COMMENTARY

Experimental evolution heals the scars of genome-scale recoding

Olivier Tenaillon[a,1]

Much of the dramatic plot of Mary Shelley's *Frankenstein* resulted from the apparent scars and imperfections of the creature her hero brought to life. Similarly, organisms highly modified by synthetic biologists suffer from scars and imperfect functioning that their creators had not intended. However, as presented in PNAS by Wannier et al. (1), synthetic biologists can use experimental evolution to rapidly heal some of these scars, as they have performed on a highly recoded *Escherichia coli* strain modified to incorporate efficiently nonstandard amino acids (nsAA).

One of the aims of synthetic biology is to expand the range of functions natural organisms can perform. While chemists have created myriads of new molecules over the last centuries, their ability to create complex molecules is surpassed by the biochemistry that evolved within living organisms during billions years of evolution. However, organisms too have their limits, and a large one is that their proteins use a limited set of amino acids. The needs for robust protein encoding and historical contingencies have limited this set to 20 canonical amino acids. However, there appear to be no short-term constraints to have a more diverse pool of amino acids. Synthetic biologists have, therefore, proposed that protein's biochemical creativity could be unleashed with the incorporation of nsAA (2, 3). These nsAA can provide a finer understanding of protein functions and structures, a precise control on protein activity, or a larger range of functions (4).

Recoding Organisms to Incorporate Artificial Amino Acids

Several steps are required to introduce a nsAA to the existing set of standard amino acids. For translation to occur, amino acids are charged to their cognate tRNA by aminoacyl-tRNA synthetases (aaRS). Then, when the codon complementary to the anticodon part of the tRNA is exposed in the ribosome, the charged tRNA is loaded and the amino acid is transferred from the charged tRNA to the elongating peptide. The uncharged tRNA is subsequently released and the ribosome moves forward to expose the next codon. Incorporating a nsAA therefore requires a combination of codon, tRNA, and aaRS that act specifically and exclusively among themselves and the nsAA. Although every step is challenging, the use of tRNA-aaRS from a diverged organism and the use of positive and negative selections to improve specificity and exclude interactions with other tRNA, aaRS, or amino acids has been quite successful (4). As a result, the machinery for the introduction of the nsAA can be carried on plasmids and easily introduced in new genotypes.

Reserving exclusively a codon for the nsAA is yet another challenge, as it requires multiple modifications of the genome. The degeneracy of the genetic code offers numerous opportunities for extending the genetic code to more than 20 amino acids. Ideally, up to 63 amino acids and a stop codon could be used. However, all 64 codons are used in the genome and one cannot simply attribute one of them to the new tRNA-aaRS without dramatically altering survival, as all proteins using that codon will be harmed to some extent by the nsAA. To avoid these collateral damages, the strategy is to replace genome-wide a given codon by an alternative codon coding for the same amino acids or stop-codon. The replaced codon can then be used exclusively for the nsAA. Fortunately, codon usage varies largely and some codons are better candidates for a genome-wide operation.

The amber stop codon, UAG, is the rarest codon in the *E. coli* genome. There is only one stop codon per protein and among the three stop codons, UAG is the least-frequently used, with just 321 instances in some genotypes. Using multiplex automatable genome engineering (MAGE) (5), a strategy based on recombineering (genetic engineering with recombination), mutagenic oligos can be used to introduce multiple mutations simultaneously in the chromosome with high efficiency. A strain without any UAG stop codon was hence produced (6). As no tRNA in the genome

[a]Infection, Antimicrobials, Modelling, Evolution, INSERM, Unité Mixte de Recherche 1137, Université Paris Diderot, Université Paris Nord, 75018 Paris, France

Author contributions: O.T. wrote the paper.

The author declares no conflict of interest.

Published under the PNAS license.

See companion article on page 3090.

[1]Email: olivier.tenaillon@inserm.fr.

