

Coevolution of interacting proteins through non-contacting and non-specific mutations

David Ding^{1,2}, Anna G. Green^{2,3}, Boyuan Wang⁴, Thuy-Lan Vo Lite⁵, Eli N. Weinstein⁶, Debora S. Marks², Michael T. Laub^{1,7,*}

¹Department of Biology, Massachusetts Institute of Technology; Cambridge, MA 02139, USA.

²Department of Systems Biology, Harvard Medical School; Boston, MA, 02115, USA.

³Department of Biomedical Informatics, Harvard Medical School; Boston, MA 02115, USA.

⁴Department of Pharmacology, UT Southwestern Medical Center; Dallas, TX 75235, USA.

⁵Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School; Boston, MA 02115, USA.

⁶Program in Biophysics, Harvard University; Boston, MA 02115, USA.

⁷Howard Hughes Medical Institute, Massachusetts Institute of Technology; Cambridge, MA 02139, USA.

Summary

Proteins often accumulate neutral mutations that do not affect current functions but can profoundly influence future mutational possibilities and functions. Understanding such hidden potential has major implications for protein design and evolutionary forecasting, but has been limited by a lack of systematic efforts to identify potentiating mutations. Here, through the comprehensive analysis of a bacterial toxin-antitoxin system, we identified all possible single substitutions in the toxin that enable it to tolerate otherwise interface-disrupting mutations in its antitoxin. Strikingly, the majority of enabling mutations in the toxin do not contact, and promote tolerance non-specifically to, many different antitoxin mutations, despite covariation in homologs occurring primarily between specific pairs of contacting residues across the interface. In addition, the enabling mutations we identified expand future mutational paths that both maintain old toxin-

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

*Corresponding author: laub@mit.edu.

Author contributions:

D.D., D.S.M. and M.T.L. conceived the project and wrote the paper. D.D. designed and performed experiments, analyzed data and built the quantitative models. A.G.G. performed covariation analysis for ~350 protein-protein interactions. B.W. helped with library transformations. T-L.V.L. created the combinatorial antitoxin mutant library. E.N.W. suggested helpful tips on Bayesian modelling. D.S.M. and M.T.L. supervised the project.

Data and materials availability:

Raw sequencing read data will be available upon publication in the Sequencing Reads Archive under accession PRJNA768258.

Structural analysis is based in PDB ID: 5CEG.

Analysis code and processed data availability: https://github.com/ddingding/coevolution_mechanism

Competing interests:

DSM is an advisor for Dyno Therapeutics, Octant, Jura Bio, Tectonic Therapeutics, and Genentech, and a co-founder of Seismic. The remaining authors declare no competing interests.

antitoxin interactions and form new ones. These non-specific mutations are missed by widely used covariation and machine learning methods. Identifying such enabling mutations will be critical for ensuring continued binding of therapeutically relevant proteins, such as antibodies, aimed at evolving targets.

Main Text

The ability of biological systems to maintain old functions and attain new ones after acquiring random mutations forms the substrate on which natural selection acts. It is unclear how this process is affected by neutral mutations that do not, on their own, change the function of the system, but can shape which future mutations are possible. Do these neutral mutations enable only few subsequent mutations to be tolerated (Fig. 1a, left) or do they broadly expand the mutations that can subsequently arise (Fig. 1a, right)? In the absence of systematic studies, the identities and hidden potential of such neutral mutations in enabling subsequent mutational trajectories remain unclear.

For interacting proteins, several types of mutations can, in principle, enable tolerance to mutations that would have otherwise been disruptive (Fig. 1b–c). One is the mutation of a residue that contacts the disruptive mutation to directly restore the protein-protein interaction (Fig. 1b). The notion that interacting proteins evolve through such restricted and specific, complementary changes in contacting residues is suggested by analyses of amino-acid covariation in natural sequences, which can be used to predict protein structures and protein-protein interactions^{1–9} by identifying specific pairs of residues in close proximity^{10–12}. Additionally, potent binding proteins have been engineered by mutating key interface residues^{13–16}. However, some interface-disrupting mutations may be tolerated by mutations elsewhere in the partner protein, either along the interface or not, that indirectly restore the interaction (Fig. 1c). There are anecdotal examples, at least within proteins, of mutations that can only be tolerated in the presence of prior, enabling mutations at non-contacting positions^{17–36}, and of mutations away from the interface regions of antibodies that affect antigen affinity and specificity^{37–39}.

For virtually all protein-protein interactions, it is not yet known how many enabling mutations exist among the set of possible neutral mutations, nor is it known how close and how specific such mutations typically are to the disruptive mutations they enable^{40–42}. Additionally, whether neutral, enabling mutations affect the evolvability of a protein - its ability to subsequently acquire new functions or binding partners - has not been systematically examined. A better understanding of how neutral mutations shape future mutational trajectories and protein binding partners promises to aid efforts to engineer protein interactions of clinical or therapeutic value, and enable better forecasting of fast evolving protein sequences.

To comprehensively study enabling mutations in protein coevolution, we examined a bacterial toxin-antitoxin system. The *Mesorhizobium opportunistum* antitoxin ParD3 normally binds and restrains the activity of its cognate, co-operonic toxin ParE3¹⁰. When co-expressed in *Escherichia coli*, ParD3-ParE3 form an inert multimeric complex and cells can grow. Any mutation that disrupts the interface will liberate toxin, which slows

cell growth. Thus, cell proliferation provides a powerful, easy-to-measure readout of the ParD3-ParE3 interaction *in vivo* (Fig. 1d). Prior analyses of amino-acid covariation in ParD-ParE homologs identified residues in each protein that most strongly covary, map to the protein-protein interface, and can dictate interaction specificity of paralogs^{10,11} (Fig. 1e). To experimentally probe the possible mutational trajectories for ParD3-ParE3, we first used deep mutational scanning⁴³⁻⁴⁶ to identify all possible single point mutations in the toxin that are neutral and thus retain binding to the antitoxin and retain toxicity if produced alone. Then, we identified all possible mutations in the toxin that also enable it to tolerate interface-disrupting mutations in the antitoxin.

We find that the majority of such enabling mutations in the toxin can restore binding to many of the otherwise disruptive single mutations in the antitoxin. Hence, we refer to these as *non-specific suppressors*. These enabling mutations often arise far from the mutations they rescue, and some are even outside the protein-protein interface. If they arise before a disrupting mutation arises, these non-specific suppressors would stabilize the toxin-antitoxin interaction and thereby increase the robustness of this interaction to subsequent mutations. Notably, very few mutant pairs – a disruptive mutation and a suppressor mutation – correspond to the pairs identified in covariation analyses, suggesting that protein coevolution may often involve residues that are not directly contacting. We also demonstrate that non-specific suppressor mutations in the toxin enable it to bind substantially more antitoxins that have multiple mutations. Importantly, this includes an increased number of antitoxin variants that have the ability to also bind a different toxin. Thus, the enabling mutations in the toxin potentiate the evolvability of this protein-protein interaction by promoting binding to antitoxins with additional functions. Collectively, our findings highlight the potential of neutral, non-specific mutations in expanding the space of future mutational possibilities in protein coevolution and likely protein design.

Results

Mutational tolerance of the ParD3 antitoxin

To examine the mutational robustness of the ParD3-ParE3 (antitoxin-toxin) complex, we first built a library of 5,796 variants containing each possible codon at each position in the 93 amino-acid ParD3 antitoxin. This library encodes all 1,784 possible single amino-acid variants and 242 different synonymous versions of wild-type ParD3. The library was transformed into and expressed in cells harboring the wild-type toxin ParE3. If a cell has an antitoxin variant that can bind and neutralize the toxin, it will grow and proliferate in the population over time; conversely, if a cell has an antitoxin variant that disrupts the interface or misfolds the antitoxin leading to liberated toxin, it will not grow and eventually be lost from the population. The growth rates of individual variants were assessed in pooled cells, using deep-sequencing to measure the change in variant frequency over 10 hours, which is ~6 generations for cells harboring wild-type ParD3 and ParE3. To infer the effect of each substitution from sequencing read counts, including their uncertainties, we developed a hierarchical Bayesian error model that considers sampling noise of reads and information from synonymous variants and replicate experiments (see Methods). The inferred growth rates were highly correlated with independently measured growth rates

(Pearson $r=0.94$; Extended Data Fig. 1a), and the change in variant frequency over time was highly reproducible between biological replicates (Pearson $r=0.92$, Extended Data Fig. 1b).

For each variant in the antitoxin library we calculated Δ growth-rate - the difference in number of doublings of the mutant, AT*, compared to wild-type antitoxin, AT (Fig. 2a). For synonymous variants of the wild-type antitoxin, the Δ growth-rate values were tightly and symmetrically distributed around 0, as expected. For variants with a stop codon that would produce a non-functional antitoxin, the Δ growth-rate values had a mean of -5.2 , also as expected. For all other variants, the Δ growth-rate values had a mean of -0.5 , with $> 98\%$ of all variants > -2.5 . Only 32 amino-acid variants produced Δ growth-rate values < -2.5 indicating that few single substitutions severely disrupt the toxin-antitoxin interface.

To examine the position-wise effect of amino-acid substitutions, we generated a 21×92 heatmap showing the Δ growth-rate value for each possible substitution at each position in ParD3 (Fig. 2b). This map revealed an apparent periodicity within α -helices with residues on the face of an α -helix that contacts the toxin exhibiting negative mean Δ growth-rate values (Extended Data Fig. 1c), whereas solvent-exposed residues on the opposite face exhibited more mutational tolerance. Helix-breaking proline substitutions did not follow this periodicity and were disruptive at most positions in α -helices. Aside from these prolines, the substitutions disruptive to growth rate were generally clustered in two regions. One was the N-terminal half of the antitoxin, which mediates oligomerization and likely promotes overall stability of the ParD3-ParE3 complex, but does not directly contact the toxin (Extended Data Fig. 1d). The second region involves the C-terminal end of α_2 and α_3 , the elements of ParD3 that bind the toxin (Fig. 1c).

The most disruptive substitutions in ParD3 arose at just two positions, tryptophan-59 and glycine-62 (Fig. 2b). These positions strongly covary with nearby residues ($< 6 \text{ \AA}$ minimum atom distance) in the toxin as measured by EVcouplings⁴ (Extended Data Fig. 1e-f) and are critical for antitoxin binding specificity¹⁰. For position 59, substituting tryptophan with another aromatic residue or some hydrophobic residues had relatively modest effects, with all others leading to substantial defects (Δ growth-rate < -2.6). For position 62, substituting glycine with anything other than alanine, valine, or serine severely disrupted growth. Notably, W59 and G62 are at the very C-terminus of α_2 in the antitoxin and both are in close proximity to the toxin. The side chain of W59 inserts into a snug hydrophobic pocket on the toxin and G62 is tightly packed against the toxin (Fig. 2c). The substitutions at positions W59 and G62 that severely diminish growth rate may either disrupt the toxin-antitoxin interface or trivially unfold the antitoxin. Below we provide evidence for the former.

Mutational robustness of the ParE3 toxin

Next, we identified all possible single substitutions in the toxin that maintain both toxin-antitoxin binding and toxicity of the toxin. We performed a deep mutational scan of the 103 amino-acid toxin ParE3 (Fig. 2d), using a library of all possible 6,426 single-codon variants (2,040 amino-acid variants). The distribution of Δ growth-rate values in the presence of wild-type antitoxin (*i.e.* the growth rate of each toxin variant, T*, relative to the wild-type toxin, T) was narrowly centered around 0. Thus, every possible single substitution either

retains binding to the antitoxin or causes the toxin to lose toxicity, with the latter including mutations that lead to either unfolding or an inability to bind the toxin's cellular target.

To identify mutations that allowed cells to grow by disrupting the toxicity of the toxin rather than retaining antitoxin binding, we repeated the experiment but in cells lacking the antitoxin (Fig. 2e, Extended Data Fig. 2). In this case, unfolded or nonfunctional toxin variants permit growth, whereas properly folded and functional toxins do not. For variants containing a synonymous mutation (n=278), the growth-rate values were symmetrically centered around 0, while variants harboring a stop codon, which are presumably non-functional, exhibited high growth-rate values (> 5) indicating significantly faster growth than cells expressing wild-type toxin. For all other variants, the distribution was bimodal, with one mode at 0 and the other at ~ 6 . The set nearest to 0 are those that minimally disrupt toxin folding and activity. We performed this experiment using 7 different induction conditions for the toxin, and then identified the 310 'most toxic' variants at 65 residue positions whose ability to inhibit growth rate was consistently comparable to the wild-type toxin (see Methods, Extended Data Fig. 2e). We also used a less stringent cutoff to define a set of 781 'toxic' mutants at 91 residue positions that are comparable to wild-type toxin under 4 fully inhibiting induction conditions (Extended Data Fig. 2e).

Substitutions that ablated ParE3 toxicity (and hence have high growth-rate values) were particularly pronounced in $\beta 2$, $\beta 3$, and $\alpha 3$, which have low solvent accessibility, suggesting that they simply unfold the toxin (Extended Data Fig. 2g–h). Notably, for toxin positions that strongly covary with positions in the antitoxin and that make direct contact in the co-crystal structure, the majority of mutations preserved toxicity of the toxin similar to the wild-type toxin (Extended Data Fig. 2i). Substitutions at these positions also retained their neutralization by the antitoxin (Extended Data Fig. 2j). Taken together, our results indicate that the interface residues of the ParE3 toxin are highly robust to mutations. This robustness does not result from overexpression of the antitoxin, as our experiments were performed with minimally neutralizing levels of antitoxin (Extended Data Fig. 2a–d).

A suppressor scan reveals distant interaction-enabling mutations

We next sought to systematically assess which mutations in the toxin enable binding to otherwise interaction-disrupting mutations in the antitoxin. To this end, we performed a 'deep suppressor scan', by screening all toxin single mutations for their ability to rescue a disruptive mutation in the antitoxin (Fig. 1d). We chose 36 single antitoxin substitutions, spanning the measured range of effects, mostly at positions that (i) strongly covary with positions in the toxin and (ii) that involve amino acids commonly found at those positions in ParD antitoxin homologs (see Methods). We examined each antitoxin mutant + toxin library ($36 \times 2,040 = 73,440$ amino-acid variant pairs) at two different induction levels of antitoxin: the minimal concentration needed to neutralize wild-type toxin, as in the antitoxin single mutant screen above, and a slightly lower concentration of antitoxin such that the toxin is almost but not fully neutralized, which enables more sensitive detection of rescuing mutations in the toxin. Of the 36 antitoxin mutants examined, 9 were deleterious with disrupted toxin binding at the higher antitoxin concentration, and 12 at the lower concentration. We focused on the 9 most deleterious antitoxin mutants, but found that

including data from all 36 in our model (see below) improved our assessment of the effect of each toxin mutation.

As an example, we first considered the antitoxin mutant ParD(W59T), which strongly disrupts the ParD3-ParE3 interface (Fig. 3a, Fig. 2b). Toxin variants with positive growth-rate values in this antitoxin background either increase binding to the W59T antitoxin mutant, or simply disrupt the toxicity of the toxin. Considering only the 310 ‘most toxic’, neutral mutations defined above that maintain the toxicity of the toxin, we identified 11 mutations in the toxin that significantly and substantially improved the growth rate of cells harboring the ParD3(W59T) mutant ($p < 0.0001$ and growth-rate values > 0.5 , see Methods; Fig. 3b, Extended Data Fig. 3). These 11 mutations represent just 3.5% (11/310) of the ‘most toxic’ toxins and 0.54% (11/2,040) of all possible single mutants. We mapped these 11 rescuing mutations onto the ParD3-ParE3 co-crystal structure (Fig. 3c), finding that they were distributed throughout the toxin and many were $> 10 \text{ \AA}$ away from the W59 residue in the antitoxin (Fig. 3d).

We repeated the same analysis for the other 8 deleterious antitoxin mutations at the higher antitoxin concentration (Fig. 3e, Extended Data Fig. 3b). In the pooled data, we detected 51 pairs of mutations in which the toxin variant significantly and substantially alleviates the growth defect of a deleterious antitoxin mutation (Fig. 3f). These 51 pairs involved 32 different toxin variants, which represents 10.3% (32/310) of the ‘most toxic’ toxin variants and 1.6% (32/2,040) of all possible single mutant toxins. For these pooled data, there was no strong correlation between the magnitude of rescue by toxin mutations and their distances to the position mutated in the antitoxin (Fig. 3f). Similar results were seen for the 12 deleterious mutations at the lower antitoxin concentration (Extended Data Fig. 3c–d). We conclude that there are many possible mutations that can relieve the deleterious effect of a given antitoxin mutation and that such mutations can arise throughout the toxin, not simply through local, directly compensating mutations.

Enabling mutations tolerate multiple deleterious mutations

Next, we sought to assess whether enabling toxin mutations are non-specific or specific, *i.e.* whether a toxin mutation increases binding irrespective of the deleterious antitoxin mutation present, or allows binding only to particular deleterious antitoxin mutations. To do this quantitatively, we built a model, similar to previous models^{42,47–55}, that tries to explain all observed single and double mutant growth rates as a sum of non-specific, independent (but unobserved) single substitution effects passed through a global nonlinearity (see Methods, Extended Data Fig. 4). This model explained 89% of the observed growth rate deviations (Extended Data Fig. 4b), with inferred mutation effects highly robust to expression conditions (Extended Data Fig. 4h), and allowed us to calculate the expected growth rate of each double mutant if the toxin and antitoxin substitutions act independently and non-specifically. We defined a toxin-antitoxin double mutant pair as specific if it showed a significant and substantial positive deviation from its double mutant expectation (>2 -fold change in growth rate, and $p < 0.0001$; Fig. 4a,c; Extended Data Fig. 5), and as non-specific if the observed growth rate was close to the expected rate (Fig. 4b,c; Extended Data Fig. 5). We then called a particular toxin mutation as specific if it showed a positive deviation in

any antitoxin background, and conversely called a toxin mutation non-specific if it produced growth rates close to the expectation in all antitoxin backgrounds.

The 51 mutation pairs in which the growth defect of a deleterious antitoxin mutation is alleviated by a toxin mutation, include 32 different toxin variants, and we found that 21 of these toxin variants act non-specifically, outnumbering the 11 specific toxin variants (Fig. 4c, Extended Data Fig. 5c–d). These 51 mutant pairs involved any toxin mutation that significantly improved the growth rate of an antitoxin mutant, regardless of how close the double mutant growth rate was to the wild-type toxin-antitoxin pair. We also examined only those pairs in which the double mutant had a growth rate close to the wild-type toxin-antitoxin pair. At each threshold considered, the number of non-specific, binding-restoring toxin mutations was the same or greater than the number of specific binding-restoring mutations (Fig. 4d). These conclusions were even more pronounced under the ‘low antitoxin’ expression conditions (Extended Data Fig. 5e–j).

