Metzger, B. P.; Thornton, J. W.; Dickinson, B. C. Contingency and chance erase necessity in the experimental evolution of ancestral proteins. *Elife* **2021**, *10*, e67336.

66. Pu, J.; Chronis, I.; Ahn, D.; Dickinson, B. C. A panel of protease-responsive RNA polymerases respond to biochemical signals by production of defined RNA outputs in live cells. *Journal of the American Chemical Society* **2015**, *137*, 15996–15999.

67. Blum, T. R.; Liu, H.; Packer, M. S.; Xiong, X.; Lee, P.-G. et al. Phage-assisted evolution of botulinum neurotoxin proteases with reprogrammed specificity. *Science* **2021**, *371*, 803–810.

68. Quenault, T.; Lithgow, T.; Traven, A. PUF proteins: repression, activation and mRNA localization. *Trends in cell biology* **2011**, *21*, 104–112.

69. Wang, X.; McLachlan, J.; Zamore, P. D.; Hall, T. M. T. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **2002**, *110*, 501–512.

70. Cheong, C.-G.; Hall, T. M. T. Engineering RNA sequence specificity of Pumilio repeats. *Proceedings of the National Academy of Sciences* **2006**, *103*, 13635–13639.

71. Filipovska, A.; Razif, M. F.; Nygård, K. K.; Rackham, O. A universal code for RNA recognition by PUF proteins. *Nature chemical biology* **2011**, *7*, 425–427.

72. Choudhury, R.; Tsai, Y. S.; Dominguez, D.; Wang, Y.; Wang, Z. Engineering RNA endonucleases with customized sequence specificities. *Nature communications* **2012**, *3*, 1147.

73. Lou, T.-F.; Weidmann, C. A.; Killingsworth, J.; Hall, T. M. T.; Goldstrohm, A. C. et al. Integrated analysis of RNA-binding protein complexes using in vitro selection and high-throughput sequencing and sequence specificity landscapes (SEQRS). *Methods* **2017**, *118*, 171–181.

74. Bhat, V. D.; McCann, K. L.; Wang, Y.; Fonseca, D. R.; Shukla, T. et al. Engineering a conserved RNA regulatory protein repurposes its biological function in vivo. 2019.

75. Koirala, D.; Lewicka, A.; Koldobskaya, Y.; Huang, H.; Piccirilli, J. A. Synthetic antibody binding to a preorganized RNA domain of hepatitis C virus internal ribosome entry site inhibits translation. *ACS chemical biology* **2019**, *15*, 205–216.

76. Crawford, D. W.; Blakeley, B. D.; Chen, P.-H.; Sherpa, C.; Le Grice, S. F. et al. An evolved RNA recognition motif that suppresses HIV-1 Tat/TAR-dependent transcription. *ACS chemical biology* **2016**, *11*, 2206–2215.

77. Rauch, S.; He, E.; Srienc, M.; Zhou, H.; Zhang, Z. et al. Programmable RNA-guided RNA effector proteins built from human parts. *Cell* **2019**, *178*, 122–134.

78. Rauch, S.; Jones, K. A.; Dickinson, B. C. Small molecule-inducible RNA-targeting systems for temporal control of RNA regulation. *ACS central science* **2020**, *6*, 1987–1996.

79. Charles, E. J.; Kim, S. E.; Knott, G. J.; Smock, D.; Doudna, J. et al. Engineering improved Cas13 effectors for targeted post-transcriptional regulation of gene expression. 2021.

80. Fukunaga, K.; Yokobayashi, Y. Directed evolution of orthogonal RNA–RBP pairs through library-vs-library in vitro selection. *Nucleic Acids Research* **2022**, *50*, 601–616.

81. Cervettini, D.; Tang, S.; Fried, S. D.; Willis, J. C.; Funke, L. F. et al. Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase–tRNA pairs. *Nature Biotechnology* **2020**, *38*, 989–999.

82. Leconte, A. M.; Dickinson, B. C.; Yang, D. D.; Chen, I. A.; Allen, B. et al. A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* **2013**, *52*, 1490–1499.

83. 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* **2013**, *110*, 9007–9012.

84. Hubbard, B. P.; Badran, A. H.; Zuris, J. A.; Guilinger, J. P.; Davis, K. M. et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nature methods* **2015**, *12*, 939–942.

85. Popa, S. C.; Inamoto, I.; Thuronyi, B. W.; Shin, J. A. Phage-Assisted Continuous Evolution (PACE): A Guide Focused on Evolving Protein–DNA Interactions. *ACS omega* **2020**, *5*, 26957–26966.

86. Inamoto, I.; Sheoran, I.; Popa, S. C.; Hussain, M.; Shin, J. A. Combining rational design and continuous evolution on minimalist proteins that target the E-box DNA site. *ACS Chemical Biology* **2020**, *16*, 35–44.

87. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nature communications* **2018**, *9*, 1911.

88. Pickar-Oliver, A.; Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nature reviews Molecular cell biology* **2019**, *20*, 490–507.

89. Nidhi, S.; Anand, U.; Oleksak, P.; Tripathi, P.; Lal, J. A. et al. Novel CRISPR–Cas systems: an updated review of the current achievements, applications, and future research perspectives. *International journal of molecular sciences* **2021**, *22*, 3327.

90. Tsai, S. Q.; Nguyen, N. T.; Malagon-Lopez, J.; Topkar, V. V.; Aryee, M. J. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nature methods* **2017**, *14*, 607–614.

