G3 Genes | Genomes | Genetics

INVESTIGATION

Reduced Protein Expression in a Virus Attenuated by Codon Deoptimization

Benjamin R. Jack,*,†,‡,[1] Daniel R. Boutz,‡ Matthew L. Paff,*,†,‡ Bartram L. Smith,*,†,‡ James J. Bull,*,†,‡ and Claus O. Wilke*,†,‡

*Department of Integrative Biology, †Center for Computational Biology and Bioinformatics, and ‡Institute for Cellular and Molecular Biology, The University of Texas at Austin, Texas 78712

ORCID IDs: 0000-0001-7777-1653 (B.R.J.); 0000-0002-7470-9261 (C.O.W.)

ABSTRACT A general means of viral attenuation involves the extensive recoding of synonymous codons in the viral genome. The mechanistic underpinnings of this approach remain unclear, however. Using quantitative proteomics and RNA sequencing, we explore the molecular basis of attenuation in a strain of bacteriophage T7 whose major capsid gene was engineered to carry 182 suboptimal codons. We do not detect transcriptional effects from recoding. Proteomic observations reveal that translation is halved for the recoded major capsid gene, and a more modest reduction applies to several coexpressed downstream genes. We observe no changes in protein abundances of other coexpressed genes that are encoded upstream. Viral burst size, like capsid protein abundance, is also decreased by half. Together, these observations suggest that, in this virus, reduced translation of an essential polycistronic transcript and diminished virion assembly form the molecular basis of attenuation.

KEYWORDS
bacteriophage T7
codon deoptimization
recoding
translation
viral attenuation

A recent methodological advance in the development of attenuated viral vaccines has been the design of genomes with hundreds of synonymous codon changes that are collectively suboptimal (Burns *et al.* 2006, 2009; Mueller *et al.* 2006, 2010; Coleman *et al.* 2008; Wimmer *et al.* 2009; Bull *et al.* 2012; Nougairede *et al.* 2013; Shen *et al.* 2015). The nature of the design varies from the replacement of common codons with rare codons to merely shuffling existing codons to create uncommon codon pairs. By maintaining wild-type protein sequences, the recoding method retains the antigenic profile of the wild type while reducing viral growth rate and virulence. Based on the premise that silent codon changes have individually small effects, the method should not only allow the tunable crippling of viral growth to an arbitrary degree but also profoundly retard the reevolution of high viral fitness. Both predictions have been supported empirically. Thus, poliovirus, influenza, several arboviruses, and a bacteriophage all exhibit quantitative fitness declines after synonymous codon changes (Burns *et al.* 2006; Mueller *et al.* 2006, 2010; Coleman *et al.* 2008; Nougairede *et al.* 2013; Shen *et al.* 2015; Bull *et al.* 2012), and fitness recovery during viral growth over hundreds of generations is at best slow (Burns *et al.* 2006; Coleman *et al.* 2008; Bull *et al.* 2012; Nougairede *et al.* 2013).

The broad success of attenuation from synonymous codon changes in different viruses and with different designs could arise from a common underlying mechanism. Yet, there is ongoing debate about how synonymous codon changes affect fitness and thus what that mechanism could be. One popular hypothesis is that codon usage controls translational efficiency, in turn affecting the rate of protein synthesis (Fredrick and Ibba 2010; Plotkin and Kudla 2010; Tuller *et al.* 2010; Shah and Gilchrist 2011; Zur and Tuller 2016). Under this hypothesis, highly expressed and functionally important genes are encoded by optimal codons to increase translational efficiency. However, a simple codon "optimality" model cannot explain the attenuation attained by merely shuffling codons; since the abundance of the different codons is not being changed by shuffling, the suboptimality must be due to something besides codon abundance. Likewise, some highly expressed genes in cyanobacteria and *Neurospora* have nonoptimal encodings (Xu *et al.* 2013; Zhou *et al.* 2013). In other cases, codon usage determines expression through transcription, not translation (Zhou *et al.* 2016). In the face of so many seemingly contradictory observations, further advances in this research program would benefit from the identification of a common mechanism for viral attenuation, or at least benefit from the demonstration that different mechanisms are involved.

doi: https://doi.org/10.1534/g3.117.041020
Manuscript received March 2, 2017; accepted for publication June 27, 2017; published Early Online July 11, 2017.

[1]Corresponding author: The University of Texas at Austin, 2500 Speedway, A4800, Austin, TX 78712. E-mail: benjamin.r.jack@gmail.com

Resolving the basis of attenuation in viral systems is partly hampered by the sequence of life-history steps between the initial effects of codon changes and the final emergence of assembled virions. The initial impact may lie in transcription or translation of one or more genes, but the effect on the number of virions will depend on which proteins are limiting during assembly. The same modification of an early life-history stage may have different fitness effects in different viruses. Ultimately, it may be necessary to interpret the genome engineering in the context of a system-wide, comprehensive model of the viral life cycle. Such is our motivation.

In the bacterial virus T7, recoding the major capsid protein (gene *10A*) with synonymous codons reduced the fitness of the phage (Bull *et al.* 2012). To engineer T7, codons in gene *10A* that were highly utilized in *Escherichia coli* (the T7 host) were replaced with codons that were underutilized in the host. The major capsid protein, which forms the head of the T7 phage particle, is the most abundant and highly expressed phage protein (Dunn and Studier 1983). In the phage genome with the most extensive set of gene *10A* synonymous codon replacements, the fitness was 35.7 doublings/hr compared to 43.2 doublings/hr in the wild type (Bull *et al.* 2012). This difference translates to a 180-fold decline in descendants produced per hour. In total, 182 codons were changed, just over half the number of codons in the major capsid gene; in addition, the first 20 codons were not altered to avoid disrupting translation initiation processes.

After adapting the recoded phage for 1000 generations, fitness increased to 38.7, a recovery of almost half the initial deficit (on a log scale). Thus, recoding gene *10A* induced a moderately stable fitness reduction. A mere nine nucleotide changes were responsible for the fitness recovery, and seven fell outside the recoded region, shedding little light on the underlying mechanism of fitness reduction.

