Co-evolution of interacting proteins through non-contacting and non-specific mutations

David Ding1,2, Anna G. Green2,3, Boyuan Wang4, Thuy-Lan Vo Lite5, Eli N. Weinstein6, Debra S. Marks2 and Michael T. Laub1,7,✉

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 homologues occurring primarily between specific pairs of contacting residues across the interface. In addition, the enabling mutations we identified expand future mutational paths for both maintain old toxin–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.

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 interactions12–14 by identifying specific pairs of residues in close proximity9–12. Additionally, potent binding proteins have been engineered by mutating key interface residues15–16. 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 positions17–18 and of mutations away from the interface regions of antibodies that affect antigen affinity and specificity19–20.

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 enable19–21. Additionally, whether neutral, enabling mutations affect the evolvability of a protein—it’s 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 co-evolution, we examined a bacterial toxin–antitoxin system. The Mesorhizobium opportunatum antitoxin ParD3 normally binds and restrains the activity of its cognate, co-operonic toxin ParE3 (ref. 10). 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 read-out of the ParD3–ParE3 interaction in vivo (Fig. 1d). Previous analyses of amino acid covariation in ParD–ParE homologues identified residues in each protein that most strongly covary, map to the protein–protein interface and can dictate interaction specificity of paralogues9,21 (Fig. 1e). To experimentally probe the possible mutational trajectories for ParD3–ParE3, we first used deep mutational scanning3–6 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 most such enabling mutations in the toxin can restore binding to many of the otherwise disruptive single mutations.
Fig. 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 versus non-local (b) and specific versus 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.

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 harbouring 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 h, which is about six generations for cells harbouring 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 (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 Δgrowth rate—the difference in number of doublings of the mutant, AT*/
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 Δgrowth 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 amino-terminal half of the antitoxin, which mediates oligomerization and probably promotes overall stability of the ParD3–ParE3 complex but does not directly contact the toxin (Extended Data Fig. 1d). The second region involves the carboxy-terminal end of α2 and α3, the elements of ParD3 that severely disrupt the toxin–antitoxin interface.

To examine the position-wise effect of amino acid substitutions, we generated a 21 × 92 heatmap showing the Δgrowth rate value for each possible substitution at each position in ParD3 (Fig. 2b). This map revealed an apparent periodicity within α-helices with 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.

For variants Δ antitoxin showing mutations that disrupt toxin neutralization (blue). The antitoxin–antitoxin oligomerization and antitoxin–toxin binding regions are indicated above. Top eight 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. WT, wild type.

Fig. 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 (WT) antitoxin or toxin, with the grey scale-coded categories indicated. b, Heatmap of Δ 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 eight 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. WT, wild type. c, Structure of ParD3–ParE3 (PDB 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 (b) in ParD3 is colour-coded, as indicated, on the structure. d, 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.
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 Å minimum atom distance) in the toxin as measured by EVcouplings (Extended Data Fig. 1e,f) and are critical for antitoxin-binding specificity. For position 59, substituting tryptophan with another aromatic residue or some hydrophobic residues had relatively modest effects, with all others leading to substantial defects (∆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 α2 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 ∆growth rate values in the presence of wild-type antitoxin (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 and Extended Data Fig. 2). In this case, unfolded or non-functional 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 harbouring a stop codon, which are presumably non-functional, exhibited high ∆growth rate values (>5) indicating substantially 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 seven 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 (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 four fully inhibiting induction conditions (Extended Data Fig. 2e).

Substitutions that ablated ParE3 toxicity (and hence have high ∆growth rate values) were particularly pronounced in β2, β3 and α3, 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 cocrystral structure, most 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). Together, our results indicate that the interface residues of the ParE3 toxin are highly robust to mutations. This robustness does not result from over-expression 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 (1) strongly covary with positions in the toxin and (2) involve amino acids commonly found at those positions in ParD antitoxin homologues (Methods). We examined each antitoxin mutant × toxin library (36 × 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 (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 (Figs. 3a and 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 harbouring the ParD3(W59T) mutant (P < 0.0001 and ∆ growth rate values > 0.5; Methods; Fig. 3b and 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 cocrystral structure (Fig. 3c), finding that they were distributed throughout the toxin and many were >10 Å away from the W59 residue in the antitoxin (Fig. 3d).

