Supplementary Online Material

Bacterial Hsp90 predominantly buffers but does not potentiate the phenotypic effects of deleterious mutations during fluorescent protein evolution.

Bharat Ravi Iyengar[1,2,3], Andreas Wagner[1,2,4,5*]

October 5, 2022

[1] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland.

[2] Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Genopode, Lausanne, Switzerland.

[3] Current address: Institute for Evolution and Biodiversity, Westfalian Wilhelms - University of Münster, Germany.

[4] The Santa Fe Institute, Santa Fe, New Mexico, USA.

[5] Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, South Africa.

* andreas.wagner@ieu.uzh.ch

Contents

(1) Hsp90 is expressed from the constructed expression plasmid.

(1.1) Plasmid construction

We constructed a plasmid for constitutive expression of GFP and inducible expression of Hsp90 in a manner similar to that described previously for another chaperone (Iyengar and Wagner, 2022). We used the pGro7 plasmid designed by Takara (Takara Bio Inc., 2017) as the template for our expression plasmid construction. This is a low-to-medium copy number plasmid with the pACYC origin of replication. It encodes chloramphenicol acetyltransferase, the transcription factor *araC* from *Salmonella typhimurium*, and the *groE* operon consisting of GroEL and GroES, which lie downstream of the *araBAD* promoter from *S. typhimurium* (Figure S1A).

To create the Hsp90 expression plasmid, we first amplified the *E.coli* Hsp90 gene *htpG* from *E.coli* genomic DNA using the primers HtpG-Gen-F and HtpG-Gen-R (Table S1), which contain XhoI and BglII restriction sites at their 5' ends, respectively. We then digested pGro7 (Takara Bio Inc., 2017) using these restriction enzymes to excise the GroEL operon, and then cloned the amplified *htpG* gene into the plasmid backbone. We named the resultant plasmid pHtpG7, and confirmed the expression of Hsp90 using polyacrylamide gel electrophoresis (SDS-PAGE; Figure S1B).

We obtained a GFP expression cassette, which consists of the rplN promoter followed by a ribosome binding site, and the GFP coding sequence, from plasmid pMSs201 (Zaslaver *et al.*, 2006). This cassette encodes the GFPmut2 variant of *Aequorea victoria* GFP, which has three amino acid mutations relative to the wild-type ancestor (Cormack *et al.*, 1996). We chose this GFP variant, because it has a single excitation peak (488nm) and undergoes fast maturation (Balleza *et al.*, 2018).

We engineered 5'-SalI and 3'-SacI restriction sites flanking the GFP coding sequence, by PCR-amplifying the template DNA (pMSs201) with the primers pMS-Sal1-GFP-F and pMS-GFP-SacI-R (Table S1). Next, we amplified this modified GFP expression cassette using the primers pMS-BglII-F and pMS-HindIII-R (Table S1), and cloned the resultant GFP expression cassette into the pHtpG7 plasmid. In this way we obtained our pHtpG7-rplN-GFP plasmid (Figure S1A), which we used for all evolution experiments. We confirmed that both pHtpG7 and pHtpG7-rplN-GFP are able to express Hsp90 upon induction with arabinose (Figure S1B).
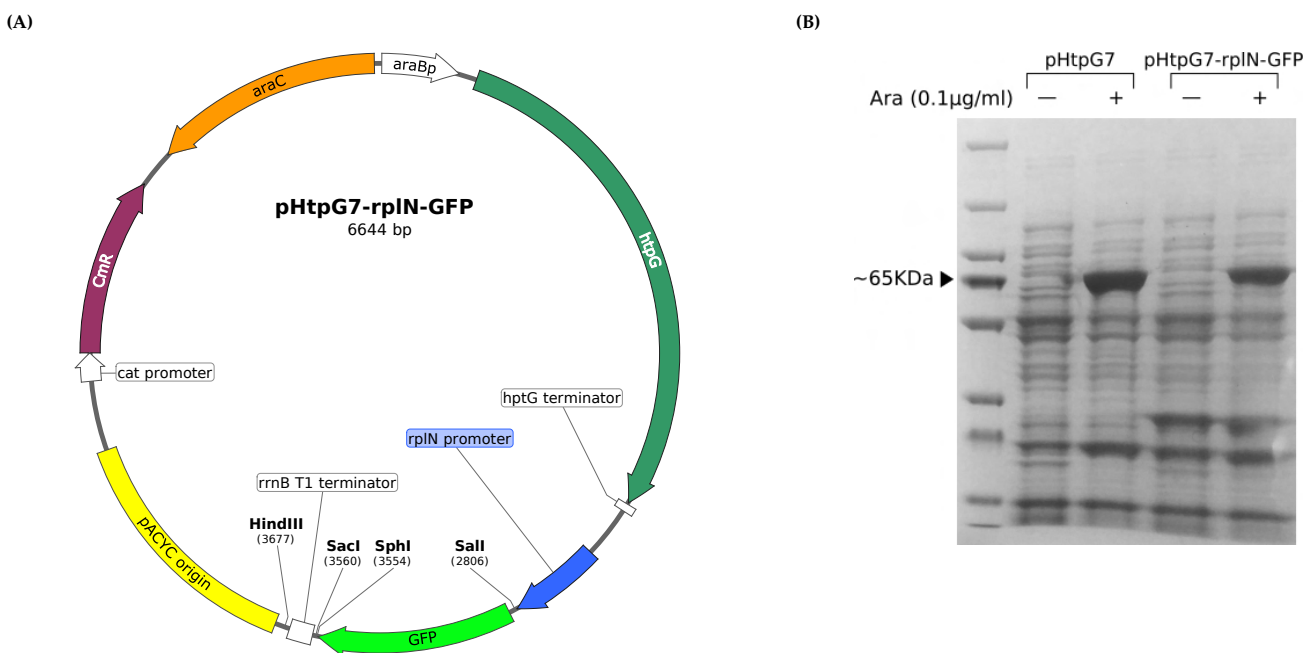

Figure S1: **(A)** Map of pHtpG7-rplN-GFP. **(B)** Image of polyacrylamide gels post-electrophoresis showing the protein profile of cells harboring pHtpG7 and pHtpG7-rplN-GFP plasmids, with and without arabinose induction. The thick band in the induction samples indicated by the 65KDa marker is HtpG (69KDa).

(1.2) Measurement of chaperone expression

We chose the *E. coli* strain BW27784 (CGSC 7881) as our host organism, because it cannot metabolize arabinose. To determine if our constructed plasmids express Hsp90, we performed SDS-PAGE of the protein extracts and observed the intensity of bands corresponding to proteins of the appropriate size. From our previous work with this expression system (Iyengar and Wagner, 2022), we know that its induction is hypersensitive to L-arabinose concentration such that we see strong chaperone expression even at a very low L-arabinose concentration of 0.002mg/ml (the concentration range recommended by Takara is 0.5 – 4 mg/ml; Takara Bio Inc., 2017). We chose a modest concentration of 0.1mg/ml, reasoning that this concentration of L-arabinose is enough to saturate plasmid-borne expression of Hsp90, such that small deviations from this chosen value during the experiments would not affect expression. To quantify Hsp90 expression, we first inoculated 30µl of overnight culture of the pHtpG7-rplN-GFP strain in two 12ml tubes, each containing 3ml of LB medium supplemented with 25µg/ml chloramphenicol (LB-chl). We added L-arabinose to one of the tubes after 1 hour of growth. We allowed both cell populations to grow for 8 hours. For each population, we then pelleted cells from 1ml of cell suspension by centrifuging at 8000g for 3 minutes. We resuspended each pellet in 300µl of lysis buffer, which consists of 50mM Tris-HCl pH 7.5, 100mM NaCl, 5% (v/v) glycerol, 1mM dithiothreitol (added fresh), 1x protein inhibitor cocktail (cOmplete, Roche; added fresh), 300µg/ml lysozyme, 3µg/ml DNAseI, and 16mM MgCl$_2$. We then incubated this suspension for 4 hours at 4°C. We lysed the cells by freezing the suspension in liquid nitrogen, followed by thawing it in a water bath, and repeated this freeze-thaw cycle ten times. Then, we centrifuged the suspension at 18000g for 30 minutes at 4°C and collected the supernatant. We quantified protein concentration using the Bradford method (Bio-Rad Quick Start^{TM} Bradford reagent). We then heated 10µg of protein sample with suitable amounts of 4x SDS-PAGE loading buffer (250 mm Tris-HCl pH 6.8, 8% w/v SDS, 0.2% w/v bromophenol blue, 40% v/v glycerol and 20% v/v 2-mercaptoethanol) at 95°C for 5 minutes. We loaded the samples on a polyacrylamide gel (4% for stacking and 12% for resolving; TruPAGE precast gel, Sigma-Aldrich), and performed electrophoresis at 180V for 45min in 1x TruPAGE TEA-Tricine SDS buffer (Sigma-Aldrich). We fixed the gel for 30 minutes in fixing/destaining solution (50% v/v methanol, 10% v/v acetic acid), and stained it overnight in Coomassie brilliant blue staining solution (0.1% w/v Coomassie brilliant blue R-250, 50% v/v methanol, 10% v/v acetic acid). Next, we destained the gel with destaining solution until the background was clean and the bands were clear. We repeated the same process for cells carrying the pHtpG7 plasmid.