91. Wang, D.; Zhang, C.; Wang, B.; Li, B.; Wang, Q. et al. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nature communications* **2019**, *10*, 4284.

92. Hiranniramol, K.; Chen, Y.; Wang, X. CRISPR/Cas9 guide RNA design rules for predicting activity. *RNA Interference and CRISPR Technologies: Technical Advances and New Therapeutic Opportunities* **2020**, 351–364.

93. Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **2014**, *156*, 935–949.

94. Kleinstiver, B. P.; Pattanayak, V.; Prew, M. S.; Tsai, S. Q.; Nguyen, N. T. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **2016**, *529*, 490–495.

95. Slaymaker, I. M.; Gao, L.; Zetsche, B.; Scott, D. A.; Yan, W. X. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **2016**, *351*, 84–88.

96. Lee, J. K.; Jeong, E.; Lee, J.; Jung, M.; Shin, E. et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature communications* **2018**, *9*, 3048.

97. Casini, A.; Olivieri, M.; Petris, G.; Montagna, C.; Reginato, G. et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nature biotechnology* **2018**, *36*, 265–271.

98. Hu, J. H.; Miller, S. M.; Geurts, M. H.; Tang, W.; Chen, L. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **2018**, *556*, 57–63.

99. Starr, T. N.; Picton, L. K.; Thornton, J. W. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* **2017**, *549*, 409–413.

100. Schreiber, S. L. The rise of molecular glues. *Cell* **2021**, *184*, 3–9.

101. Dewey, J. A.; Azizi, S.-A.; Lu, V.; Dickinson, B. C. A System for the Evolution of Protein–Protein Interaction Inducers. *ACS synthetic biology* **2021**, *10*, 2096–2110.

102. Siau, J. W.; Nonis, S.; Chee, S.; Koh, L. Q.; Ferrer, F. J. et al. Directed co-evolution of interacting protein–peptide pairs by compartmentalized two-hybrid replication (C2HR). *Nucleic Acids Research* **2020**, *48*, e128–e128.

103. Salinas, V. H.; Ranganathan, R. Coevolution-based inference of amino acid interactions underlying protein function. *elife* **2018**, *7*, e34300.

104. McClune, C. J.; Alvarez-Buylla, A.; Voigt, C. A.; Laub, M. T. Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. *Nature* **2019**, *574*, 702–706.

105. Kawecki, T. J.; Lenski, R. E.; Ebert, D.; Hollis, B.; Olivieri, I. et al. Experimental evolution. *Trends in ecology & evolution* **2012**, *27*, 547–560.

106. Blount, Z. D.; Lenski, R. E.; Losos, J. B. Contingency and determinism in evolution: Replaying lifes tape. *Science* **2018**, *362*, eaam5979.

107. Wörsdörfer, B.; Woycechowsky, K. J.; Hilvert, D. Directed evolution of a protein container. *Science* **2011**, *331*, 589–592.

108. Terasaka, N.; Azuma, Y.; Hilvert, D. Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. *Proceedings of the National Academy of Sciences* **2018**, *115*, 5432–5437.

109. Tetter, S.; Terasaka, N.; Steinauer, A.; Bingham, R. J.; Clark, S. et al. Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* **2021**, *372*, 1220–1224.

110. Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D. et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **2016**, *353*, 389–394.

111. Butterfield, G. L.; Lajoie, M. J.; Gustafson, H. H.; Sellers, D. L.; Nattermann, U. et al. Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **2017**, *552*, 415–420.

112. Hu, D.; Tateno, H.; Hirabayashi, J. Directed evolution of lectins by an improved error-prone PCR and ribosome display method. *Lectins: Methods and Protocols* **2014**, 527–538.

113. Lawrie, J.; Waldrop, S.; Morozov, A.; Niu, W.; Guo, J. Engineering of a small protein scaffold to recognize sulfotyrosine with high specificity. *ACS chemical biology* **2021**, *16*, 1508–1517.

114. Zhou, H.; Rauch, S.; Dai, Q.; Cui, X.; Zhang, Z. et al. Evolution of a reverse transcriptase to map N 1-methyladenosine in human messenger RNA. *Nature methods* **2019**, *16*, 1281–1288.

115. Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **2010**, *464*, 441–444.

116. Hankore, E. D.; Zhang, L.; Chen, Y.; Liu, K.; Niu, W. et al. Genetic incorporation of noncanonical amino acids using two mutually orthogonal quadruplet codons. *ACS synthetic biology* **2019**, *8*, 1168–1174.

117. Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* **2006**, *3*, 357–359.

118. Hammerling, M. J.; Fritz, B. R.; Yoesep, D. J.; Kim, D. S.; Carlson, E. D. et al. In vitro ribosome synthesis and evolution through ribosome display. *Nature communications* **2020**, *11*, 1108.

119. Liu, F.; Bratulić, S.; Costello, A.; Miettinen, T. P.; Badran, A. H. Directed evolution of rRNA improves translation kinetics and recombinant protein yield. *Nature Communications* **2021**, *12*, 5638.

120. Morrison, D.; Rothenbroker, M.; Li, Y. DNAzymes: selected for applications. *Small Methods* **2018**, *2*, 1700319.

121. Tou, C. J.; Schaffer, D. V.; Dueber, J. E. Targeted diversification in the S. cerevisiae genome with CRISPR-guided DNA polymerase I. *ACS Synthetic Biology* **2020**, *9*, 1911–1916.

