Evolution of new genes under intermittent selection

Joakim Näsvall[1*] and Hind Abdalaal[1,2]

[1] Dept. of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden
[2] Current address: Neobe Therapeutics, Stevenage, England
[*] Corresponding author. Email: joakim.nasvall@imbim.uu.se

Supplementary Results and Discussion

Supplementing the growth medium with tryptophan causes growth defects that can be overridden by guanosine

The HisA variants with the highest TrpF activity displayed faster growth in the absence of tryptophan compared to its presence (Figures S3 and S4). This observation might be linked to the availability of phosphoribosylpyrophosphate (PRPP), a precursor molecule common to the de novo synthesis of histidine, tryptophan, purines, pyrimidines, and pyridine nucleotides.

In wild-type Salmonella enterica, histidine feedback-inhibits the first enzyme (HisG) of its biosynthetic pathway, making HisG the committing and rate-determining step[1]. Conversely, a mutation in any other enzyme besides HisG is expected to decrease histidine synthesis, leading to increased flux into the pathway due to derepression of the operon and reduced HisG inhibition. Consequently, a HisA mutation could potentially increase PRPP consumption for histidine biosynthesis, limiting its availability for other pathways.

Tryptophan addition slightly reduces the PRPP pool in wild-type *S. enterica*, possibly through indirect effects on other pathways[2]. A similar phenomenon was observed in *S. enterica* mutants with a transposon insertion in *trpC*, exhibiting a filamentous growth phenotype. This phenotype was reversed by the addition of histidine, tryptophan, mutations in *hisF* or *hisH*, but not by nucleotides[3]. This effect was attributed to excessive PRPP channeling into tryptophan synthesis, leaving insufficient PRPP for histidine biosynthesis.

We hypothesized that the observed growth inhibition of HisA mutants upon tryptophan addition might be related to its effect on the PRPP pool. Reduced PRPP synthesis due to tryptophan addition, coupled with increased PRPP channeling into the histidine pathway due to the HisA mutation, could potentially lead to insufficient nucleotide synthesis.

Given the critical role of purines (adenosine and guanosine) in energy homeostasis and as building blocks for RNA and DNA, compared to the less demanding pyrimidines (cytosine, uridine, and thymidine)[4], we reasoned that if the growth reduction stemmed from limited PRPP availability, supplementing the medium with guanosine could rescue the growth defect. Notably, guanosine addition improved the growth of HisA mutants in tryptophan-containing medium but had no significant effect on wild-type HisA strains (including *trpA* mutants).

Regardless of the precise mechanism underlying the growth reduction in HisA mutants upon tryptophan addition, it appears unrelated to a direct limitation in histidine or tryptophan biosynthesis. Therefore, to isolate the specific effects on HisA activity, guanosine was added to the growth media when assaying the growth of reconstructed mutants in the presence of tryptophan (Figures 3, 4, S2, S3, S4, S5).

Two populations evolved TrpF-activity or TrpF-independent growth in still unidentified ways

Two populations (designated "1-7" and "2*mutS*22" in Tables S1 and S2) evolved the ability for TrpF-independent growth through unidentified mechanisms.

Identification of the specific mutations responsible for this phenotype in these populations proved challenging. While both populations shared four mutation targets (in *yafS*, *yeaG*, *trpD*, and large duplications between rRNA operons), these alterations were also present in other populations, making them unlikely candidates for TrpF-independent growth.

The involvement of *trpD* mutations in the phenotype was specifically excluded. We reintroduced the wild-type allele of *trpD* into a clone isolated from population "2*mutS*22" using transduction. Despite this genetic modification, the transductants retained their ability for TrpF-independent growth, regardless of the *trpD* allele they possessed.

Furthermore, although isolated clones from population "2*mutS*22" functioned effectively as both a donors and recipients in transduction experiments, the TrpF-independent growth phenotype could not be transferred to the ancestral strain through this method. This suggests a polygenic basis for the phenotype, where mutations in multiple genes contribute, each with an effect too subtle for successful selection during single-gene transduction.

How can a frameshift mutation in *trpA* both create a new function and retain the original function?

The isolated frameshift mutation in *trpA* (dup173_179; P62Fs) is a seven-nucleotide duplication that disrupts the reading frame after codon 60. This leads to the misincorporation of 23 incorrect amino acids before translation terminates at a UGA stop codon (Figure S6a). Given the early location of the mutation, the resulting 83-amino acid peptide is highly unlikely to retain any functional activity.