We repeated the same analysis for the other eight deleterious antitoxin mutations at the higher antitoxin concentration (Fig. 3e and 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; that is, 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, 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 non-linearity (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. 4c and Extended Data Fig. 5) and as non-specific if the observed growth rate was close to the expectation (Fig. 4b,c and Extended Data Fig. 5). We then called a particular toxin mutation 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 and 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.

**Fig. 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 analysed 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 colour-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 space-filled, from set of most toxic toxin variants) are distributed across the toxin in the ParD3–ParE3 structure (PDB 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 versus effect size of rescue. e, Schematic indicating that all toxin single mutants were screened against nine deleterious antitoxin mutants (G62L/D/Y, W59A/L/T/V, K63D and F73K). f, Same as d but for significantly beneficial toxins from all nine suppressor scans. Experiments performed using $1.2 \times 10^{-4}$% w/v (arabinose) and 10 mM (IPTG) in the arabinose-titratable strain BW27783.
Fig. 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 (Δgrowth rate) is shown when combined with each antitoxin variant indicated on the x axis (large dots represent mean posterior T*/AT* Δgrowth rate; error bars indicate 95% posterior highest density interval). The mean posterior Δgrowth rate for each AT* combined with wild-type T is shown (small black dots) along with the Δgrowth rate for T*/AT* expected under the independent, non-linear model (green lines). Purple and orange indicate T*/AT* pairs (T*/AT* − 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 AT*, the growth rate relative to the wild-type pair T*/AT* expected under the independent, non-linear model (green lines). Purple and orange indicate T*/AT* pairs (T*/AT* − T/AT) where the toxin substitution is specific or non-specific, respectively. For each toxin variant (n = 32) beneficial to at least one deleterious AT*, the growth rate relative to the wild-type pair T*/AT* expected under the independent, non-linear model (green lines). Purple and orange indicate T*/AT* pairs (T*/AT* − T/AT) where the toxin substitution is specific or non-specific, respectively. d–f, For each toxin variant (n = 32) beneficial to at least one deleterious AT*, the growth rate relative to the wild-type pair T*/AT* expected under the independent, non-linear model (green lines). Purple and orange indicate T*/AT* pairs (T*/AT* − T/AT) where the toxin substitution is specific or non-specific, respectively. For each toxin variant (n = 32) beneficial to at least one deleterious AT*, the growth rate relative to the wild-type pair T*/AT* expected under the independent, non-linear model (green lines). Purple and orange indicate T*/AT* pairs (T*/AT* − T/AT) where the toxin substitution is specific or non-specific, respectively.
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 homologues of ParD3–ParE3, which show strong covariation scores, even compared to other complexes (Figs. 1e and 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 non-linear 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 (EVmutation and variational auto-encoder) could not predict which toxin mutations would alleviate the growth defect of a particular antitoxin mutation, with no correlation between the measured Δgrowth rate values and the scores produced by these models (Fig. 5e and Extended Data Fig. 6). Notably, the homologues 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 three interface positions in the ParD3 antitoxin critical to binding specificity (D60, K63 and E79) (Fig. 6a). We picked eight 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 ten 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 and 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, <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 (refs. 16,17), 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 four-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 co-evolve 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 may 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 co-evolution driven mainly by directly contacting specifically compensating pairs of mutations as it involves distant residues and because it does not involve 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 proteins15–18,23,24 and some between proteins19–23, it had been unclear in the absence of systematic studies how likely non-specific versus specific suppressors are to arise and whether each class of suppressors would map to directly contacting pairs of residues or not24.

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 favourable interaction (Supplementary Table 1 and Supplementary Data), 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 proteins25–28. The non-specific toxin mutations probably 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 antibodies29,30. Notably, the non-specific suppressors we identified were relatively rare and not predictable based on naturally occurring homologues of ParE and ParD, which exhibit strong covariation. Models based on natural sequence homologies and covariation are powerfully able to predict protein and protein complex structures31–35 and can help in protein design36,37. 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 homologues 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.