Here, we apply new methods to continue exploration of the T7 attenuation. Using the same strains and designs as the prior study (Bull *et al.* 2012), our purpose is to develop a comprehensive model of the way silent codon changes cause reduced fitness. As part of this effort, we propose and test three mechanistic models that could explain the fitness reduction in recoded T7. In testing those models, we apply proteomic methods, RNA sequencing, and various phenotypic measures in a systems approach to understanding the basis of attenuation. We find that recoding gene *10A* reduces protein abundances of gene *10* and also of several downstream genes. From there, we address the impact of protein abundances on viral fitness components (burst size and lysis time), ultimately connecting these measurements to a model that describes actual fitness.

MATERIALS AND METHODS

Gene 10 nomenclature

Gene *10* is translated in two forms, A and B. Form A is 344 amino acids and is formally denoted the major capsid protein. Form B (the minor capsid protein) is not essential and results from a ribosomal frameshift at the end of A and is 397 amino acids. In the engineering, all codon changes were within *10A* and thus also within *10B*. Moreover, since most peptide fragments coming from the minor and major capsid proteins ambiguously mapped to both proteins, abundances of 10A and 10B were not differentiated using our proteomic methods and were combined. We followed a similar procedure for our RNA-sequencing analyses. To simplify notation, we merely refer to the recoded gene as *10* and the affected A and B proteins as capsid protein.

Bacteriophage T7 strains and E. coli hosts

The host for all experiments was IJ1133 [*E. coli* K-12, F-ΔlacX74 thiΔ (mcrC-mrr)102::Tn10]. T7 strains used in this study come from previous work (Bull *et al.* 2012). An isolate of $T7_{61}$ (a population adapted to grow optimally on IJ1133 specifically through serial passage) was first deleted of its gene *10*, then recombined over a plasmid carrying a different gene *10* engineered to contain a low fraction (0.1) of preferred codons, with 182 codon changes. The recombinant, denoted here as the recoded strain, could be identified by its ability to grow without complementation. The evolved strain was initiated from the recoded phage and adapted over 800–1000 generations [strain L1 from Bull *et al.* (2012)]. The wild-type strain in this study was derived from the recoded strain after recombination over a plasmid containing wild-type gene *10*, then grown out for 6 hr of serial transfer on IJ1133. Fitness of this "wild-type" strain was approximately the same as that of the ancestral population ($T7_{61}$).

Burst size and lysis time

Lysis time and burst size assays were performed as previously described (Heineman and Bull 2007; Bull *et al.* 2011). The initial infection steps were identical for both assays. Briefly, 10^8 phage (MOI = 0.1) were added to a 10 ml culture of exponentially growing cells (37° with agitation), incubated for 3 min, and subsequently diluted 10^4-fold to prevent further adsorption. For lysis times, phage were plated at various time points between 4 and 18 min (after initial infection) to monitor changes in titer; lysis time was taken as the time of the first significant increase in titer.

To determine burst size, initial density of phage-infected cells was determined by plating phage before and after treatment with chloroform 5 and 6 min after initial infection. Cells infected with phage at the time of chloroform treatment do not produce viable phage, so only free phage will form plaques, allowing for the determination of phage-infected cells at these times. Final phage titers were obtained at 15, 16, and 17 min by plating chloroform-treated samples. Burst size was then calculated as the phage titer at the end time points divided by the number of initial phage-infected cells.

RNA sequencing

E. coli was grown in LB broth to a concentration of 10^8 cells/ml at 37° with agitation, then infected with phage at an MOI of 2.5. At 1, 5, and 9 min postinfection, 2 ml of bacterial suspension were removed from the phage-infected cultures and pelleted in a microcentrifuge. Pellets were either flash frozen in liquid nitrogen or immediately used for downstream processes (RNA extraction or protein preparation for proteomics). RNA was isolated using Trizol reagent, following the manufacturer's protocol. Library preparation and sequencing was performed by the University of Texas Genome Sequencing and Analysis Facility using Illumina NextSequation 500 (SR75).

Since gene *10B* is a readthrough product of gene *10A*, we excluded the gene *10B* transcript from the reference transcriptome. RNA-sequencing reads were quantified using Kallisto (Bray *et al.* 2016) and *E. coli* K-12 (NCBI: U00096.3) and T7 (NCBI: NC_001604.1) reference genomes. For analyses of the recoded and evolved T7 strains, the gene *10* sequence was replaced with the recoded sequence (supplementary file S1 in Bull *et al.* 2012,) in the reference genome. A population of T7-infected *E. coli* has no core set of stably expressed genes with which to normalize during differential expression analysis. Therefore, all transcript abundance estimates (transcripts per million) were normalized to the total cellular transcript abundance (including both T7 and *E. coli* transcripts). Differential expression analysis was only possible within genes across treatments, but not between genes in the same treatment.

Proteomics

Proteomics was performed as previously described in Houser *et al.* (2015). In brief, T7-infected *E. coli* cell pellets (prepared as described in the *RNA sequencing* section above) were resuspended in 50 mM Tris-HCl, pH 8.0,

10 mM DTT. 2,2,2-trifluoroethanol (Sigma-Aldrich) was added to 50% (v/v) final concentration and samples were incubated at 56° for 45 min. Following incubation, iodoacetamide was added to a concentration of 25 mM and samples were incubated at room temperature in the dark for 30 min. Samples were diluted 10-fold with 2 mM $CaCl_2$, 50 mM Tris-HCl, pH 8.0. Samples were digested with trypsin (Pierce) at 37° for 5 hr. Digestion was quenched by adding formic acid to 1% (v/v). Tryptic peptides were bound, washed, and eluted from HyperSep C18 SpinTips (Thermo Fisher Scientific). Eluted peptides were dried by speed-vac and resuspended in Buffer C (5% acetonitrile, 0.1% formic acid) for analysis by LC-MS/MS.

For LC-MS/MS analysis, peptides were subjected to separation by C18 reverse phase chromatography on a Dionex Ultimate 3000 RSLCnano UHPLC system (Thermo Fisher Scientific). Peptides were loaded onto an Acclaim C18 PepMap RSLC column (Dionex; Thermo Fisher Scientific) and eluted using a 5–40% acetonitrile gradient over 250 min at 300 nl/min flow rate. Eluted peptides were directly injected into an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) by nano-electrospray and subjected to data-dependent tandem mass spectrometry, with full precursor ion scans (MS1) collected at 60,000 resolution. Monoisotopic precursor selection and charge-state screening were enabled, with ions of charge >+1 selected for collision-induced dissociation. Up to 20 fragmentation scans (MS2) were collected per MS1. Dynamic exclusion was active with 45-sec exclusion for ions selected twice within a 30-sec window.