For the P62Fs mutant to acquire a new function, the frameshift mutation in *trpA* requires efficient frameshift suppression. This suppression would allow translation to resume in a new reading frame, generating a full-length protein with an altered amino acid sequence. Two potential scenarios are possible for both restoring the original function and generating the new function. In the first scenario, a single suppression event would yield a bifunctional protein capable of both the original and the new activity. Alternatively, suppression might occur through distinct mechanisms, resulting in two separate peptides, each responsible for one specific function.

Intriguingly, several populations harboring the P62Fs mutation in *trpA* also displayed mutations in genes encoding ribosomal components: rRNA methyltransferase RsmD (YhhF) and ribosomal proteins S2 and L9 (Table S2). Mutations in L9 have been documented to enhance ribosomal frameshifting[5]. While a direct role for RsmD in frameshifting has not been established, it has been shown to contribute to frameshift suppression when combined with another frameshift suppressor mutation[6]. Notably, both populations with RsmD mutations also possess mutations in S2. This co-occurrence suggests a potential synergistic effect of S2 and RsmD mutations in suppressing the *trpA* frameshift mutation.

+1 ribosomal frameshifting occurs when the ribosome and peptidyl-tRNA "slip" one nucleotide forward on the mRNA. This is more likely to happen at "slippery sites" where the interaction between the P-site codon and anticodon can be re-established in the alternative reading frame if broken[7]. A common slippery sequence is CCC-U, where a proline tRNA (anticodon cmo^{5}UGG) can pair with both the original (CCC) and the +1 frameshifted (CCU) codons[7].

In the frameshifted *trpA* sequence, the codon pair CCC-UAC (Pro-Tyr) is encountered two codons into the new reading frame. If +1 slippage of peptidyl-tRNAPro occurs at this site, the result is an insertion of two amino acids (arginine and tryptophan) between Gly61 and Pro62 (61-62insRW) of the TrpA protein.

Interestingly, these same two positions (Gly61 and Pro62) are mutated in one of the other evolutionary paths observed for *trpA* (T24S G61S P62S). A constructed *trpA*(61-62insRW) mutant was able to grow on minimal media without the *trpF* gene, although at a slower rate than the P62Fs mutant. This suggests that another

suppression mechanism might be necessary for optimal activity of one or both functions. One possibility, although we have no evidence for it, is a "hop" where the P-site codon-anticodon interaction is broken at one site and re-formed at another. This could potentially result in restoration of the wild-type sequence in a fraction of the "full-length" protein.

Additional Mutations in Evolved Populations Potentially Affecting TrpF-Independent Growth

Beyond the likely frameshift suppressor mutations identified in populations with the *trpA*(Pro62Fs) mutation, many evolved lineages harboured additional mutations. These mutations could potentially influence growth in the conditions of the evolution experiment in several different ways.
Several populations contained mutations known to provide selective advantages in minimal glucose environments (e.g., *pykF*, *rpoS*, *spoT*) [8–10]. This suggests the evolved populations adapted not only to overcome the tryptophan deficiency but also for efficient growth under the specific experimental conditions.

Metabolic adjustments affecting the tryptophan or histidine pathways: Some mutations might directly modify the impact of the causative mutations. These could allow for more growth without restoring tryptophan synthesis. Potential benefits include reduced wasteful consumption of precursors like PRPP and chorismate by the non-functional tryptophan pathway in a Δ*trpF* mutant, or potentially mitigating a potential toxic effect by the accumulation of the TrpF substrate. However, none of these potential benefits were experimentally verified through reconstruction and testing of the specific mutations.

Most populations displayed mutations affecting other genes within the *his* and *trp* operons. For instance, *hisG* mutations, potentially reducing the flow of PRPP into the histidine pathway, were identified in six out of ten sequenced non-mutators and four out of eight sequenced mutators from evolution experiment 2. Notably, these *hisG* mutations always co-occurred with *trpA* mutations (Table S2).

More than half of the whole-genome sequenced populations harboured mutations affecting TrpD and/or TrpE, seemingly independent of the presence or absence of TrpF-activating mutations. The nature of some of these mutations suggests negative impacts on protein function (e.g., frameshifts, large deletions, premature stop codons). However, since these mutations were present in populations with both *hisA* and *trpA* mutations, they must still allow for some level of functional TrpD and TrpE protein production. Reduced activity of these enzymes could potentially limit the flow of precursors into the tryptophan pathway, impacting growth or viability of the Δ*trpF* mutant under tryptophan-limiting conditions.