Published online March 7, 2018.

matches UAG, incorporating the nsAA as a UAG codon then requires just the addition of a tRNA-aaRS system in which the anticodon of the tRNA is modified to match UAG. However, UAG exposed in the ribosome is attracting the release factor RF1 (encoded by *prfA*), which will improve translation termination. RF1 may then compete with the introduced UAG-tRNA and lead to truncated proteins. One simple solution is therefore to delete *prfA* from the genome.

The Genomic Scars and Their Healing

Unfortunately, the introduction of multiple mutations in the genome is not a scarless method. First, recombineering, which is based on the use of a phage recombinase (7) to promote recombination of linear DNA in the genome, may promote recombination between repeated parts of the chromosome. Second, recombination of the mutagenic oligonucleotides in the chromosome generates transiently a mismatch at the base of interest. Such mismatch can be recognized by the mismatch repair system (MMR), which aborts the recombination process with high efficiency (8). Consequently, to boost the chances of recombination, the strains used to do MAGE are deficient in MMR. Their mutation rate is therefore 100-fold higher as replication errors are not properly corrected. Hence, the production of a modified genome comes along with the accumulation of many alternative mutations whose effect on fitness may be important. Finally, the intended modifications of the genome, such as the deletion of RF1 that is also recruited for UAA termination, may have deleterious fitness effects.

The construction of the strain lacking amber codons and RF1 resulted in the accumulation of 355 alternative mutations and had about 60% decrease in growth rate (6). To alleviate this burden, one strategy is to use evolution. Experimental evolution with multiple replicates and whole-genome sequencing is a powerful tool not only to optimize the fitness of strains in a given environment, but also to identify the molecular bases of that adaptation (9). Observing similar changes in independent replicates uncovers the filtering action of natural selection, and therefore points to the molecular targets of adaptation.

Adaption of a highly modified strain may proceed through different pathways. First, one hopes, using experimental evolution, to reverse or to compensate for the unwanted scars of the modification process. For point mutations, the exact reversion of the mutation is possible, but in most cases, mutations elsewhere in the affected protein or pathway are recruited, as they are much more numerous than exact reversion. The compensation, though, may be partial. For example, compensation of antibiotic resistance mutations is often limited (10). The compensation of gene complete or partial deletion is also more difficult to fulfill, as it requires the existence of alternative genes or pathways to compensate for the irreversible loss of the gene (11). Aware of these potential limitations, Wannier et al. (1) used a mixed strategy and performed experimental evolution on a strain in which three identified deleterious mutations were precisely reverted.

Second, the experimental evolution can reduce the cost associated with the intended modifications: here, the inactivation of *prfA* and the lack of amber codons. Through the analysis of multiple replicates, it appeared that some specific mutations affected *prfB* and *prfC*, which encode for two other release factors, RF2 and RF3. RF2 acts as RF1 but on the UGA and UAA codon, while RF3 favors the release of RF1 and -2 from the ribosome. These mutations were found at high frequency during the adaptation of the modified strain but not recovered during the evolution nonmodified strains in the same media. The pattern of occurrence and the specificity of the function prove that these mutations compensate for the change in RF1 balance that the lack of the *prfA* or the lack of amber codons imposes. When introduced in isolation, these mutations could decrease growth rate of the recoded strain by up to 30%. It is worth noting, however, that in other cases, the beneficial mutations that respond to the modifications imposed on the genome could be troublesome. For example, engineered strains may involve unstable and costly metabolic modifications. In that case, adaptation may simply damage the engineered properties.

Third, adaptation to the laboratory conditions may also be selected for. Wannier et al. (1) recovered multiple mutations affecting biofilm formations, or the RNA polymerase. These mutations are not only recovered during adaptation of the nonmodified strain, but also in other experimental evolution. Although these mutations can be seen as a positive, they also have a down side: they can be environment-specific. In other words, some of the mutations recruited in one media may be costly in another one. Particularly, mutations in *rpoB* that are rarely seen in nature are often found during adaptation to minimal media, but can be deleterious in rich media.

In Wannier et al., the use of experimental evolution was a success that resulted in the production of an *E. coli* strain lacking amber codons, which has a 30-min division time in minimal media and a high efficiency of nsAA incorporation.