We assigned each peptide to a protein or protein group (in the case of ambiguous peptides which map to multiple proteins) using Proteome Discoverer (Thermo Fisher Scientific) and REL606 and T7 reference proteomes (NCBI: NC_012967, NC_001604.1) concatenated with a database of contaminant proteins (http://www.biochem.mpg.de/5111795/maxquant). We selected the top three most-abundant peptides by peak area for each protein. We averaged these peptide peak areas across technical replicates to obtain a protein-abundance estimate (Silva *et al.* 2006). All protein-abundance estimates were normalized to the total *E. coli* protein content of the sample.

To determine if gene *10* C-terminal peptides were more common than N-terminal peptides in the recoded T7 strain compared to the wild-type strain, we compared the fit of the following two models [given here in notation from the lme4 R package (Bates *et al.* 2015)]:

$$\text{count} \sim \text{strain} + \text{location} + (1|\text{peptide}) \quad (1)$$

$$\text{count} \sim \text{strain} + \text{location} + \text{strain:location} + (1|\text{peptide}), \quad (2)$$

where count is the number of peptides of type peptide, strain is the strain of T7, and location is the location of peptide within gene *10*. We assume peptide to be a random effect, given by the (1|peptide) term. The term strain:location indicates interaction between strain and location. Thus, we compare a model in which the location of a peptide and the strain interact, and a model in which there is no such interaction. If N-terminal peptides were less prevalent than C-terminal peptides in the recoded T7 strain, we would expect Equation 2 to provide a better fit than Equation 1.

Models of translational coupling

Biophysical model: Secondary structure near the ribosome binding site (RBS) can inhibit translation initiation. On polycistronic transcripts, this secondary structure can be disrupted by ribosomes completing translation of an upstream gene, thus increasing translation initiation rates. Estimates of this relative increase in translation initiation of gene *11* due to translation of gene *10* was predicted using the Operon Calculator (Tian and Salis 2015).

Mathematical model: To model the effects of translational coupling on protein production, we first assume a polycistronic transcript with three genes *a*, *b*, and *c*. We write the effective initiation rate a_i of gene *a* as

$$a_{\text{in}} = \min\{i_a, \tau_a\}, \quad (3)$$

where i_a is the aggregate initiation rate of *a* and τ_a is the translation elongation rate of gene *a*. We assume that if the initiation rate ever exceeds the elongation rate, ribosomes will quickly back up on the transcript and make elongation the rate-limiting step of translation. Thus, in our model, the elongation rate can never be exceeded by the aggregate translation initiation rate. For gene *a*, the aggregate initiation rate i_a is simply the *de novo* initiation rate because there are no genes upstream of *a*.

The rate at which ribosomes complete translation of gene *a*, equivalent to the production rate of protein A, is defined as

$$\dot{A} = a_{\text{in}} \quad (4)$$

at steady state. In this context, the steady state assumption means that all three protein products are being produced continuously at the equilibrium rate.

For the effective translation initiation rate b_{in} of gene *b*, we similarly write

$$b_{\text{in}} = \min\{i_b, \tau_b\}, \quad (5)$$

where i_b is the aggregate translation initiation rate due to upstream-dependent reinitiation and *de novo* initiation, and τ_b is the translation elongation rate of *b*. We define the aggregate translation initiation rate

$$i_b = b_{\text{reinit}} + b_{de\ novo}, \quad (6)$$

where b_{reinit} is the rate of upstream translating ribosomes reinitiating on gene *b*, and $b_{de\ novo}$ is the rate of ribosomes initiating *de novo* on gene *b*.

Lastly we define reinitiation and *de novo* initiation rates on gene *b* as follows:

$$b_{\text{reinit}} = q_b\dot{A}, \quad (7)$$

$$b_{de\ novo} = z_b\dot{A} + w_b, \quad (8)$$

where $\dot{A}$ is the rate of ribosomes flowing from upstream translation of gene *a*, and q_b represents the proportion of that ribosome flow reinitiating on gene *b*. We assume that the rate of *de novo* initiation depends, in part, on upstream ribosomes relaxing secondary structure around the RBS of gene *b*. Thus, the rate of *de novo* initiation is given by the upstream ribosome flow $\dot{A}$ scaled by some constant z_b (facilitated binding), and by a constant rate w_b that does not depend on upstream ribosome flow.

We can simplify the effective initiation rate to

$$b_{\text{in}} = \min\{\dot{A}(q_b + z_b) + w_b, \tau_b\}. \quad (9)$$

We simplify the effective initiation rate further by defining a coupling constant,

$$y_b = q_b + z_b, \quad (10)$$

which accounts for both the effects of facilitated binding and ribosome reinitiation. The final effective initiation rate of *b* is defined as

$$b_{\text{in}} = \min\{y_b\dot{A} + w_b, \tau_b\}. \quad (11)$$

Similar to the protein production rate of gene *a*, we define the protein production of gene *b*