To probe the spatial distribution of specific and non-specific toxin mutants, we used the ParD3-ParE3 co-crystal structure to assess whether each lies at the protein interface and to measure the distance of each from the antitoxin mutation they rescue (Fig. 4e–f). Specific rescuing toxin mutations were all within 10 Å of the antitoxin mutation they rescued with nearly half within ~5 Å. While similar spatial closeness between positively epistatic residues have been observed previously⁴², our specific rescuing pairs did not necessarily arise from biochemically obvious compensation, such as charge swapping or preservation of size complementarity. In contrast to the specific rescuing mutations, the non-specific rescuing toxin substitutions were mostly far (> 6 Å when considering minimum atom distance) from the antitoxin substitutions they rescued (includes 31 of the 34 pairs involving a non-specific rescuing mutation, which involves 20 of the 21 non-specific toxin mutations) (Fig. 4e–f, Extended Data Fig. 5h), with 8 of 21 non-specific toxin mutations at positions more than 6 Å from any antitoxin residue indicating that they are not toxin-antitoxin interface residues (Extended Data Fig. 5h, Supplementary Table 1, Supplementary Data 1). Taken together, our analyses indicate that non-specifically enabling mutations are more frequent than specifically enabling mutations and that many of these non-specific toxin mutations arise at sites spatially distant from the mutation they rescue.

Natural sequence variation cannot predict enabling mutations

We next asked whether the outcome of our ‘deep suppressor scan’ could have been predicted from the features of naturally occurring homologs of ParD3-ParE3, which show strong covariation scores, even compared to other complexes¹ (Fig. 1e; Fig. 5a–b). The top covarying positions involve nearby pairs of interface residues (28 of top 29 covarying pairs of residues are within 6 Å minimum atom distance, Extended Data Fig. 1f). However, the distribution of covariation scores for position pairs identified as enabling in our suppressor scan was indistinguishable from that of randomly selected pairs (Fig. 5c). There was also no correlation between the effect of each toxin mutation inferred in our nonlinear model and the frequency at which the mutant amino acid is found at that position in natural sequences (Fig. 5d). Similarly, pairwise frequencies or enrichments thereof were not predictive of binding-restoring pairs of residues, with more than half (29/51) of the binding-restoring

toxin and antitoxin variant pairs not observed in natural sequences (Extended Data Fig. 6a–b).

We also found that models trained on the sequence alignment (EVmutation⁵⁶, variational autoencoder⁵⁷) could not predict which toxin mutations would alleviate the growth defect of a particular antitoxin mutation, with no correlation between the measured growth-rate values and the scores produced by these models (Fig. 5e, Extended Data Fig. 6). Notably, the homologs in our alignment differ, relative to our ParD3-ParE3 complex, at 80% of positions, on average (Extended Data Fig. 6h). These natural sequences may effectively be too sparse of a dataset to predict the immediately available compensating mutations, either specific or non-specific, available in a particular sequence background.

Non-specific suppressors expand mutational trajectories

The non-specific enabling mutations identified in the ParE3 toxin promote tolerance to many different single mutations in the antitoxin. To ask whether they also expand subsequent mutational trajectories containing multiple antitoxin mutations, we performed a ‘combinatorial mutation scan’ using a previously developed library containing all 8,000 possible combinations of residue variants at 3 interface positions in the ParD3 antitoxin critical to binding specificity¹¹ (D60, K63, E79) (Fig. 6a). We picked 8 of our non-specific suppressors in the toxin and two double mutants that combine two of these suppressors (Fig. 6b). We verified that each of these 10 toxin variants are as, or almost as, toxic as wild-type toxin (Extended Data Fig. 7a). We then assayed each, and the wild-type toxin, for neutralization by the combinatorial library of 8,000 ParD3 antitoxin variants (Fig. 6c, Extended Data Fig. 7b). In each case, we calculated a normalized fitness between 0 (comparable to a truncated antitoxin) and 1 (comparable to wild-type antitoxin).

Notably, each non-specific enabling mutation in the toxin led to a substantial increase in the number of antitoxin variants that could neutralize it (Fig. 6d). For the wild-type toxin ParE3, less than 20% of antitoxin variants achieve half-maximal neutralization (fitness > 0.5). In contrast, each non-specific rescuing mutation led to many more neutralizing antitoxin variants, often with > 60% of the library exhibiting fitness values > 0.5. The independent effect of each toxin mutant inferred from our initial suppressor scan was almost perfectly correlated with the number of antitoxin variants that could bind (Fig. 6e). We conclude that the global, non-specific mutations identified in our suppressor scan expand the subsequent mutational robustness of the toxin-antitoxin complex.

Whereas the wild-type antitoxin ParD3 can only neutralize its cognate toxin ParE3, we identified 25 promiscuous variants in the antitoxin combinatorial library that could also interact with the non-cognate toxin, ParE2^{10,11}, which shows ~40% sequence identity with toxin ParE3 (fitness > 0.9, Fig. 6f). We asked whether the non-specific enabling mutations in ParE3 increased the number of promiscuous antitoxin variants. Indeed, the number of antitoxin variants that could neutralize both ParE3(V5L) as well as ParE2 was 101, a 4-fold increase relative to the number that neutralize wild-type ParE3 and ParE2 (Fig. 6f–g). All of these promiscuous antitoxin variants are accessible from the wild-type ParD3 antitoxin via trajectories comprising single mutational steps (Fig. 6h). Even larger increases in the number of promiscuous antitoxin variants were obtained when considering other toxin ParE3 global

suppressor variants (Fig. 6i). We conclude that the non-specific suppressors identified in ParE3 both improve its ability to maintain an interaction with the wild-type partner ParD3 and promote the evolution of new toxin-antitoxin interactions.

Discussion

Collectively, our results indicate that interacting proteins may coevolve by first acquiring neutral, non-specific mutations in one protein that dramatically expand the number of mutations tolerated in the partner that would have otherwise disrupted the interaction. These enabling mutations are often found at positions far from the disruptive substitutions they promote tolerance to, and many lie outside of the protein-protein interface. These mutations not only promote maintenance of the specific cognate interaction, but expand the number of subsequent mutational trajectories that include partners with additional functions. Collectively, these findings contrast with models of coevolution driven mainly by directly contacting, specifically compensating pairs of mutations as it involves distant residues and because it does not invoke a broken, or disrupted, intermediate state. The non-specific mutations we identified, in fact, allow the inverse: a neutral, non-specifically enabling mutation could occur first, and then permit the mutation in its partner that would otherwise have disrupted the interface. Although anecdotal examples of non-specific suppressors have previously been reported within proteins^{17-34,58,59} and some between proteins³⁷⁻³⁹, it had been unclear in the absence of systematic studies how likely non-specific vs. specific suppressors are to arise and whether each class of suppressors would map to directly contacting pairs of residues or not⁴¹.

Our systematic ‘suppressor scan’ identified, for ParD3-ParE3, all possible compensatory mutations after introducing a handful of deleterious mutations. The 21 non-specific suppressors we identified arose at 15 different positions in the toxin, and are mostly far from the deleterious antitoxin substitution they rescue, sometimes without contacting the antitoxin at all. The mechanistic basis of these 21 non-specific suppressors identified in ParE3 is not yet clear. Some of these non-specific mutations may promote binding by forming new points of favorable interaction (Supplementary Table 1, Supplementary Data 1), thereby enhancing complex formation and tolerance to subsequent mutations. They may also promote or stabilize existing points of interaction. In either case, the ‘excess’ binding energy may then permit a wide range of subsequent mutations that would have otherwise destabilized the complex. Such a model has also been proposed for non-specific suppressors of destabilizing mutations within proteins^{20,34}. The non-specific toxin mutations likely increase binding to both wild-type and mutant antitoxins. Alternatively, the non-specific mutations could increase binding only to mutant antitoxins while retaining similar binding to the wild-type antitoxin, but such a scenario is less parsimonious as it invokes multiple specific mutation dependencies and is less biochemically plausible, especially for the non-specific mutations that do not lie on the ParD3-ParE3 interface.

Here, we have illuminated the mutational possibilities given the molecular binding and folding constraints of this system, and show that our main conclusions are robust to the two expression levels examined. Which mutational trajectories are taken by this or other protein-protein interactions in natural settings is as yet unclear, and may depend on the particulars

of the host bacteria and expression levels of each component, as well as population genetic factors such as the strength of selective pressures and population sizes.

Our strategy for identifying non-specific suppressors could be valuable in the design of therapeutic proteins aimed at an evolving target. For instance, systematically finding and engineering distal, non-specific suppressors of different magnitude into antibodies could help render them broadly neutralizing and tolerant to subsequent mutations in the target antigen. Examples of such mutations have been reported for certain broadly neutralizing antibodies^{37,39}. Notably, the non-specific suppressors we identified were relatively rare and not predictable based on naturally occurring homologs of ParE and ParD, which exhibit strong covariation. Models based on natural sequence homologs and covariation are powerfully able to predict protein and protein complex structures^{1-4,7-9}, and can help in protein design^{9,60,61}. However, our results indicate that such unsupervised models are currently insufficient to accurately predict the immediate mutational trajectories of our proteins. One possibility is that evolution has not fully explored the sequence variation that can be achieved in systematic mutational studies like ours. Alternatively, or in addition, the covariation signal measured in sequence alignments may represent an average signal across all homologs found in nature, but that the exact mutational steps possible for each extant sequence are highly idiosyncratic. If so, systematic experimental methods, such as the suppressor scan performed here, will be critical to future protein engineering efforts.

Materials and Methods

Bacterial strains, vectors and media

All strains used are listed in Supplementary Table 2. *E. coli* strains were grown at 37 °C in M9L medium (1x M9 salts, 100 µM CaCl₂, 0.4% glycerol, 0.1% casamino acids, 2 mM MgSO₄, 10% v/v LB). Antibiotics were used as follows: 50 µg/ml carbenicillin, 20 µg/ml chloramphenicol in liquid media, and 100 µg/ml carbenicillin, 30 µg/ml in agar plates. The toxins (ParE3 or ParE2) were carried as before⁸ on the pBAD33 vector (chlor^R marker, ML3482 for wild-type ParE3, ML3303 for wild-type ParE2) with expression repressed or induced with 1% glucose and L-arabinose at indicated concentrations, respectively, and the antitoxin ParD3 was carried on the pEXT20 vector (carb^R marker, ML3483) with expression induced by IPTG¹⁰. Toxin and antitoxin libraries containing all possible single mutants were each cloned under a bicistron RBS design, in which a short leader peptide is engineered upstream of, and co-operonic with, the toxin or antitoxin. This design substantially reduces expression effects that would otherwise arise by variant 5' regions of the toxin or antitoxin forming secondary structure with its ribosome binding site^{62,63}. Consistent with this desired effect, we confirmed that synonymous variants throughout the toxin and antitoxin behaved comparably.

For the single mutant and suppressor scan, we used the arabinose-titratable wild-type *E. coli* strain BW27783⁶⁴. For the combinatorial antitoxin library experiments, we used the previously optimized TOP10 *E. coli* background¹⁰.

Library construction

The toxin and antitoxin single mutant libraries were each constructed using a previously described 2-step overlap-extension PCR protocol⁶⁵. For the toxin library, we used pBAD33-*parE3* as a template. To introduce mutations at a given amino-acid position/codon, we used a pair of mutagenic primers containing NNNs at the position to be mutated (forward and complementary reverse mutagenesis primers; see Supplementary Table 3). The reverse mutagenesis primer was used with the primer DDP115 specific to the 5' end of *parE3* and the complementary forward mutagenesis primer was used with the primer DDP116 specific to the 3' end of *parE3* (PCR cycling was: 30 sec. at 98°C; 20 cycles of: 10 sec. at 98°C, 20 sec. at 55°C, 1 min. at 72°C; 2 min. at 72°C, hold at 4°C; using Phusion kit (NEB) or KAPA). The products of these two PCRs were then combined, diluted 1:100, and amplified with DDP115 and DDP116 using the same thermocycling protocol to create full-length *parE3* harboring all possible nucleotides at a single codon position. For codon positions that failed to yield a desired, full-length PCR product on a 1% agar gel, we added 3% DMSO, 1M betaine and/or 6% 1,2-propanediol. This same process was repeated for each possible position in *parE3* and the final PCR products combined in approximately equimolar concentrations using the Qubit kit (ThermoFischer). The same overall process was followed to create the antitoxin mutant library, but used DDP141 and DDP142 as the flanking primers and pEXT20-*parD3* as the template. We then amplified the pBAD33-*parE3* and pEXT20-*parD3* vectors using primers DDP508 and 509 or DDP540 and 541 (following KAPA kit thermocycling recommendations), respectively. These PCR products were digested with HindIII-HF and SacI-HF (NEB), and then subjected to a PCR clean up kit (Qiagen or Zymo). The PCR products that comprised the toxin and antitoxin library inserts were also digested with HindIII-HF and SacI-HF, subjected to a PCR clean up kit (Select-a-size DNA Clean&Concentrate with >150 base-pair size cutoff to manufacturer specifications), and then ligated into the amplified vectors using T4 DNA ligase (NEB) at 16 °C for 16 hours with a 1:3 molar ratio of insert to vector with 50 ng vector per 20 µl ligation reactions. These ligation reactions were scaled based on downstream needs. Ligations were dialyzed on Millipore VSWP 0.025 µm membrane filters for 90 minutes before transformation.

Single mutant deep mutational scanning library preparations

For the antitoxin single mutant library, we transformed in replicate ~40 µl of electrocompetent BW27783 (made electrocompetent as described previously⁶⁶) harboring the wild-type toxin on pBAD33 with 1–2 µg of dialyzed antitoxin single mutant ligation reactions, recovered in 1 ml SOC for 1 hour at 37 °C. We plated 10 µl of the recovered transformants in a 1:10 serial dilution series on selective agar plates (carbenicillin/chloramphenicol/1% w/v glucose) to check transformation efficiency (all libraries were propagated with >10⁶ transformants). We minimized the number of generations that the libraries were propagated within cells due to leaky toxin expression, and grew cells at 37 °C until they reached OD₆₀₀ ~ 0.3, at which point glucose was removed by washing 4 times with M9L (spinning cells down at 8000 *g* for 5 min), and ready for the growth rate measurements described below.

For the toxin single mutant library in the presence of wild-type antitoxin, we followed the process outlined above for the antitoxin library with the exception that we transformed

electrocompetent BW27783 cells containing wild-type antitoxin on pEXT20 with the toxin single mutant ligation reactions. For the toxin single mutant library in the absence of antitoxin, we transformed electrocompetent BW27783 cells containing an empty pEXT20 plasmid with the toxin single mutant ligation reactions, and followed the same process as above.

Suppressor library preparation

For the suppressor scan, we screened all toxin single substitutions against a set of 36 antitoxin single substitutions. 29 of the 36 antitoxin substitutions are found at positions that are within the top 15 toxin-antitoxin covarying pairs of residues with the toxin, 5 of the 36 are found within the top 7 antitoxin-antitoxin covarying pairs of positions (E26, I33, W41, V17, K44), and 2 did not fall in either of these categories (A16K, D60I). For the suppressor scan, we measured the bulk growth of cells pooled in a single flask containing the toxin single mutant library plasmids with up to 15 different antitoxin mutant backgrounds (see github repository for pooling of variants). To read out both the toxin single mutant and antitoxin single mutant background by sequencing the toxin gene only, we cloned the toxin mutant library into separate vectors containing one of 15 different 4-nucleotide barcodes just 3' to the toxin gene and restriction site. Barcoded vector backbones of pBAD33 were generated using primers (reverse primer DDP 508 and forward primers DDP239–254, using KAPA kit and KAPA recommended cycling protocols) that contain and therefore introduce the barcodes. Barcodes were chosen to be at least 3 nucleotides different from any other barcode (Supplementary Table 3). These barcoded vector backbones were then separately ligated with the toxin single mutant library insert and prepared as described above. We transformed this DNA ligation reactions separately into TOP10 cells, grew them to $OD_{600} \sim 0.5$ at 37 °C in M9L/carb/chlor/1% glucose and then minipreped plasmids. We then transformed each barcoded toxin library in replicate into electrocompetent BW27783 cells containing one of these antitoxin single mutant plasmids, and kept track of which barcode corresponds to which antitoxin single mutant. The antitoxin single mutant plasmids were generated by following the above overlap extension PCR protocol with mutagenesis primers described in Supplementary Table 3 and flanking primers DDP141 and DDP142 using the pEXT20-*parD3* plasmid as template. This allowed us to sequence the full-length toxin gene and adjacent 10 nucleotides to determine both the toxin mutation, as well as the antitoxin mutant background from pooled cell samples.

Each of the transformed libraries was recovered in 1 ml SOC, transformation efficiency checked as described above, and then grown in 50 ml M9L/carb/chlor/1% glucose to $OD_{600} \sim 0.3$, resuspended in 5 ml M9L/carb/chlor/1% glucose/20% glycerol, aliquoted into 1 ml tubes, and flash frozen in liquid nitrogen. On the day of growth rate measurements, aliquots of the toxin single mutant libraries in each of the different antitoxin backgrounds were thawed, equally mixed based on their OD_{600} measurement at the time of freezing, and recovered in 50 ml M9L/carb/chlor/1% glucose at 30 °C for 3 hours. Subsequently, glucose was removed by washing 4 times with M9L, and cells were ready for growth rate measurement.

Combinatorial antitoxin library preparation

For the combinatorial mutant scanning, we followed the previously optimized expression conditions¹⁰ under which wild-type ParD3 shows differential neutralization of cognate and non-cognate toxins. TOP10 *E. coli* cells containing the combinatorial library were made electrocompetent⁶⁶. A plasmid containing wild-type ParE3, wild-type ParE2, or one of 10 ParE3 variants, generated using overlap-extension PCR⁶⁵ (as above using wild-type toxin plasmid as template and mutagenesis primers DDP627–642), was then transformed into these cells, transformation efficiency checked, and cells grown up in 50 ml M9L/carb/chlor/1% glucose for snap-freezing. Cells were then thawed, recovered at 30 °C for 3 hours and glucose washed off as above.

Library growth rate measurements

On the day of growth rate measurement, cells prepared as detailed above were resuspended at an OD₆₀₀ of ~0.03 in 250 ml M9L/carb/chlor/IPTG, which induces antitoxin expression. After 100 minutes, arabinose was added to induce toxin expression (t=0 minutes). Cell growth was followed by measuring OD₆₀₀ every 30–60 minutes, and when cultures reached an OD₆₀₀ of ~0.3, they were diluted 1:10 with fresh, pre-warmed M9L/carb/chlor/IPTG/arabinose to keep cells in exponential growth throughout the duration of the experiment. Samples were taken at the time of toxin induction, and 10 hours after toxin induction, which corresponds to ~6 generations of cells harboring wild-type ParE3 and ParD3. This results in a dynamic range of $2^6 = 64$ -fold change. Each library growth experiment was performed in duplicate, using separate library transformations for each replicate.