We found that both pHtpG7 and pHtpG7-rplN-GFP express Hsp90 in response to L-arabinose, which was evident through a strong electrophoretic band corresponding to a molecular weight of ∼69KDa (Figure S1B). In the absence of arabinose this band was missing, suggesting that leaky expression from the plasmid is negligible compared to arabinose induced expression.

(2) Ancestral GFP (GFPmut2) is a client of Hsp90
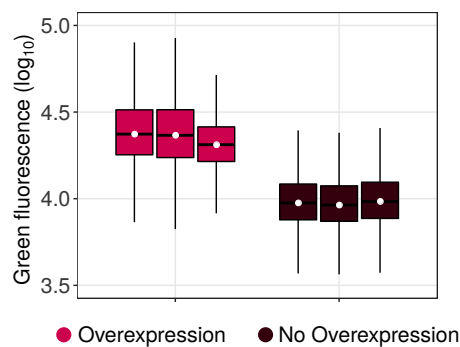
Figure S2: Box plots of the distribution of logarithmically (base 10) transformed green fluorescence (arbitrary units) of ancestral GFP, with (red) and without (brown) Hsp90 overexpression. In each boxplot, the boxes extend from the first quartile to the third quartile, and the whiskers have a length equal to 1.5 times the value of inter-quartile range (1.5 x IQR). The three grouped boxplots correspond to three biological replicate measurements. The black horizontal line and the white circle in the center of each box denote the median of the corresponding distribution. Fluorescence of ancestral GFP is significantly higher when Hsp90 is overexpressed (one-tailed Mann-Whitney U-test, $P < 10^{-15}$).

(3) Most incoming mutations are deleterious

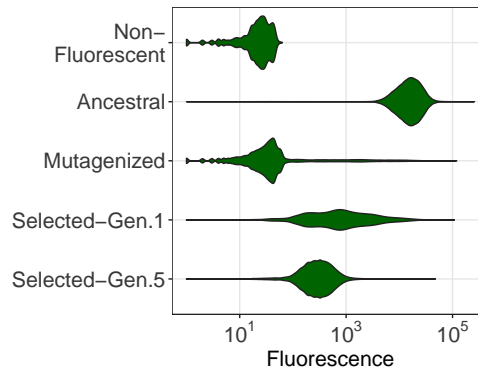Figure S3: Violin plots of the green fluorescence distribution (arbitrary units) of a non-fluorescent *E.coli* cell population, of an isogenic cell population expressing ancestral GFP, of a cell population after mutagenesis of ancestral GFP (round 1 H$^-$, replicate 1), of a cell population that survived weak selection for green fluorescence (generation 1 H$^-$, replicate 1), and of a cell population that survived five rounds of directed evolution with weak selection (generation 5 H$^-$, replicate 1). All these fluorescence measurements are from live cells (flow cytometry) that do not overexpress Hsp90. Since the fluorescence distribution of a post-selection population is approximately symmetric (bottom violin plot), we display distribution statistics with boxplots in the main text.

(4) Hsp90 overexpression alters the spectrum of genotypes that accumulate in the population.

(4.1) Method: principal component analysis

We performed principal component analysis (PCA; Bratulic *et al.*, 2017) to visualize different genotypes accumulated in a population. To this end, we randomly sampled 200 sequences without replacement from every replicate population after the end of the final round of evolution. We converted the amino-acid sequence of each genotype to a numerical sequence, assigning a numerical code to each amino acid. Specifically, we assigned the numbers 1 to 20 to amino acids in the following order: W, F, Y, I, V, L, M, C, D, E, G, A, P, H, K, R, S, T, N, and Q. This ordering of amino acid, in contrast to an alphabetical order, ensures that chemically similar amino acids have a small numerical difference between them (Kim *et al.*, 2009). We assigned the number −10 to the stop codon, because effects of nonsense mutations are dramatically different from those of missense mutations.

We then performed PCA on a matrix containing all these numerical sequences, using the `prcomp` function from the R package stats (v3.4.4; R Core Team, 2018). The rows of this matrix harbor individual sequences (genotypes). Its columns correspond to individual positions in the sequence.

(A)   **(B)**
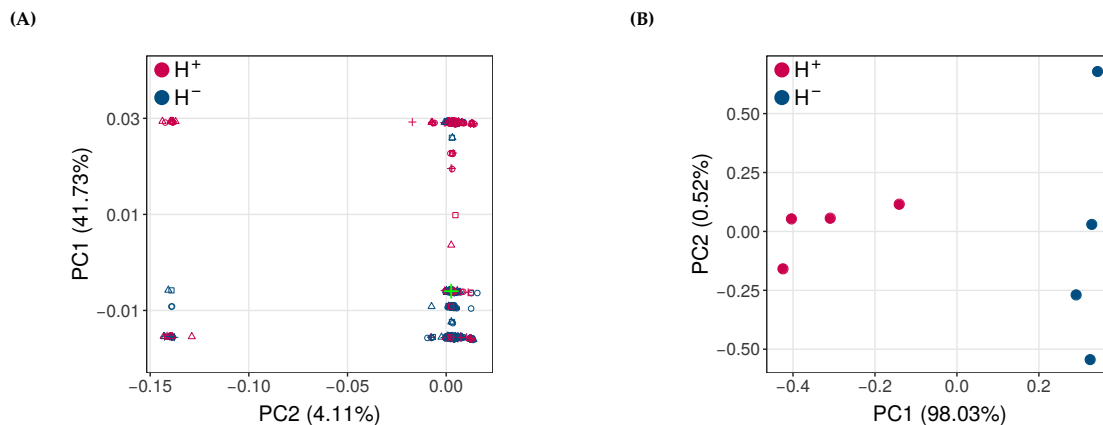
Figure S4: Hsp90 overexpression alters the spectrum of genotypes accumulated in phase 1 populations. **(A)** Principal component analysis (PCA) of genotypes in H$^+$ (blue) and H$^-$ (red) populations at the end of phase 1 evolution. Different replicate populations are denoted by different marker symbols (triangle, circle, square and cross). The ancestral genotype is represented by a green cross. **(B)** PCA based on the frequencies of individual mutations in H$^+$ (blue) and H$^-$ (red) populations at the end of phase 1 evolution.

We also performed PCA on a matrix harboring frequencies of all single amino acid mutations from each population at the end of evolution. In this matrix, each row contains allele frequency data from a different population, and each column corresponds to a different observed mutant.