122. Zhong, Z.; Wong, B. G.; Ravikumar, A.; Arzumanyan, G. A.; Khalil, A. S. et al. Automated continuous evolution of proteins in vivo. *ACS synthetic biology* **2020**, *9*, 1270–1276.

123. Gould, S. J. *Wonderful life: the Burgess Shale and the nature of history*; WW Norton & Company, 1989.

124. Jablonski, D. Approaches to macroevolution: 1. General concepts and origin of variation. *Evolutionary Biology* **2017**, *44*, 427–450.

125. Ramsey, G.; Pence, C. *Chance in Evolution, Chance in Evolution*; The University of Chicago Press, 2016.

126. Travisano, M.; Mongold, J. A.; Bennett, A. F.; Lenski, R. E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **1995**, *267*, 87–90.

127. Mayr, E. How to carry out the adaptationist program? *The American Naturalist* **1983**, *121*, 324–334.

128. Kimura, M. *The neutral theory of molecular evolution*; Cambridge University Press, 1983.

129. Lobkovsky, A. E.; Koonin, E. V. Replaying the tape of life: quantification of the predictability of evolution. *Frontiers in genetics* **2012**, *3*, 246.

130. Monod, J. *Chance and Necessity*; First Vintage Books, 1972.

131. Morris, S. *The Runes of Evolution*; Templeton Press, 2015.

132. Beatty, J. In *The Oxford Handbook of Philosophy of Biology*; Ruse, M., Ed.; Oxford University Press, 2009; pp 1–22.

133. Desjardins, E. Historicity and experimental evolution. *Biology & philosophy* **2011**, *26*, 339–364.

134. Gould, S. J.; Lewontin, R. C. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. 1979; pp 581–598.

135. Bloom, J. D.; Gong, L. I.; Baltimore, D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **2010**, *328*, 1272–1275.

136. Bridgham, J. T.; Ortlund, E. A.; Thornton, J. W. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **2009**, *461*, 515–519.

137. Gong, L. I.; Suchard, M. A.; Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife* **2013**, *2*, e00631.

138. Harms, M. J.; Thornton, J. W. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* **2014**, *512*, 203–207.

139. McKeown, A. N.; Bridgham, J. T.; Anderson, D. W.; Murphy, M. N.; Ortlund, E. A. et al. Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* **2014**, *159*, 58–68.

140. Natarajan, C.; Hoffmann, F. G.; Weber, R. E.; Fago, A.; Witt, C. C. et al. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* **2016**, *354*, 336–339.

141. Ortlund, E. A.; Bridgham, J. T.; Redinbo, M. R.; Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **2007**, *317*, 1544–1548.

142. Risso, V. A.; Manssour-Triedo, F.; Delgado-Delgado, A.; Arco, R.; Barroso-delJesus, A. et al. Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Molecular biology and evolution* **2015**, *32*, 440–455.

143. Starr, T. N.; Flynn, J. M.; Mishra, P.; Bolon, D. N.; Thornton, J. W. Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proceedings of the National Academy of Sciences* **2018**, *115*, 4453–4458.

144. Wu, N. C.; Thompson, A. J.; Xie, J.; Lin, C.-W.; Nycholat, C. M. et al. A complex epistatic network limits the mutational reversibility in the influenza hemagglutinin receptor-binding site. *Nature communications* **2018**, *9*, 1264.

145. Baier, F.; Hong, N.; Yang, G.; Pabis, A.; Miton, C. M. et al. Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. 2019.

146. Blount, Z. D.; Barrick, J. E.; Davidson, C. J.; Lenski, R. E. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* **2012**, *489*, 513–518.

147. Bollback, J. P.; Huelsenbeck, J. P. Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics* **2009**, *181*, 225–234.

148. Counago, R.; Chen, S.; Shamoo, Y. In vivo molecular evolution reveals biophysical origins of organismal fitness. *Molecular cell* **2006**, *22*, 441–449.

149. Kacar, B.; Ge, X.; Sanyal, S.; Gaucher, E. A. Experimental evolution of Escherichia coli harboring an ancient translation protein. *Journal of molecular evolution* **2017**, *84*, 69–84.

150. Meyer, J. R.; Dobias, D. T.; Weitz, J. S.; Barrick, J. E.; Quick, R. T. et al. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **2012**, *335*, 428–432.

151. Salverda, M. L.; Dellus, E.; Gorter, F. A.; Debets, A. J.; Van Der Oost, J. et al. Initial mutations direct alternative pathways of protein evolution. *PLoS genetics* **2011**, *7*, e1001321.

152. Spor, A.; Kvitek, D. J.; Nidelet, T.; Martin, J.; Legrand, J. et al. Phenotypic and genotypic convergences are influenced by historical contingency and environment in yeast. *Evolution* **2014**, *68*, 772–790.

153. Van Ditmarsch, D.; Boyle, K. E.; Sakhtah, H.; Oyler, J. E.; Nadell, C. D. et al. Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. *Cell reports* **2013**, *4*, 697–708.

154. Wichman, H.; Badgett, M.; Scott, L.; Boulianne, C.; Bull, J. Different trajectories of parallel evolution during viral adaptation. *Science* **1999**, *285*, 422–424.

155. Wünsche, A.; Dinh, D. M.; Satterwhite, R. S.; Arenas, C. D.; Stoebel, D. M. et al. Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. *Nature ecology & evolution* **2017**, *1*, 0061.