For the single mutant and suppressor scans, performed in the arabinose titratable strain BW27783, toxin was induced by adding the lowest concentration of arabinose (1.2×10^{-4} %) that produced complete cessation of growth using the wild-type toxin (Extended Data Fig. 2b). For this concentration of arabinose, 10 mM IPTG was the minimal concentration necessary for the wild-type antitoxin to provide a nearly complete restoration of growth (Extended Data Fig. 2c). To assess whether our results are sensitive to the chosen induction levels, we performed the entire suppressor scan (73,440 amino acid variants) also at a 'low' antitoxin concentration that almost, but not fully, neutralizes the toxin (8×10^{-5} % arabinose, 17.5 mM IPTG, Extended Data Fig. 2d). For the toxin single mutant scan done in the absence of wild-type antitoxin, we performed the growth measurement at 7 different arabinose induction levels (0.2 , 5.3×10^{-3} , 8×10^{-4} , 1.2×10^{-4} , 8×10^{-5} , 4×10^{-5} , 2×10^{-5} % arabinose) for which the highest 4 reach full growth inhibition (Extended Data Fig. 2b,e), each in replicate. For the combinatorial mutant scans, we performed the experiments with previously optimized inducer concentrations of 0.2% arabinose (w/v) and 100 μ M IPTG induction in the TOP10 strain background.

Sample sequencing and preparation

For each replicate, samples taken at 0 and 600 minutes after toxin induction were miniprep (Zymo Research Plasmid Miniprep Kit). We then performed a high-input (400 ng plasmid DNA), low cycle (12 rounds) PCR reaction using KAPA-HiFi (cycling conditions) to amplify the toxin or antitoxin library of interest. The primers introduced Illumina adapter sequences, sequencing primer homology regions, and Illumina multiplex

indices for each sample (forward primer DDP543 and reverse primers DDP178–193 + DDP569–580 for the toxin in single mutant and suppressor screens, forward primer DDP544 and reverse primers DDP545–568 for antitoxin single mutant library, forward primers DDP643–645 with reverse primers DDP646–693 for combinatorial antitoxin library). These primers also introduce variable numbers of random nucleotide or YRYR nucleotides (Y corresponding to random pyrimidines, R corresponding to random purines) as the very first bases to be sequenced on the forward primers in order to allow for Illumina cluster definition, and stagger our homopolymer-like amplicon.

We gel purified amplicons of interest (~500 nucleotides) by running samples with a loading dye for 30 minutes at 180 V on a Novex 8% TBE gel. We sheared the excised gel band by spinning it through a bottom-pierced 0.5 ml tube placed within a 1.5 ml tube, added 500 μ l of 10 mM Tris buffer (pH=8), and then froze the sample at -20°C for 15 minutes followed by incubation at 70°C for 10 minutes to solvate the DNA from the gel. We spun each sample through a 0.22 μ m spin-x cellulose acetate column to separate the gel from the supernatant, and then performed an isopropanol precipitation of the DNA by adding 32 μ l 5M NaCl, 2 μ l glycoblue, and 550 μ l 100% isopropanol. The mixture was then chilled at -80°C and centrifuged at 4°C at 14,000 *g*. Finally, the sample was washed with ice cold 70% ethanol, air-dried, and resuspended in 10 μ l water.

Each sample was then run on a fragment analyzer, and qPCR was used to quantify the DNA concentration. Finally, samples were pooled and sequenced on a MiSeq, or HiSeq 2500, with 250 or 300 bp paired end reads. Sequencing was performed with variable 20–30% PhiX spike in.

Analysis of sequencing data and growth rate measurements

Raw fastq paired-end sequencing reads were merged using FLASH 1.2.11⁶⁷. Merged reads were quality filtered based on their phred-score using vsearch 2.13.0⁶⁸, with the following arguments: `vsearch --fastq_filter {0} --fastq_truncqual 20 --fastq_maxns 3 --fastq_maxee 0.5 --fastq_ascii 33 --fastaout {1}.fasta`

Toxin mutant reads were subsequently split into separate files based on their four-nucleotide barcode indicating the antitoxin background, if applicable. Subsequently, reads were filtered for those that (i) spanned the full-length of the toxin or antitoxin gene, (ii) had no mutations or indels in the immediately flanking 10 bp upstream (which includes the restriction site, RBS, and the stop codon for the upstream bicistronic peptide), as well as the downstream 6 bp (which includes the restriction cloning site). The majority of reads (~77%) contained a single codon mutation. Custom analysis scripts for raw read processing, Bayesian inference of growth rates (see below) and nonlinear modeling (see below) are available at: https://github.com/ddingding/coevolution_paper.

Hierarchical Bayesian inference of mutant growth rates

We used a Bayesian model (Extended Data Fig. 8) that allowed us, given a plausible data generating process (likelihood function) that captures how growth rates could give rise to our observed codon-level read count data before and after selection, and vague priors, to get posterior probabilities for our growth rates of interest, namely how likely different values

of a growth rate for a particular amino acid substitution are given the data observed. This model takes into account sampling noise of reads, synonymous mutant observations per amino acid, and biological replicate experiments to infer amino-acid variant growth rates and their uncertainties. Our model allowed for calibrated uncertainty inference, as well as unbiased inference of amino acid substitution effects compared to the widely-used log-read ratio statistic (Extended Data Fig. 9).

For our likelihood model, we extended previous Bayesian approaches for mutant effect inference from read data^{69,70} and built a hierarchical Bayesian model. As done previously, read counts for each codon at a particular time-point were emitted from a Poisson distribution with an inferred Poisson rate parameter λ . The post-selection Poisson rate parameter is the pre-selection Poisson rate multiplied by the exponentiated growth rate for each codon mutant (following exponential growth rates of the form $N_t = N_0 * e^{(\text{growth rate} * t)}$).

We found that previously used models, in which all synonymous codon variants share the exact same growth rate, were insufficient to explain the observed variance in read ratios for synonymous variants (Extended Data Fig. 10a–c). This finding motivated us to expand from a non-hierarchical generative model to a multilevel model allowing for partial pooling of synonymous variant growth rates to inform inference of their shared, amino-acid-level variant growth rates. For this model, the growth rate for a particular codon mutant was drawn from a normal distribution whose mean is the growth rate of the amino-acid variant. In this way, each synonymous mutant's growth rate informs the amino-acid variant growth rate, while still capturing the observed variation in synonymous mutant growth rates.

Because the posterior of our model is not analytically tractable, we used Stan^{71,72} to perform inference (using 2 MCMC chains, 10,000 steps each, discarding the first half of MCMC chains as 'warmup'). This gave us 10,000 discrete samples for each amino-acid variant growth rate of interest, approximating the true continuous posterior density. We used the mean of these 10,000 samples as our best guess for the true growth rate of that particular amino-acid variant, with the distribution of these 10,000 samples reflecting the posterior uncertainty in the inferred growth rate.

Growth rate inference validation

We validated our Bayesian model in three ways, as summarized below.

Inference of synthetic growth rates: We generated three different synthetic datasets (pre- and post-selection read counts for each codon variant) by assuming three different distributions of true amino-acid mutant growth rates (from the range of antitoxin single mutant growth rate distributions). We compared the Bayesian inference to the classically used log read ratio (+1 pseudo-count), and found that our model removed the bias introduced in the raw log read ratio when few reads are observed post-selection (Extended Data Fig. 9). The mean posterior growth rate for each amino acid mutant still correlated well with the true synthetic growth rates at low growth rates. As desired, our model assigns these low growth rate values with few observed post-selection reads a larger 95% highest posterior density interval.

Calibrated uncertainty: Our Bayesian model allowed us to calculate the associated uncertainty, *i.e.* the posterior distribution for each amino-acid variant. We compared differing uncertainty intervals for each variant (95%, 90%, 80%, 50% posterior highest density intervals) with the true growth rates used to simulate the observed count data, and found that the percentage of true growth rates falling within the posterior highest density intervals corresponded to the percentage of the highest density interval (Extended Data Fig. 9d).

Posterior predictive checks: Posterior predictive checks allowed us to assess whether a given model was complex enough to capture the observed data. After model parameter inference, replicate data were generated from the model using the inferred distribution of parameters, and compared to the observed data (Extended Data Fig. 10d–g). We chose multiple different test quantities (log read ratios for each codon and averaged across amino-acid variants, the standard deviation of synonymous mutant log read ratios for each amino-acid variant as well as synonymous wild-type toxin mutants) to compare 10,000 simulated replicate datasets generated from the model to the true observed data, and for each quantity calculated their posterior predictive p-value (*i.e.* the fraction of simulated data above the observed data along the test statistic). Based on these test quantities, the model developed plausibly generates the observed data, demonstrating sufficient complexity.

Calling significantly beneficial toxin mutants

Toxin mutants were called as significantly beneficial to a given antitoxin mutation if they grew at least 0.5 log₂-fold better than that antitoxin mutant combined with the wild-type toxin measured in the same flask, with all 10,000 posterior samples exceeding the growth rate of the antitoxin mutant with the wild-type toxin, *i.e.* the 99.99% highest posterior density interval does not overlap zero difference in growth rate. For the ‘high’ antitoxin expression condition (see main text), we sought beneficial toxins in 9 deleterious antitoxin backgrounds (F73K, K63D, W59A/L/V/T, G62D/Y/L). For the ‘low’ antitoxin expression condition, we sought beneficial toxins in 12 deleterious antitoxin backgrounds (the former plus E26R, E79H, E79K).

We verified that the toxins harboring beneficial mutations were still as, or nearly as, toxic as the wild-type toxin by measuring the growth rate of the toxin single mutant library at 7 different toxin induction levels as indicated above (Extended Data Fig. 2). For the 2,040 possible amino acid substitutions in the toxin, we called 310 as fully toxic as they did not show any significant growth rate differences (the 95% highest posterior density interval overlapped with the growth rate of the wild-type across all 7 expression conditions. Individual, low-throughput validation of four such toxin mutants revealed no distinguishable growth rate differences to the wild-type toxin (Extended Data Fig. 7a) at both the minimum expression level under which the wild-type toxin produces maximal growth inhibition (1.2×10^{-4} % arabinose) in the absence of an antitoxin, as well as when reduced to a level that produced half-maximal inhibition with the wild-type toxin (6×10^{-5} % arabinose). We also called as toxic a less stringently defined set of 781 mutants that were not significantly different from the wild-type toxin across the four highest expression conditions, which are conditions that lead to full growth inhibition by the wild-type toxin (Extended Data Fig. 2).

Combinatorial mutant scan analysis

Following previous analyses^{10,73}, we calculated for each antitoxin amino-acid variant (combining synonymous mutants) a mean log-read ratio for 600 and 0 minutes post-induction of the toxin, with a pseudo-count added to both time-points. We then scaled the growth rates to between 0 and 1 based on the mean growth rate of truncated antitoxin mutations and the wild-type antitoxin, respectively. In the case of the antitoxin library in the background of the non-cognate toxin ParE2, we used the rewired antitoxin ParD3 ILK for the maximum growth scaling.

Nonlinear modeling and calling of specific vs. non-specific

We pooled all of the observed growth rate data for toxin mutants combined with the wild-type antitoxin, one of the antitoxin single mutants, or no antitoxin. The predictor for each growth rate is the one-hot encoded amino acid sequence, indicating the site-wise presence of a particular toxin substitution or antitoxin substitution at each position. We fit weights associated with each unobserved, single substitution effect, passed through a global sigmoid function by using the Adam optimizer⁷⁴ to minimize the mean squared error of predicted growth rates relative to the observed growth rates. This was done using Tensorflow 2⁷⁵, run for 4,000 steps until no decreases in fitting accuracy were observed. Weights were initialized with the ordinary least square weights and Adam learning rates were grid-scanned among [1, 0.1, 0.01, 0.001]. The linear fitting was done similarly using the Statsmodels python package⁷⁶, with an identical model missing the sigmoid transformation.

Generally speaking, relative growth rate comparisons should be most robust when comparing mutants grown in the same flask because a particular mutant with one fixed absolute growth rate might decrease in relative fraction in a flask with faster growing mutants, but increase in a flask with slower growing mutants. In our case, the distribution of growth rates in each flask were similar enough, such that a nonlinear model trained on relative growth rates pooled across flasks could fit and explain 89% of the observed growth rate variance, giving us confidence that this issue is a negligible factor in our analysis. While inferred mutation effects are robust to expression levels (Extended Data Fig. 4h), the observed single mutation growth rate effects (Extended Data Fig. 4i) or deviation of double mutant growth rate from expected growth rates (Extended Data Fig. 4j) correlate less well. This is expected since single mutation effects can be buffered, depending on the expression conditions (Extended Data Fig. 4g).

To call a particular double mutant combination as having an observed growth rate that deviates substantially and significantly from that expected by the model, we required the double mutant growth rate to be at least 2-fold greater than or less than the independent expectation, with a 99.99% highest posterior density interval that did not overlap the independent expectation (*i.e.*, all 10,000 samples were greater or less than the expectation). The most positively deviating toxin mutation in each deleterious antitoxin background were mostly in direct contact and were biochemically rationalizable. However, most of these positively deviating double substitutions did not improve the growth rate over the antitoxin single substitution growth rate, but deviate because their independent expectations are highly detrimental.

In contrast to previous studies^{42,47–55}, which have also successfully used nonlinear, independent models to quantify double mutant expectations, we demonstrated that the inferred, unobserved mutation effects in our nonlinear model were robust with respect to the details of expression conditions. In particular, the inferred mutant effects for all toxin mutants were highly correlated for the ‘high’ and ‘low’ induction levels of antitoxin ($r=0.98$, Extended Data Fig. 4h).

Low-throughput toxicity measurement and orthogonal growth rate validation

To assess whether non-specific suppressor mutations in the toxin maintained toxicity, we measured the growth rate of cells containing these toxin mutations in the absence of the antitoxin in the arabinose titratable strain at both full ($1.2 \times 10^{-4}\%$ arabinose) and half-maximal induction levels ($6 \times 10^{-5}\%$ arabinose). We diluted saturated overnight cultures 1:400 into the wells of a 96-well plate containing M9L supplemented with 10 μM IPTG and arabinose (concentration as indicated, Extended Data Fig. 7a), as well as carbenicillin and chloramphenicol. We ran a maximum of 8 samples per 96-well plate, such that each sample is measured at least 10 times in each plate. Row A was kept blank. Each sample was staggered diagonally across the plate reader to minimize plate reader position biases. For example, sample 1 was loaded into wells B1, C2, D3, E4, F5, G6 H7, B9, C10, D11, E12. Each plate also contained samples corresponding to a wild-type toxin combined with a wild-type antitoxin and the wild-type toxin combined with an empty vector for within plate growth rate comparisons. We used the plate reader Biotek Synergy H1 with orbital shaking 365 rpm at 37 °C, with 180 μl of media and 70 μl mineral oil on top to prevent evaporation. Assayed toxin and antitoxin variants (T*+AT*) are: T+AT(G62L), T+AT(W59T), T+AT(F73K), T+AT(K63L), T(V5L)+AT(G62L), T(V5L)+AT(W59T), T(E37D)+AT(G62L), T(G81Y)+AT(G62L), T(P8A)+AT(G62L), T(R52L)+AT(G62L), T(P8N)+AT(W59T), T(V75G)+AT(W59T), T(G94)+AT(W59T), T(W85P)+AT(W59T), T(V5L)+no antitoxin, T(A66F)+no antitoxin.

For orthogonal growth rate validations, we diluted overnight cultures 1:50 into 10 ml of fresh M9L/carb/chlor/1% glucose/10 μM IPTG to pre-induce antitoxin, grew cells 2–3 hours at 37 °C to $\text{OD}_{600} \sim 0.5$, then washed 4 times with M9L. We then diluted these cells 1:200 into 96-well plates containing M9L/carb/chlor/10 μM IPTG/ $1.2 \times 10^{-4}\%$ arabinose to measure growth rates. We calculated the growth rate as the normalized fold change in OD_{600} across time. Error bars in Extended Data Fig. 1a are calculated using the error propagation formula for independent variables and a first order Taylor expansion.

Coevolution analysis

We performed covariation analysis similar to before⁴. Briefly, we generated JackHMMR⁷⁷ alignments using our wild-type toxin (ParE3, uniprot ID: F7YBW7) and antitoxin (ParD3, uniprot ID: F7YBW8) as query sequences at a range of bitscore cutoffs (ranging from 0.1 to 0.9) from the uniref100 database. For each bitscore, we concatenated toxin and antitoxin sequences based on their genome distance (< 1000 nucleotides). We selected the bitscore cutoff with the highest number of true positive (minimum atom distance $< 6\text{\AA}$) between-protein covarying pairs, resulting in the bitscore choice of 0.3. Alignment quality filtering was done similar to previous studies, calculating covariation scores only

for residue positions with at least 80% coverage across sequences, discarding sequences if they did not span 80% of the full-length concatenated query sequence (196 amino acids), and down-weighting sequences if their sequence identity exceeded 80%. The final alignment contained 1650 concatenated sequences, with an effective number of sequences of 1088 after down-weighting.

We chose to highlight the top 10 covarying residues from this analysis, which were all close ($< 6 \text{ \AA}$ minimum atom distance) in distance. Using previous calibration sets¹, the 90% precision cutoff (pairs of residues $< 6 \text{ \AA}$ distance/all selected pairs) in covariation score resulted in 29 covarying pairs of residues between protein, of which 28 are indeed within 6 \AA minimum atom distance.

For covariation score comparison against other complexes (Fig. 3b), we looked at the top N (with N being 1, 2, 3, 4, 5, or 10) covariation scores between proteins, compared against the corresponding top N covariation scores for other complexes from ref¹.

Network visualizations

Force-directed graphs were constructed using the python package networkx⁷⁸. Network clusters were defined using Louvain clustering. Sequence motifs were visualized for each cluster using Logomaker⁷⁹.