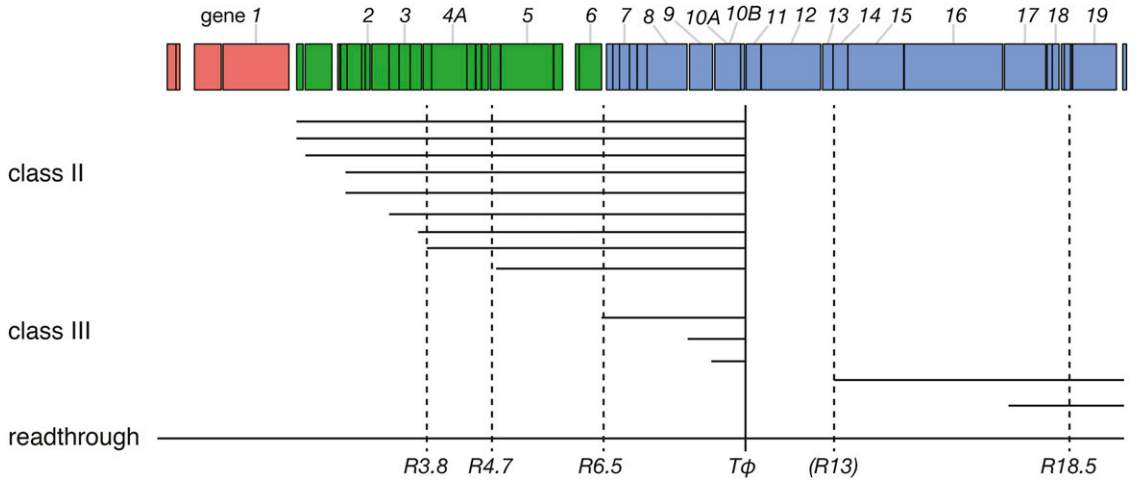

Figure 1 T7 produces many polycistronic transcripts. The bar across the top shows the T7 genome, and each class is shown in a different color. Horizontal lines represent transcripts. Dashed vertical lines represent RNase cleavage sites, where R*3.8*, R*4.7*, R*6.5*, and R*18.5* are strong cleavage sites. R*13* is a weak RNase cleavage site. The solid vertical line represents the terminator Tφ. Genes *11* and *12* are only ever expressed as a product of readthrough of Tφ, indicated by the readthrough transcript. Only transcripts containing class-III genes are shown. Not all read-through products are shown. Redrawn from Dunn and Studier (1983).

$$\dot{B} = b_{\text{in}} \tag{12}$$

at steady state.

The effective initiation rate of gene c is similar to that of gene b

$$c_{\text{in}} = \min\{y_c\dot{B} + w_c, \tau_c\}, \tag{13}$$

$$y_c = q_c + z_c, \tag{14}$$

where $\dot{B}$ is the rate of ribosomes flowing from the end of upstream gene b, and q_c represents the proportion of that ribosome flow reinitiating on gene c. The rate of *de novo* initiation dependent on upstream ribosome flow (facilitated binding) is given by z_c, and w_c is the *de novo* initiation rate independent of upstream ribosome flow. Again, y_c is a coupling constant that incorporates the effects of both facilitated binding and ribosome reinitiation.

Statistical software and plots

All statistical tests were conducted using the R language (R Core Team 2014). All plots were generated using the ggplot2 package (Wickham 2009).

Data availability

Raw RNA reads are available at NCBI GEO (ID: GSE96573) (Barrett *et al.* 2013). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno *et al.* 2016) partner repository with the data set identifier PXD006502. All processed data and scripts are available at https://github.com/benjaminjack/phage_attenuation.

RESULTS

Codon deoptimization reduces capsid protein abundances

As codon deoptimization is thought to affect translational efficiency, we propose three models in which recoding gene *10* (capsid protein) affects protein abundances in T7. In model 1, codon deoptimization slows translation of gene *10* and reduces the abundance of the capsid protein only. In model 2, deoptimization depletes the ribosome pool by creating high ribosomal densities on capsid protein transcripts, thus reducing translation of all viral proteins late in the infection cycle (Vind *et al.* 1993; Birch *et al.* 2012; Raveh *et al.* 2016). In model 3, codon deoptimization has intermediate effects between models 1 and 2: translation is impaired for gene *10* and downstream genes but not for upstream genes. The expectation in model 3 arises because T7 produces many polycistronic transcripts, leading to translational coupling of gene *10* and genes immediately downstream. Translational coupling has been observed in bacterial operons but has not been considered in the context of codon deoptimization (Oppenheim and Yanofsky 1980; Schümperli *et al.* 1982; Aksoy *et al.* 1984; Tian and Salis 2015). All three models assume that deoptimizing gene *10* will reduce the abundance of at least the capsid protein.

To differentiate between our three proposed models, we compared the T7 proteome during infection among wild-type, attenuated, and evolved phages using mass spectrometry-based protein quantitation. T7 is thought to encode 58–60 proteins, but only 19 are essential, and many have no known function (Dunn and Studier 1983; Molineux 2006) (*e.g.*, some are thought to be homing endonucleases, selfish elements). All genes are encoded on the same strand, and expression order is linear; genes are numbered in order with essential genes having integral numbers (*1–19*) and nonessential genes having fractional numbers. The genome is divided into three expression groups. Class-I genes are the first to enter and are expressed from promoters at the entering end of the genome, transcribed by the host RNA polymerase (RNAP). The phage RNAP gene (numbered gene *1*) is the last of the class-I genes and the first essential gene. All other genes are expressed from phage promoters, but nearly all transcripts are polycistronic as there is only one terminator for phage RNAP (immediately after gene *10*), and there are only 17 phage promoters (Figure 1). Gene *1.1* is the first class-II gene (Dunn and Studier 1983). [Note that genes *1.1–1.3* are sometimes also considered to be part of class I, because they are transcribed by both *E. coli* and T7 RNAP (Molineux 2006).] Gene *6.5* is the first class-III gene.

Because the wild-type phage lyses the cell at ~11 min after infection (Heineman and Bull 2007; Bull *et al.* 2011), the proteome of infected

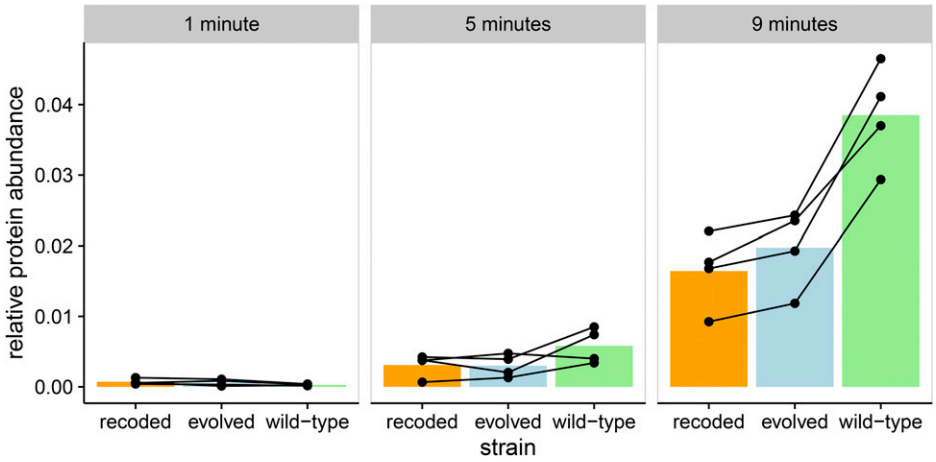

Figure 2 Recoding gene *10* reduces capsid protein abundances. We measured protein abundance at 1, 5, and 9 min after infection. In the recoded (attenuated) strain, protein abundance for capsid protein after 9 min of infection is half of that of the wild type ($p < 0.05$, paired *t*-test). The evolved strain also has significantly lower levels of capsid protein after 9 min. Each point represents a single measurement, and lines connect biological replicates.