Methods

Bacterial strains, vectors and media. All strains used are listed in Supplementary Table 2. E. coli strains were grown at 37°C in M9 liquid medium (1× 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−1 of carbenicillin, 20 μg ml−1 of chloramphenicol in liquid media and 100 μg ml−1 of carbenicillin, 30 μg ml−1 in agar plates. The toxins (ParE3 or ParE2) were carried as before on the pBAD33 vector (chlorR marker, M33482 for wild-type ParE3, M33030 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 pEXET20 vector (carbR marker, M34833) with expression induced by isopropyl-β-D-thiogalactopyranoside (IPTG). Toxin and antitoxin libraries containing all possible single mutants were created under a bicistron ribosome binding site 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 site19,20. Consistent with this desired effect, we confirmed that systematic variants throughout the toxin and antitoxin are fully tolerated.

For the single-mutant and suppressor scan, we used the arabino-β-tiratitable wild-type E. coli strain BW27783 (ref. 7). For the combinatorial antitoxin library experiments, we used the previously optimized TOP10 E. coli background41.

Library construction. The toxin and antitoxin single-mutant libraries were each constructed using a previously described two-step overlap-extension PCR protocol3. 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; Supplementary Table 3). The reverse mutagenesis primer was used with the primer DDP115 specific to the 5′ end of parE3 and the forward mutagenesis primer was used with the primer DDP116 specific to the 3′ end of parE3 (PCR cycling was: 30 s at 98°C, 20 cycles of 10 s at 98°C, 20 s 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 primer pairs using the same thermocycling protocol to create full-length PCR products 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, 1 M 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 pEXET20-parD3 as the template. We then amplified the pBAD33-parE3 and pEXET20-parD3 vectors using primers DDP508 and S09 or DDP540 and S41 (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 then digested with HindIII-HF and SacI-HF kit (Select-a-Size DNA Clean & Concentrator with >150 base pair (bp) size cutoff to manufacturer (Zymo Research) specifications) and then ligated into the amplified vectors using T4 DNA ligase (NEB) at 16 °C for 16 h with a 1.3 molar ratio of insert to vector with 50 ng of vector per 20 μl ligation reactions. These ligation reactions were scaled on the basis of downstream needs. Ligationss were dialyzed on Millipore VSWP 0.025 μm membrane filters for 90 min 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 previously10) harboring the wild-type antitoxin on pBAD33 with the toxin single-mutant ligation reactions, recovered in 1 ml of Super Optimal broth with Catabolite repression (SOC) for 1 h 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% v/v glucose) to check transformation efficiency (all libraries were composed of >90% of 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 optical density OD_{560}~0.3, at which point glucose was removed by washing four times with M9L (spinning cells down at 8,000g 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. Twenty-nine of the 36 toxin substitutions are found at positions that are within the top 15 toxin–antitoxin covarying parirs of residues with the toxin, five of the 36 are found within the top seven toxin–antitoxin covarying pairs of positions (E26, I33, W41, V17 and K44) and two did not fall in either of these categories (A16K and 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 backgounds (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, we cloned the toxin mutant library into separate vectors containing one of 15 different four-nucleotide barcodes just 3′ to the toxin gene and restriction site. Barcoded vector backbones of pBAD33 were generated using primers (reverse primer DDP508 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 three nucleotides different from any other barcode (Supplementary Table 3). These barcoded vector backbones were then separately ligated with the barrary insert and prepared as described above. We transformed the DNA ligation reactions separately into TOP10 cells, grew them to OD_{560}~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-plasmids and kept track of each barcoded toxin library in replicate into electrocompetent BW27783 cells containing wild-type toxin (8 × 10^{−9}% arabinose, 17.5 mM IPTG; Extended Data Fig. 3d). For the toxin single-mutant scan done in the absence of wild-type antitoxin, we performed the growth measurement at seven different arabinose induction levels (0.2, 5.3 × 10^{−5}, 8 × 10^{−4}, 1.2 × 10^{−3}, 8 × 10^{−3}, 4 × 10^{−2} and 2 × 10^{−2}% arabinose) for which the highest four reach full growth inhibition (Extended Data Fig. 2c, 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 μM IPTG induction in the TOP10 strain background.

Sample sequencing and preparation. For each replicate, samples taken at 0 and 600 min after toxin induction were minipreped (Zymo Research Plasmid Miniprep Kit). We then performed a high-input (400 ng of 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 adaptor sequences, sequencing primer homology regions and Illumina multiplex indices for each sample (forward primer DDP5643 and reverse primers DDP178–193 + DDP569–580 for the toxin in single-mutant and suppressor screens, forward primer DDP554 and reverse primers DDP65–568 for antitoxin single-mutant library, forward primer DDP464–645 with reverse primers DDP464–693 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 to allow for Illumina cluster definition and stagger our homopolymer-like amplicons.