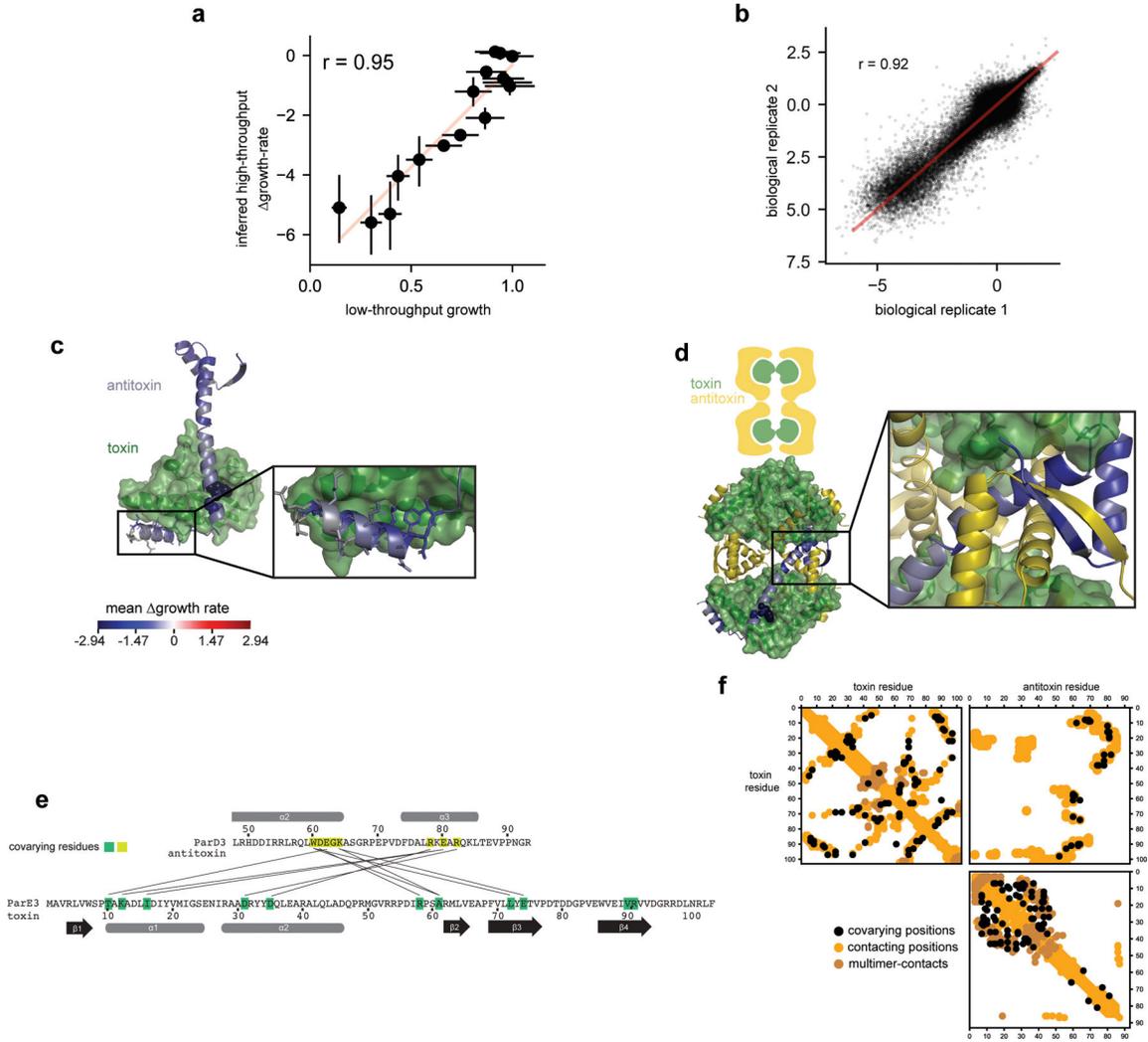
EVmutations and VAE prediction calculations

EVmutations and DeepSequence (variational autoencoder) predictions were performed as previously described^{4,57}. Briefly, both models are trained in an unsupervised fashion on the multiple sequence alignment of a given protein family. EVmutation is a pairwise undirected graphical model, or Potts model, that explicitly captures both site-wise preferences as well as the pair-wise residue dependencies between residues. DeepSequence is a variational auto-encoder, and through being a nonlinear latent-variable model, can implicitly capture higher-order dependencies between residues. Both models enable mutation effect prediction by calculating the log-odds ratio between the wild-type and mutated sequences.

Structural analysis

We used the wild-type toxin ParE3-ParD3 (PDB: 5CEG) for structural analysis and distance calculations. Minimum atom distances refer to the minimal atom centroid distance between any two atoms in two different amino acids. Minimum atom distance of a toxin residue to any antitoxin atom were calculated for the minimum of particular toxin residue from toxin chain B or chain D in PDB:5CEG.

Extended Data



Extended Data Fig. 1. Orthogonal validation of growth rate inference, structural explanation for antitoxin mutation effects, and covariational signal between toxin-antitoxin ParE3/ParD3.

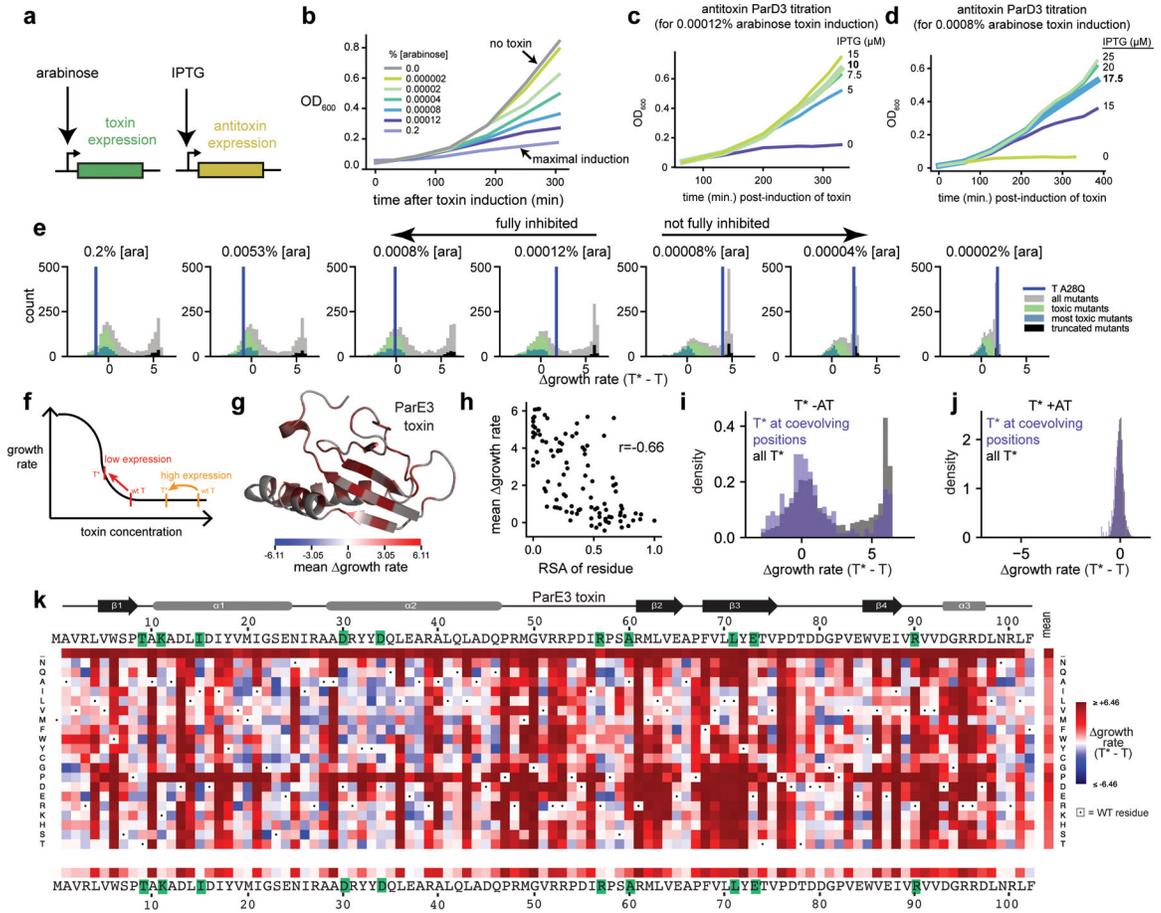
a, Comparison of growth rates inferred by high-throughput vs. individual growth measurement. X-axis error bars indicate $\pm 2x$ standard deviation derived from $n=10$ or $n=11$ technical plate reader replicates (see Methods). Y-axis error bars indicate 95% posterior highest density interval. The Pearson correlation coefficient (r) is indicated.

b, Raw log read ratio reproducibility between replicates (+1 pseudocount) for all single and double mutants. The Pearson correlation coefficient (r) is indicated.

c, Mean mutation effect of residues in the C-terminal α -helix 3 of the ParD3 antitoxin indicates that residues facing the toxin are more susceptible to mutations that disrupt the ParD3-ParE3 interaction, producing negative growth rate values.

d, Mean mutation effect in the N-terminal oligomerization region of the antitoxin highlights residues susceptible to disrupting the ParD3-ParE3 interaction when mutated. Cartoon illustrates arrangement of ParE3-ParD3 octamer observed in the co-crystal structure (PDB: 5CEG). One of the 4 antitoxin monomers is colored by the mean mutation effect.

e, Top 10 toxin-antitoxin covarying residue pairs indicated for reference.
 f, The 90% precision cutoff yields 29 toxin-antitoxin covarying residue pairs (black in upper, right quadrant) of which 28 pairs fall within toxin-antitoxin interface residues that are $< 6\text{\AA}$ minimum atom distance (ochre dots) in the ParE3-D3 crystal structure (PDB ID: 5CEG).



Extended Data Fig. 2. Titration of toxin and antitoxin expression levels, and sensitive identification of toxin substitutions which do not disrupt toxicity.

a, Cartoon illustration of the expression system. IPTG induces antitoxin, arabinose induces toxin.

b, Growth rate of cells harboring wild-type toxin ParE3 without antitoxin at different arabinose induction levels in arabinose titratable *E. coli* strain BW27783.

c, d, Growth rate of cells harboring wild-type toxin-antitoxin ParE3/ParD3 under different antitoxin induction levels modulated with IPTG and 0.00012% arabinose induction (c) or 0.0008% arabinose induction (d).

e, Distribution of growth rates (T^*-T) for all toxin single substitutions under different arabinose inducer concentrations, with positive growth rate (T^*-T) values indicating loss of toxin function. The set of ‘most toxic’ toxin substitutions ($n=310$) is colored in light blue, the set of ‘toxic’ substitutions ($n=781$) is colored in green (see Methods). Other classes of substitutions are indicated. The dynamic range (difference between 0 and the truncated toxin mutants) shrinks, as expected, for lower expression levels that do not fully inhibit

growth with the wild-type toxin, and a higher fraction of mutants show loss of toxicity (higher growth rates) under lower expression conditions. The toxin substitution A28Q is highlighted (dark blue) as an example that shows no growth rate difference relative to wild-type toxin at high expression conditions, but is not as toxic as wild-type toxin at lower expression conditions.

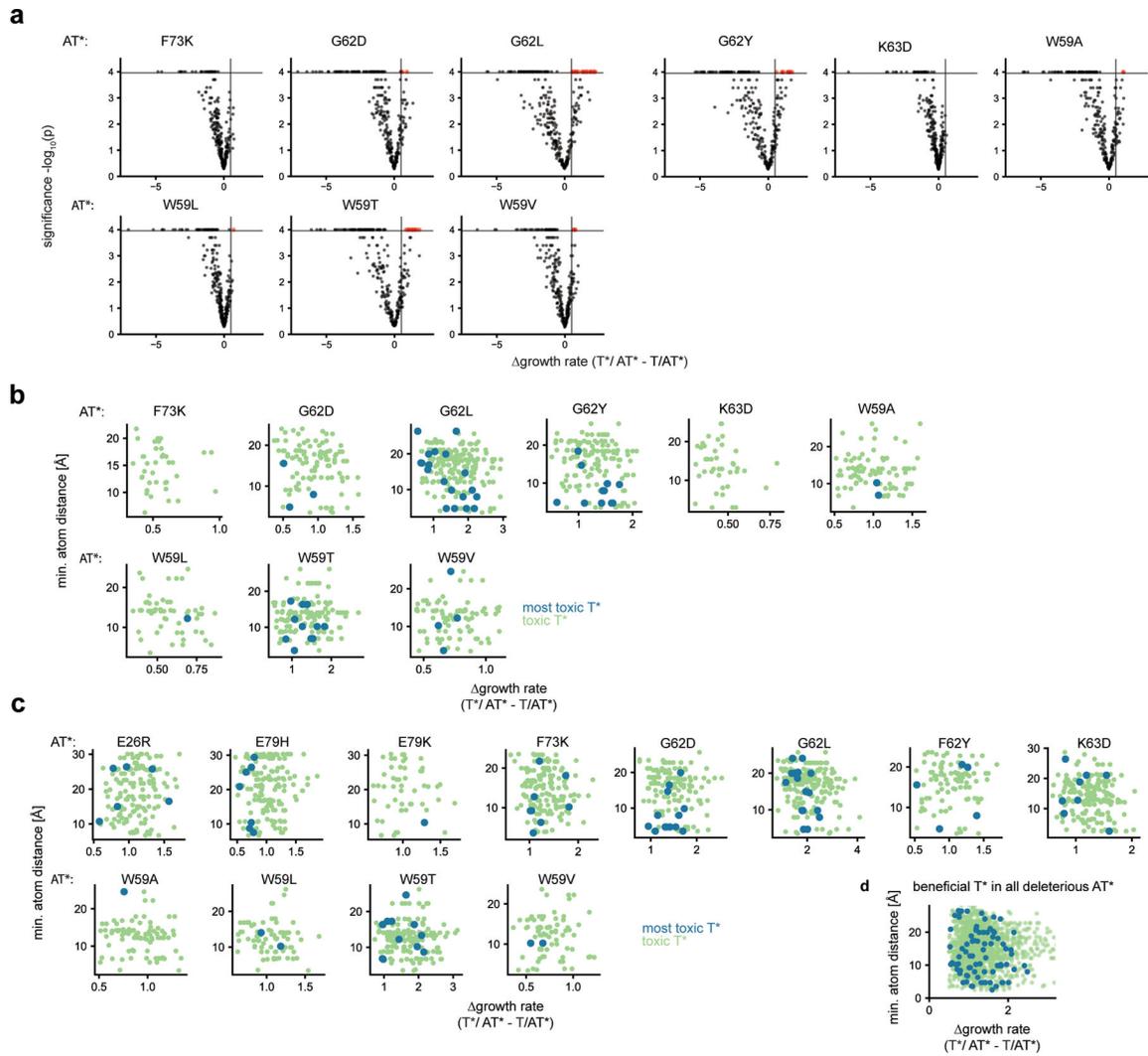
f, Schematic illustrating loss of toxicity detection using growth rate measurements in different expression regimes.

g, Mean growth rates(T^*-T) of residue positions mapped onto the ParE3 toxin structure. Values shown for 0.00012% [arabinose] inducer.

h, The mean growth rates(T^*-T) of a residue are correlated with the relative solvent accessibility of the residue (Pearson $r = -0.66$). Values shown for 0.00012% [arabinose] inducer.

i,j, Distribution of growth rate(T^*-T) for all toxin substitutions (black) or top 10 coevolving residue substitutions (purple) in the toxin in absence of antitoxin (g) or presence of antitoxin (h). Values shown for 0.00012% [arabinose] inducer, and antitoxin is induced with 10 μ M IPTG.

k, The growth rate(T^*-T) values of each substitution at any position along the toxin ParE3. Green highlights the top 10 covarying positions between toxin and antitoxin in natural homologs. Values shown for 0.00012% [arabinose] inducer.



Extended Data Fig. 3. Volcano plot visualizing significant and substantial beneficial toxin variants in different antitoxin backgrounds, and beneficial toxin variants in various antitoxin backgrounds under ‘high’ and ‘low’ antitoxin expression conditions.

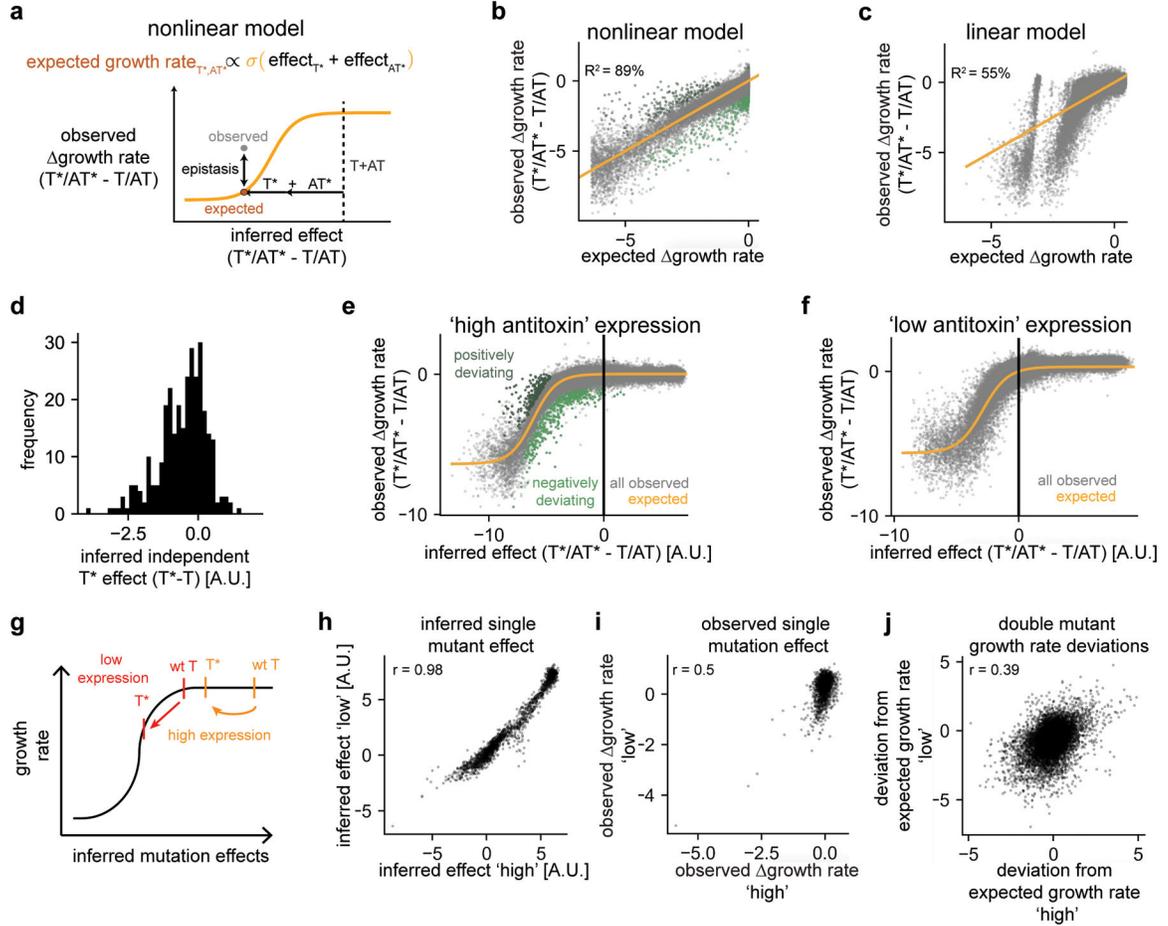
a, For each deleterious antitoxin variant background, the mean posterior change in the number of doublings, Δ growth rate($T^*/AT^* - T/AT^*$), of the most toxic toxin mutants are plotted vs. their significance ($-\log_{10}(p(\Delta$ growth rate $<0))$) of deviation from the AT* single mutation. This is based on 10,000 discrete samples of the posterior Δ growth rate($T^*/AT^* - T/AT^*$) values inferred from the hierarchical Bayesian inference model (see Methods). Vertical line: $+0.5 \Delta$ growth rate, horizontal line: $p(\Delta$ growth rate $>0) = 0.0001$. Red indicates significant and substantial beneficial toxin substitution using this cutoff. Experiments performed under ‘high antitoxin’ expression conditions.

b, The minimum atom distance from a given deleterious antitoxin residue to each beneficial toxin is plotted vs. Δ growth rate($T^*/AT^* - T/AT^*$). Experiments performed under ‘high antitoxin’ expression conditions.

c, The minimum atom distance from a given deleterious antitoxin residue to each beneficial toxin is plotted vs. Δ growth rate($T^*/AT^* - T/AT^*$). Experiments performed under ‘low antitoxin’ expression conditions.

d, Distance vs. growth rate ($T^*/AT^* - T/AT^*$) of beneficial toxin variants for all deleterious antitoxin variant backgrounds. Experiments performed under ‘low antitoxin’ expression conditions.