(4.2) Result

Although H^+ and H^- populations showed similar fluorescence intensities after evolution (Figure S8), H^+ populations had lower fluorescence than H^- populations when Hsp90 was no longer overexpressed (Figure S3). This suggests that different GFP variants accumulate when Hsp90 is overexpressed during evolution. To validate this hypothesis, we sequenced the GFP coding regions from each population at every round of phase 1 evolution to an average coverage of 2450 single molecule reads, and calculated the frequencies of single amino acid mutations and protein variants (genotypes) with multiple mutations.
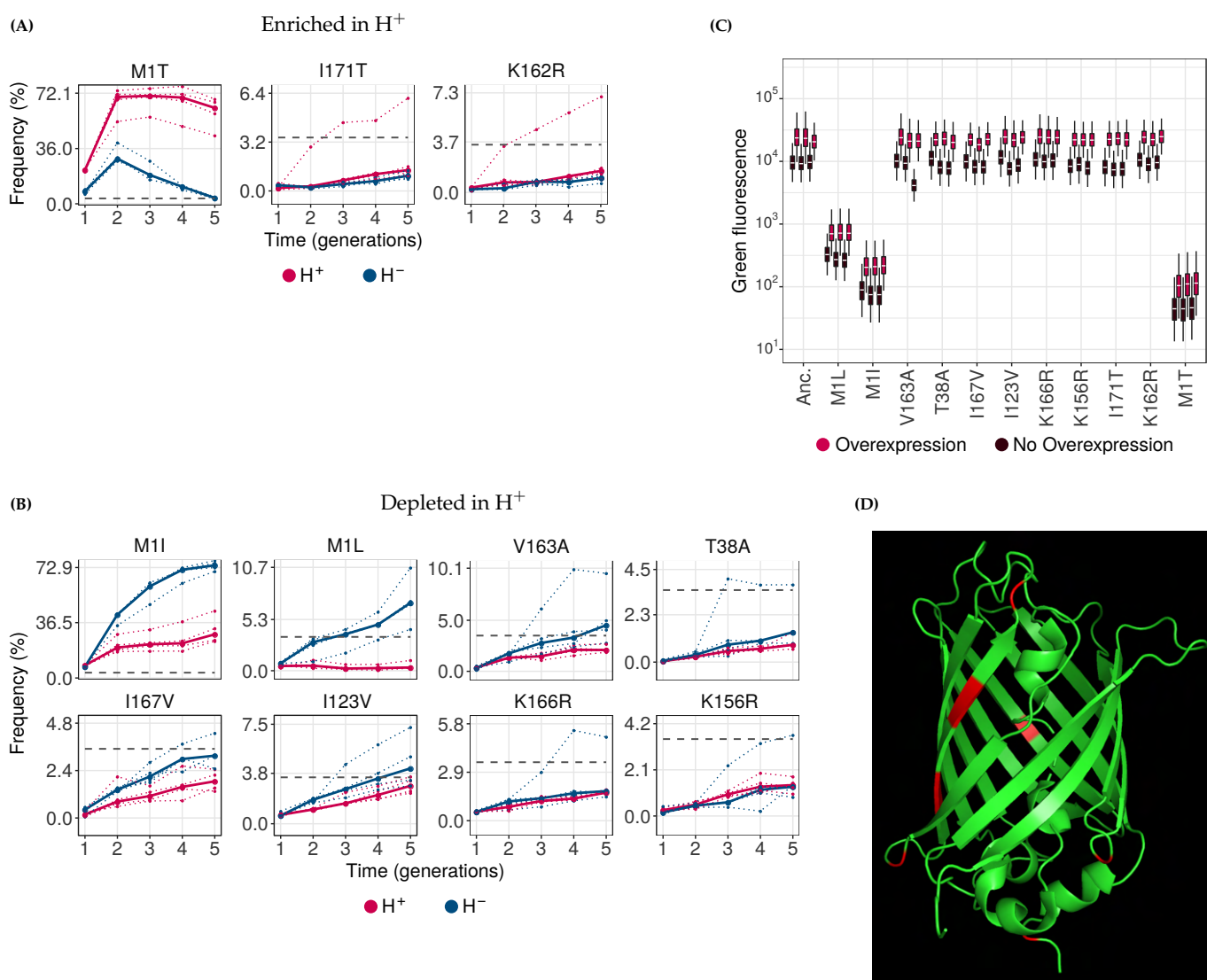

Figure S5: Frequency of single amino acid mutations during phase 1 evolution. The figure shows only those mutations whose frequencies exceed 3.5% (horizontal dashed grey line) in at least one replicate population by the end of phase 1 evolution, and that are **(A)** significantly enriched (have higher frequency), or **(B)** significantly depleted (have lower frequency) in H^+ (red) relative to H^- (blue) populations. Vertical axes show mutation frequencies, and horizontal axes denote time in generations, i.e. round of evolution. Dotted lines indicate the frequency of mutations in individual replicates, whereas solid lines denote the median frequency. Dashed horizontal lines indicate the frequency of 3.5% that we used as a cutoff during the analysis of differentially enriched mutations. Boxplots in panel **(C)** show the distribution of absolute fluorescence values (arbitrary units) of the 11 differentially enriched mutants (as well as ancestral GFP [Anc.]) with (red) and without (brown) Hsp90 overexpression. In each boxplot, the boxes extend from the first quartile to the third quartile, and the whiskers have a length equal to 1.5 times the value of inter-quartile range (1.5 x IQR). The three grouped boxplots correspond to three biological replicate measurements. The white horizontal line in the center of each box denotes the median of the corresponding distribution. Panel **(D)** Shows the location of the 11 differentially enriched mutations (highlighted in red) in the tertiary structure of GFPmut2 (PDB: 6GO9, Lolli *et al.*, 2018). We generated this image using PyMOL (Schrödinger, LLC, 2015).

To visualize the distribution of the accumulated genotypes for H$^+$ and H$^-$ populations at the end of phase 1 evolution, we randomly sampled 300 sequences from each population, and displayed the location of these sequences in genotype space using principal component analysis (PCA), a commonly used dimensionality reduction method (Bratulic *et al.*, 2017). The analysis shows that the variants accumulating in H$^+$ populations occupy different locations in genotype space than those accumulating in H$^-$ populations (Figure S4A). This finding is corroborated by a complementary principal component analysis, which we conducted on the frequency spectrum of individual amino acid mutations (Figure S4B).

These observations are also corroborated by our analysis of differentially enriched individual variants using generalized linear models (Figure S5). This analysis shows that specific variants accumulate preferentially in H$^+$ or H$^-$ populations.

(5) Monte-Carlo simulations for expected mutation frequencies:

Here we describe a Monte-Carlo method that we have also used before to estimate the frequency of a mutation expected under mutation pressure alone (Iyengar and Wagner, 2022). We applied this method here to amino-acid variants that attain a minimal threshold frequency in at least one replicate population at the end of evolution. We chose this threshold frequency to be 3.5% for phase 1 (Figure S5), and 5% for phase 2 (Figure S11) to keep the number of mutations manageable for further analysis. No individual population had more than 36 mutations exceeding these thresholds. For each of these mutations, we performed the Monte Carlo simulations described below to test the null hypothesis that mutation pressure alone may be responsible for its frequency. Rejection of this null hypothesis implies that selection must be involved in explaining their frequency. (Our experimental populations are very large, such that genetic drift is negligible on the time scales of our experiment.)

Our numerical analysis consists of two parts. In the first part, we compute the probability μ that a specific amino acid variant arises in the population. In the second part, the Monte Carlo simulation proper, we simulate how the frequency of this variant changes over time due to mutation pressure alone. We explain this procedure with the mutation S147P, which occurs at a frequency of 0.2% in one of the HtpG overexpressing populations (replicate 3) at the final (5th) round of phase 1. The frequency of this mutation after five additional rounds of evolution in phase 2, is 95%. S147P is encoded by the codon CCA, which requires a T→C change at the 439th position of GFP coding sequence (CDS) i.e. the first position of the ancestral codon TCA (TCA→CCA).

For an S147P mutation to occur, three events must take place. We calculate their probability as follows.

- At least one mutation must occur in the GFP coding sequence. Because mutations are rare, we model the probability P_{mut} of this event with a Poisson distribution, such that

$$P_{mut} = 1 - e^{-\lambda}$$

  Here λ denotes the average number of mutations in each individual per round of mutagenesis. Because we had calibrated our mutagenesis protocol such that this number lies between 1 and 2 (main text), we use a value of $\lambda = 1.5$, which leads to $P_{mut} = 0.777$.

- One of these mutations must occur at the 439th nucleotide position, whereas all other mutations must occur outside codon 147. If only one mutation occurs in the GFP coding sequence, the probability of this event (P_{pos}) is equal to the probability of choosing this one position from the GFP coding sequence of length 717nt, which is $1/717 = 0.0014$. If two mutations occur in the coding sequence then P_{pos} would be $1/717 \times 715/716 = 0.001393$. For more than two mutations, P_{pos} would be $1/717 \times 715/716 \times 714/717 = 0.001391$. Since the difference between the values of P_{pos} for the three cases is very small, we can conveniently approximate its value to be $P_{pos} = 1/717 = 0.0014$. Because this value is an overestimate, our statistical inference would be conservative.

- This mutation must cause a T→C change. From our estimation of mutation rates by Sanger sequencing experiments (main text), we know that this probability (P_{sub}) is 0.75.

The probability (μ) that the mutation S147P occurs (i.e. all the above-mentioned events occur) in any one generation is the product of the above three probabilities: $P_{\text{mut}} \times P_{\text{pos}} \times P_{\text{sub}} = 8.18 \times 10^{-4}$. It is relevant for this analysis that every amino acid variant that exceeded our threshold frequency is caused by a single nucleotide change, such that we do not have to take multiple mutations in a single codon into consideration.

We next turn to the second part of our numerical analysis, where we used the probability μ that a specific variant arises to calculate how the expected mean frequency of this variant changes over time. To this end, we used a discrete time stochastic model of a population whose individuals mutate at a rate μ, such that the number of unmutated individuals becomes progressively smaller. Our simulations neglect back-mutations to the wild-type allele, which will slightly overestimate the allele frequencies caused by mutation pressure. In consequence, our analysis below will be statistically conservative. That is, it may accept some variants as having a frequency consistent with mutation pressure alone, while they may actually be affected by selection.

Specifically, for each mutant whose frequency exceeded our threshold, we performed the following simulation with a starting population of $N_0 = 10^5$ individuals. Each individual has a probability μ (as described earlier in this section) of acquiring a given mutation. The number of individuals mutated in the first round of evolution is thus given by a random variable that is binomially ($B(\mu, N_0)$) distributed. In our simulations, we generated a pseudorandom number M_0 from this distribution, and computed the number of unmutated individuals after the first round of evolution as $N_1 = N_0 - M_0$. In the second round, the number of individuals experiencing the mutation is a random variable with binomial distribution, $B(\mu, N_1)$. We also generated an instance M_1 of this random variate numerically, and calculated the number of unmutated individuals after the second round as $N_2 = N_1 - M_1$. We repeated this procedure for three more rounds to obtain the frequency of the remaining wild-type alleles, and thus also the frequency of the mutant alleles at the end of phase 1. We repeated the procedure for an additional five rounds to predict mutant allele frequencies for the end of phase 2 evolution. For each mutation, we performed 10^5 such simulations and calculated the fraction of simulations in which the estimated frequency exceeded the threshold frequency 3.5% for phase 1 and 5% for phase 2 evolution.