We gel purified amplicons of interest (~500 nucleotides) by running samples with a loading dye for 30 min at 180 V on a Novex 8% TBE gel. We sheared the excised gel band by spinning it through a bottom-tered 0.5 ml tube placed within a 1.5 ml tube, added 300 μl of 10 mM Tris buffer (pH 8) and then froze the sample at −20 °C for 15 min followed by incubation at 70 °C for 10 min to solvate the DNA from the gel. We spun each sample through a 0.22 μm spin-on culture acetate column to separate the gel from the supernatant and then performed an isopropanol precipitation of the DNA by adding 32 μl of 5 M NaCl, 2 μl of glycolde and 550 μl of 100% isopropanol. The mixture was then chilled at −80 °C to centrifuged at 4 °C at 14,000 g. Finally, the sample was washed with ice-cold 70% ethanol, air-dried and resuspended in 10 μl of water.

Each sample was then run on a fragmental analyser and quantitative PCR 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 v1.2.11 (ref. 1). Merged reads were quality filtered on the basis of their phred-score using vsearch v2.13.0 (ref. 7), with the following arguments: vsearch --fastq_filter [0] --fastq_truncqual 20 --fastq_maxlen 80 --fastq_minlen 50 --fastq_maxee 0.5 --fastq_ascii 33 --fastq_maxn 5 --fastq_maxq 20 --fastq_maxh 0.5. Venn diagrams were generated using the Venny online tool (http://bioinfogp.cnb.csic.es/tools/venny/).

Toxin mutant reads were subsequently split into separate files on the basis of their four-nucleotide barcode indicating the antitoxin background, if applicable. Subsequently, reads were filtered for those that (1) spanned the full-length of the toxin or antitoxin gene and (2) had no mutations or indels in the immediately flanking 10 bp upstream (which includes the restriction site, ribosome binding site and the stop codon for the upstream bicistronic peptide), as well as the
Hierarchical Bayesian inference of mutant growth rates. We used a Bayesian model (Extended Data Fig. 8) that allowed us to generate 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 data68–69 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 \( \lambda \). The postselection Poisson rate parameter is the preselection Poisson rate multiplied by the exponentiated growth rate for each codon mutant (following exponential growth rates of the form \( N(t) = N(0) \times e^{growth \ rate \ \times \ t} \)). We found that we 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, we extended our model from log-read rate (that was the only way to infer growth rate) to 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 Stan70 to perform inference (using two Markov Chain Monte Carlo chains, 10,000 steps each, discarding the first half of Markov Chain Monte Carlo 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 postselection read counts for each codon variant) by assuming three different distributions of true amino acid mutant growth rates (from the range of antibiotic single-mutant growth rate distributions). We compared the Bayesian inference to the classically used log-read ratio (\( +1 \) pseudocount) and found that our model was more robust to noise introduced in the raw-read ratio data and were 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 postselection reads a larger 95% highest posterior density interval.

Calibrated uncertainty. Our Bayesian model allowed us to calculate the associated uncertainty; that is, the posterior distribution for each amino acid variant. We compared differing uncertainty intervals for each variant (95%, 90%, 80% and 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 toxic mutants) to compare 10,000 simulated growth replicate datasets generated from the model to the true observed data and for each quantity calculated their posterior predictive \( P \) value (the fraction of simulated data above the observed data along the test statistic). On the basis of these test quantities, the model developed plausibly generates the observed data, demonstrating sufficient complexity.

Calling significantly beneficial toxic mutants. Toxic mutants were called as significantly beneficial to a given antibiotic mixture if they grew at least 0.3-log, fold better than the antibiotic mutant combined with the wild-type toxic measured in the same flask, with all 10,000 posterior samples exceeding the growth rate of the antibiotic mutant with the wild-type toxic; that is, the 99.99% highest posterior density interval does not overlap zero difference in growth rate. For the 'high' antitoxin expression condition (main text), we sought beneficial toxins in nine deleterious antitoxin backgrounds (E73K, K65D, 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 and E79K).