Values for (b-d) shown for double mutants with growth rate effect size $>+0.5$ and $p(\text{growth rate} > 0) < 0.0001$.



Extended Data Fig. 4. A non-specific, non-linear model can explain most of the observed single and double mutant growth rates.

a, Schematic of nonlinear, non-specific model: double mutant expected growth rates (brown) are based on the independent (non-specific) sum of underlying toxin and antitoxin mutant effects, passed through a sigmoid function (yellow).

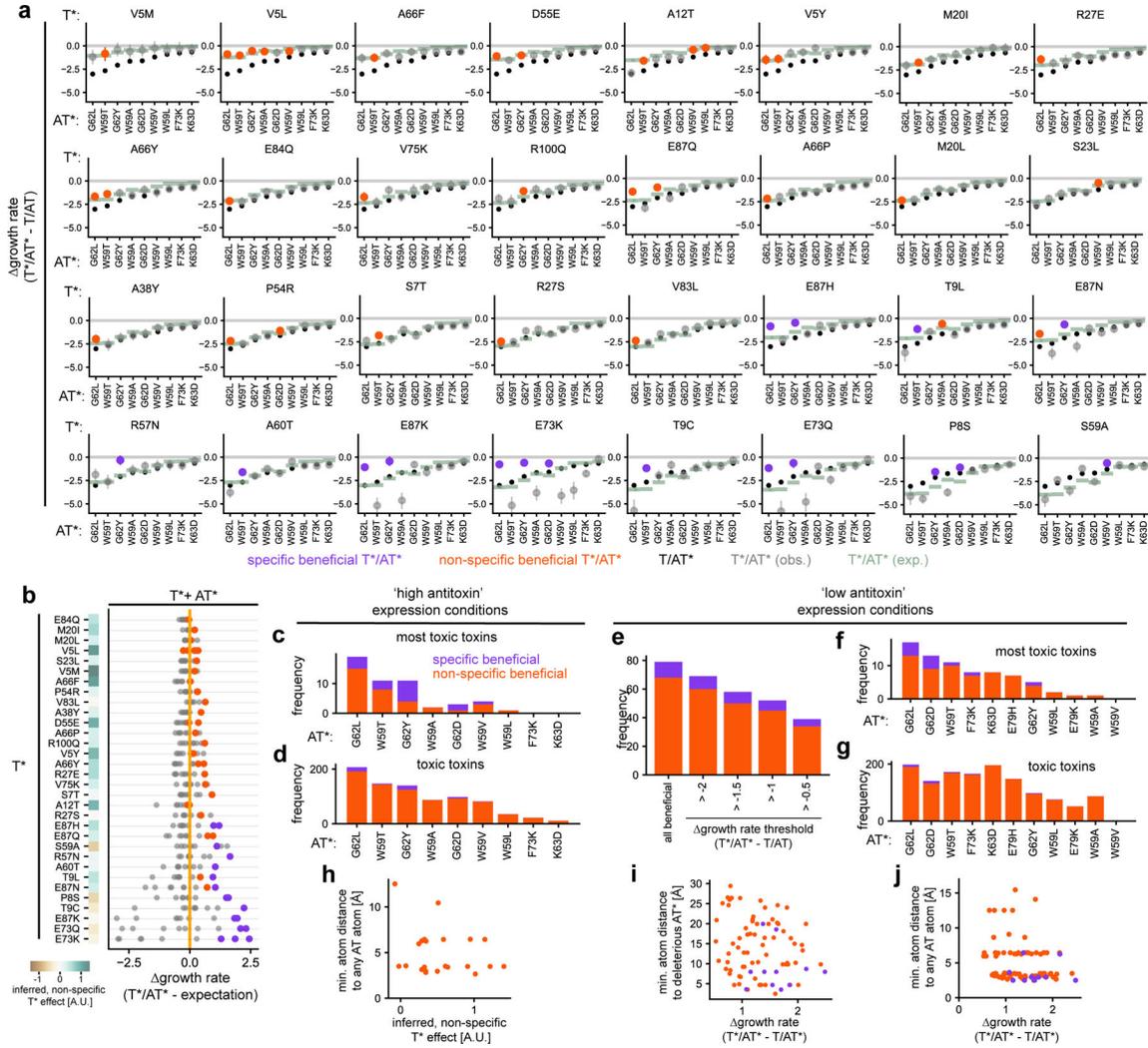
b,c, Residuals for non-linear, non-specific model (b) or linear non-specific model of the same structure without a non-linearity (c) showing unbiased residuals for the nonlinear model, but a complete misfit of the linear model. Model built using ‘high antitoxin’ expression levels. Explained variance (R^2) is indicated. Significant and substantially positively (dark green) or negatively (green) deviating mutations are shown in (b) (see Methods).

d, Inferred independent toxin single substitution effects among the set of most toxic toxin mutants demonstrating a tail of independently beneficial toxin variants. Experiment performed under ‘high antitoxin’ expression levels.

e,f, Nonlinear independent model fit to growth rates measured under ‘high antitoxin’ (e) or ‘low antitoxin’ (f) expression conditions. The wild-type toxin -antitoxin pair is inferred to be differently close to the sigmoid ‘cliff’ between expression conditions.

g, Cartoon illustrating different detection of single mutant effects depending on expression conditions.

h-j, Correlation of inferred single mutant effects (h), observed single mutant growth rate ($T^*/AT^* - T/AT$) effects (i), and double mutant deviations of observed from expected growth rates (j) from separate inference under ‘high antitoxin’ (x-axis) or ‘low antitoxin’ (y-axis) expression conditions.



Extended Data Fig. 5. Deviation of observed from expected double mutant growth rates reveals toxin variants with specific or with only non-specific beneficial effects, and fraction of specific vs. non-specific toxin variants.

a, For each beneficial toxin mutation (indicated above each plot) combined with each antitoxin variant indicated on the x-axis, the plot shows the growth-rate relative to the wild-type toxin-antitoxin pair (mean posterior growth rate $(T^*/AT^* - T/AT)$). Grey dots represent T^*/AT^* , error bars indicate 95% posterior highest density interval. The growth

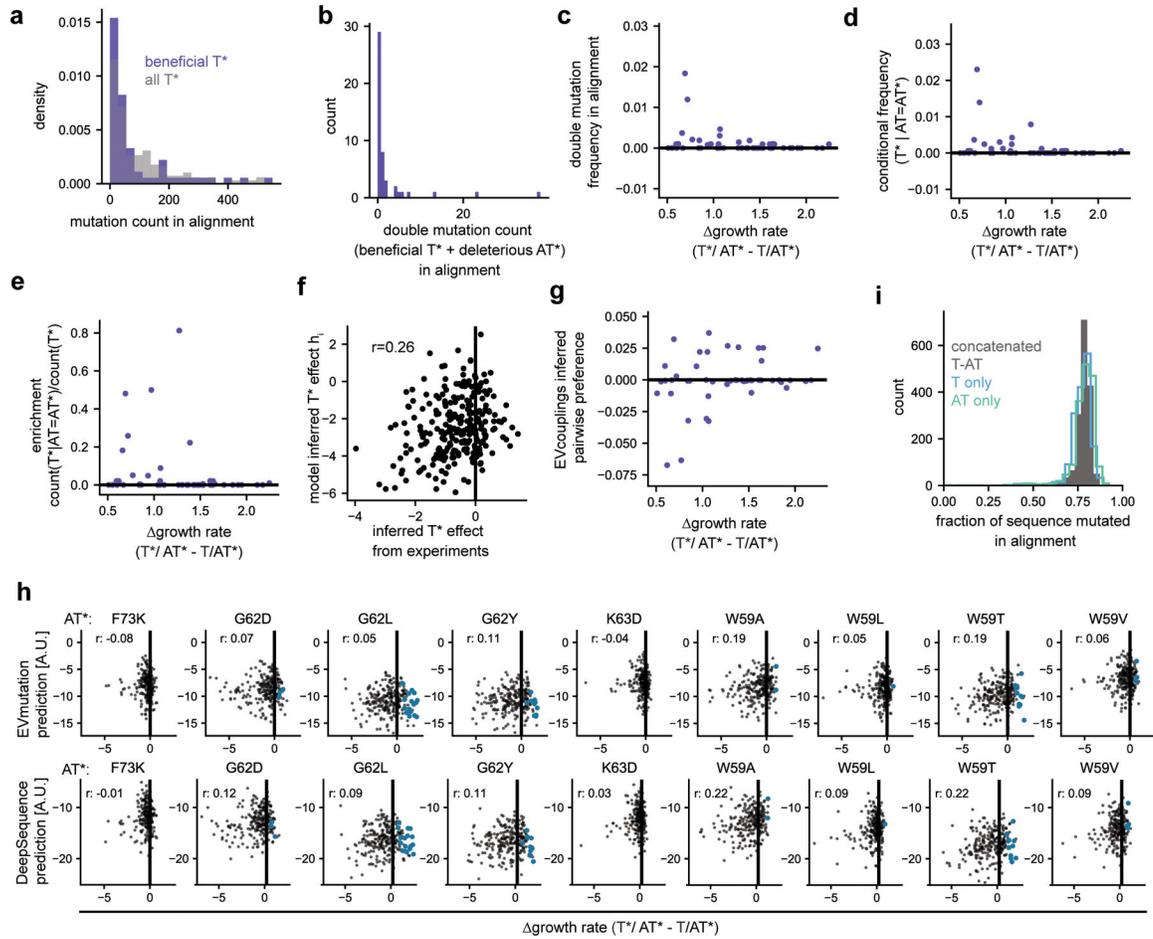
rate for each antitoxin mutant combined with wild-type toxin (T/AT*) is shown (black dots) along with the growth rate for T*/AT* expected under the non-specific, nonlinear model (green dots).

b, Deviation of the observed (dots) from the expected double mutant growth rates (orange line) highlights classification of specific and non-specific toxin variants. Beneficial toxin substitutions (rows, n=32) ordered by their range of growth rate deviations across deleterious antitoxin variants as in panel b.

c-g, Specific vs. non-specific enabling toxin variants under 'high' antitoxin expression for all enabling toxin variants grouped by deleterious antitoxin for the more stringent set of 310 'most toxic' toxins (c) and less stringent set of 781 'toxic' toxins (d). Orange and purple indicate mutant pairs involving non-specific and specific, respectively, rescuing mutations in the toxin. Enabling toxin variants under 'low' antitoxin expression at different absolute growth rate cutoffs relative to the wild-type toxin/antitoxin growth rate (e), or grouped by 'most toxic' (f) or 'toxic' (g) toxin variants.

h, Inferred non-specific toxin variant effect vs. minimum atom distance to any antitoxin atom for 21 non-specifically rescuing toxin variants (orange).

i, j, For specific and non-specific beneficial toxin mutants, the change in growth rate in a deleterious antitoxin mutant background, growth rate (T*/AT* - T/AT*), is plotted vs. minimum atom distance to the deleterious antitoxin mutation it rescues (i) or any antitoxin atom (j) in the 'low antitoxin' expression condition.



Extended Data Fig. 6. Natural sequence statistics, EVcouplings or DeepSequence models are not predictive of beneficial toxin substitution effects.

a, Distribution of number of specific and non-specific beneficial toxin substitutions (purple) vs. all possible toxin variants (grey) observed in natural sequences.

b, Frequency distribution of beneficial toxin and deleterious antitoxin mutant pairs in natural sequences, with 29/51 pairs never observed.

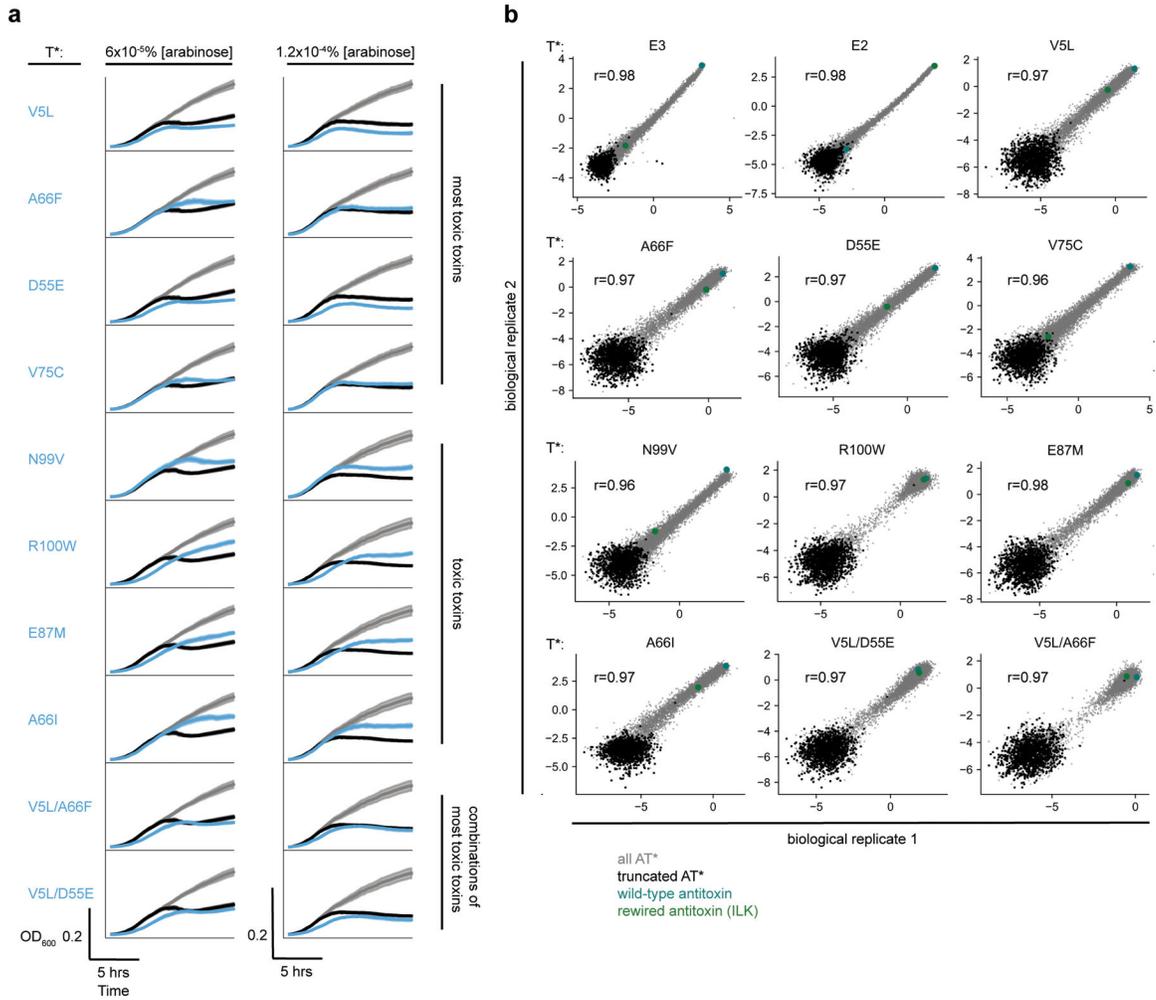
c-e, Effect size of toxin variant rescue vs. frequency of variant pair in natural sequences (c), conditional frequency of toxin variant given natural sequences containing the particular deleterious antitoxin substitution (d), or enrichment of beneficial toxin variant in natural sequences containing the deleterious antitoxin substitution (e).

f-g, EVcouplings model inferred site-wise toxin mutant preferences (h_i) vs. toxin mutant effect inferred in suppressor scan with the Pearson correlation coefficient indicated (f), or EVcouplings pairwise T^*/AT^* variant preference (J_{ij}) vs. effect size of beneficial toxin mutation effect in a deleterious antitoxin variant background (g).

h, Scatterplot of observed beneficial toxin effect in deleterious antitoxin mutant backgrounds (AT^*), vs EVmutation (top row) or DeepSequence (variational autoencoder) mutation effect predictions (bottom row). Pearson correlation (r) is indicated.

i, Distribution of natural sequence identity fractions across the alignment. Different histograms illustrate fraction mutated for homologs containing the full concatenated toxin

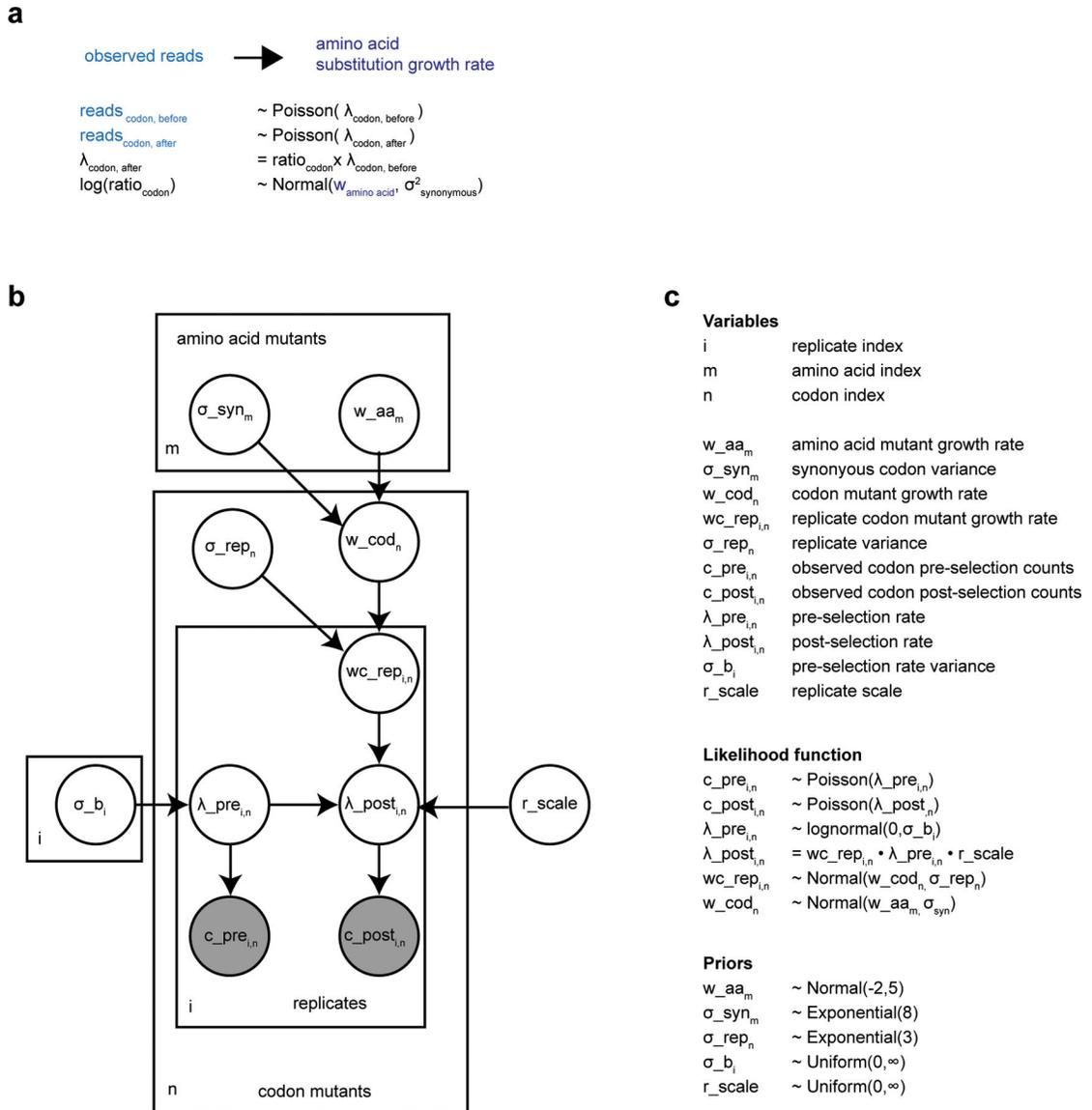
and antitoxin (grey), the toxin homologs only (blue), or the antitoxin homologs only (turquoise).