For each variant whose frequency exceeded the threshold in our experimental populations, not a single one among 10^5 simulation reached this threshold. Thus, if we consider the null-hypothesis that the observed frequency of any one variant can be explained by mutation pressure alone, our simulations reject this null-hypothesis at a p-value of $P < 10^{-5}$. Applying a Bonferroni correction to the number of such tests we performed (<100 tests, i.e., variants, for each of the two thresholds), we reject the null-hypothesis at a Bonferroni-corrected p-value of $P < 0.001$. In sum, the frequency of every mutation we consider here cannot be explained by mutation pressure alone. Since our populations are so large that we can neglect genetic drift at the time scale of this experiment, selection or hitchhiking must be invoked to explain their frequency.

(6) Start codon mutations increase the growth rate of host cells

To measure whether start codon mutations (M1I, M1L, M1T) confer a growth advantage in the absence of Hsp90 overexpression, we prepared 1:20 dilutions of an overnight culture of each mutant, as well as of the strain expressing ancestral GFP. We inoculated 10µl of each diluted suspension into 1.4 ml of fresh LB+Chl, and aliquoted 200µl of this inoculated medium into six wells (replicates) of a 96 well plate. Next, we measured optical density (OD at 600nm) every 10 minutes during 20 hours on a Tecan Spark plate reader with temperature being maintained at 37°C, and with the plate shaken constantly between measurements.

All mutants reached stationary phase after 13 hours of growth. Because the growth data fitted a logistic growth equation poorly (not shown), we calculated the maximum growth rate, another commonly used estimate of growth (Papkou *et al.*, 2020). To this end, we first logarithmically (base 2) transformed the measured OD values for each mutant. Using a sliding window of six time points (corresponding to 1h of growth), we calculated the rate of change of $\log_2$-OD between consecutive time points using a linear model (R stats package v3.4.4; R Core Team, 2018). Next, we calculated the maximum value of this slope for all time points, and for each of the six replicates for each mutant and the ancestor. We then compared the median maximum growth rate (of the six replicates) for each mutant to that
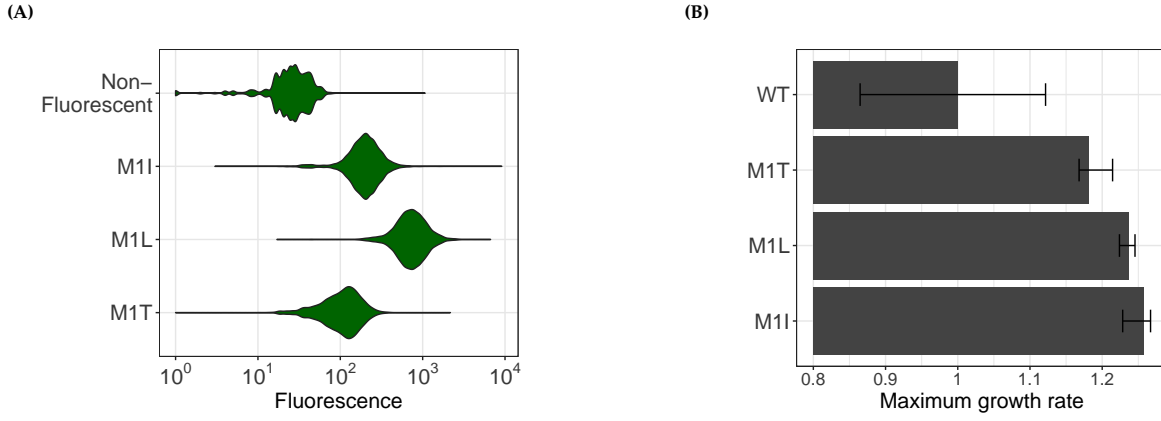
Figure S6: Start codon mutations increase population growth rate. **(A)** Violin plots of the green fluorescence distribution (arbitrary units) of a non-fluorescent *E.coli* cell population, and of isogenic populations that express each of the three GFP start codon mutants, M1I, M1L and M1T. The fluorescence of all the three mutants is significantly higher than that of non-fluorescent cells (one tailed Mann-Whitney U-test, $P < 10^{-15}$). **(B)** Maximum growth rate(arbitrary units; horizontal axis) of *E.coli* hosts carrying different mutations (vertical axis) in the GFP gene. WT refers to ancestral GFP. Start codon mutations M1I, M1L and M1T improved the growth rate significantly (Mann-Whitney U-test, false discovery rate corrected $P = 0.0043$).

of ancestral GFP using a one-tailed Mann-Whitney U-test. All the three start codon mutations conferred significantly higher growth relative to that of ancestral GFP (Mann-Whitney U-test, false discovery rate corrected $P = 0.0043$).

(7) Hsp90 overexpression does not affect genetic diversity during phase 1 evolution.

(7.1) Method: estimation of genetic diversity

As also described in our previous work on a different chaperone (Iyengar and Wagner, 2022), we used three measures of genetic diversity, all of which are based on the observed number of amino acid changes in our evolving GFP sequences. The first is the average number of amino acid mutations in a population relative to the ancestral GFP sequence. The second is the average pairwise distance between two genotypes, defined as the Hamming distance between their amino acid sequences. This metric is analogous to a widely used nucleotide diversity metric (Nei and Li, 1979), except that we apply it to amino acid sequences.

$$\pi = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \pi_{ij} \tag{1}$$

Here π denotes the average pairwise distance between any two genotypes in a population, π_{ij} denotes the distance between the ith and the jth genotypes, and n is the total number of genotypes in the population.

The third metric is the Shannon entropy H of individual allele frequencies p_i in a population (Vajapeyam, 2014), defined as

$$H = \sum_{i} -p_i \log_2(p_i) \tag{2}$$

This diversity measure is largest if all alleles have equal frequencies, and it decreases as an allele frequency distribution becomes peaked at one or few alleles that occur at much higher frequencies than the other alleles.

We used linear models (LM) implemented in the base R package stats (v3.4.4; R Core Team, 2018) to analyze the effect of Hsp90 overexpression on the changes in genetic diversity between different populations. We represented diversity as the response variable with time (rounds of evolution) and state of chaperone overexpression ($+$ or $-$) as interacting predictor variables. Specifically, we used the following expression to define the model: diversity $\sim$
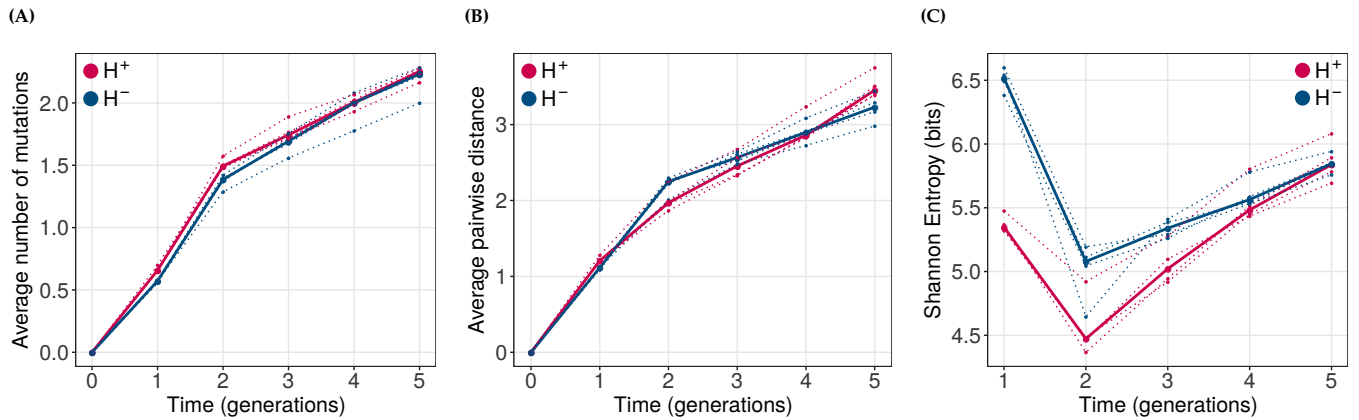
Figure S7: Hsp90 expression does not significantly affect genetic diversity during phase 1 evolution. Genetic diversity metrics **(A)** average distance from the ancestral GFP, **(B)** average pairwise distance between genotypes and **(C)** Shannon entropy are shown on the vertical axes. Horizontal axes denote time in generations of evolution, where generation zero corresponds to the ancestral GFP sequence. H^+ and H^- populations are color coded as red and blue, respectively.