156. Zheng, J.; Payne, J. L.; Wagner, A. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* **2019**, *365*, 347–353.

157. Arendt, J.; Reznick, D. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in ecology & evolution* **2008**, *23*, 26–32.

158. Gompel, N.; Prud'homme, B. The causes of repeated genetic evolution. *Developmental biology* **2009**, *332*, 36–47.

159. Orgogozo, V. Replaying the tape of life in the twenty-first century. *Interface focus* **2015**, *5*, 20150057.

160. Storz, J. F. Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics* **2016**, *17*, 239–250.

161. Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature Reviews Genetics* **2004**, *5*, 366–375.

162. Chipuk, J. E.; Moldoveanu, T.; Llambi, F.; Parsons, M. J.; Green, D. R. The BCL-2 family reunion. *Molecular cell* **2010**, *37*, 299–310.

163. Danial, N. N.; Korsmeyer, S. J. Cell death: critical control points. *Cell* **2004**, *116*, 205–219.

164. Kale, J.; Osterlund, E. J.; Andrews, D. W. BCL-2 family proteins: changing partners in the dance towards death. *Cell Death & Differentiation* **2018**, *25*, 65–80.

165. Petros, A. M.; Olejniczak, E. T.; Fesik, S. W. Structural biology of the Bcl-2 family of proteins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **2004**, *1644*, 83–94.

166. Chen, L.; Willis, S. N.; Wei, A.; Smith, B. J.; Fletcher, J. I. et al. Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. *Molecular cell* **2005**, *17*, 393–403.

167. Chen, T. S.; Palacios, H.; Keating, A. E. Structure-based redesign of the binding specificity of anti-apoptotic Bcl-xL. *Journal of molecular biology* **2013**, *425*, 171–185.

168. Dutta, S.; Gullá, S.; Chen, T. S.; Fire, E.; Grant, R. A. et al. Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. *Journal of molecular biology* **2010**, *398*, 747–762.

169. Lomonosova, E.; Chinnadurai, G. BH3-only proteins in apoptosis and beyond: an overview. *Oncogene* **2008**, *27*, S2–S19.

170. Certo, M.; Moore, V. D. G.; Nishino, M.; Wei, G.; Korsmeyer, S. et al. Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer cell* **2006**, *9*, 351–365.

171. Banjara, S.; Suraweera, C. D.; Hinds, M. G.; Kvansakul, M. The Bcl-2 family: ancient origins, conserved structures, and divergent mechanisms. *Biomolecules* **2020**, *10*, 128.

172. Lanave, C.; Santamaria, M.; Saccone, C. Comparative genomics: the evolutionary history of the Bcl-2 family. *Gene* **2004**, *333*, 71–79.

173. Pu, J.; Disare, M.; Dickinson, B. C. Evolution of C-terminal modification tolerance in full-length and split T7 RNA Polymerase biosensors. *ChemBioChem* **2019**, *20*, 1547–1553.

174. Kimura, M. DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **1986**, *312*, 343–354.

175. Goodsell, D. S.; Olson, A. J. Structural symmetry and protein function. *Annual review of biophysics and biomolecular structure* **2000**, *29*, 105–153.

176. Nguyen, V.; Wilson, C.; Hoemberger, M.; Stiller, J. B.; Agafonov, R. V. et al. Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **2017**, *355*, 289–294.

177. Somero, G. N. Proteins and temperature. *Annual review of physiology* **1995**, *57*, 43–68.

178. ZAvodszky, P.; Kardos, J.; Svingor, Á.; Petsko, G. A. Adjustment of conformational flexibility is a key event in the thermal adaptation of proteins. *Proceedings of the National Academy of Sciences* **1998**, *95*, 7406–7411.

179. Echave, J.; Spielman, S. J.; Wilke, C. O. Causes of evolutionary rate variation among protein sites. *Nature Reviews Genetics* **2016**, *17*, 109–121.

180. Kimura, M.; Ohta, T. On some principles governing molecular evolution. *Proceedings of the National Academy of Sciences* **1974**, *71*, 2848–2852.

181. Perutz, M.; Kendrew, J.; Watson, H. Structure and function of haemoglobin: II. Some relations between polypeptide chain configuration and amino acid sequence. *Journal of Molecular Biology* **1965**, *13*, 669–678.

182. Blount, Z. D.; Borland, C. Z.; Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proceedings of the National Academy of Sciences* **2008**, *105*, 7899–7906.

183. Breen, M. S.; Kemena, C.; Vlasov, P. K.; Notredame, C.; Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **2012**, *490*, 535–538.

184. Pollock, D. D.; Thiltgen, G.; Goldstein, R. A. Amino acid coevolution induces an evolutionary Stokes shift. *Proceedings of the National Academy of Sciences* **2012**, *109*, E1352–E1359.

185. Quandt, E. M.; Gollihar, J.; Blount, Z. D.; Ellington, A. D.; Georgiou, G. et al. Fine-tuning citrate synthase flux potentiates and refines metabolic innovation in the Lenski evolution experiment. 2015.

186. Sailer, Z. R.; Harms, M. J. High-order epistasis shapes evolutionary trajectories. *PLoS computational biology* **2017**, *13*, e1005541.