Extended Data Fig. 7. Non-specific suppressor toxin ParE3 variants are as or almost as toxic as wild-type ParE3, and reproducibility of antitoxin combinatorial variant log read ratios.

a, Growth rates of ParE3 non-specific suppressor toxin variants (blue) compared to wild-type toxin ParE3 without antitoxin (black) and wild type toxin and antitoxin (grey) under fully inhibitory toxin expression conditions (0.00012% [arabinose]) or half-maximal inhibitory expression conditions (0.00006% [arabinose]). Dark lines represent the mean OD600, shaded regions show standard deviation of the replicates (n=10 or n=11).

b, Raw log read ratio reproducibility between biological replicates (+1 pseudocount) for the combinatorial antitoxin library (8000 amino acid variants) in different toxin mutant backgrounds. Specific classes of antitoxin mutants, and Pearson correlation coefficients (r) are indicated.

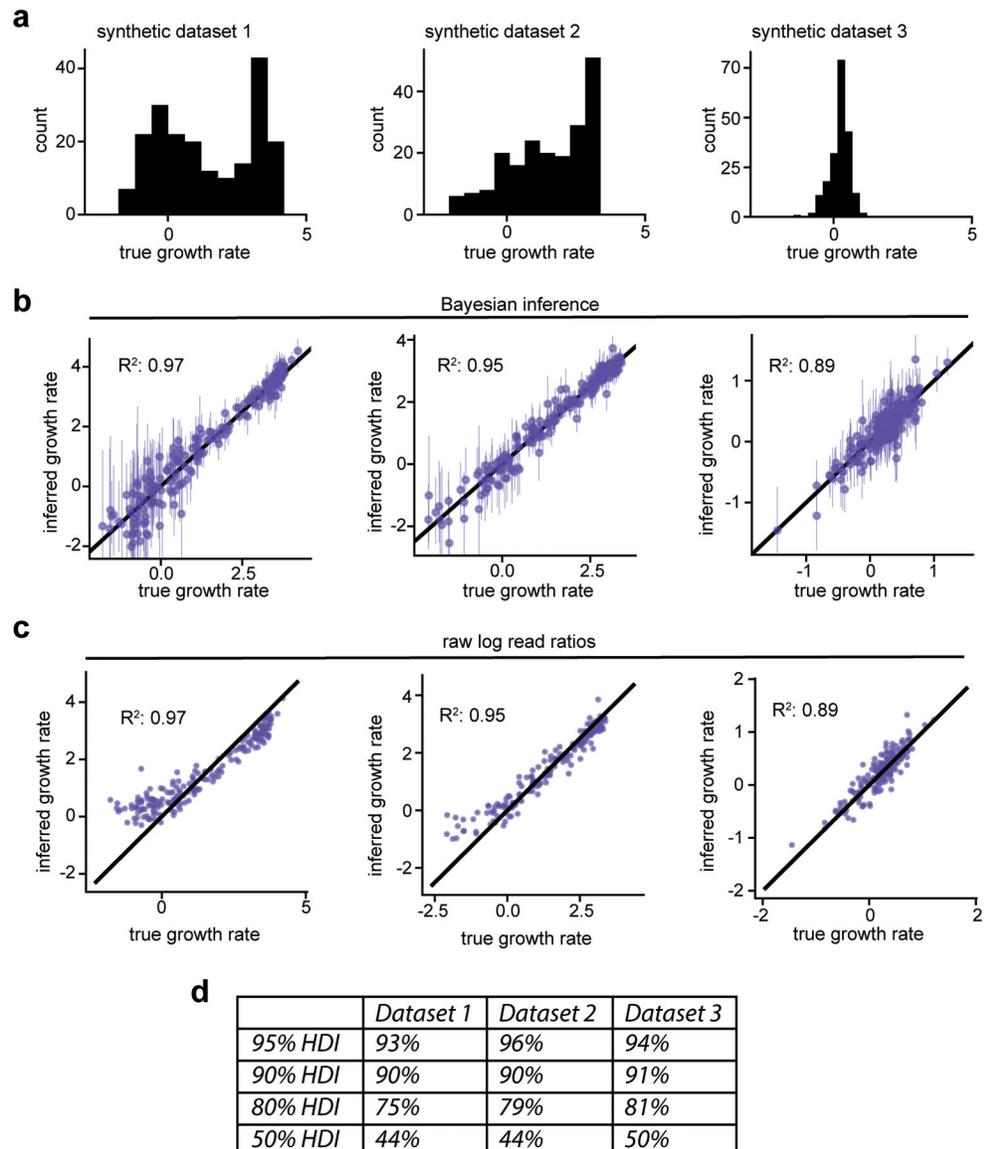


Extended Data Fig. 8. Bayesian hierarchical model.

a, Simplified description of the Bayesian hierarchical model. Pre- and post-selection reads for each codon are drawn from a Poisson distribution. The log-ratios of these Poisson parameters are not fixed between synonymous codons but are instead drawn from a normal distribution, whose mean forms the amino acid mutant growth rate of interest. This model allows for different synonymous codons to inform each other as well as the amino acid mutant growth rate without being completely fixed.

b, Full plate diagram description of the hierarchical Bayesian model capturing both replicates. Replicate index *i* takes values 1 or 2, amino acid index *m* takes on values ranging from 1–2040 (20*102) for the toxin or 1–1840 (92 * 20) for the antitoxin, codon index *n* takes on values ranging from 1–6426 (63*102) for the toxin or 1–5796 (63*92) for the antitoxin. Circles indicate random variables, grey circles represent observed random variables.

c. Description of variables, likelihood function and priors used. The likelihood function incorporates maximum entropy distributions for the observed variables, and the priors incorporate computationally tractable, vague priors for the amino acid substitution growth rates. The relative priors on the standard deviation of replicate σ_{rep_n} vs. synonymous variant σ_{syn_m} reflect our prior belief that replicate experiment noise is larger than synonymous mutant noise. σ_{b_i} and r_{scale} have improper priors.



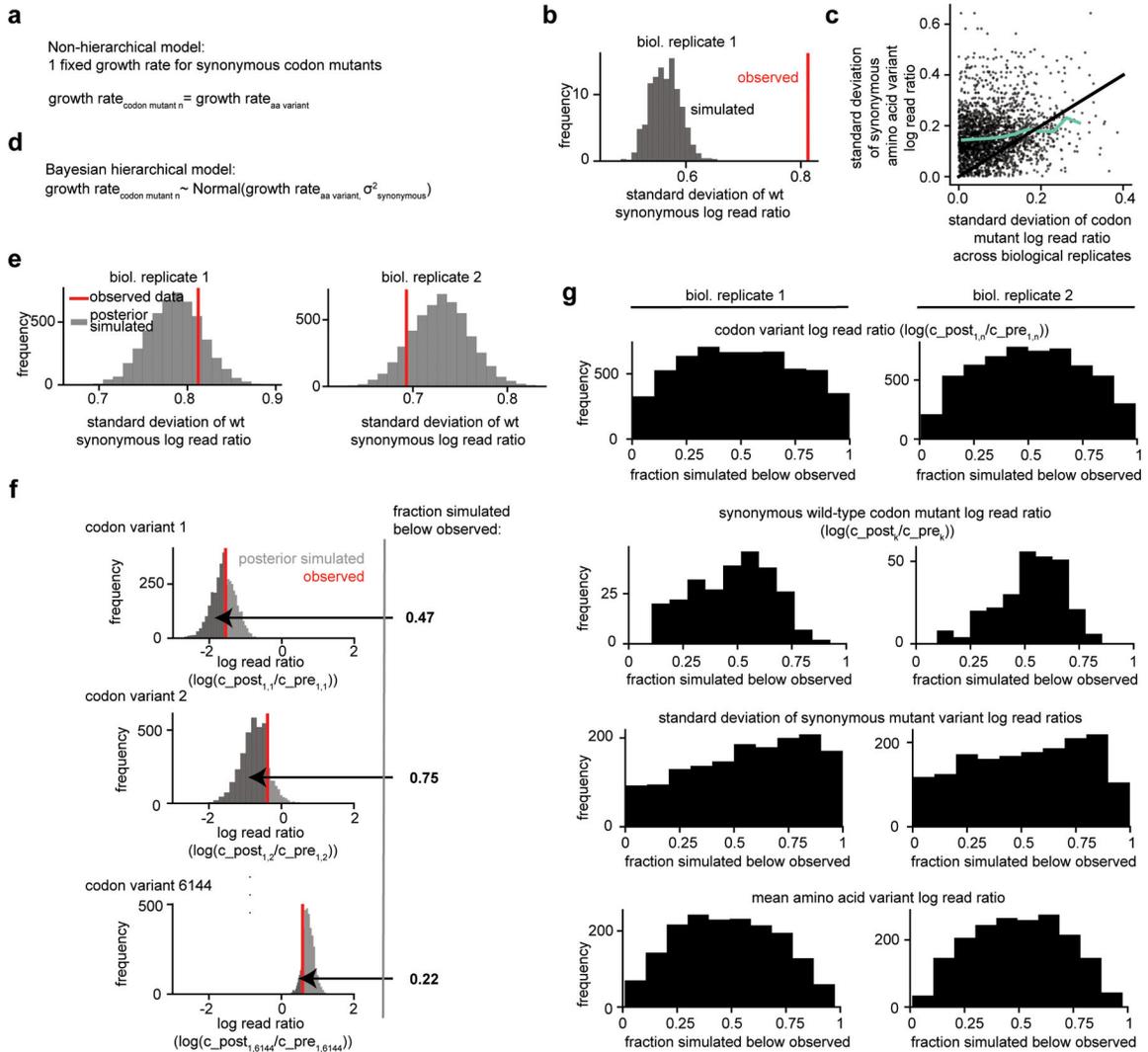
Extended Data Fig. 9. Validation of Bayesian growth rate inference on synthetic datasets.

a, Three different true synthetic growth rate distributions used for simulating pre- and post-selection codon variant read count data. Synthetic growth rate distributions were chosen from observed toxin single mutant growth rate distributions in 3 different antitoxin backgrounds, spanning the range of distributions observed.

b,c, Inferred growth rates using the Bayesian hierarchical model (b) show less bias and incorporate uncertainty estimates compared to mean log read ratio summary of pre-and

post-selection read counts (+1 pseudocount) (c). Error bars in panel b reflect the 95% highest density posterior intervals, with the measure of centre being the mean posterior growth rate.

d. Model uncertainties accurately reflect deviations of inferred true growth rates. Percentage of true synthetic amino acid growth rates falling into a certain highest density interval among all 2040 simulated toxin amino acid variants.



Extended Data Fig. 10. Posterior predictive checks show that the Bayesian hierarchical model can capture observed data statistics for both replicate experiments, whereas a non-hierarchical model cannot.

a,b, A non-hierarchical model, in which all synonymous codon variants have the same growth rate (a), cannot explain the observed data. (b) The observed standard deviation of log read ratios for synonymous wild-type toxin codon variants (red) (n=278) fall outside of the non-hierarchical model's expectations (grey).

c. The synonymous amino acid mutant standard deviations within a replicate (y-axis) are higher than codon mutant standard deviations between replicates (x-axis). Light green indicates binned average.

d. Bayesian hierarchical model allows for growth rate variation between synonymous codon mutants by drawing these from a Gaussian distribution.

e-g. Observed data statistics fall within the hierarchical Bayesian model's expected values.

(e) The observed standard deviation of synonymous wild-type toxin codon mutant log read ratios (red) fall within the model simulated values ($\text{stdev}(\log(c_{\text{post}1_k}/c_{\text{pre}1_k})$ or $\text{stdev}(\log(c_{\text{post}2_k}/c_{\text{pre}2_k})$ for biological replicate 1 or 2 respectively), see model code). Compare to panel (b) for the non-hierarchical model. (f) For each codon mutant, the hierarchical Bayesian model allows for simulating pre- and post-selection read counts ($\log(c_{\text{post}i,n}/c_{\text{pre}i,n})$, see ED Fig. 9), including log read ratios, using the posterior parameter distribution. For each codon mutant, we calculate the p-value statistic (ie. the fraction of simulated samples falling below the observed log read ratio). (g) Distribution of posterior simulated p-values for various statistics, demonstrating that no observed data statistic is biased to fall outside of the posterior simulated statistics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank members of the Laub and Marks labs, Alexandra Batchelor, Conor McClune, John Ingraham, Armin Schoech, and Ivana Cvijovic for helpful discussions. We thank Andrew Murray, Nicholas Gauthier, Tatsuo Okubo, Sam Sinai, and Nour Youssef for feedback on the manuscript, and Michael Stiffler for sharing protocols before publication. This work was supported by the Howard Hughes Medical Institute (M.T.L.), National Institutes of Health grant R01CA260415 (D.S.M.), Chan Zuckerberg Initiative CZI2018–191853 (D.S.M.), Ashford PhD fellowship (D.D.), Boehringer Ingelheim Funds PhD fellowship (D.D.), Fanny and John Hertz Fellowship (E.N.W.), National Institutes of Health NLM Training Grant T15LM007092 (A.G.G.), National Institutes of Health grant T32GM007753 (T-L.V.L.), Jane Coffin Childs Memorial Fund for Medical Research fellowship (B.W.) and National Institutes of Health grant K99GM135536 (B.W.).

References and Notes

1. Green AG et al. Large-scale discovery of protein interactions at residue resolution using co-evolution calculated from genomic sequences. *Nat. Commun* 12, 1396 (2021). [PubMed: 33654096]
2. Cong Q, Anishchenko I, Ovchinnikov S & Baker D Protein interaction networks revealed by proteome coevolution 6 (2019).
3. Ovchinnikov S, Kamisetty H & Baker D Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* 2014, 1–21 (2014).
4. Hopf TA et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* 3, e03430 (2014).
5. Kondrashov AS, Sunyaev S & Kondrashov FA Dobzhansky-Muller incompatibilities in protein evolution. *Proc. Natl. Acad. Sci* 99, 14878–14883 (2002). [PubMed: 12403824]
6. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). [PubMed: 34265844]
7. Morcos F et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U. S. A* 108, E1293–301 (2011). [PubMed: 22106262]
8. Sulkowska JI, Morcos F, Weigt M, Hwa T & Onuchic JN Genomics-aided structure prediction. *Proc. Natl. Acad. Sci* 109, 10340–10345 (2012). [PubMed: 22691493]

9. Cheng RR, Morcos F, Levine H & Onuchic JN Toward rationally redesigning bacterial two-component signaling systems using coevolutionary information. *Proc. Natl. Acad. Sci* 111, E563–E571 (2014). [PubMed: 24449878]
10. Aakre CD et al. Evolving New Protein-Protein Interaction Specificity through Promiscuous Intermediates. *Cell* 163, 594–606 (2015). [PubMed: 26478181]
11. Lite TLV et al. Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife* 9, 1–57 (2020).
12. McClune CJ, Alvarez-Buylla A, Voigt CA & Laub MT Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. *Nature* 574, 702–706 (2019). [PubMed: 31645757]
13. McMahon C et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Mol. Biol* 25, 289–296 (2018). [PubMed: 29434346]
14. Schoof M et al. An ultrapotent synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive Spike. *Science* eabe3255 (2020) doi:10.1126/science.abe3255.
15. Damen LAA et al. Construction and evaluation of an antibody phage display library targeting heparan sulfate. *Glycoconj. J* 37, 445–455 (2020). [PubMed: 32468289]
16. Zupancic JM et al. Directed evolution of potent neutralizing nanobodies against SARS-CoV-2 using CDR-swapping mutagenesis. *Cell Chem. Biol* S2451945621002646 (2021) doi:10.1016/j.chembiol.2021.05.019.
17. Aramli LA & Teschke CM Single amino acid substitutions globally suppress the folding defects of temperature-sensitive folding mutants of phage P22 coat protein. *J. Biol. Chem* 274, 22217–22224 (1999). [PubMed: 10428787]
18. Baroni TE et al. A global suppressor motif for p53 cancer mutants. *Proc. Natl. Acad. Sci. U. S. A* 101, 4930–4935 (2004). [PubMed: 15037740]
19. Berroteran RW & Hampsey M Genetic analysis of yeast Iso-1-cytochrome c structural requirements: Suppression of Gly6 replacements by an Asn52 → Ile replacement. *Arch. Biochem. Biophys* 288, 261–269 (1991). [PubMed: 1654826]
20. Bloom JD, Labthavikul ST, Otey CR & Arnold FH Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A* 103, 5869–5874 (2006). [PubMed: 16581913]
21. Bloom JD & Glassman MJ Inferring Stabilizing Mutations from Protein Phylogenies: Application to Influenza Hemagglutinin. *PLoS Comput. Biol* 5, (2009).
22. Gong LI, Suchard MA & Bloom JD Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* 2, e00631 (2013). [PubMed: 23682315]
23. Brown NG, Pennington JM, Huang W, Ayvaz T & Palzkill T Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM β -lactamases. *J. Mol. Biol* 404, 832–46 (2010). [PubMed: 20955714]
24. Fane B, Villafane R, Mitraki A & King J Identification of global suppressors for temperature-sensitive folding mutations of the P22 tailspike protein. *J. Biol. Chem* 266, 11640–11648 (1991). [PubMed: 1828803]
25. Huang W & Palzkill T A natural polymorphism in β -lactamase is a global suppressor. *Proc. Natl. Acad. Sci. U. S. A* 94, 8801–8806 (1997). [PubMed: 9238058]
26. Hudson WH et al. Distal substitutions drive divergent DNA specificity among paralogous transcription factors through subdivision of conformational space 113, 1–6 (2015).
27. Joyet P, Declerck N & Gaillardin C Hyperthermostable variants of a highly thermostable alpha-amylase. *Bio/Technology* 10, 1579–1583 (1992).
28. Marciano DC et al. Genetic and Structural Characterization of an L201P Global Suppressor Substitution in TEM-1 β -Lactamase. *J. Mol. Biol* 384, 151–164 (2008). [PubMed: 18822298]
29. McKeown AN et al. Evolution of DNA Specificity in a Transcription Factor Family Produced a New Gene Regulatory Module. *Cell* 159, 58–68 (2014). [PubMed: 25259920]
30. Poteete AR, Rennell D, Bouvier SE & Hardy LW Alteration of T4 lysozyme structure by second-site reversion of deleterious mutations. *Protein Sci* 6, 2418–2425 (1997). [PubMed: 9385644]