`rounds * condition`. We tested the goodness of fit to the data using the `anova` function from the stats R package (v3.4.4; R Core Team, 2018).

Because replicates differed considerably in genetic variation during phase 2 evolution, we analogously used linear mixed effect models to fit the data (R package lme4, v1.1-21; Bates *et al.*, 2015) and tested the goodness of fit using `anova` function from the R package lmerTest (v3.1-0; Kuznetsova *et al.*, 2017).

(7.2) Result

If a chaperone buffers the effect of destabilizing mutations, then more mutations might accumulate when it is over-expressed, thereby increasing the genetic diversity in these populations. To test this hypothesis, we calculated the average number of amino acid variants per GFP gene in a population, which describes how much the population has diverged from the ancestor (Figure S7A). Using a linear model, we found that genetic diversity increased during phase 1 in all populations (ANOVA, $P < 10^{-15}$). However, Hsp90 did not significantly affect this rate of increase (LM: ANOVA, $P = 0.978$). Moreover, the diversities of H^+ and H^- populations at the end of evolution were not significantly different (two-tailed Mann-Whitney U-test, $P = 0.88$). However, the average number of mutations were higher in H^+ populations in the first two generations of evolution (one-tailed Mann-Whitney U-test, $P = 0.0143$). Analogous analyses with the average pairwise distance between genotypes, also did not reveal any significant effect of the chaperone on genetic diversity (two-tailed Mann-Whitney U-test, $P \geq 0.2$; Figure S7B). Shannon entropy was lower in H^+ populations than in H^- populations during the first three generations (one-tailed Mann-Whitney U-test, $P \leq 0.029$) but it did not differ significantly between the two kind of populations in the next two generations (two-tailed Mann-Whitney U-test, $P > 0.3$; Figure S7C). Overall, we conclude that although Hsp90 buffers the effect of fluorescence-reducing mutations (Main text: Fig 2), it does not lead to a consistent increase in genetic diversity.

(8) Hsp90 buffers the effect of mutations during directional evolution towards a new color.

Phenotype altering mutations often destabilize a protein (Tokuriki *et al.*, 2008; Wang *et al.*, 2002; Studer *et al.*, 2014; Fromer and Shifman, 2009). By mitigating this destabilization, some chaperones can facilitate adaptive evolution (Tokuriki and Tawfik, 2009; Wyganowski *et al.*, 2013). To find out whether Hsp90 affects adaptive evolution towards a new phenotype, we carried out a second phase of directed evolution in which we selected GFP variants that fluoresce in cyan. We note that green and cyan fluorescence intensities are correlated, such that an apparently higher cyan fluorescence could result from a stabilizing mutation that does not affect the fluorescence spectrum (Figure S8A). Also, destabilizing mutations may survive selection, whether they result in a color shift or not, if one selects only on absolute fluorescence. To avoid this problem, we did not select for absolute fluorescence. Instead, we selected for variants with higher cyan fluorescence relative to green fluorescence in comparison to ancestral GFP, irrespective of their absolute fluorescence intensities (Figure S8A). We started phase 2 evolution with populations from the end
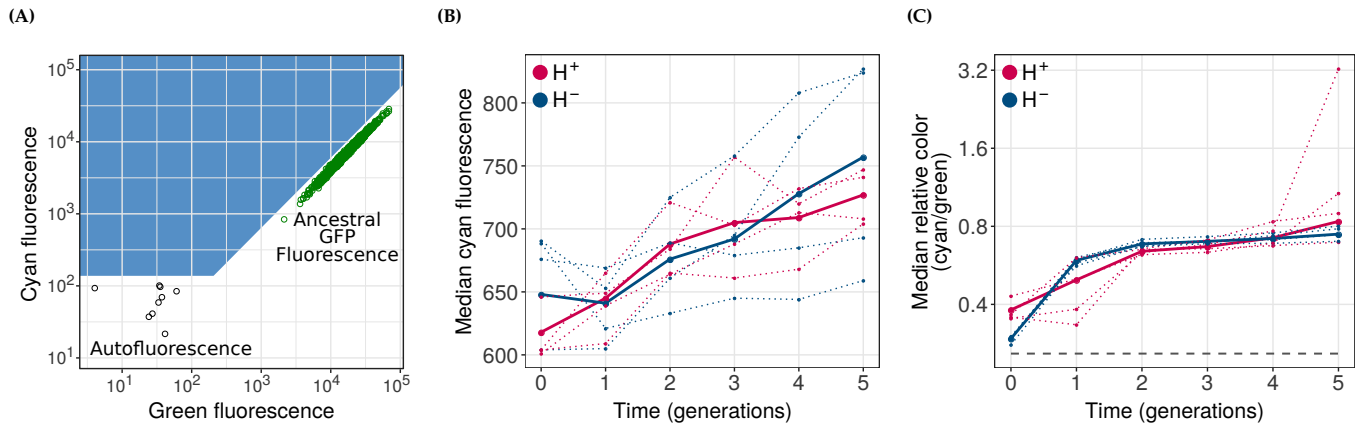
Figure S8: Hsp90 overexpression does not facilitate evolution of the cyan phenotype. **(A)** Strategy for selection of cells exhibiting a shift towards cyan fluorescence. Green circles denote the fluorescence of cells expressing wild type GFP. Black circles denote the autofluorescence of cells. We selected cells mapping to the blue area during directed evolution in phase 2. **(B)** Cyan fluorescence (vertical axis) and **(C)** relative color (cyan to green ratio; vertical axis) of H$^+$ (red) and H$^-$ (blue) populations, during phase 2 of directed evolution. In panel **(C)**, the horizontal dashed grey line denotes the relative color of ancestral GFP. In all panels, dotted lines denote the median values of individual replicate populations, and solid lines denote the median values when data from all populations are pooled. The horizontal axis denotes the time in generations (round of evolution), where generation zero refers to populations at the end of phase 1 evolution.

of phase 1 (ancestor or generation 0 of phase 2), and subjected them to five additional rounds of directed evolution towards the new cyan phenotype.

We measured the distribution of cyan and green fluorescence for all replicate populations in every generation of phase 2 evolution (Figure S8B). We found that median cyan fluorescence increased in all populations (LM:ANOVA, $P < 10^{-6}$), while median green fluorescence decreased (LM:ANOVA, $P < 10^{-4}$, data not shown). However, Hsp90 did not have a significant effect on this rate of increase of cyan fluorescence (LM:ANOVA, $P = 0.82$). At the end of evolution we observed no significant difference between the cyan fluorescence intensities of H$^+$ and H$^-$ populations, based on a comparison of median fluorescence (two-tailed Mann-Whitney U-test, $P = 1$). Since our definition of a new phenotype revolves around a color shift, we also measured another aspect of phenotype, cyan fluorescence relative to green fluorescence (relative color; Figure S8C). Like absolute cyan fluorescence, median relative color also significantly increased during evolution (LM:ANOVA, $P < 0.005$) but Hsp90 affected neither the rate of change of relative color nor its final value at the end of evolution significantly (two-tailed Mann-Whitney U-test, $P = 0.34$).

A chaperone can interact with different protein variants in a population and affect their phenotypes, despite causing no discernible overall difference in the phenotype of populations that evolved under its expression relative to populations that did not. For example it can buffer the effects of some deleterious mutants, causing such mutants to fluoresce more intensely under its overexpression. Conversely, it can potentiate the effect of other deleterious mutations, which can cause a reduction in fluorescence of these mutations, leading to their elimination from the population during selection. To understand which of these two effects dominate in our populations, we measured the fluorescence distribution of populations that evolved under Hsp90 overexpression (H$^+$), with the chaperone's overexpression turned on or off (Figure S9A). We found that Hsp90 overexpression significantly increased cyan fluorescence in all four replicate populations (one-tailed Mann-Whitney U-test, $P < 10^{-15}$), which means that the chaperone causes a net buffering of deleterious mutations.

According to a previous study (Geiler-Samerotte *et al.*, 2016), yeast Hsp90 potentiates the effect of newly acquired mutations, whereas it buffers the effect of mutations that have evolved under its expression for many generations. To determine whether an analogous phenomenon exists also for *E.coli* Hsp90 during adaptive GFP evolution towards a new phenotype, we measured the distribution of cyan fluorescence of populations evolved without Hsp90 overexpression (fifth generation H$^-$ phase 2 populations), both in the presence and the absence of chaperone overexpression (Figure S9B). We observed that cyan fluorescence significantly increased when we overexpressed the chaperone in these populations (Mann-Whitney U-test, $P < 10^{-15}$). This shows that even during evolution towards a new phenotype, Hsp90 predominantly buffered mutational effects, irrespective of whether these mutants evolved under

chaperone overexpression during phase 2.

By comparing the median fluorescence intensities of different populations at the end of phase 2 evolution, we also found that H^+ populations fluoresce less intensely than H^- populations in the absence of Hsp90 overexpression (one tailed Mann-Whitney U-test, $P = 0.014$; Figure S9C). This suggests that Hsp90 overexpression enables strongly fluorescence-reducing mutations to persist in the populations by buffering their mutational effects.