187. Shah, P.; McCandlish, D. M.; Plotkin, J. B. Contingency and entrenchment in protein evolution under purifying selection. *Proceedings of the National Academy of Sciences* **2015**, *112*, E3226–E3235.

188. Gompel, N.; Prud'homme, B.; Wittkopp, P. J.; Kassner, V. A.; Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* **2005**, *433*, 481–487.

189. Shubin, N.; Tabin, C.; Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **2009**, *457*, 818–823.

190. Jensen, J. D.; Payseur, B. A.; Stephan, W.; Aquadro, C. F.; Lynch, M. et al. The importance of the neutral theory in 1968 and 50 years on: a response to Kern and Hahn 2018. *Evolution* **2019**, *73*, 111–114.

191. Kern, A. D.; Hahn, M. W. The neutral theory in light of natural selection. *Molecular biology and evolution* **2018**, *35*, 1366–1371.

192. Beatty, J.; Carrera, I. When what had to happen was not bound to happen: history, chance, narrative, evolution. *Journal of the Philosophy of History* **2011**, *5*, 471–495.

193. Chandler, C. H.; Chari, S.; Dworkin, I. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends in genetics* **2013**, *29*, 358–366.

194. Zhu, X.; Guan, Y.; Signore, A. V.; Natarajan, C.; DuBay, S. G. et al. Divergent and parallel routes of biochemical adaptation in high-altitude passerine birds from the Qinghai-Tibet Plateau. *Proceedings of the National Academy of Sciences* **2018**, *115*, 1865–1870.

195. Hawkins, N. J.; Bass, C.; Dixon, A.; Neve, P. The evolutionary origins of pesticide resistance. *Biological Reviews* **2019**, *94*, 135–155.

196. Karageorgi, M.; Groen, S. C.; Sumbul, F.; Pelaez, J. N.; Verster, K. I. et al. Genome editing retraces the evolution of toxin resistance in the monarch butterfly. *Nature* **2019**, *574*, 409–412.

197. Menéndez-Arias, L. Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiviral research* **2010**, *85*, 210–231.

198. Yokoyama, S.; Tada, T.; Zhang, H.; Britt, L. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proceedings of the National Academy of Sciences* **2008**, *105*, 13480–13485.

199. Kryazhimskiy, S.; Rice, D. P.; Jerison, E. R.; Desai, M. M. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **2014**, *344*, 1519–1522.

200. 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* **2014**, *10*, 216–222.

201. Abascal, F.; Zardoya, R.; Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **2005**, *21*, 2104–2105.

202. Hoehler, D.; Haag, J.; Kozlov, A. M.; Stamatakis, A. A representative performance assessment of maximum likelihood based phylogenetic inference tools. 2022.

203. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **1997**, *25*, 3389–3402.

204. Miller, H. C.; Biggs, P. J.; Voelckel, C.; Nelson, N. J. De novo sequence assembly and characterisation of a partial transcriptome for an evolutionarily distinct reptile, the tuatara (Sphenodon punctatus). *BMC genomics* **2012**, *13*, 1–13.

205. Wyffels, J.; King, B. L.; Vincent, J.; Chen, C.; Wu, C. H. et al. SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes. *F1000Research* **2014**, *3*.

206. Zerbino, D. R.; Achuthan, P.; Akanni, W.; Amode, M. R.; Barrell, D. et al. Ensembl 2018. *Nucleic acids research* **2018**, *46*, D754–D761.

207. Hughes, L. C.; Ortí, G.; Huang, Y.; Sun, Y.; Baldwin, C. C. et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences* **2018**, *115*, 6249–6254.

208. Smith, J. J.; Timoshevskaya, N.; Ye, C.; Holt, C.; Keinath, M. C. et al. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nature genetics* **2018**, *50*, 270–277.

209. Takechi, M.; Takeuchi, M.; Ota, K. G.; Nishimura, O.; Mochii, M. et al. Overview of the transcriptome profiles identified in hagfish, shark, and bichir: current issues arising from some nonmodel vertebrate taxa. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **2011**, *316*, 526–546.

210. Delsuc, F.; Philippe, H.; Tsagkogeorga, G.; Simion, P.; Tilak, M.-K. et al. A phylogenomic framework and timescale for comparative studies of tunicates. *Bmc Biology* **2018**, *16*, 1–14.

211. Reich, A.; Dunn, C.; Akasaka, K.; Wessel, G. Phylogenomic analyses of Echinodermata support the sister groups of Asterozoa and Echinozoa. *PloS one* **2015**, *10*, e0119627.

212. Riesgo, A.; Farrar, N.; Windsor, P. J.; Giribet, G.; Leys, S. P. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Molecular biology and evolution* **2014**, *31*, 1102–1120.

213. Moroz, L. L.; Kocot, K. M.; Citarella, M. R.; Dosung, S.; Norekian, T. P. et al. The ctenophore genome and the evolutionary origins of neural systems. *Nature* **2014**, *510*, 109–114.

214. Rech de Laval, V.; Deleage, G.; Aouacheria, A.; Combet, C. BCL2DB: database of BCL-2 family members and BH3-only proteins. *Database* **2014**, *2014*, bau013.

215. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **2018**, *46*, W296–W303.

216. Finnigan, G. C.; Hanson-Smith, V.; Stevens, T. H.; Thornton, J. W. Evolution of increased complexity in a molecular machine. *Nature* **2012**, *481*, 360–364.