Two populations with *hisA* mutations displayed mutations in *hisH*, the preceding gene in the *his* operon. As the specific mutations are amino acid substitutions located upstream of the *hisA* start codon, they could potentially affect HisH function or even influence *hisA* translation initiation, potentially increasing HisA (and thus its TrpF) activity.

The evolution experiments were conducted in a medium lacking guanosine. Since our later findings revealed that *hisA* mutations could be rescued by guanosine supplementation, it's possible that any mutations affecting flux through the histidine or tryptophan pathways (*hisG*, *trpD*, *trpE*, and possibly *hisH*) might act by effects on the PRPP pool, and that some of these would not have been seen if the evolution experiments were done in medium with added guanosine.

Regulation of Histidine Biosynthesis: One population (DA65453) contained a mutation affecting $tRNA^{His}$ (the *hisR* gene), and two additional populations (DA59062, DA59064) had different mutations in the promoter of the *argX-hisR-leuT-proM* tRNA operon that likely reduce the expression of the tRNAs. Reduced expression or function of $tRNA^{His}$ is expected to lead to over-expression of the *his*-operon by affecting the regulation of transcription attenuation in the *his*-operon leader region, mimicking histidine starvation[1]. The mutations affecting *hisR* are thus likely to cause increased expression of HisA (and thus increase expression of the evolving TrpF activity).

Supplementary Figures

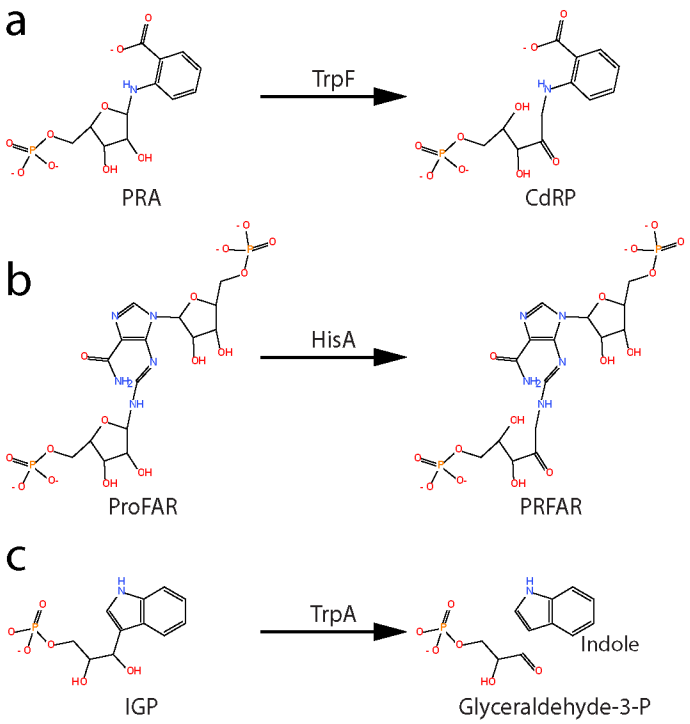

Figure S1. Reactions catalyzed by the native activities of TrpF, HisA, and TrpA. **(a)** TrpF (Phosphoribosyl anthranilate synthase) catalyzes the isomerization of N-(5'-phosphoribosyl)-anthranilate (PRA) to 1-(2-carboxyphenylamino)-1'-deoxyribulose-5'-phosphate (CdRP), the third step in tryptophan biosynthesis. **(b)** HisA (ProFAR isomerase) catalyzes the isomerization of N'-[(5'-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) to N'-[(5'-phosphoribulosyl) formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR), the fourth step in histidine biosynthesis. **(c)** TrpA – Tryptophan synthase subunit A; together with tryptophan synthase subunit B (TrpB) catalyzes the two-step conversion of (3-indolyl)-glycerol-phosphate (IGP) to L-tryptophan, the final step in tryptophan biosynthesis. Without TrpB, Indole and glyceraldehyde-3-phosphate is released.