hosts was sampled at 1, 5, and 9 min after phage addition to the culture; infection of cells is neither immediate nor synchronous upon phage addition to the culture, so these times are approximate postinfection values. By 9 min after infection, ~50 of the known or predicted T7 proteins were detected. All samples recovered ~4000 *E. coli* proteins. Since proteins have a much longer half-life than transcripts, and T7 has no known mechanism of degrading *E. coli* proteins (Molineux 2006), we assumed that *E. coli* protein abundances remained constant over the 9 min infection, and normalized the phage protein abundances to that of *E. coli* (Houser *et al.* 2015). Thus, we report all phage protein abundances as a proportion of *E. coli* protein content.

Abundances of the major capsid protein (a product of gene *10*) were of primary interest, as gene *10* is the most highly expressed phage gene and is also the one deoptimized. Gene *10* comprises two protein products: the major capsid protein (gp10A) and the much less abundant minor capsid protein (gp10B). The minor capsid protein is produced after a frameshift and stop codon readthrough of *10A* and, except for the C-terminal ~53 amino acids, is identical in sequence to the major capsid protein. Thus, our proteomic methods have limited ability to distinguish between the two protein products, so we combine abundance estimates into a single capsid protein measurement (see *Materials and Methods*). By 9 min after infection, capsid protein abundances in the attenuated strain were about half of those in the wild type ($p < 0.05$, paired *t*-test; Figure 2). The capsid protein abundance for the evolved strain was intermediate.

Recoding of gene *10* could have reduced capsid protein abundance by reducing rates of translation elongation, thereby increasing the likelihood of ribosome stalling and fall-off. If ribosome fall-off were the dominant mechanism by which protein abundance was reduced, attenuation should have been accompanied by an excess of short peptides from the N-terminal end. Alternatively, if translation was slowed down without ribosome fall-off, a uniform distribution of peptides should be observed across the capsid protein. When mapping individual peptides recovered from the mass spectrometry proteomics, no systematic change was observed in the distribution of peptides across the protein (Figure 3). Thus, the recoded phage strains produced complete capsid protein, but in smaller quantities than that of the wild type.

Codon deoptimization reduced some other class-III protein abundances

If recoding a highly expressed gene saturates the pool of ribosomes by slowing translation of the recoded gene, the rate of translation of all T7 genes could ultimately decline (model 2, above). Models of T7 replication in *E. coli*, and limited experimental data, are consistent with protein synthesis being the rate-limiting step of T7 replication (although the evidence is at best weak and indirect) (Endy *et al.* 1997; You *et al.* 2002). Moreover, depletion of free ribosomes is common in *E. coli* transformed with highly expressed heterologous genes (Vind *et al.* 1993; Scott *et al.* 2010; Reuveni *et al.* 2011; Raveh *et al.* 2016).

The data allow us to measure other T7 protein abundances over time. 9 min postinfection, abundances of class-III proteins gp11, gp12, gp13, gp14, and gp15 in the recoded strain were lower than that of the wild type

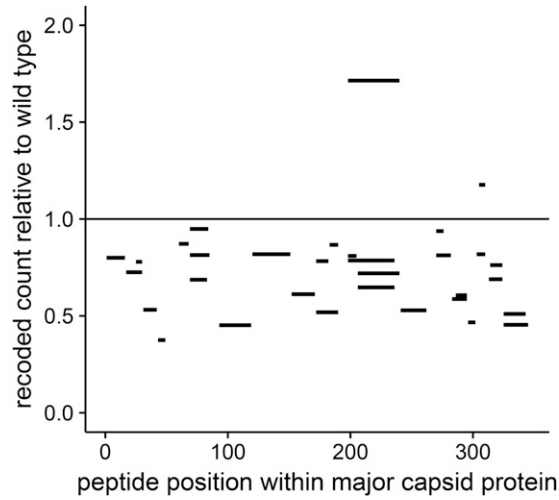

Figure 3 Peptide abundances are uniformly distributed across the capsid protein in the recoded T7 strain. Shown are the abundances of individual peptides within the capsid protein of the recoded strain, relative to the wild type. Data are from 9 min after infection. If translation consistently terminated before the stop codon in recoded gene *10*, we would expect the relative peptide abundance to systematically vary with the location of the peptide. Therefore, we tested a model that includes an interaction between strain and peptide location, and a model that includes no interaction (see *Materials and Methods*). We found that a model in which an interaction is included between the strain and peptide location fits the peptide count data no better than a model without interaction (log-likelihood = −1682.2, log-likelihood = −1682.4, respectively). Thus, in the recoded strain, there is no evidence of early translation termination.