217. Popgeorgiev, N.; Sa, J. D.; Jabbour, L.; Banjara, S.; Nguyen, T. T. M. et al. Ancient and conserved functional interplay between Bcl-2 family proteins in the mitochondrial pathway of apoptosis. *Science advances* **2020**, *6*, eabc4149.

218. Schrodinger, L. The PyMOL Molecular Graphics System.

219. Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. Creating new fluorescent probes for cell biology. *Nature reviews Molecular cell biology* **2002**, *3*, 906–918.

220. Venkatesan, K.; Rual, J.-F.; Vazquez, A.; Stelzl, U.; Lemmens, I. et al. An empirical framework for binary interactome mapping. *Nature methods* **2009**, *6*, 83–90.

221. Nossal, N. G. Protein-protein interactions at a DNA replication fork: bacteriophage T4 as a model. *The FASEB journal* **1992**, *6*, 871–878.

222. Dalrymple, B. P.; Kongsuwan, K.; Wijffels, G.; Dixon, N. E.; Jennings, P. A. A universal protein–protein interaction motif in the eubacterial DNA replication and repair systems. *Proceedings of the National Academy of Sciences* **2001**, *98*, 11627–11632.

223. Patrone, J. D.; Kennedy, J. P.; Frank, A. O.; Feldkamp, M. D.; Vangamudi, B. et al. Discovery of protein–protein interaction inhibitors of replication protein A. *ACS medicinal chemistry letters* **2013**, *4*, 601–605.

224. Gallie, D. R. Protein-protein interactions required during translation. *Plant molecular biology* **2002**, *50*, 949–970.

225. Link, A. J.; Niu, X.; Weaver, C. M.; Jennings, J. L.; Duncan, D. T. et al. Targeted identification of protein interactions in eukaryotic mRNA translation. *Proteomics* **2020**, *20*, 1900177.

226. Arkin, M. R.; Whitty, A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein–protein interactions. *Current opinion in chemical biology* **2009**, *13*, 284–290.

227. Schrum, A. G.; Gil, D. Robustness and specificity in signal transduction via physiologic protein interaction networks. *Clinical & experimental pharmacology* **2012**, *2*, S3–001.

228. Westermarck, J.; Ivaska, J.; Corthals, G. L. Identification of protein interactions involved in cellular signaling. *Molecular & Cellular Proteomics* **2013**, *12*, 1752–1763.

229. Zhang, X.; Wang, Y.; Wang, J.; Sun, F. Protein-protein interactions among signaling pathways may become new therapeutic targets in liver cancer. *Oncology reports* **2016**, *35*, 625–638.

230. Stelzl, U.; Worm, U.; Lalowski, M.; Haenig, C.; Brembeck, F. H. et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **2005**, *122*, 957–968.

231. Rual, J.-F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A. et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **2005**, *437*, 1173–1178.

232. Luck, K.; Kim, D.-K.; Lambourne, L.; Spirohn, K.; Begg, B. E. et al. A reference map of the human binary protein interactome. *Nature* **2020**, *580*, 402–408.

233. Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *Journal of cell biology* **2012**, *196*, 801–810.

234. Rhee, H.-W.; Zou, P.; Udeshi, N. D.; Martell, J. D.; Mootha, V. K. et al. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **2013**, *339*, 1328–1331.

235. Qin, W.; Cho, K. F.; Cavanagh, P. E.; Ting, A. Y. Deciphering molecular interactions by proximity labeling. *Nature methods* **2021**, *18*, 133–143.

236. Pfeiffer, C. T.; Paulo, J. A.; Gygi, S. P.; Rockman, H. A. *Methods in Cell Biology*; Elsevier, 2022; Vol. 169; pp 237–266.

237. Kenworthy, A. K. Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. *Methods* **2001**, *24*, 289–296.

238. Seegar, T.; Barton, W. Imaging protein-protein interactions in vivo. *JoVE (Journal of Visualized Experiments)* **2010**, e2149.

239. Sahl, S. J.; Hell, S. W.; Jakobs, S. Fluorescence nanoscopy in cell biology. *Nature reviews Molecular cell biology* **2017**, *18*, 685–701.

240. Strong, M.; Eisenberg, D. The protein network as a tool for finding novel drug targets. *Systems Biological Approaches in Infectious Diseases* **2007**, 191–215.

241. Kuzmanov, U.; Emili, A. Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome medicine* **2013**, *5*, 1–12.

242. Lage, K. Protein–protein interactions and genetic diseases: the interactome. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **2014**, *1842*, 1971–1980.

243. Dervishi, I.; Gozutok, O.; Murnan, K.; Gautam, M.; Heller, D. et al. Protein-protein interactions reveal key canonical pathways, upstream regulators, interactome domains, and novel targets in ALS. *Scientific reports* **2018**, *8*, 1–19.

244. Mo, X.; Niu, Q.; Ivanov, A. A.; Tsang, Y. H.; Tang, C. et al. Systematic discovery of mutation-directed neo-protein-protein interactions in cancer. *Cell* **2022**, *185*, 1974–1985.

245. Cheng, S.-S.; Yang, G.-J.; Wang, W.; Leung, C.-H.; Ma, D.-L. The design and development of covalent protein-protein interaction inhibitors for cancer treatment. *Journal of hematology & oncology* **2020**, *13*, 1–14.

246. Lu, S.; Zhao, Y.; Yu, W.; Yang, Y.; Gao, J. et al. Comparison of nonhuman primates identified the suitable model for COVID-19. *Signal transduction and targeted therapy* **2020**, *5*, 157.