Since phase 2 involves evolution towards a new phenotype, we would expect that the populations accumulate some activity altering mutations that are "beneficial" to the new phenotype. In the main text we show that Hsp90 overexpression increases fluorescence of all differentially enriched mutations in phase 1 populations, irrespective of the effect of these mutations on fluorescence (Main text: Fig 3B). In section 10, we show that Hsp90 also increases the cyan fluorescence of many beneficial (color shifting) mutations in phase 2 (Figure S13).
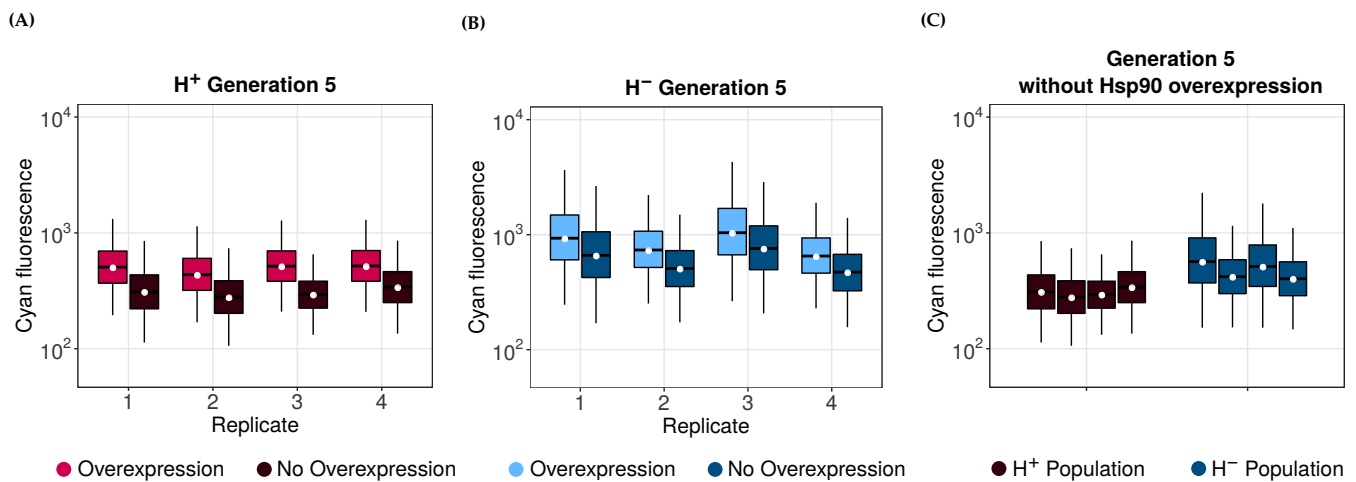

Figure S9: Hsp90 buffers the effect of fluorescence-reducing mutations in phase 2. Box plots of the distribution of logarithmically (base 10) transformed cyan fluorescence (arbitrary units) of different populations, with and without Hsp90 overexpression. In each boxplot, the boxes extend from the first quartile to the third quartile, and the whiskers have a length equal to the inter-quartile range (1 x IQR). The black horizontal line and the white circle in the center of each box denote the median of the corresponding distribution. **(A)** Fluorescence distribution of H^+ populations after the end of phase 2 evolution (generation 5) where the grouped red and brown boxes denote the fluorescence distributions of the same population with and without Hsp90 expression, respectively. **(B)** Fluorescence distribution of H^- populations at the end of phase 2 evolution, with (light blue) and without (dark blue) Hsp90 expression. **(C)** Fluorescence distribution of H^+ (brown) and H^- (dark blue) populations in the absence of Hsp90 overexpression.
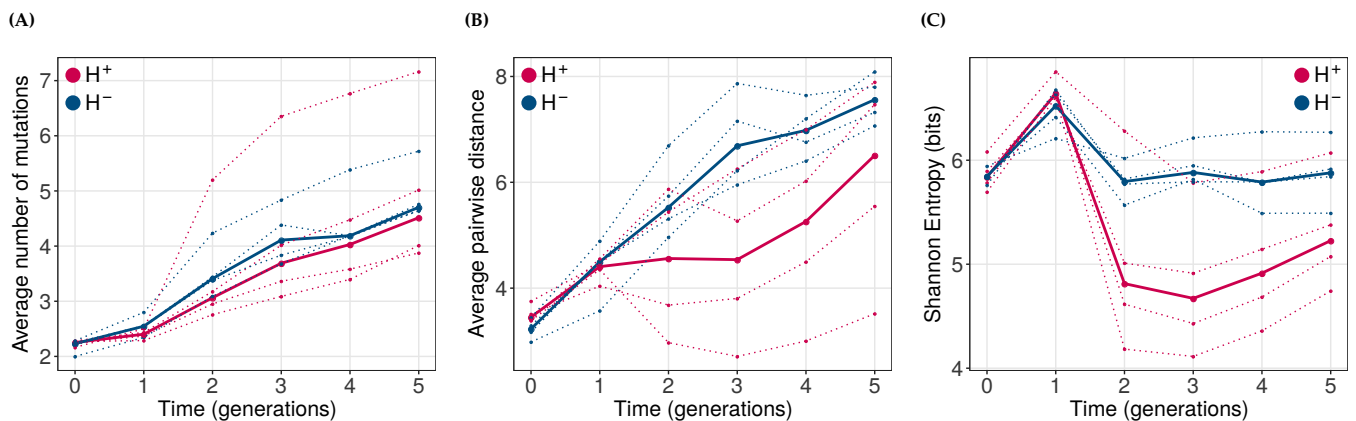

Figure S10: H^+ (red) populations diverge more slowly than H^- (blue) populations during phase 2 evolution. Genetic diversity metrics **(A)** average distance from the ancestral GFP, **(B)** average pairwise distance between genotypes and **(C)** Shannon entropy are shown on the vertical axes. Horizontal axes denote time in generations of evolution, where generation zero corresponds to the ancestral populations, i.e., populations at the end of phase 1.

(9) Hsp90 overexpressing populations diverge more slowly during phase 2 evolution.

We subjected the GFP coding region from each phase 2 population to SMRT sequencing, which yielded an average coverage of 2350 single molecule reads per population. First, we asked if the populations continued their divergence from the ancestor during phase 2 by analyzing the average number of mutations per variant in a population. This was indeed the case for all populations (Linear Mixed Model [LMM]: ANOVA, $P < 10^{16}$; Figure S10A).

Next, we analyzed how genetic diversity changes during phase 2 evolution. We first used the average pairwise distance between the genotypes as a measure of genetic diversity and found that the genetic diversity of the populations increased further during phase 2 evolution (LMM: ANOVA, $P < 0.0003$; Figure S10B). However, H^+ populations diverged more slowly than H^- populations (LMM: ANOVA, $P < 10^{-10}$). This slower increase in diversity was especially pronounced for two replicate H^+ populations. Despite this difference, at the end of evolution, the genetic diversity in H^+ and H^- populations was not significantly different (two-tailed Mann-Whitney U-test, $P = 0.49$). We then also used the Shannon entropy as a complementary metric for genetic diversity (Figure S10C). It increased for all populations it increased during the first generation of evolution, decreased in the second generation, and slowly increased during subsequent generations. In three out of four replicate H^+ populations, the Shannon entropy was lower than in all H^- populations from the second generation onwards. However, the difference in Shannon entropy between H^+ and H^- populations was also not statistically significant at the end of evolution (two-tailed Mann-Whitney U-test, $P = 0.2$). Taken together, these analyses of genetic diversity suggest that H^+ populations diverge more slowly than H^- populations, but only modestly so. We do not know whether Hsp90 overexpression in phase 2 or the deleterious mutations accumulated during phase 1 are more important for the slower increase of genetic diversity in H^+ populations.

(10) Hsp90 overexpression causes a different set of genotypes to accumulate during phase 2 evolution.