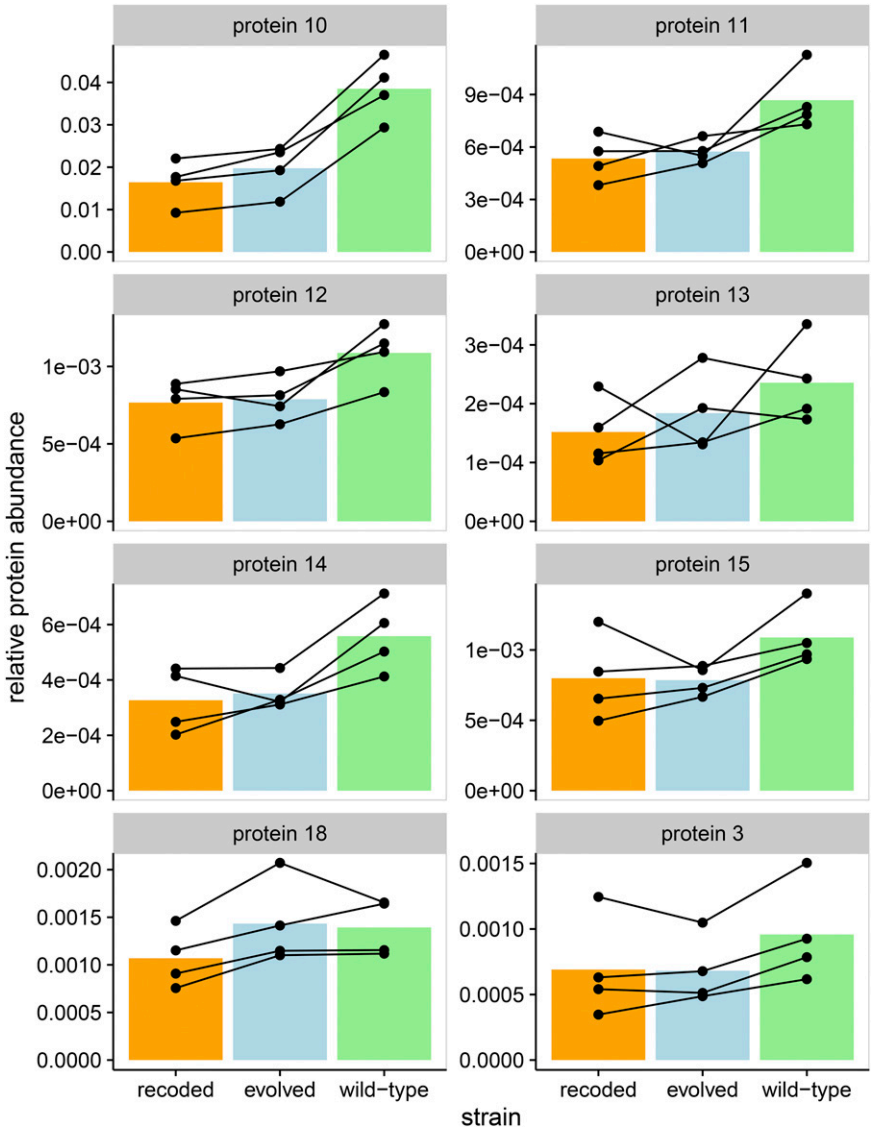

Figure 4 Recoding of gene *10* reduces abundances of capsid protein (gp10), of five other proteins encoded immediately downstream of *10* (gp11–15), and of T7 endonuclease (gp3). The bacteriophage T7 genome contains many polycistronic transcripts. Genes *11* and *12* are always transcribed with gene *10*, following a read through of the Tφ terminator. Genes *13*, *14*, and *15* may also share the same transcript as gene *10*, although these will be less common because of an RNase cleavage site between genes *12* and *13*, and a promoter before gene *13*. The panels show the relative protein abundances corresponding to these genes, in addition to gene *3*, for all wild-type, recoded, and evolved strains at the 9-min time point. All of these genes, with the exception of gene *3*, are class-III genes, expressed late in the T7 life cycle. In addition to lower abundance of capsid protein (gp10), all protein products from the five genes immediately downstream of *10* are also suppressed (FDR < 0.1, FDR-corrected paired *t*-test). These proteins are tail tubular proteins (gp11 and gp12), probable virion-associated protein (gp13), and two internal virion proteins (gp14 and gp15). Gene *3*, which codes for the T7 endonuclease, also has a reduced abundance in the recoded strain. Protein abundances for the evolved strain fall somewhere between that of the wild-type and the recoded strains. Each point represents a single measurement, and lines connect biological replicates.

[false discovery rate (FDR) < 0.1, FDR-corrected *t*-test] (Figure 4 and Table 1). These proteins are mostly or all structural: tail tubular proteins (gp11 and gp12), two internal virion proteins (gp14 and gp15), and a protein of unknown function but required for incorporation of other essential proteins in the virion (gp13). The T7 endonuclease (gp3), encoded by a class-II gene, also showed reduced abundance in the recoded strain. Abundances of these class-III proteins in the evolved strain again fall somewhere between that of the wild-type and recoded strains. Together, these results demonstrate that recoding gene *10* reduced the protein abundances of the capsid protein and several proteins encoded downstream of the capsid protein. The false discovery threshold we employed allows for one or two of those downstream proteins identified to be false positives, but the majority are likely real.

Whereas proteins encoded downstream of the recoded gene *10* showed decreased abundances, those encoded immediately upstream of gene *10* were not obviously affected. Gene *9*, immediately upstream of *10*, encodes the highly expressed scaffold protein but this protein showed no difference in abundance between wild-type and recoded strains (Figure 5). Under the ribosome-depletion hypothesis (model 2), we expected all genes expressed at the same time as gene *10* to be suppressed, thus including gene *9*. Indeed, because of the high levels of expression of *9* and the consequent ease of measuring it with our proteomics methods, we reject the ribosome-depletion model as the basis of attenuation.

E. coli mRNA transcripts declined after bacteriophage T7 infection

A reduction in capsid protein abundance in the attenuated virus could be explained by fewer transcripts from the gene. Of course, a reduction in transcription is not expected from a change in codon usage, and indeed, the design left the 5′ end of the gene unaltered specifically to keep transcription and translation initiation unchanged. Nonetheless, RNA sequences were obtained from phage-infected *E. coli* at 1, 5, and 9 min after infection to see if the decline in protein was accompanied by a decline in transcripts. Transcript abundances were normalized to the total RNA abundance within a sample, excluding rRNA and tRNA. Both phage and *E. coli* transcripts were included.

Table 1 Proteins in which abundance differs significantly (FDR <0.1, FDR-corrected paired *t*-test) between the wild-type and recoded strains of T7

Protein	Mean Difference	*p*-Value	False Discovery Rate
gp3	2.7×10^{-4}	0.00014	0.0061
gp10	2.2×10^{-2}	0.00051	0.0110
gp13	8.4×10^{-5}	0.00180	0.0260
gp12	3.2×10^{-4}	0.00590	0.0510
gp14	2.3×10^{-4}	0.00560	0.0510
gp11	3.3×10^{-4}	0.00740	0.0530
gp15	2.9×10^{-4}	0.01400	0.0880
gp18	3.2×10^{-4}	0.01600	0.0880

Most differentially expressed genes are class-III genes, with the exception of gp3, a class-II gene. The mean differences in abundance, unadjusted *p*-values, and FDRs are shown. Since eight genes fall below a FDR of 0.1, we expect approximately one of these genes to be a false positive.