247. Fry, D. C. Targeting protein-protein interactions for drug discovery. *Protein-Protein Interactions: Methods and Applications* **2015**, 93–106.

248. Ran, X.; Gestwicki, J. E. Inhibitors of protein–protein interactions (PPIs): an analysis of scaffold choices and buried surface area. *Current opinion in chemical biology* **2018**, *44*, 75–86.

249. Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. *Nature Reviews Drug Discovery* **2016**, *15*, 533–550.

250. Laraia, L.; McKenzie, G.; Spring, D. R.; Venkitaraman, A. R.; Huggins, D. J. Overcoming chemical, biological, and computational challenges in the development of inhibitors targeting protein-protein interactions. *Chemistry & biology* **2015**, *22*, 689–703.

251. Al-Mugotir, M.; Kolar, C.; Vance, K.; Kelly, D. L.; Natarajan, A. et al. A simple fluorescent assay for the discovery of protein-protein interaction inhibitors. *Analytical biochemistry* **2019**, *569*, 46–52.

252. Wadsworth, P. A.; Folorunso, O.; Nguyen, N.; Singh, A. K.; DAmico, D. et al. High-throughput screening against protein: protein interaction interfaces reveals anticancer therapeutics as potent modulators of the voltage-gated Na+ channel complex. *Scientific reports* **2019**, *9*, 16890.

253. Taylor, I. R.; Dunyak, B. M.; Komiyama, T.; Shao, H.; Ran, X. et al. High-throughput screen for inhibitors of protein–protein interactions in a reconstituted heat shock protein 70 (Hsp70) complex. *Journal of Biological Chemistry* **2018**, *293*, 4014–4025.

254. Wang, X.; Ni, D.; Liu, Y.; Lu, S. Rational design of peptide-based inhibitors disrupting protein-protein interactions. *Frontiers in chemistry* **2021**, *9*, 682675.

255. Choi, J.; Yun, J. S.; Song, H.; Kim, N. H.; Kim, H. S. et al. Exploring the chemical space of protein–protein interaction inhibitors through machine learning. *Scientific reports* **2021**, *11*, 13369.

256. Mourez, M.; Collier, R. J. Use of phage display and polyvalency to design inhibitors of protein-protein interactions. *Protein-Protein Interactions: Methods and Applications* **2004**, 213–227.

257. Chang, H.-N.; Liu, B.-Y.; Qi, Y.-K.; Zhou, Y.; Chen, Y.-P. et al. Blocking of the PD-1/PD-L1 interaction by ad-peptide antagonist for cancer immunotherapy. *Angewandte Chemie International Edition* **2015**, *54*, 11760–11764.

258. Maculins, T.; Garcia-Pardo, J.; Skenderovic, A.; Gebel, J.; Putyrski, M. et al. Discovery of protein-protein interaction inhibitors by integrating protein engineering and chemical screening platforms. *Cell Chemical Biology* **2020**, *27*, 1441–1451.

259. Nim, S.; Jeon, J.; Corbi-Verge, C.; Seo, M.-H.; Ivarsson, Y. et al. Pooled screening for antiproliferative inhibitors of protein-protein interactions. *Nature chemical biology* **2016**, *12*, 275–281.

260. Yang, X. Lennard KR He C. Walker MC Ball AT Doigneaux C. Tavassoli A. van der Donk WA Nat. *Chem. Biol* **2018**, *14*, 375–380.

261. Wong, J. H. A yeast two-hybrid system for the screening and characterization of small-molecule inhibitors of proteinprotein interactions identifies a novel putative Mdm2-binding site in p53. *BMC biology 15*, 117.

262. Leanna, C. A.; Hannink, M. The reverse two-hybrid system: a genetic scheme for selection against specific protein/protein interactions. *Nucleic acids research* **1996**, *24*, 3341–3347.

263. Osher, E. L.; Castillo, F.; Elumalai, N.; Waring, M. J.; Pairaudeau, G. et al. A genetically selected cyclic peptide inhibitor of BCL6 homodimerization. *Bioorganic & Medicinal Chemistry* **2018**, *26*, 3034–3038.

264. Lennard, K. R.; Gardner, R. M.; Doigneaux, C.; Castillo, F.; Tavassoli, A. Development of a cyclic peptide inhibitor of the p6/UEV protein–protein interaction. *ACS Chemical Biology* **2019**, *14*, 1874–1878.

265. Ismail, M.; Martin, S. R.; George, R.; Houghton, F.; Kelly, G. et al. Characterisation of a cyclic peptide that binds to the RAS binding domain of phosphoinositide 3-kinase p110α. *Scientific Reports* **2023**, *13*, 1889.

266. Jewel, D.; Kelemen, R. E.; Huang, R. L.; Zhu, Z.; Sundaresh, B. et al. Virus-assisted directed evolution of enhanced suppressor tRNAs in mammalian cells. *Nature methods* **2023**, *20*, 95–103.

267. García-García, J. D.; Van Gelder, K.; Joshi, J.; Bathe, U.; Leong, B. J. et al. Using continuous directed evolution to improve enzymes for plant applications. *Plant Physiology* **2022**, *188*, 971–983.

268. Rix, G.; Watkins-Dulaney, E. J.; Almhjell, P. J.; Boville, C. E.; Arnold, F. H. et al. Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nature communications* **2020**, *11*, 5644.