Even though Hsp90 overexpression affected genetic diversity only modestly during phase 2, it might still cause different kinds of genotypes to accumulate in evolving populations. Our phenotypic analysis also shows that although H^+ and H^- populations did not differ significantly in their cyan fluorescence at the end of phase 2 evolution, H^+ populations fluoresced less than H^- populations when Hsp90 was no longer overexpressed (Figure S9C). This suggests that Hsp90 expression can cause the accumulation of different genotypes in a population. To visualize the differences in accumulated genotypes in H^+ and H^- populations, we randomly sampled 200 sequences from each population at the end of phase 2 evolution (generation 5), and plotted these sequences in genotype space, using principal component analysis (PCA). We found that the genotypes accumulated in H^+ populations occupy regions of genotype space that are different from those of genotypes in H^+ populations (Figure S11A). A complementary PCA with frequencies
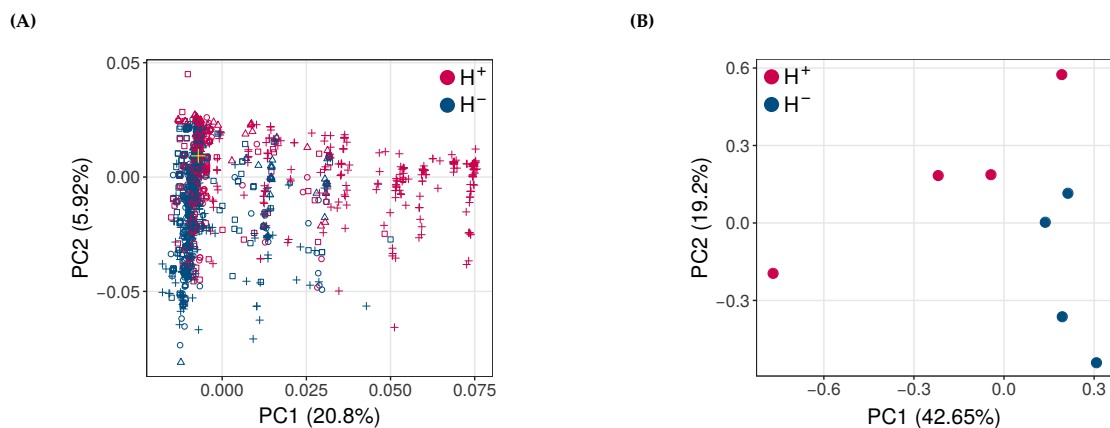

Figure S11: Hsp90 overexpression causes different kinds of genotypes to accumulate during phase 2 evolution. **(A)** Principal component analysis (PCA) of genotypes in H^+ (blue) and H^- (red) populations at the end of phase 2 evolution. Different replicate populations are denoted by different marker symbols (triangle, circle, square and cross). The ancestral genotype is represented by a green cross. **(B)** PCA based on the frequencies of individual mutations in H^+ (blue) and H^- (red) populations at the end of phase 2 evolution.

of single mutation showed a similar separation of allele frequencies between H⁺ and H⁻ populations (Figure S11B). Both PCAs also suggested a greater divergence between replicates in H⁺ than in H⁻ populations, which is consistent with our analysis of overall genetic diversity from Figure S10. Our more detailed analysis of individual mutations corroborates these observations from PCA, because it shows that some mutations accumulate preferentially in H⁺ or H⁻ populations (Section 11, Figure S12).

(11) Hsp90 overexpression affects the evolutionary dynamics of mutations during phase 2.

We next focused on single mutations that attained a high frequency at the end of phase 2 evolution. We limited our analysis to mutations that exceeded a frequency of 5% in at least one replicate population, in order to focus
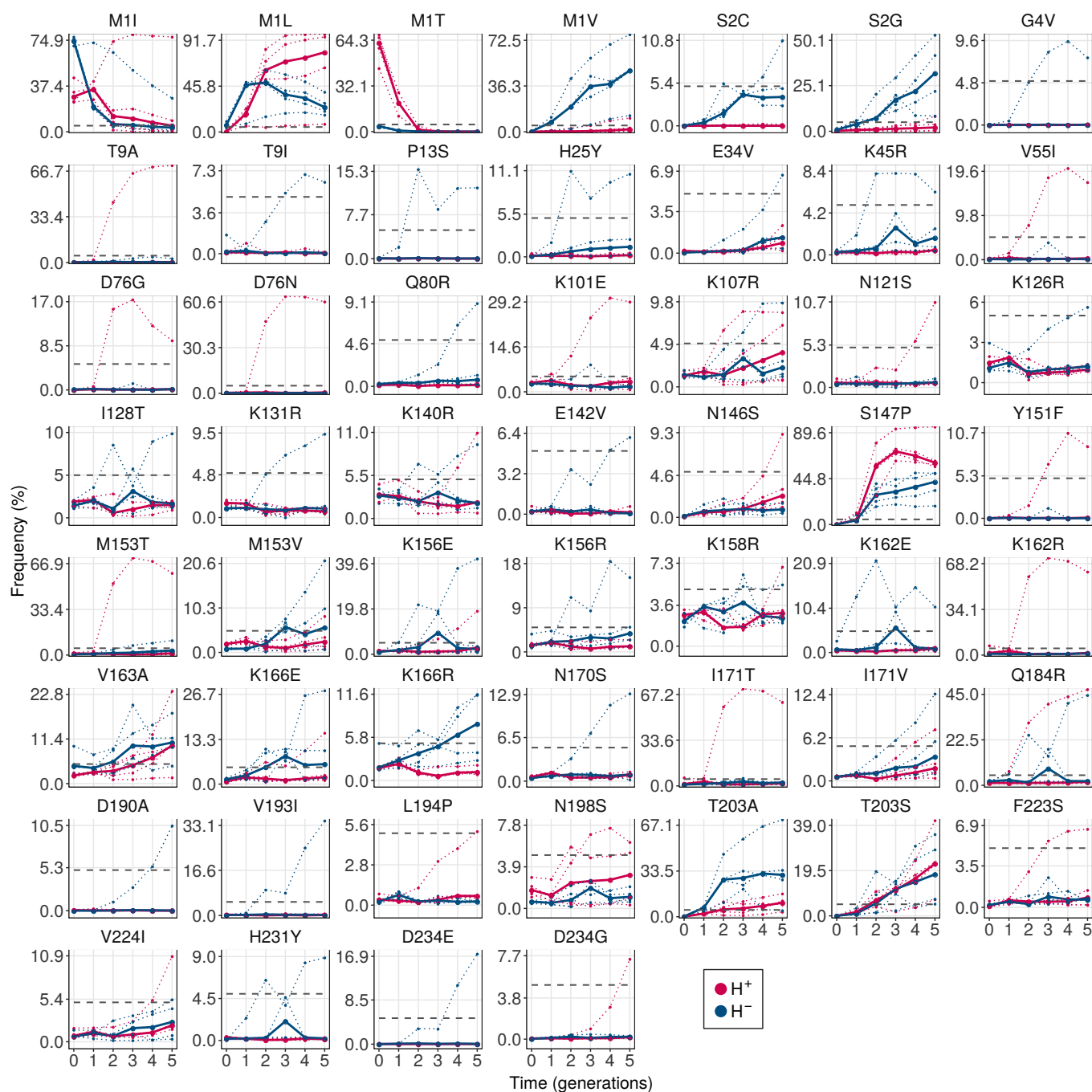

Figure S12: Frequency of single amino acid variants during phase 2 evolution. The figure shows only those variants whose frequencies exceeded 5% (horizontal dashed grey line) in at least one replicate population by the end of evolution, and that are differentially enriched between H⁺ (red) and H⁻ (blue) populations. In addition, we show data from variant M1I to show that while it is abundant in phase 1 H⁻ populations, it is lost in phase 2. Vertical axes (on a logarithmic scale) denote mutation frequency, and horizontal axes denote time in generations, i.e., round of evolution. Dotted lines indicate the frequency of mutations in individual replicates, whereas solid lines indicate the median frequency.

on mutations that are most likely to affect the phenotype substantially. This threshold is higher than the expected frequency of any mutation due to mutation pressure alone (Monte-Carlo simulations, $N = 10^5, P < 10^5$). Next, we used generalized linear models (GLM) to identify mutations whose abundances differed significantly between H$^+$ and H$^-$ populations (GLM:LRT, $P < 0.05$), and found 52 such differentially enriched mutations. This corroborates our findings from PCA that different mutations accumulate in populations when Hsp90 is overexpressed than when it is not.

Since the median cyan fluorescence of the evolved populations was still close to the cellular autofluorescence ($\sim 10^2$ arbitrary units of fluorescence), we reasoned that our populations continued to harbor deleterious mutations during phase 2. These mutations could either have been carried over from phase 1, or they could have risen to high frequency during phase 2. The most abundant phase 1 deleterious mutations were M1I, M1T and M1L. We analyzed the frequencies of these mutations during phase 2 evolution to determine if they persisted at high frequencies during phase 2. M1I and M1T, which were highly abundant in H$^-$ and H$^+$ populations during phase 1, respectively, gradually disappeared from the corresponding phase 2 populations. The mutation M1L whose frequency was significantly lower in H$^+$ populations than in H$^-$ populations during phase 1, became significantly more abundant in H$^+$ populations relative to H$^-$ populations during phase 2. These observations indicate that low cyan fluorescence in phase 2 populations is not due to a carry-over of deleterious mutations from phase 1. The evolutionary dynamics of deleterious mutations in phase 2 is different from that of phase 1.

(12) Hsp90 enhances the fluorescence of some but not all beneficial variants during phase 2 evolution.