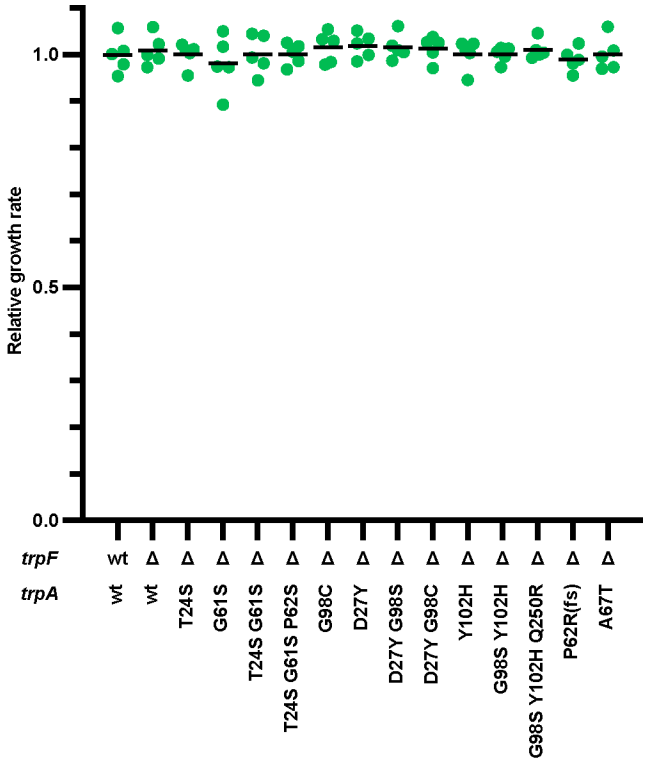

Figure S2. Growth rates of *trpA* mutants prior to tryptophan depletion. Exponential growth rates were determined in the early rapid growth phase ($0.095 > OD_{600} < 0.184$) in M9 glucose medium supplemented with tryptophan (5 µM) and guanosine (3 mM), and set relative to the *trpF*(wt) *trpA*(wt) wild-type strain grown in the same experiment. All strains were grown as five biological replicates, and none of the strains grew differently from the *trpF*(wt) *trpA*(wt) strain according to a two-tailed Student's t-test with equal variance.