31. Shortle D & Lin B Genetic analysis of staphylococcal nuclease: identification of three intragenic 'global' suppressors of nuclease-minus mutations. *Genetics* 110, 539–55 (1985). [PubMed: 3896923]
32. Tsai AYM, Itoh M, Streuli M, Thai T & Saito H Isolation and characterization of temperature-sensitive and thermostable mutants of the human receptor-like protein tyrosine phosphatase LAR. *J. Biol. Chem* 266, 10534–10543 (1991). [PubMed: 1645351]
33. Yang R et al. Second-site suppressors of HIV-1 capsid mutations: Restoration of intracellular activities without correction of intrinsic capsid stability defects. *Retrovirology* 9, 1–14 (2012). [PubMed: 22214232]
34. Zheng J, Guo N & Wagner A Selection enhances protein evolvability by increasing mutational robustness and foldability. *Science* 370, (2020).
35. Ortlund EA, Bridgham JT, Redinbo MR & Thornton JW Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science* 1544, 1544–1549 (2007).
36. Starr TN, Picton LK & Thornton JW Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* 549, 409–413 (2017). [PubMed: 28902834]
37. Klein F et al. Somatic Mutations of the Immunoglobulin Framework Are Generally Required for Broad and Potent HIV-1 Neutralization. *Cell* 153, 126–138 (2013). [PubMed: 23540694]
38. Angelini A et al. Directed evolution of broadly crossreactive chemokine-blocking antibodies efficacious in arthritis. *Nat. Commun* 9, 1461 (2018). [PubMed: 29654232]
39. Madan B et al. Mutational fitness landscapes reveal genetic and structural improvement pathways for a vaccine-elicited HIV-1 broadly neutralizing antibody. *Proc. Natl. Acad. Sci* 118, e2011653118 (2021). [PubMed: 33649208]
40. Ivankov DN, Finkelstein AV & Kondrashov FA A structural perspective of compensatory evolution. *Curr. Opin. Struct. Biol* 26, 104–112 (2014). [PubMed: 24981969]
41. Starr TN & Thornton JW Epistasis in protein evolution. *Protein Sci* 25, 1204–1218 (2016). [PubMed: 26833806]
42. Diss G & Lehner B The genetic landscape of a physical interaction. *eLife* 7, 1–31 (2018).
43. Fowler DM & Fields S Deep mutational scanning: a new style of protein science. *Nat. Methods* 11, 801–807 (2014). [PubMed: 25075907]
44. Fowler DM et al. High-resolution mapping of protein sequence-function relationships. *Nat. Methods* 7, 741–746 (2010). [PubMed: 20711194]
45. McLaughlin RN, Poelwijk FJ, Raman A, Gosal WS & Ranganathan R The spatial architecture of protein function and adaptation. *Nature* 491, 138–42 (2012). [PubMed: 23041932]
46. Whitehead TA et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol* 30, 543–548 (2012). [PubMed: 22634563]
47. Otwinowski J, McCandlish DM & Plotkin JB Inferring the shape of global epistasis. *Proc. Natl. Acad. Sci. U. S. A* 115, E7550–E7558 (2018). [PubMed: 30037990]
48. Poelwijk FJ Context-Dependent Mutation Effects in Proteins. *Methods Mol. Biol* 1851, 123–134 (2019). [PubMed: 30298395]
49. Schmiedel JM & Lehner B Determining protein structures using deep mutagenesis. *Nat. Genet* 51, 1177–1186 (2019). [PubMed: 31209395]
50. Tareen A, Posfai A, Ireland WT, Mccandlish DM & Kinney JB MAVE-NN: learning genotype-phenotype maps from multiplex assays of variant effect. *bioRxiv* 1–19 (2020).
51. Atwal GS & Kinney JB Learning Quantitative Sequence–Function Relationships from Massively Parallel Experiments. *J. Stat. Phys* 162, 1203–1243 (2016).
52. Sarkisyan KS et al. Local fitness landscape of the green fluorescent protein. *Nature* 1–11 (2016) doi:10.1038/nature17995.
53. Pokusaeva VO et al. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLoS Genet* 15, 1–30 (2019).
54. Rollins NJ et al. Inferring protein 3D structure from deep mutation scans. *Nat. Genet* 51, 1170–1176 (2019). [PubMed: 31209393]
55. Poelwijk FJ, Socolich M & Ranganathan R Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nat. Commun* 10, 1–11 (2019). [PubMed: 30602773]

56. Hopf TA et al. Quantification of the effect of mutations using a global probability model of natural sequence variation 1–26 (2015).
57. Riesselman AJ, Ingraham JB & Marks DS Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* 15, 816–822 (2018). [PubMed: 30250057]
58. Hecht MH & Sauer RT Phage lambda repressor revertants. Amino acid substitutions that restore activity to mutant proteins. *J. Mol. Biol* 186, 53–63 (1985). [PubMed: 2934554]
59. Ortlund E. a. Crystal Structure of an Ancient. *Science* 1544, 1544–1549 (2008).
60. Russ WP et al. An evolution-based model for designing chorismate mutase enzymes. *Science* 369, 440–445 (2020). [PubMed: 32703877]
61. Jiang X-L, Dimas RP, Chan CTY & Morcos F Coevolutionary methods enable robust design of modular repressors by reestablishing intra-protein interactions. *Nat. Commun* 12, 5592 (2021). [PubMed: 34552074]
62. Mutalik VK et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* 10, 354–360 (2013). [PubMed: 23474465]
63. McClune CJ, Alvarez-Buylla A, Voigt CA & Laub MT Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. *Nature* 574, 702–706 (2019). [PubMed: 31645757]
64. Khlebnikov A, Datsenko KA, Skaug T, Wanner BL & Keasling JD Homogeneous expression of the PBAD promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity *araE* transporter. *Microbiology* 147, 3241–3247 (2001). [PubMed: 11739756]
65. Stiffler MA, Subramanian SK, Salinas VH & Ranganathan R A protocol for functional assessment of whole-protein saturation mutagenesis libraries utilizing high-throughput sequencing. *J. Vis. Exp* 2016, 1–11 (2016).
66. Warren DJ Preparation of highly efficient electrocompetent *Escherichia coli* using glycerol/mannitol density step centrifugation. *Anal. Biochem* 413, 206–207 (2011). [PubMed: 21362398]
67. Magoc T & Salzberg SL FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963 (2011). [PubMed: 21903629]
68. Rognes T, Flouri T, Nichols B, Quince C & Mahé F VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584 (2016). [PubMed: 27781170]
69. Bloom JD Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics* 16, 168 (2015). [PubMed: 25990960]
70. Bank C, Hietpas RT, Wong A, Bolon DN & Jensen JD A Bayesian MCMC Approach to Assess the Complete Distribution of Fitness Effects of New Mutations: Uncovering the Potential for Adaptive Walks in Challenging Environments. *Genetics* 196, 841–852 (2014). [PubMed: 24398421]
71. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2.26 (2021).
72. Riddell A, Hartikainen A & Carter M PyStan (3.0.0) (2021).
73. Lite TLV et al. Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife* 9, 1–57 (2020).
74. Kingma D & Ba J Adam: A Method for Stochastic Optimization *Int. Conf. Learn. Represent.* (2014).
75. Abadi M et al. Tensorflow: A system for large-scale machine learning. in 12th USENIX Symposium on Operating Systems Design and Implementation 265–283 (2016).
76. Seabold S & Perktold J statsmodels: Econometric and statistical modeling with python. in 9th Python in Science Conference (2010).
77. HMMER, <http://hmmer.org/>.
78. Hagberg AA, Schult DA & Swart PJ Exploring Network Structure, Dynamics, and Function using NetworkX. in Proceedings of the 7th Python in Science Conference (eds. Varoquaux G, Vaught T & Millman J) 11–15 (2008).
79. Tarean A & Kinney JB Logomaker: Beautiful sequence logos in Python. *Bioinformatics* 36, 2272–2274 (2020). [PubMed: 31821414]

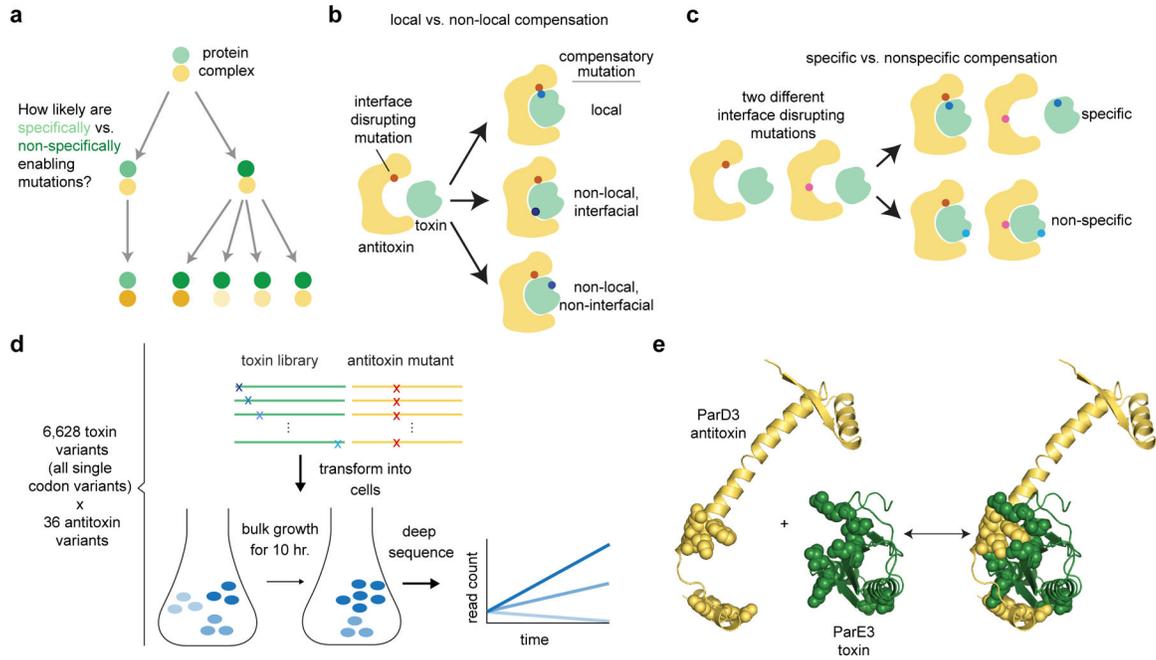


Figure 1: Comprehensive identification of neutral and enabling mutations for the toxin-antitoxin system ParE3-ParD3.

a, Schematic of possible future mutational trajectories enabled by specific or non-specific neutral mutations. Yellow and green circles represent interacting proteins, shaded circles indicate mutated proteins, and arrows represent single mutations to sequences that retain binding. Specific mutations allow only particular subsequent mutations (left), whereas non-specific mutations enable tolerating many different subsequent mutations in the partner protein (right).

b-c, Schematic examples of local vs. non-local (b) and specific vs. non-specific (c) compensatory mutations that rescue interface-disrupting mutations. Dots represent mutations.

d, Schematic summary of experimental pipeline for identifying enabling mutations across the antitoxin-toxin interface. A library of all possible toxin single mutants is transformed into cells with a given, interface-disrupting mutation in the antitoxin (top). Cells are then grown in bulk and the abundance of each toxin variant over time is measured by sequencing. These changes are used to infer growth rates. Dots represent mutations.

e, A single chain of the ParD3 antitoxin (yellow) in complex with the ParE3 toxin (green), from PDB:5CEG (right), with isolated antitoxin (left) and toxin (middle). The top 10 covarying positions are spacefilled.

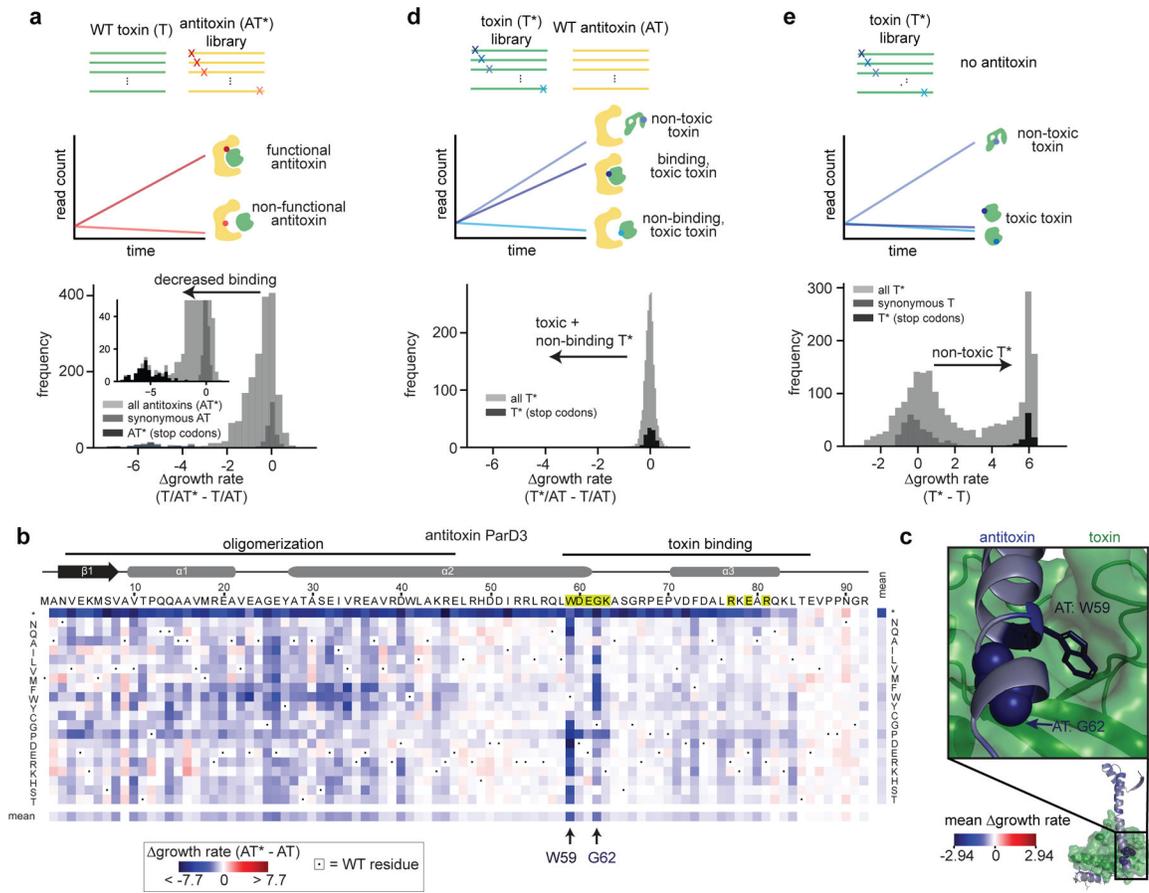


Figure 2: Deep mutational scanning reveals mutational tolerance and interface-disrupting substitutions in ParE3-ParD3.

a, Comprehensive single mutant scan of all possible antitoxin variants. The histograms indicate the change in growth rate for each mutant (with a blow-up histogram inset), relative to the wild-type antitoxin or toxin, with the greyscale-coded categories indicated.

b, Heatmap of growth-rate values for each possible antitoxin single mutant showing mutations that disrupt toxin neutralization (blue). The antitoxin-antitoxin oligomerization and antitoxin-toxin binding regions are indicated above. Top 8 positions that covary with positions in the toxin are shaded in yellow on the primary sequence. The substituted residue (or stop codon indicated with *) is listed on the far left and far right. Mean effects for each row and column are shown below and to the right.

c, Structure of ParD3-ParE3 (PDB ID: 5CEG) highlighting antitoxin residue W59 with its pocket in the toxin and antitoxin residue G62 where single substitutions disrupt the interaction most. The mean effect at each position (see panel b) in ParD3 is color-coded, as indicated, on the structure.

d,e, Same as (a) but for all toxin variants in presence of wild-type antitoxin (d) or absence of antitoxin (e).

Experiments performed using 1.2×10^{-4} % w/v [arabinose], and 10 mM [IPTG] in the arabinose-titratable strain BW27783.