In a previous study (Iyengar and Wagner, 2022), we showed that the variants S147P, T203A and T203S cause a color shift towards cyan fluorescence. During phase 2 evolution, these variants rose to a high frequency in different (Figure S12). Specifically, these mutations exceeded a frequency of 40% in at least one population at the end of phase 2, although their allele frequencies varied across populations, To understand the effect of Hsp90 on these mutations, we engineered them into the ancestral GFP gene using site directed mutagenesis. We also engineered the double mutants S147P+T203A and S147P+T203S. We then measured the fluorescence of the three single and two double mutants, with and without Hsp90 overexpression.

These measurements showed that Hsp90 significantly increased the fluorescence of all color shifting mutants (Mann-Whitney U-test, $P < 10^{-15}$) except that of the double mutant S147P-T203S (Figure S13).

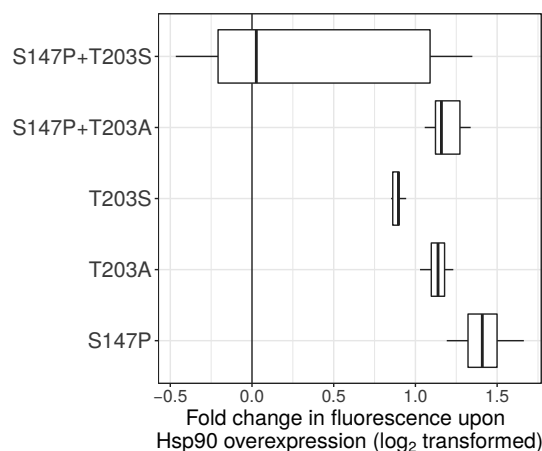We did not discover any other color shifting mutations in this study.

Figure S13: Box plots of median fluorescence for all color shifting mutations with Hsp90 overexpression relative to that without Hsp90 overexpression (fold-change; horizontal axis). We calculated the fold-change for every possible pair (overexpression versus no-overexpression) of biological replicate measurements ($3 \times 3 = 9$ pairs). Data in individual boxes are based on these nine pairs. In each plot, the boxes extend from the first quartile to the third quartile, and the whiskers have a length equal to the inter-quartile range (1 x IQR). The vertical bar inside each box denotes the median fold change in fluorescence due to Hsp90 overexpression.

(13) List of primers

Primer name	Sequence
pMS-BglII-F	ATAGCCGCGCTGCCTCGTCCTGCAGTTC
pMS-HindIII-R	TTCCAAGCTTCCTAGGTCTAGGGCGGCGGATTTGTC
pMS-Sal1-GFP-F	gagCTCGAGTCGACATTTAAGAAggagaTATACATATGAGTAAAGGAGAA
pMS-GFP-SacI-R	GCTCTAGAGCTcGCATGCCTGCAGG
pHtpG-Mut-F	TCGAGTCGACATTTAAGAAGGAGATATACAT
pHtpG-Mut-R	GCCTCTAGAGCTcGCATGC
HtpG-ORF-seq-F	CAAGGTTCTGGGTGGCTCTCG
HtpG-ORF-seq-R	GGGCGGCGGATTTGTCCTAC
HtpG-Gen-F	TGACTCGAGTTAAAATGGCATTATTGAGGTAGA
HtpG-Gen-R	TAGAGATCTAGAAAAATGCCGGATGACAC
HtpG-ORF-PacBio-F	gcagtcgaacatgtagctgactcaggtcacCAAGGTTCTGGGTGGCTCTCG
HtpG-ORF-PacBio-R	tggatcacttgtgcaagcatcacatcgtagGGGCGGCGGATTTGTCCTAC

Table S1: List of primers used for cloning and mutagenesis PCR

Mutation	Forward primer	Reverse primer
I123V	AATAGAGTCGAGTTAAAAGGTATTGATTTTAAAGAAGATG	TAACTCGACTCTATTAACAAGGGTATCACCTTCAAACTTG
I167V	TTCAAAGTTAGACACAACATTGAAGATGGAAGCGTTCAAC	GTGTCTAACTTTGAAGTTAACTTTGATTCCATTCTTTTGT
I171T	CACAACACTGAAGATGGAAGCGTTCAACTA	ATCTTCAGTGTTGTGTCTAATTTTGAAGTTAACTTTGATT
K156R	GCAGACAGACAAAAGAATGGAATCAAAGTTAACTTCAAAA	CTTTTGTCTGTCTGCCATGATGTATACATTGTGTGAGTTA
K162R	GGAATCAGAGTTAACTTCAAAATTAGACACAACATTGAAG	GTTAACTCTGATTCCATTCTTTTGTTTGTCTGCCATGATG
K166R	AACTTCAGAATTAGACACAACATTGAAGATGGAAGCGTTC	TCTAATTCTGAAGTTAACTTTGATTCCATTCTTTTGTTTG
M1I	ATACATATAAGTAAAGGAGAAGAACTTTTCACTGGAGTTG	TTTACTTATATGTATATCTCCTTCTTAAATGTCGACTCGA
M1L	ATACATCTAAGTAAAGGAGAAGAACTTTTCACTGGAGTTG	TTTACTTAGATGTATATCTCCTTCTTAAATGTCGACTCGA
M1T	ATACATACGAGTAAAGGAGAAGAACTTTTCACTGGAGTTG	TTTACTCGTATGTATATCTCCTTCTTAAATGTCGACTCGA
T38A	GATGCAGCATACGGAAAACTTACCCTTAAATTTATTTGCA	TCCGTATGCTGCATCACCTTCACCC
V163A	ATCAAAGCTAACTTCAAAATTAGACACAACATTGAAGATG	GAAGTTAGCTTTGATTCCATTCTTTTGTTTGTCTGCCATG
S147P	TATAACCCACACAATGTATACATCATGGCAGACAAACAAA	ATTGTGTGGGTTATAGTTGTATTCCAATTTGTGTCCAAGA
T203S	CTGTCCTCACAATCTGCCCTTTCGAAAGATCCCAACGAAA	AGATTGTGAGGACAGGTAATGGTTGTCTGGTAAAAGGACA
T203A	CTGTCCGCACAATCTGCCCTTTCGAAAGATCCCAACGAAA	AGATTGTGCGGACAGGTAATGGTTGTCTGGTAAAAGGACA

Table S2: List of primers used for site directed mutagenesis

References

Balleza E, Kim JM, Cluzel P. 2018. Systematic characterization of maturation time of fluorescent proteins in living cells. Nature Methods. 15:47–51.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 67:1–48.

Bratulic S, Toll-Riera M, Wagner A. 2017. Mistranslation can enhance fitness through purging of deleterious mutations. Nature Communications. 8:15410.

Cormack BP, Valdivia RH, Falkow S. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 173:33 – 38. Flourescent proteins and applications.

Fromer M, Shifman JM. 2009. Tradeoff between stability and multispecificity in the design of promiscuous proteins. PLoS Computational Biology. 5:1–16.

Geiler-Samerotte KA, Zhu YO, Goulet BE, Hall DW, Siegal ML. 2016. Selection transforms the landscape of genetic variation interacting with hsp90. PLoS Biology. 14:1–21.

Iyengar BR, Wagner A. 2022. GroEL/S Overexpression Helps to Purge Deleterious Mutations and Reduce Genetic Diversity during Adaptive Protein Evolution. Molecular Biology and Evolution. 39. msac047.

Kim Y, Sidney J, Pinilla C, Sette A, Peters B. 2009. Derivation of an amino acid similarity matrix for peptide:mhc binding and its application as a bayesian prior. BMC Bioinformatics. 10:394.

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software. 82:1–26.

Lolli G, Raboni S, Pasqualetto E, Benoni R, Campanini B, Ronda L, Mozzarelli A, Bettati S, Battistutta R. 2018. Insight into gfpmut2 ph dependence by single crystal microspectrophotometry and x-ray crystallography. The Journal of Physical Chemistry B. 122:11326–11337.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences. 76:5269–5273.

Papkou A, Hedge J, Kapel N, Young B, MacLean RC. 2020. Efflux pump activity potentiates the evolution of antibiotic resistance across s. aureus isolates. Nature Communications. 11:3970.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Schrödinger, LLC. 2015. The PyMOL molecular graphics system, version 1.8.

Studer RA, Christin PA, Williams MA, Orengo CA. 2014. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. Proceedings of the National Academy of Sciences. 111:2223–2228.

Takara Bio Inc.. 2017. Chaperone plasmid set (cat.#3340). https://www.takarabio.com/documents/User%20Manual/3340/3340_e.v1701Da.pdf.

Tokuriki N, Stricher F, Serrano L, Tawfik DS. 2008. How protein stability and new functions trade off. PLoS Computational Biology. 4:1–7.

Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature. 459:668–673.

Vajapeyam S. 2014. Understanding Shannon's Entropy metric for Information.

Wang X, Minasov G, Shoichet BK. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. Journal of Molecular Biology. 320:85 – 95.

Wyganowski KT, Kaltenbach M, Tokuriki N. 2013. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. Journal of Molecular Biology. 425:3403 – 3414.

Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U. 2006. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. Nature Methods. 3:623–628.