We verified that the toxins harboring beneficial mutations were still as, or nearly as, toxic as the wild-type toxic by measuring the growth rate of the toxic single-mutant library at seven different toxic induction levels as indicated above (Extended Data Fig. 2). For the 'high' antitoxin expression condition, we required the double-mutant growth rate to be at least twofold greater than or less than the independent expectation, with a 99.99% highest posterior density interval that did not overlap the independent expectation (that is, 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 antibiotic single-mutant growth rate but deviate because their independent expectations are highly detrimental.

In contrast to previous studies46,49, which have also successfully used non-linear, independent models to quantify double-mutant interactions, we demonstrated that the inferred, unobserved mutation effects in our non-linear 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 toxin in the toxin mutagenesis toxicity, we measured the growth rate of cells containing these toxin mutations in the absence of the antitoxin in arabinose-titratable strain at both full (1.2 × 10⁴ % arabinose) and half-maximal induction levels (6 × 10³ % arabinose). We diluted saturated overnight cultures 1:100 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 carbendazim and chloramphenicol. We ran a maximum of eight samples per 96-well plate, such that each sample is measured at least ten 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 and 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 r.p.m. at 37 °C, with 180 μl of media and 70 μl of mineral oil on top to prevent evaporation. Assayed toxin and antitoxin variants (T* + AT*) are: T + AT(G62L), T + AT(W59T), T + AT(73K), 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 and 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 for 2–3 h at 37 °C to OD₆₀₀ ~0.5, then washed four times with M9L. We then diluted these cells 1:200 into 96-well plates containing M9L/carb/chlor/10 μM IPTG/1.2 × 10⁻⁶ % arabinose to measure growth rates. We calculated the growth rate as the normalized fold change in OD₆₀₀ across time. Error bars in Extended Data Fig. 6 are calculated using the error propagation formula for independent variables and a first-order Taylor expansion.

Evolutionary analysis. We performed coevolution analysis similar to before. Briefly, we generated JackHMMR (http://hmmer.org/) 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 on the basis of their genome distance (<1,000 nucleotides). We selected the bitscore cutoff with the highest number of true positive (minimum amino domain <0 Å) between protein-covarying pairs, resulting in the bitscore choice of 0.3. Alignment quality filtering was done similar to previous studies, calculating coevolution 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 (19 amino acids) and down-weighting sequences if their sequence identity exceeded 80%. The final alignment contained 1,650 concatenated sequences, with an effective number of sequences of 1,088 after down-weighting.

We chose to highlight the top ten covarying residues from this analysis, which were all close (<6 Å minimum atom distance) in distance. Using previous calibration sets, the 90% precision cutoff (pairs of residues <6 Å distance/all selected pairs) in covariation score resulted in 29 covarying pairs of residues between protein, of which 8 are indeed within 6 Å 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 ref. 1.

Network visualizations. Force-directed graphs were constructed using the Python package NetworkX. Network clusters were defined using Louvain clustering.

Extended data

The following references are cited in the text:

1. Green, A. G. et al. Large-scale discovery of protein interactions at residue resolution using co-evolution calculated from genomic sequences. Nat. Commun. 12, 1396 (2021).
2. Cong, Q., Anishchenko, L., Ovchinnikov, S. & Baker, D. Protein interaction networks revealed by proteome coevolution. Science 365, 185–189 (2019).
3. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. eLife 3, e02030 (2014).
4. Hopf, T. A. et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. eLife 3, e03430 (2014).
5. Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky-Muller incompatibilities in protein evolution. Proc. Natl Acad. Sci. USA 99, 14878–14883 (2002).
6. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
7. Morcos, E. et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc. Natl Acad. Sci. USA 108, E1293–E1301 (2011).
8. Sulikowska, J. I., Morcos, F., Weigt, M., Hwa, T. & Onuchic, J. N. Genomics-aided structure prediction. Proc. Natl Acad. Sci. USA 109, 10340–10345 (2012).
9. Cheng, R. R., Morcos, F., Levine, H. & Onuchic, J. N. Toward rationally redesigning bacterial two-component signaling systems using coevolutionary information. Proc. Natl Acad. Sci. USA 111, E563–E571 (2014).
10. Aakre, C. D. et al. Evolving new protein–protein interaction specificity through promiscuous intermediates. Cell 163, 594–606 (2015).
11. Lité, T. L. V. et al. Uncovering the basis of protein–protein interaction specificity with a combinatorially complete library. eLife 9, e09240 (2020).
12. McClure, C. J., Alvarez-Buylla, A., Voigt, C. A. & Laub, M. T. Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. Nature 574, 702–706 (2019).
13. McMahon, C. et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. Nat. Struct. Mol. Biol. 25, 289–296 (2018).
14. Schoof, M. et al. An ultraprotein synthetic nanobody neutralizes SARS-CoV-2 by stabilizing inactive Spike. Science 370, 1473–1479 (2020).
15. Damen, L. A. A. et al. Construction and evaluation of an antibody phage display library targeting heparan sulfate. Glycoconj. J. 37, 445–455 (2020).
16. Zupancic, J. M. et al. Directed evolution of potent neutralizing nanobodies against SARS-CoV-2 using CDR-swapping mutagenesis. Cell Chem. Biol. 28, 1379–1388 (2021).
17. Aramli, L. A. & Teschke, C. M. 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).
18. Baroni, T. E. et al. A global suppressor motif for P53 cancer mutants. Proc. Natl Acad. Sci. USA 101, 4930–4935 (2004).
19. Berrooter, R. W. & Hampsley, M. Genetic analysis of yeast Iso-1-ctychrome c structural requirements: suppression of G6y replacements by an Asn52→Ile replacement. Arch. Biochem. Biophys. 288, 261–269 (1991).
20. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. Proc. Natl Acad. Sci. USA 103, 5869–5874 (2006).
21. Bloom, J. D. & Glassman, M. J. Inferring stabilization mutations from protein phylogenies: application to influenza hemagglutinin. PLoS Comput. Biol. 5, e1000349 (2009).
22. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. eLife 2, e00631 (2013).
23. Brown, N. G., Pennington, J. M., Huang, W., Ayaz, T. & Palzkill, T. Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM β-lactamases. J. Mol. Biol. 404, 832–846 (2010).
24. Fane, B., Villafane, R., Mitra, A. & King, J. Identification of global suppressors for temperature-sensitive folding mutations of the P22 tailspike protein. J. Biol. Chem. 266, 11460–11468 (1991).
25. Huang, W. & Palek, T. A natural polymorphism in $\beta$-lactamase is a global suppressor. Proc. Natl Acad. Sci. USA 94, 8801–8806 (1997).
26. Hudson, W. H. et al. Distal substitutions drive divergent DNA specificity among paralogous transcription factors through subdivision of conformational space. Nat. 113, 326–331 (2015).
27. Joyet, P., Declerck, N. & Gaillard, C. Hypermutable variants of a highly thermostable alpha-amylase. Biotechnol. 10, 1579–1583 (1992).
28. Marcianno, D. C. et al. Genetic and structural characterization of an L201P global suppressor substitution in TEM-1 $\beta$-lactamase. J. Mol. Biol. 384, 151–164 (2008).
29. McKeown, A. N. et al. Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. Cell 159, 58–68 (2014).
30. Poteete, A. R., Rennell, D., Bouvier, S. E. & Hardy, L. W. Alteration of T4 lysozyme structure by second-site reversion of deleterious mutations. Protein Sci. 6, 2418–2435 (1997).
31. Shortle, D. & Lin, B. Genetic analysis of staphylococcal nuclease: identification of three intragenic 'global' suppressors of nuclease-minus mutants. Genetics 110, 539–555 (1985).
32. Tsai, A. Y. M., 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).
33. Yang, R. et al. Second-site suppressors of HIV-1 capsid mutations: restoration of intracellular activities without correction of intrinsic capsid stability. Retrovirology 9, 30 (2012).
34. Zhang, J., Guo, N. & Wagner, A. Selection enhances protein evolvability by increasing mutational robustness and foldability. Science 370, eabb5962 (2020).
35. Ortlund, E. A., Bridgham, J. T., Redinho, M. R. & Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. Science 317, 1544–1548 (2007).
36. Starr, T. N., Picton, L. K. & Thornton, J. W. Alternative evolutionary histories in the sequence space of an ancient protein. Nature 549, 409–413 (2017).
37. Klein, F. et al. Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. Cell 153, 726–738 (2013).
38. Angelini, A. et al. Directed evolution of broadly crossreactive chemokine-blocking antibodies efficacious in arthritis. Nat. Commun. 9, 4611 (2018).
39. Madan, B. et al. Mutational fitness landscapes reveal genetic and structural improvement pathways for a vaccine-elicted HIV-1 broadly neutralizing antibody. Proc. Natl Acad. Sci. USA 118, e201653318 (2021).
40. Ivanov, D. N., Finkelstein, A. V. & Kondrashov, F. A. A structural perspective of compensatory evolution. Curr. Opin. Struct. Biol. 26, 104–112 (2014).
41. Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. Protein Sci. 25, 1204–1218 (2016).
42. Duss, G. & Lehrer, B. The genetic landscape of a physical interaction. eLife 7, e32472 (2018).
43. Fowlser, D. M. & Fields, S. Deep mutational scanning: a new style of protein bioRxiv https://doi.org/10.1101/2020.07.14.201475 (2020).
44. Fowlser, D. M. & Fields, S. High-resolution mapping of protein sequence–function relationships. Nat. Methods 7, 741–746 (2010).
45. McLaughlin, R. N., Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. Nature 491, 138–142 (2012).
46. Whitehead, T. A. et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat. Biotechnol. 30, 543–548 (2012).
47. Otwinowski, J., McCandlish, D. M. & Plotkin, J. B. Inferring the shape of global epistasis. Proc. Natl Acad. Sci. USA 115, E7550–E7558 (2018).
48. Poelwijk, F. J. Context-dependent mutation effects in proteins. Methods Mol. Biol. 1851, 123–134 (2019).
49. Schmiedel, J. M. & Lehner, B. Determining protein structures using deep mutagenesis. Nat. Genet. 51, 1177–1186 (2019).
50. Tareen, A., Posfai, A., Ireland, W. T., Mc candlish, D. M. & Kinney, J. B. MAVE-NN: learning genotype–phenotype maps from multiplex assays of variant effect. Preprint at bioRxiv https://doi.org/10.1101/2020.07.03.201475 (2020).
51. Atwal, G. S. & Kinney, J. B. Learning quantitative sequence–function relationships from massively parallel experiments. J. Stat. Phys. 162, 1203–1204 (2016).
52. Sarkissian, K. S. et al. Local fitness landscape of the green fluorescent protein. Nature 533, 397–401 (2016).
53. Pooksesva, V. O. et al. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. PLoS Genet. 15, e1008079 (2019).
54. Rollins, N. J. et al. Inferring protein 3D structure from deep mutation scans. Nat. Genet. 51, 1170–1176 (2019).
**Competing interests**
D.S.M. is an advisor for Dyno Therapeutics, Octant, Jura Bio, Tectonic Therapeutics and Genentech and a cofounder of Seismic. The remaining authors declare no competing interests.