Gene *10* transcript abundance increased over time, being highest at the final 9-min time point. Yet no significant heterogeneity was observed among the wild-type, recoded, and evolved strain transcripts of gene *10* (Figure 6A). This observation seems to rule out a transcriptional cause of capsid protein reduction.

As validation of the RNA-sequencing methods, transcription analyses were extended to other properties of the T7 and *E. coli* transcriptomes. Gene *1* (T7 RNAP) transcript abundance decreased over time as a proportion of total transcripts (Figure 6B), consistent with previously observed class-I gene expression timing (Molineux 2006). Likewise, the relative abundance of T7 transcripts to *E. coli* increased sharply, consistent with established mechanisms by which T7 shuts off host transcription and degrades the host genome (Molineux 2006). At 1 min after infection, T7 transcripts comprised <1% of the transcript pool of infected *E. coli* (Figure 7). By 5 min after infection, T7 transcripts made up >75% of the transcript pool. By 9 min postinfection, this proportion reached ~95%. Although it is not possible to assess changes in absolute transcript abundances (see *Materials and Methods*), the data require some combination of host-cell transcripts being degraded rapidly, or T7 synthesizing transcripts so rapidly that they quickly dwarf the pool of *E. coli* transcripts. No strain-specific trends were detected.

Downstream effects of recoding support a model of translational coupling

The proteomics suggest that translation of some downstream genes are specifically depressed by the recoding of *10*. This effect on downstream genes might involve translational coupling, in which a stalling of translation over *10* delays translation of genes further down on the same transcript. Translational coupling often occurs when multiple genes are encoded on a single transcript with little intergenic space, such as in bacterial operons and viruses (Lesage *et al.* 1992; Hellmuth *et al.* 1991; Schümperli *et al.* 1982; Oppenheim and Yanofsky 1980; Aksoy *et al.* 1984; Torgov *et al.* 1998). Translational coupling is a plausible process for sets of T7 genes because most transcripts include multiple genes (Dunn and Studier 1983). The T7 class-III promoters precede genes *6.5* (the first class-III gene), *9*, *10*, *13*, and *17*; some transcripts with *9* and *10* will thus include earlier genes, but many will not (Figure 1). Although a phage-specific terminator between *10* and *11* aborts most (but clearly not all) *10* transcripts before *11*, all transcripts with *11* and *12* necessarily include *10*. Thus, translational coupling would operate for *11* and *12* if many of the ribosomes on those genes first translated *10*. Translational coupling beyond *12* is less plausible, however. The combination of an RNAse-III site between *12* and *13* and a promoter before *13* will mean that many or most transcripts with *13* do not include *12*. So we expect substantially higher levels of translational coupling of *11* and *12* with *10*, but far less between *10* and *13*.

Polycistronic transcripts are necessary for translational coupling, but not sufficient. It must also be the case that downstream genes contain structured, inaccessible RBSs that only initiate translation in the presence of upstream translating ribosomes (Tian and Salis 2015; Rex *et al.* 1994; Qu *et al.* 2011). Ribosomes that reach the stop codon of one gene often expose the RBS of the next gene on the transcript and then reinitiate translation on that downstream gene (Spanjaard and van Duin 1989). Moreover, this exposure of the RBS also facilitates binding from the free ribosome pool (Rex *et al.* 1994). Support for the coupling model was evaluated from secondary structure predictions and *in silico* predictions of ribosome binding (Tian and Salis 2015). Due to limitations of software-based methods, we only tested the coupling of genes *10* and *11*. With translational coupling, translation initiation rates are predicted to be ~6 times greater for gene *11* than they would if gene *11* occurred on a single-gene transcript. This supports a model in which at least genes *10* and *11* are translationally coupled.

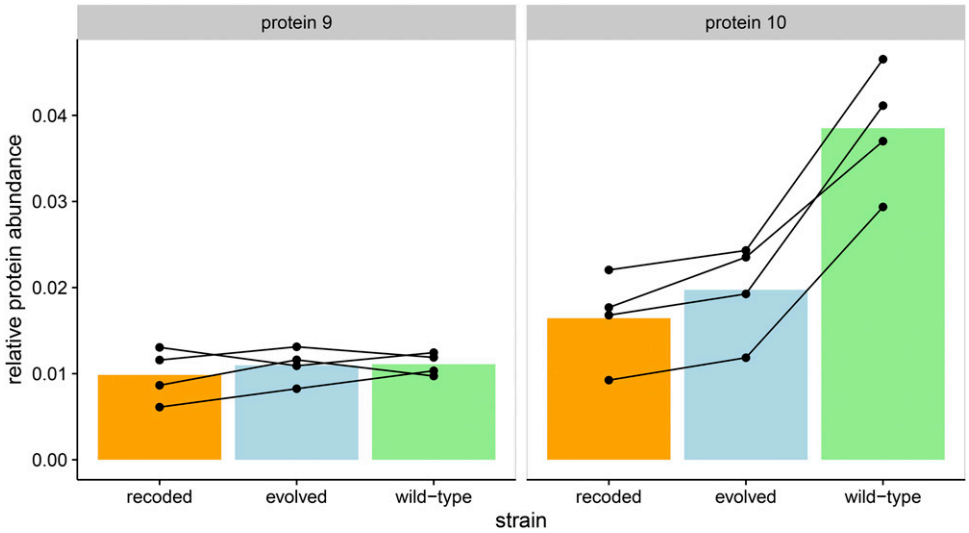

Figure 5 Gene *9*, encoding the highly expressed scaffold protein, shows no detectable difference in protein abundance between wild-type and recoded strains. Genes *9* (left) and *10* (right) are both class-III genes that are expressed at approximately the same time in the T7 life cycle. All abundances are from 9 min postinfection. Each point represents a single measurement, and lines connect biological replicates.