269. Dewey, J. A.; Dickinson, B. C. *Methods in enzymology*; Elsevier, 2020; Vol. 641; pp 413–432.

270. Pu, J.; Dewey, J. A.; Hadji, A.; LaBelle, J. L.; Dickinson, B. C. RNA polymerase tags to monitor multidimensional protein–protein interactions reveal pharmacological engagement of Bcl-2 proteins. *Journal of the American Chemical Society* **2017**, *139*, 11964–11972.

271. Pu, J.; Kentala, K.; Dickinson, B. C. Multidimensional control of Cas9 by evolved RNA polymerase-based biosensors. *ACS chemical biology* **2018**, *13*, 431–437.

272. Lamei, H.; Zhixing, G.; Wang, F.; Liwu, F. KRAS mutation: from undruggable to druggable in cancer. *Signal Transduction and Targeted Therapy* **2021**, *6*.

273. Zhu, C.; Guan, X.; Zhang, X.; Luan, X.; Song, Z. et al. Targeting KRAS mutant cancers: from druggable therapy to drug resistance. *Molecular Cancer* **2022**, *21*, 159.

274. Nomura, T. K.; Heishima, K.; Sugito, N.; Sugawara, R.; Ueda, H. et al. Specific inhibition of oncogenic RAS using cell-permeable RAS-binding domains. *Cell Chemical Biology* **2021**, *28*, 1581–1589.

275. Grimm, S.; Lundberg, E.; Yu, F.; Shibasaki, S.; Vernet, E. et al. Selection and characterisation of affibody molecules inhibiting the interaction between Ras and Raf in vitro. *New biotechnology* **2010**, *27*, 766–773.

276. Karlsson, G.; Jensen, A.; Stevenson, L.; Woods, Y.; Lane, D. et al. Activation of p53 by scaffold-stabilised expression of Mdm2-binding peptides: visualisation of reporter gene induction at the single-cell level. *British journal of cancer* **2004**, *91*, 1488–1494.

277. Yang, Y.; Liu, M.; Wang, T.; Wang, Q.; Liu, H. et al. An Optimized Transformation Protocol for Escherichia coli BW3KD with Supreme DNA Assembly Efficiency. *Microbiology Spectrum* **2022**, *10*, e02497–22.

278. Jing, L.; Liu, J.; Cui, D.; Li, Y.; Liu, Z. et al. Screening and production of an affibody inhibiting the interaction of the PD-1/PD-L1 immune checkpoint. *Protein Expression and Purification* **2020**, *166*, 105520.

279. Woldring, D. R.; Holec, P. V.; Stern, L. A.; Du, Y.; Hackel, B. J. A gradient of sitewise diversity promotes evolutionary fitness for binder discovery in a three-helix bundle protein scaffold. *Biochemistry* **2017**, *56*, 1656–1671.

280. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S. et al. ColabFold: making protein folding accessible to all. *Nature methods* **2022**, *19*, 679–682.

281. Chang, L.; Perez, A. Ranking peptide binders by affinity with AlphaFold. *Angewandte Chemie* **2023**, *135*, e202213362.

282. Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; Van Heel, A. J. et al. New developments in RiPP discovery, enzymology and engineering. *Natural product reports* **2021**, *38*, 130–239.

283. ChimeraX, U.; Pettersen, E.; Goddard, T.; Huang, C.; Meng, E. et al. Structure visualization for researchers, educators, and developers. *Protein Sci Jan;30(1):70-82*.

284. Boeke, J. D.; Model, P.; Zinder, N. D. Effects of bacteriophage f1 gene III protein on the host cell membrane. *Molecular and General Genetics MGG* **1982**, *186*, 185–192.

285. Zhou, P.; Li, M.; Shen, B.; Yao, Z.; Bian, Q. et al. Directed coevolution of β-carotene ketolase and hydroxylase and its application in temperature-regulated biosynthesis of astaxanthin. *Journal of agricultural and food chemistry* **2019**, *67*, 1072–1080.

286. Gaudelli, N. M.; Komor, A. C.; Rees, H. A.; Packer, M. S.; Badran, A. H. et al. Programmable base editing of A T to G C in genomic DNA without DNA cleavage. *Nature* **2017**, *551*, 464–471.

287. Richter, M. F.; Zhao, K. T.; Eton, E.; Lapinaite, A.; Newby, G. A. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nature biotechnology* **2020**, *38*, 883–891.

288. Wals, K.; Ovaa, H. Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Frontiers in chemistry* **2014**, *2*, 15.

289. Melo Czekster, C.; Robertson, W. E.; Walker, A. S.; Soll, D.; Schepartz, A. In vivo biosynthesis of a β-amino acid-containing protein. *Journal of the American Chemical Society* **2016**, *138*, 5194–5197.

290. Zhuo, Z.; Yu, Y.; Wang, M.; Li, J.; Zhang, Z. et al. Recent advances in SELEX technology and aptamer applications in biomedicine. *International journal of molecular sciences* **2017**, *18*, 2142.

291. Hivert, V.; Leblois, R.; Petit, E. J.; Gautier, M.; Vitalis, R. Measuring genetic differentiation from Pool-seq data. *Genetics* **2018**, *210*, 315–330.

292. Thornton, J. BCL2.ChanceAndContingency. *Github f9048f1*.

293. Weir, B. S.; Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *evolution* **1984**, 1358–1370.