Fourth, nonadaptive mutations may also be recruited during adaptation of asexual organisms, such as *E. coli*. Beneficial mutations may invade the population even if they occur on genomes loaded with neutral or slightly deleterious mutants, as long as their benefit overweighs the combined costs of these mutations. Furthermore, as beneficial mutations, some of these passenger mutations may have high costs in alternative environments. While the fraction of passenger mutations is very limited in low-mutation rate strains (12), it is high in the high-mutation rate strains (13). The use of a high mutation rate is therefore a double-edged sword. On the one hand, it improves slightly the rate of adaptation (14); on the other hand, it multiplies the number of passenger mutations and also blurs the genomic signature of adaptation, as beneficial mutations are lost in a sea of passenger mutations (13).

In Wannier et al. (1), the use of experimental evolution was a success that resulted in the production of an *E. coli* strain lacking amber codons, which has a 30-min division time in minimal media and a high efficiency of nsAA incorporation. Nevertheless, for future constructions, limiting the production of collateral mutations could be worthwhile. To this end, the use of inducible MMR deficiency (15) during recombineering seems promising. Furthermore, the evolution with a sexual system may limit the burden of deleterious passenger mutations and fasten adaptation (16, 17).

Overall, experimental evolution is a powerful way to optimize highly modified genomes and make them as good as their native counterparts. However, the fate of a scarless or even handsome Frankenstein creature could have been somehow more terrifying than the tragic destiny to which his scars led him. Fortunately, one of the aims of recoding the genetic code is to create synthetic organisms that are fully dependent upon artificial amino acids and, therefore, confined to the laboratory (18).

1 Wannier TM, et al. (2018) Adaptive evolution of genomically recoded *Escherichia coli*. *Proc Natl Acad Sci USA* 115:3090–3095.
2 Chin JW (2017) Expanding and reprogramming the genetic code. *Nature* 550:53–60.
3 Mukai T, Lajoie MJ, Englert M, Söll D (2017) Rewriting the genetic code. *Annu Rev Microbiol* 71:557–577.
4 Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. *Annu Rev Biochem* 79:413–444.
5 Wang HH, et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460:894–898.
6 Lajoie MJ, et al. (2013) Genomically recoded organisms expand biological functions. *Science* 342:357–360.
7 Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97:6640–6645.
8 Matic I, Rayssiguier C, Radman M (1995) Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species. *Cell* 80:507–515.
9 Achaz G, Rodriguez-Verdugo A, Gaut BS, Tenaillon O (2014) The reproducibility of adaptation in the light of experimental evolution with whole genome sequencing. *Ecological Genomics, Advances in Experimental Medicine and Biology*, eds Landry CR, Aubin-Horth N (Springer, Dordrecht, The Netherlands), pp 211–231.
10 Schrag SJ, Perrot V, Levin BR (1997) Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*. *Proc Biol Sci* 264:1287–1291.
11 Blank D, Wolf L, Ackermann M, Silander OK (2014) The predictability of molecular evolution during functional innovation. *Proc Natl Acad Sci USA* 111:3044–3049.
12 Tenaillon O, et al. (2016) Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 536:165–170.
13 Couce A, et al. (2017) Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proc Natl Acad Sci USA* 114:E9026–E9035.
14 Wiser MJ, Ribeck N, Lenski RE (2013) Long-term dynamics of adaptation in asexual populations. *Science* 342:1364–1367.
15 Nyerges Á, et al. (2016) A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc Natl Acad Sci USA* 113:2502–2507.
16 McDonald MJ, Rice DP, Desai MM (2016) Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* 531:233–236.
17 Peabody V GL, Li H, Kao KC (2017) Sexual recombination and increased mutation rate expedite evolution of *Escherichia coli* in varied fitness landscapes. *Nat Commun* 8:2112.
18 Rovner AJ, et al. (2015) Recoded organisms engineered to depend on synthetic amino acids. *Nature* 518:89–93.