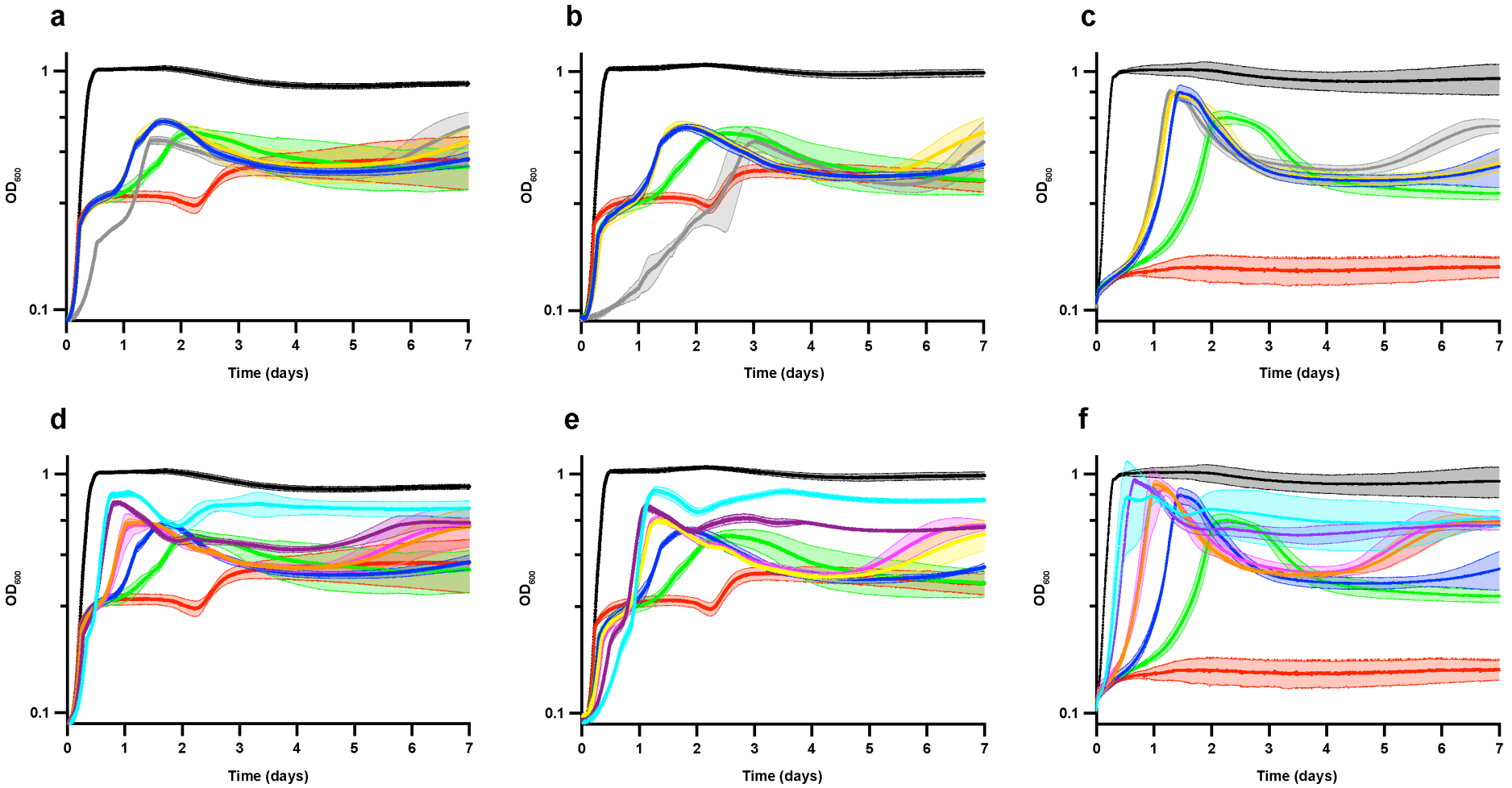

Figure S3. Growth of re-constructed *hisA* mutants. In all graphs the ancestral (*ΔtrpF*; red) and a wild-type (*trpF*+; black) strain is included for comparison. The curves show the average OD_{600} with standard deviation of at least four biological replicates. **(a, d)** Cultures grown after 1000x dilution in 2x M9 + 0.4 % glucose supplemented with 5 µM tryptophan + 3 mM guanosine. **(b, e)** Cultures grown after 1000x dilution in 2x M9 + 0.4 % glucose supplemented with 5 µM tryptophan (no guanosine). **(c, f)** Cultures grown after 100x dilution in 2x M9 + 0.4 % glucose without supplementation. **(a - c)** *hisA* mutations found at 120 and 170 generations in population 1-1. Green; *hisA*(Q18R), grey; *hisA*(H17R Q18R), yellow; *hisA*(Q18R A22T), blue; *hisA*(Q18R K107E). **(d - f)** *hisA* mutations found at 320 and 420 generations in population 1-1. Green; *hisA*(Q18R; included for comparison), blue; *hisA*(Q18R K107E), pink; *hisA*(Q18R I62M K107E); orange; *hisA*(Q18R K107E V168A), purple; *hisA*(Q18R R23L K107E V168A), cyan; *hisA*(Q18R A22T R23L K107E V168A).