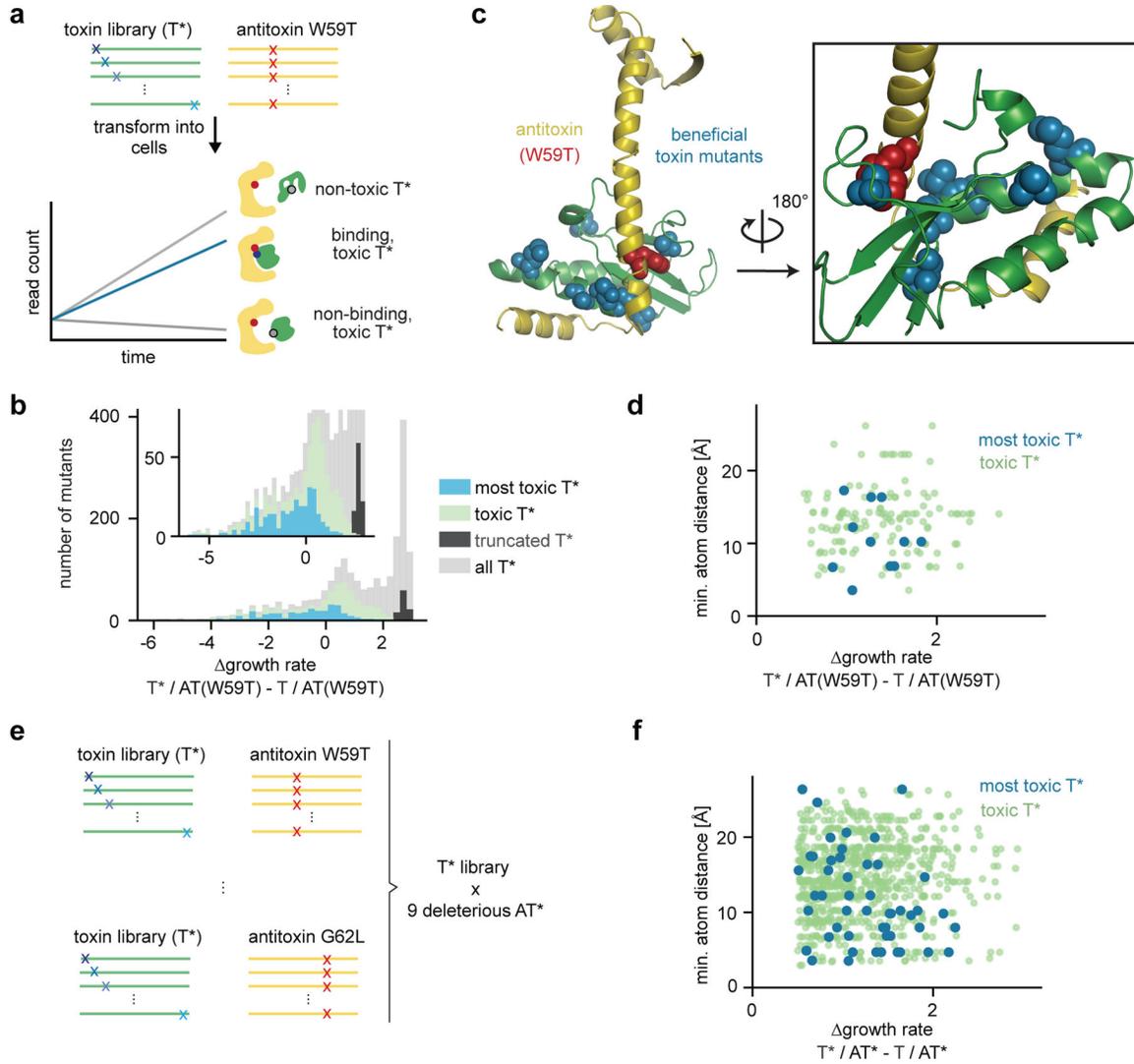


Figure 3. Beneficial, interaction-restoring mutations can be far from the deleterious mutation they rescue.

a, Schematic overview of ‘suppressor scanning’. Cells expressing antitoxin ParD3(W59T) and a library of all possible toxin single substitutions were grown and analyzed as in Fig. 2c to identify toxin variants (T*) that can rescue the growth defect of ParD3(W59T).

b, Distribution of growth rates for each toxin variant (T*) relative to the wild-type toxin (T) when co-expressed with antitoxin ParD3(W59T) reveals toxin variants alleviating the growth defect of the antitoxin W59T mutation (with a blow-up inset). Various categories of toxin variants are color-coded as indicated (right), including toxin variants that maintain toxicity at different thresholds (blue: 310 most toxic toxin variants, green: 781 toxic toxin variants).

c, The significantly beneficial toxin variants for the deleterious antitoxin W59T (blue spacefilled, from set of most toxic toxin variants) are distributed across the toxin in the ParD3-ParE3 structure (PDB ID: 5CEG). Red indicates the deleterious antitoxin residue W59.

d, Plot of distance between W59 in ParD3 and each significantly beneficial toxin mutant from the set of toxic (green) or most toxic (blue) toxin variants vs. effect size of rescue.

e, Schematic indicating that all toxin single mutants were screened against 9 deleterious antitoxin mutants (G62L/D/Y, W59A/L/T/V, K63D,F73K).

f, Same as (d) but for significantly beneficial toxins from all 9 suppressor scans.

Experiments performed using 1.2×10^{-4} % w/v [arabinose], and 10mM [IPTG] in the arabinose-titratable strain BW27783.

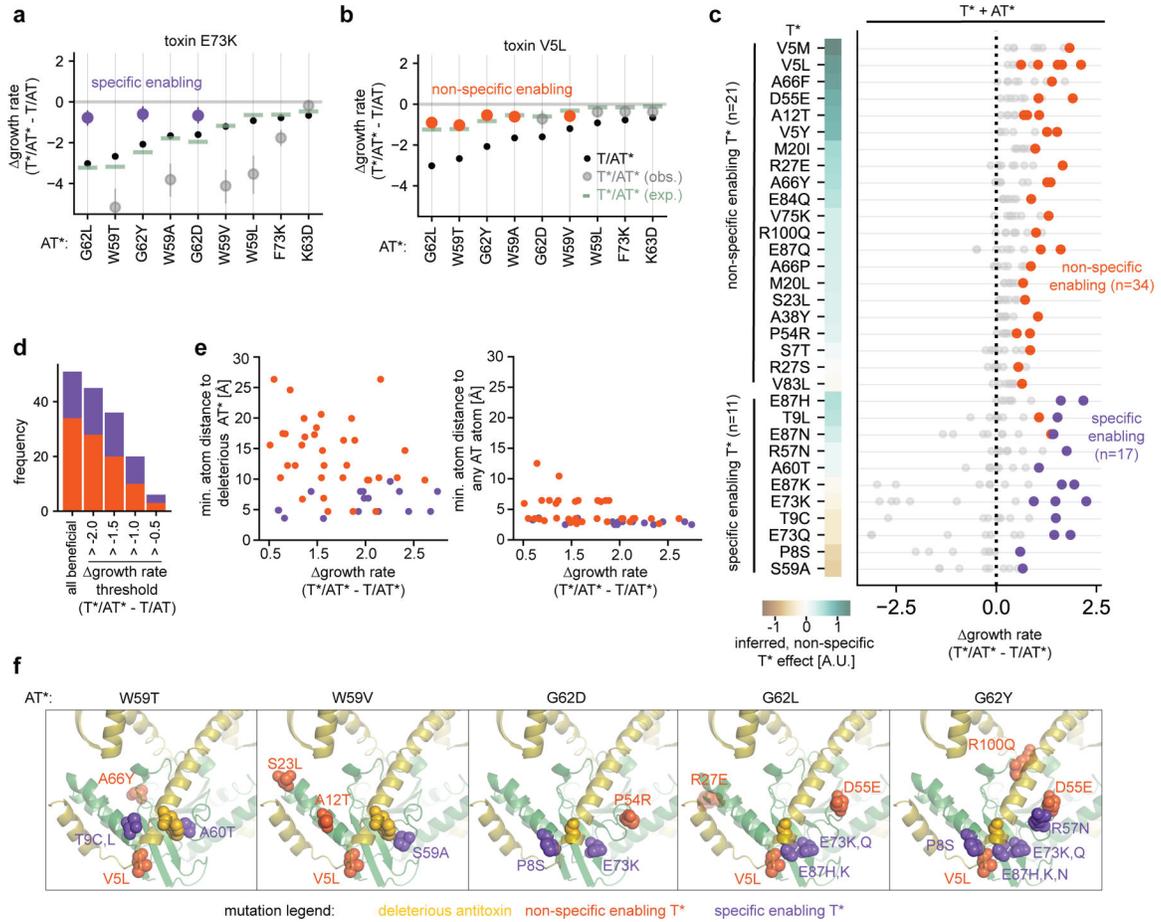


Figure 4. Non-specific enabling mutations outnumber specific mutations, and can be far from the deleterious mutation as well as the interface.

a-b, For a specifically enabling toxin mutation E73K (a) or non-specifically enabling toxin mutation V5L (b), the growth-rate relative to the wild-type pair T/AT (Δ growth rate) is shown when combined with each antitoxin variant indicated on the x-axis (large dots represent mean posterior T^*/AT^* growth rate; error bars indicate 95% posterior highest density interval). The mean posterior Δ growth rate for each AT* combined with wild-type T is shown (small black dots) along with the Δ growth rate for T^*/AT^* expected under the independent, nonlinear model (green lines). Purple and orange indicate T^*/AT^* pairs (T^*/AT^*) where the toxin substitution is specific or non-specific, respectively.

c, For each toxin variant (n=32) beneficial to at least one deleterious antitoxin mutant (51 pairs), the plot shows their effect size of rescue (Δ growth rate) combined with each of the 9 deleterious AT*. The effect of each T* inferred by the nonlinear model is indicated on the heatmap.

d, The number of non-specific and specific pairs of rescue at different Δ growth rate thresholds relative to the wild-type T/AT.

e, Minimum atom distance of rescuing toxin mutation to deleterious antitoxin mutation (left) or to any antitoxin residue (right) vs. effect size of rescue (Δ growth rate). For color codes of toxin-antitoxin double mutant pairs, see panels (a)-(b).

f, Specific (purple) and non-specific (orange) enabling toxin mutations in each antitoxin mutation background they rescue are shown (spacefilling) on the ParD3-ParE3 structure (PDB ID: 5CEG). Antitoxin is yellow; toxin is green.

Experiments performed using 1.2×10^{-4} % w/v [arabinose], and 10 mM [IPTG] in the arabinose-titratable strain BW27783.

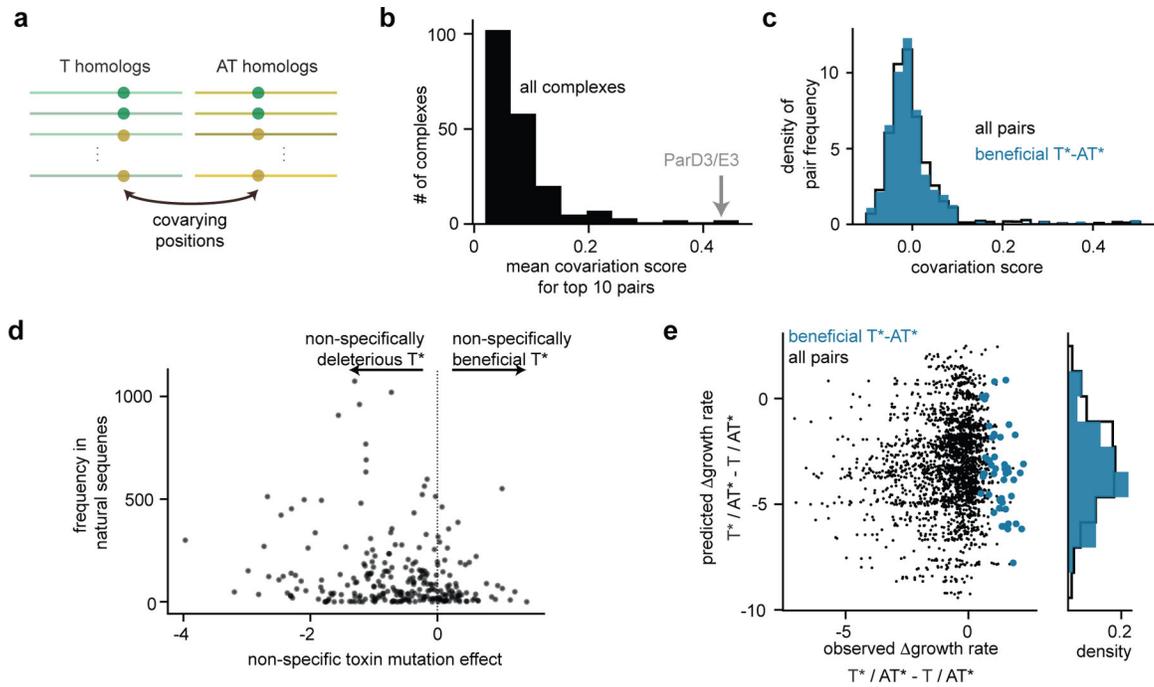


Figure 5. Natural sequences and models trained on these provide insufficient information to predict enabling mutations.

a, Schematic of identifying covarying pairs of residues in natural sequences.

b, Mean covariation score of top 10 pairs of residues in ParD/ParE homologs (grey) compared to ~350 complexes with covariation signal (black).

c, Distribution of covariation scores for beneficial pairs of residues (blue, see Fig. 2e) is similar to the null distribution for all possible pairs of positions (black).

d, Amino-acid mutant frequency in toxin homologs at a particular site is not correlated with the non-specific enabling effect of that amino acid variant, as inferred from suppressor scanning.

e, Predicted vs. observed rescue (Δ growth rate of each double mutant (T^*/AT^*) relative to the antitoxin single mutant effect (T/AT^*)). Predictions are made using EVmutation. Blue indicates observed significant rescue.

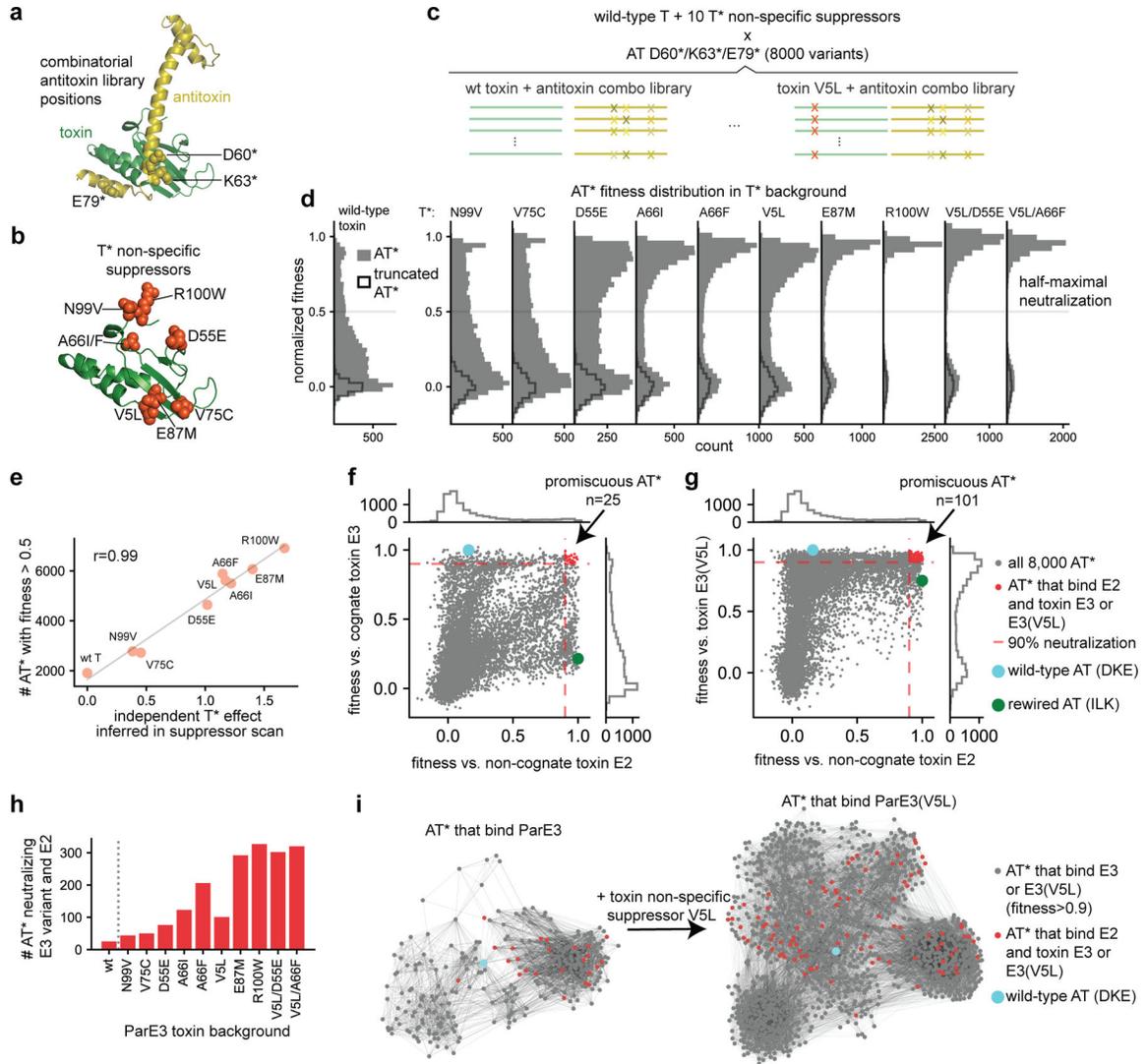


Figure 6. Non-specifically enabling mutations expand mutational paths to maintain old and evolve new interactions.

a-c, For the antitoxin ParD3, a library of all 8,000 possible variant combinations at three key specificity-determining residues (shown on the structure of ParD3-ParE3 (PDB ID: 5CEG) in (a)) was screened for binding to wild-type toxin ParE3 or ParE3 harboring non-specifically enabling variants (purple) shown in (b). Schematic of the experiment is shown in (c).

d, The fitness distribution of 8,000 antitoxin variants screened against the wild-type toxin ParE3 (left) or 10 different variants of ParE3 (right) reveals that non-specifically enabling mutations in the toxin allow binding to more combinatorial antitoxin variants than the wild-type toxin.

e, Correlation of inferred independent effect from suppressor scan (see Fig. 4c) with the number of combinatorial antitoxin variants neutralized above half-maximal for each toxin single variant.

f,g, Scatterplots showing fitness values for 8,000 antitoxin variants (dots) screened against the wild-type toxin ParE3 (y-axis) and the non-cognate toxin ParE2 (x-axis) (f), or against

toxin ParE3 with a non-specifically enabling mutation (ParE3(V5L)) (y-axis) and ParE2 (x-axis) (g). Various classes of AT* are color-coded as shown on the right, including the AT* variants that have gained the ability to bind the non-cognate ParE2 (red), as well as the rewired antitoxin (green) that binds the non-cognate toxin ParE2 but not the cognate ParE3. h, Differences in the number of promiscuous antitoxin variants that neutralize both the non-cognate toxin ParE2 as well as the wild-type type toxin ParE3 (left) or a toxin variant harboring a non-specific suppressor indicated on the x-axis (fitness > 0.9). i, Force-directed graph of antitoxin variants that bind toxin ParE3 with (left) or without (right) the non-specifically enabling mutation V5L. Nodes represent individual antitoxin sequences (fitness > 0.9), edges correspond to single mutational steps. Color codes as in panels f-g.