**Additional information**
Extended data is available for this paper at https://doi.org/10.1038/s41559-022-01688-0.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-022-01688-0.

Correspondence and requests for materials should be addressed to Michael T. Laub.

Peer review information Nature Ecology & Evolution thanks Hsin-Hung Chou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2022
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 $\alpha$-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 $\Delta$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 coloured 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 $<6\text{ Å}$ 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 harbouring wild-type toxin ParE3 without antitoxin at different arabinose induction levels in arabinose-titratable E. coli strain BW27783. c, d, Growth rate of cells harbouring 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 Δgrowth rates(T*−T) for all toxin single substitutions under different arabinose inducer concentrations, with positive Δgrowth rate(T*−T) values indicating loss of toxin function. The set of ‘most toxic’ toxin substitutions (n = 310) is coloured in light blue, the set of ‘toxic’ substitutions (n = 781) is coloured 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 Δgrowth 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, h, Distribution of Δgrowth rates(T*−T) for residue positions mapped onto the ParE3 toxin structure. Values shown for 0.00012% [arabinose] inducer. h, The mean Δgrowth rates(T*−T) of residue positions mapped onto the ParE3 toxin structure. Values shown for 0.00012% [arabinose] inducer. i, Distances of Δgrowth rate(T*−T) for all toxin substitutions (black) or top 10 coevolving residue substitutions (purple) in the toxin in absence of antitoxin (h). Values shown for 0.00012% [arabinose] inducer, and antitoxin is induced with 10 µM IPTG. k, The Δgrowth 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 homologues. 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, Δgrowth rate(T*/AT* - T/AT*), of the most toxic toxin mutants are plotted vs. their significance (-log10(p(Δgrowth rate<0))) of deviation from the AT* single mutation. This is based on 10,000 discrete samples of the posterior Δgrowth rate values inferred from the hierarchical Bayesian inference model (see Methods). Vertical line: +0.5 Δgrowth rate, horizontal line: p(Δgrowth 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. Δgrowth 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. Δgrowth rate(T*/AT* - T/AT*). Experiments performed under 'low antitoxin' expression conditions. 