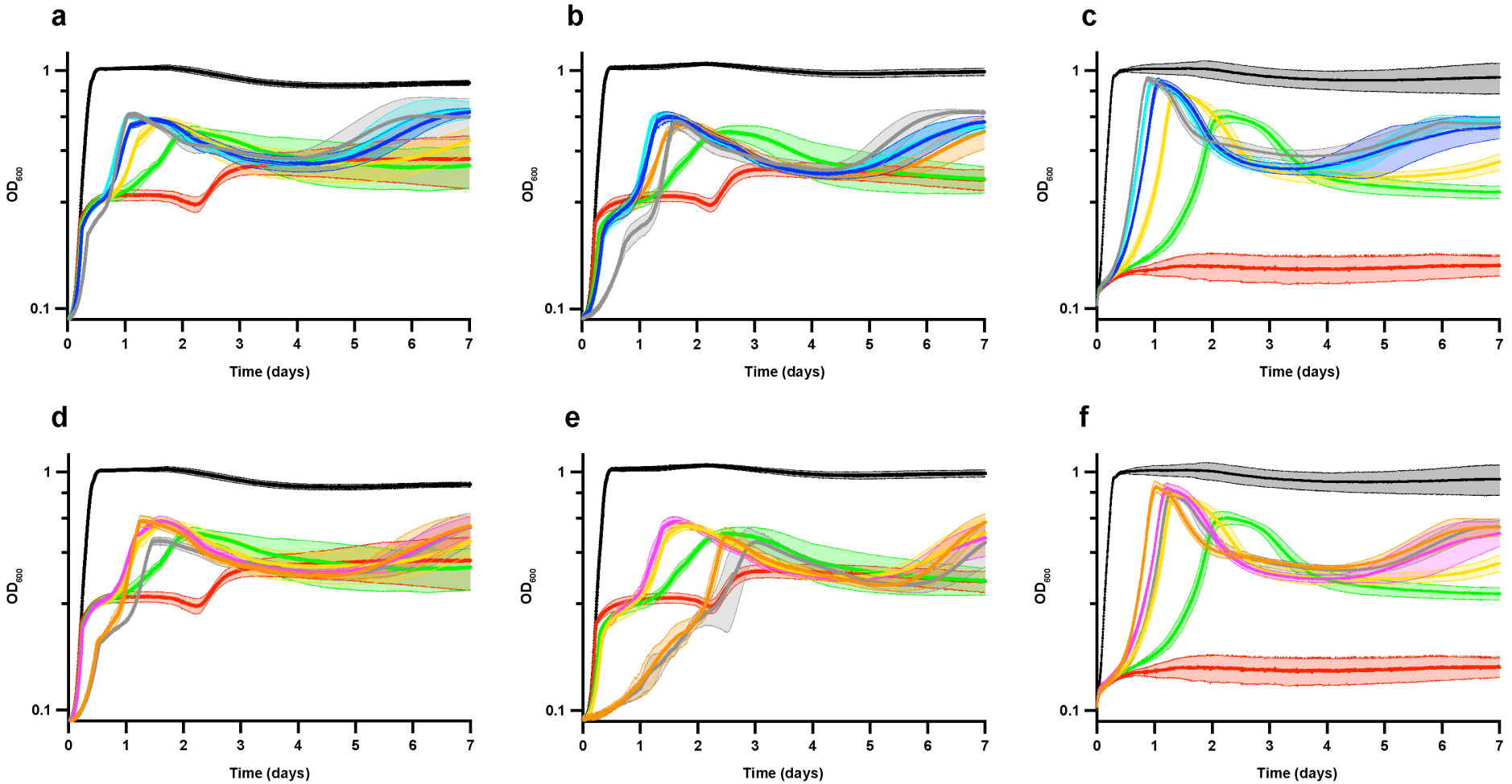

Figure S4. Growth of re-constructed *hisA* mutants, continued. In all graphs the ancestral (*ΔtrpF*; red) and a wild-type (*trpF*+; black) strain is included for comparison. The curves show the average OD_{600} with standard deviation of at least four biological replicates **(a, d)** Cultures grown after 1000x dilution in 2x M9 + 0.4 % glucose supplemented with 5 µM tryptophan + 3 mM guanosine. **(b, e)** Cultures grown after 1000x dilution in 2x M9 + 0.4 % glucose supplemented with 5 µM tryptophan (no guanosine). **(c, f)** Cultures grown after 1000x dilution in 2x M9 + 0.4 % glucose without supplementation. **(a - c)** *hisA* mutations found in population 1-3. Green; *hisA*(Q18R), yellow; *hisA*(Q18R A22T), cyan; *hisA*(Q18R A22T V168A), grey; *hisA*(Q18R A22T I62M R131H), blue; *hisA*(Q18R A22T V168A F199L). **(d - f)** *hisA* mutations found in the *mutS* populations. Green; *hisA*(Q18R), yellow; *hisA*(Q18R A22T), orange; *hisA*(Q18R A127G), grey; *hisA*(H17R Q18R), pink; *hisA*(Q18R T52S).

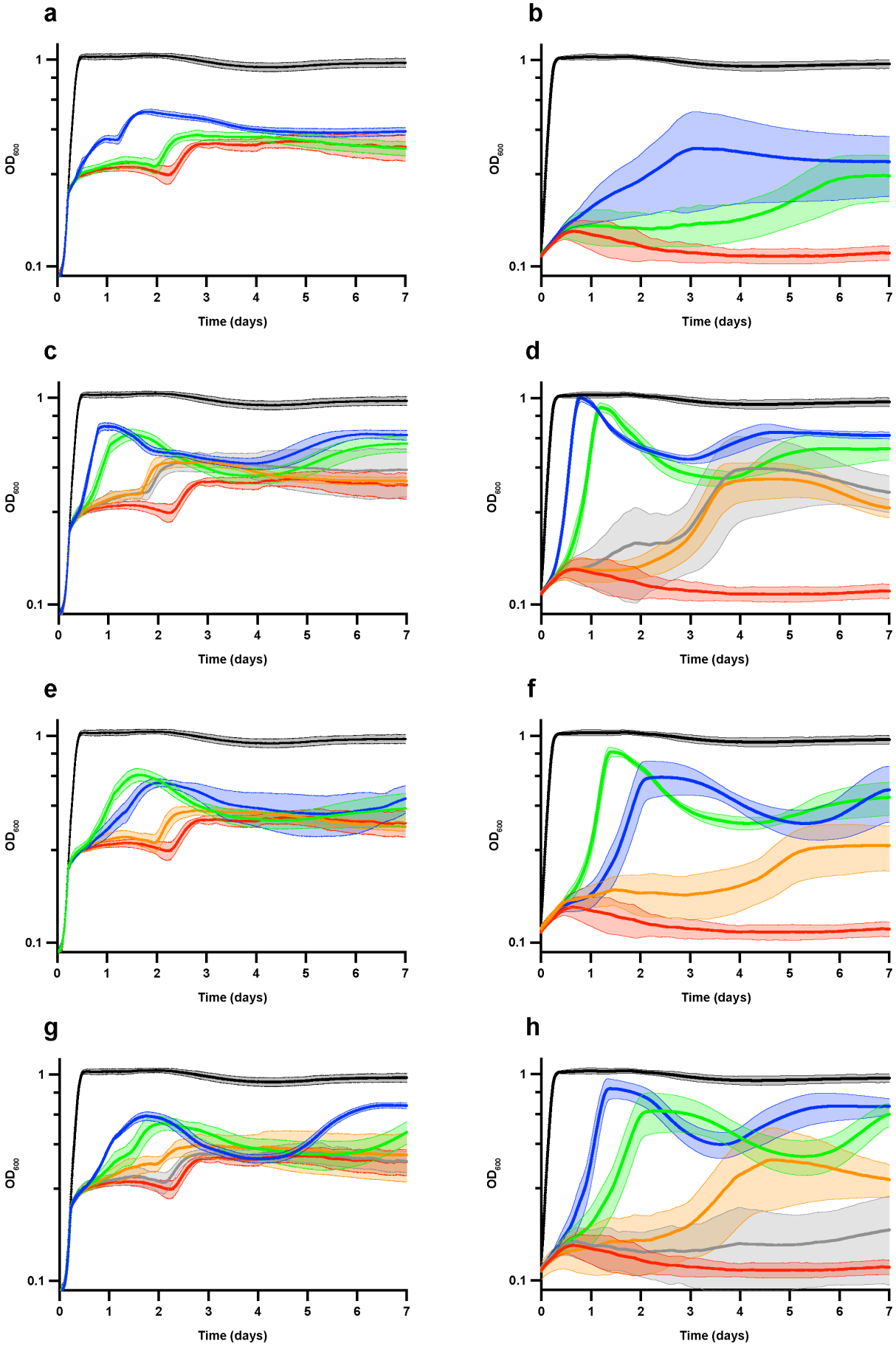
a
b
c
d
e
f
g
h
OD600
1
0.1
0
1
2
3
4
5
6
7
Time (days)

Figure S5. Growth of re-constructed *trpA* mutants. In all graphs the ancestral (*ΔtrpF*; red) and a wild-type (*trpF*+; black) strain is included for comparison. The curves show the average OD_{600} with standard deviation of five biological replicates. **(a, c, e, g)** Cultures grown after a 1000x dilution in 2x M9 + 0.4% glucose supplemented with 5 µM tryptophan + 3 mM guanosine. **(b, d, f, h)** Cultures grown after a 100x dilution in 2x M9 + 0.4% glucose without supplementation. **(a, b)** *trpA* mutations found in the non-mutator populations. Green; *trpA*(A67T), blue; *trpA*(dup173-179). **(c, d)** *trpA* mutations found in mutator populations 2*mutS*4, 2*mutS*11, and 2*mutS*23. Grey; *trpA*(D27Y), orange; *trpA*(G98C), green; *trpA*(D27YG98S), blue; *trpA*(D27Y G98C). (e, f) *trpA* mutations found in mutator population 2*mutS*14. Orange; *trpA*(Y102H), blue; *trpA*(G98S Y102H), green; *trpA*(G98S Y102H Q250R). (g, h) *trpA* mutations found in mutator population 2*mutS*16. Grey; *trpA*(T24S), orange; *trpA*(G61S), green; *trpA*(T24S G61S), blue; *trpA*(T24S G61S P62S).