d, Distance vs. Δgrowth 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 Δgrowth rate effect size > +0.5 and p(Δgrowth rate>0) < 0.0001.
Extended Data Fig. 4  | A non-specific, nonlinear 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 nonlinear, 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 Δgrowth 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 $\Delta$growth rate $T^* / AT^*$ - $T / AT$). Grey dots represent $T^* / AT^*$, error bars indicate 95% posterior highest density interval. The $\Delta$growth rate for each antitoxin mutant combined with wild-type toxin ($T / AT^*$) is shown (black dots) along with the $\Delta$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, $\Delta$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 (r_i), 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 auto-encoder) 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 homologues containing the full concatenated toxin and antitoxin (grey), the toxin homologues only (blue), or the antitoxin homologues 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 vs. synonymous variant σ_syn reflect our prior belief that replicate experiment noise is larger than synonymous mutant noise. σ_b 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.

| HDI  | Dataset 1 | Dataset 2 | Dataset 3 |
|------|-----------|-----------|-----------|
| 95%  | 93%       | 96%       | 94%       |
| 90%  | 90%       | 90%       | 91%       |
| 80%  | 75%       | 79%       | 81%       |
| 50%  | 44%       | 44%       | 50%       |
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 (stdev(log(c_post1k/c_pre1k)) or stdev(log(c_post2k/c_pre2k)) 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_posti,n/c_prei,n)), 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.
Corresponding author(s): Michael T Laub
Last updated by author(s): Jan 16 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

- https://github.com/ddingding/coevolution_mechanism
- FLASH 1.2.11
- vsearch 2.13.0

Data analysis

- https://github.com/ddingding/coevolution_mechanism
- networkx 2.3
- tensorflow 2
- pystan 2.19.1.1
- evcouplings 0.1.1
- DeepSequence
- statsmodels 0.10.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Portfolio guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw sequencing reads generated in this manuscript are available here: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA768258/
Structural analysis based on PDB ID: 5CEG.
Natural sequence alignments were generated based on the uniref100 database.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences  ☐ Behavioural & social sciences  ☑ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Study description | This study measure the growth effect of mutations in the toxin-antitoxin ParE3-ParD3. |
|-------------------|---------------------------------------------------------------------------------------|
| Research sample   | The samples are lab-made DNA mutant libraries of naturally found toxin-antitoxins ParE3/ParD3 (uniprot ID F7YBW7/F7YBW8) from Mesorhizobium opportunistum. These are expressed in Escherichia coli strains Top10 or BW27783. This enables genetically manipulable interrogation of mutation effects for this protein-protein interaction system. |
| Sampling strategy | No sample size calculation was performed as we did not have any particular prior beliefs on expected mutation effect sizes. We sampled all possible single mutants. |
| Data collection   | We collected high-throughput sequencing data for all possible single mutants in the toxin, but did not collect any environmental samples. |
| Timing and spatial scale | We measured the growth rate effects of toxin-antitoxin mutants expressed in E. coli over a 10 hour period. There is no spatial scale in our experiments as all mutants are grown in the same flask. The experiments were performed in a room at MIT. |
| Data exclusions   | No data were excluded for single mutants, instead we report on all possible single mutant effects. |
| Reproducibility  | Library growth rate measurements were performed in two biological replicates. |
| Randomization    | We measured all single mutants so did not randomize any mutants. |
| Blinding         | This is not a randomized control trial and no blinding has been performed. |

Did the study involve field work?  ☐ Yes  ☑ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.
| Materials & experimental systems | Methods |
|-------------------------------|---------|
| Involved in the study         | Involved in the study |
| Antibodies                    | ChIP-seq |
| Eukaryotic cell lines         | Flow cytometry |
| Palaeontology and archaeology | MRI-based neuroimaging |
| Animals and other organisms   |         |
| Human research participants   |         |
| Clinical data                 |         |
| Dual use research of concern  |         |