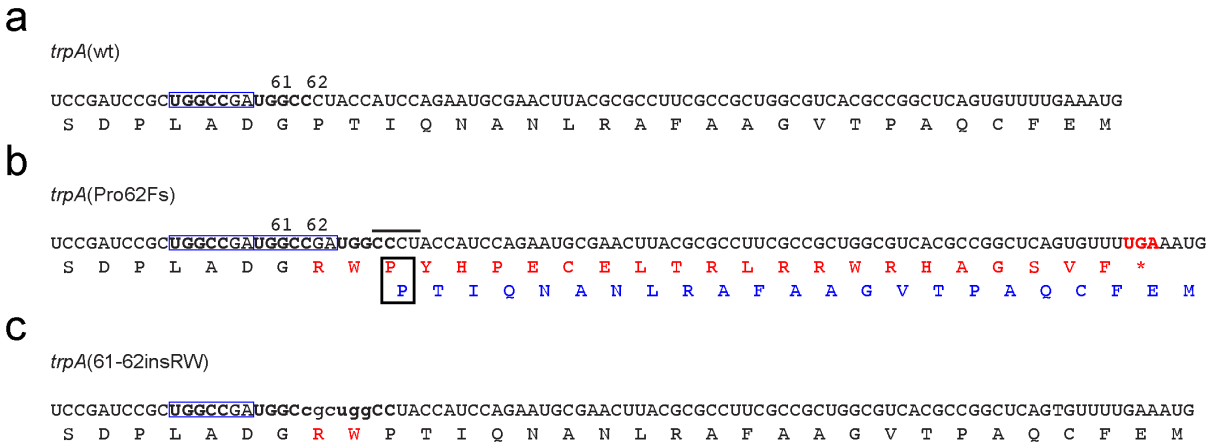

Figure S6. Intrinsic frameshift suppression generates a bifunctional enzyme. **(a)** Codons 55 – 84 in wild-type *trpA*. Top, mRNA sequence; bottom, amino acid sequence. Bold text: five nucleotide repeats involved in generation of the Pro62Fs mutant. Blue box: The 7 bp sequence that is duplicated in the Pro62Fs mutant. **(b)** The corresponding sequence in the *trpA*(Pro62Fs) mutant. Red text in amino acid sequence: misincorporated amino acids if translation continues in the original reading frame. Red text in mRNA: A UGA stop codon encountered in the original frame in the Pro62Fs mutant. Black line: Proline codon in the original frame (CCC) and the +1 frame (CCU). Black box: The hypothetical mechanism of frameshift suppression results in insertion of a single proline at the indicated CCC-U sequence. Blue text: correct amino acids if translation shifts into the +1 frame at the indicated proline codon. **(c)** Constructed *trpA*(61-62insRW) mutant. Lower case letters: Six nucleotides were inserted to generate a two amino acid insertion corresponding to the hypothetical peptide produced if the Pro62Fs mutant is suppressed by +1 frameshifting at the proline codon indicated in (b).

Supplementary References

1. Winkler, M. E. & Ramos-Montañez, S. Biosynthesis of Histidine. *EcoSal Plus* **3**, 10.1128/ecosalplus.3.6.1.9 (2009).
2. Sadler, W. C. & Switzer, R. L. Regulation of Salmonella phosphoribosylpyrophosphate synthetase activity in vivo. Deductions from pool measurements. *J. Biol. Chem.* **252**, 8504–8511 (1977).
3. Henry, T., García-del Portillo, F. & Gorvel, J. P. Identification of Salmonella functions critical for bacterial cell division within eukaryotic cells. *Mol. Microbiol.* **56**, 252–267 (2005).
4. Buckstein, M. H., He, J. & Rubin, H. Characterization of Nucleotide Pools as a Function of Physiological State in Escherichia coli. *J. Bacteriol.* **190**, 718–726 (2008).
5. Leipuviene, R. & Björk, G. R. Alterations in the two globular domains or in the connecting alpha-helix of bacterial ribosomal protein L9 induces +1 frameshifts. *J. Bacteriol.* **189**, 7024–7031 (2007).
6. Arora, S., Bhamidimarri, S. P., Weber, M. H. W. & Varshney, U. Role of the Ribosomal P-Site Elements of m2G966, m5C967, and the S9 C-Terminal Tail in Maintenance of the Reading Frame during Translational Elongation in Escherichia coli. *J. Bacteriol.* **195**, 3524–3530 (2013).
7. Näsvall, S. J., Nilsson, K. & Björk, G. R. The ribosomal grip of the peptidyl-tRNA is critical for reading frame maintenance. *J. Mol. Biol.* **385**, 350–367 (2009).
8. Knöppel, A. *et al.* Genetic Adaptation to Growth Under Laboratory Conditions in Escherichia coli and Salmonella enterica. *Front. Microbiol.* **9**, 756 (2018).
9. Phillips, K. N., Castillo, G., Wünsche, A. & Cooper, T. F. Adaptation of Escherichia coli to glucose promotes evolvability in lactose. *Evolution* **70**, 465–470 (2016).
10. King, T. & Ferenci, T. Divergent roles of RpoS in Escherichia coli under aerobic and anaerobic conditions. *FEMS Microbiol. Lett.* **244**, 323–327 (2005).