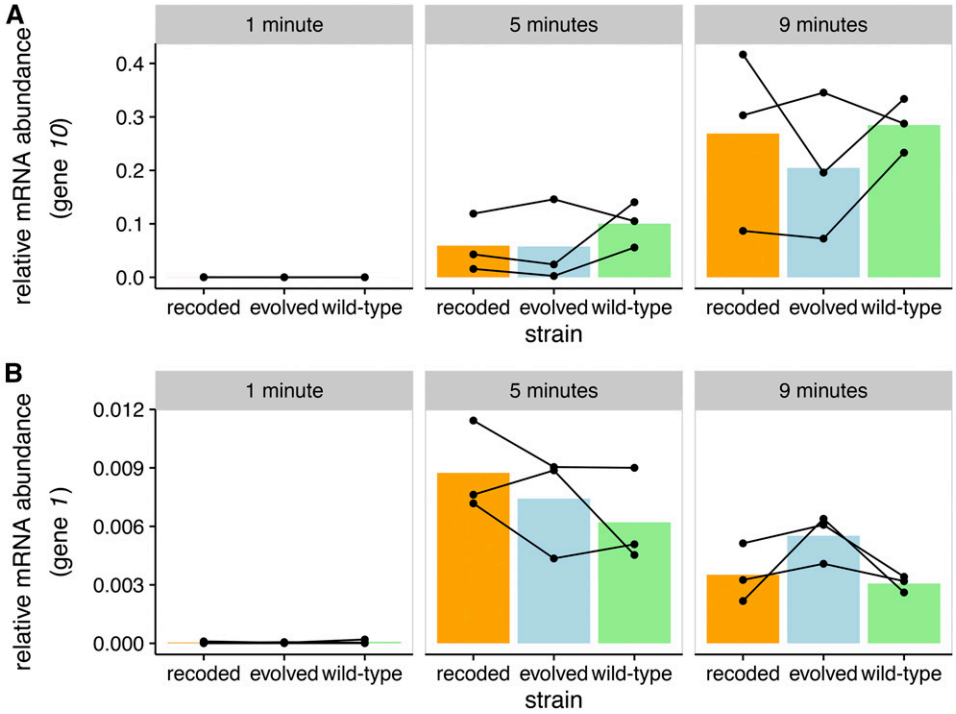

Figure 6 Transcript abundances of individual T7 genes change over time. Three biological replicates each are shown for wild-type, attenuated, and evolved strains. Transcript abundances are presented relative to the total mRNA content in the sample. Each column represents a different time point after infection. Both gene *10* and gene *1* follow the expected expression patterns of class-I and class-III genes, respectively. Each point represents a single measurement, and lines connect biological replicates. (A) Transcripts of gene *10* increase over time. There are no detectable differences in gene *10* mRNA abundances between the attenuated and wild-type T7 strains. (B) Transcripts of gene *1* increase from 1 to 5 min, then decrease from 5 to 9 min. Again, there are no detectable differences in gene *1* mRNA abundances between the attenuated and wild-type T7 strains.

As the inference of translational coupling here is indirect and tentative, additional insight was sought from a mathematical model. The model assumed three genes on a single transcript in the order *a*, *b*, and *c* (Figure 8A, Equations 3–14 in *Materials and Methods*). All genes were assumed to be the same length. In the model, protein production rates (*i.e.*, the rate at which ribosomes complete translation) depended on translation initiation and translation elongation rates. The model allowed varying degrees of coupling, with a coupling constant y, which accounts for both ribosome reinitiation and facilitated binding. The model also allowed for translation initiation independent of any coupling effects, given by w. (In the results presented here, $y = y_b = y_c$ and $w = w_b = w_c$. See Table 2 and *Materials and Methods* for all parameters). In a system with no coupling, translation initiation of gene *b* did not depend on translation of gene *a*, and likewise between genes *b* and *c*. In a strongly coupled system, translation initiation of gene *b* depended almost entirely on the rate at which ribosomes complete translation of gene *a*. Translation of gene *c* similarly depended on gene *b* in a strongly coupled system. Thus, we explored how relative rates of protein production depend on both coupling between genes and elongation rates within genes.

To simulate codon deoptimization of one gene in the transcript, we varied the translation elongation rate of gene *b* only (Figure 8). We considered reductions in translation rate within a two- to threefold range. This reduction in translation elongation is compatible with prior studies of codon deoptimization in *E. coli* (Kudla *et al.* 2009). We also varied the coupling constant y. Under strong translational coupling ($y = 1.2$), the translation initiation rates of a gene depended mostly

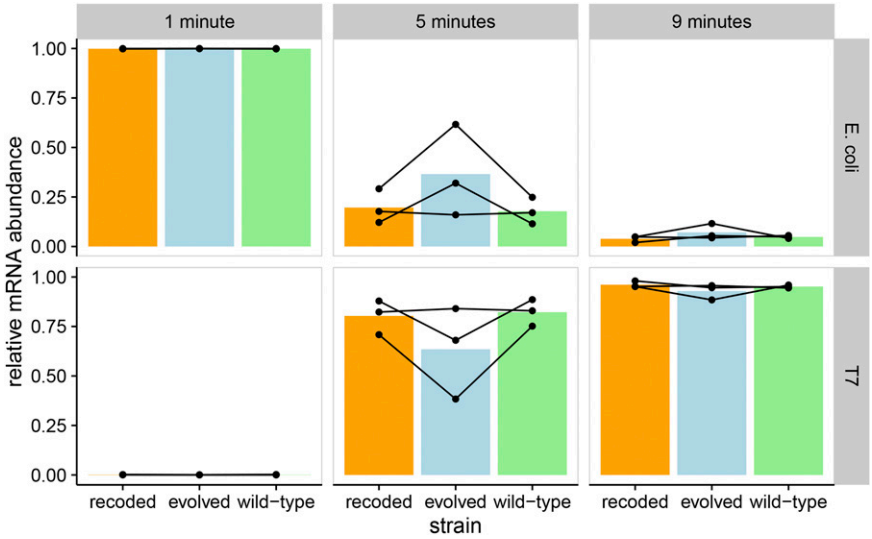

Figure 7 T7 transcripts increase relative to *E. coli* transcripts over time. T7 transcripts ultimately constitute most of the mRNA content of infected *E. coli*. The top row shows total *E. coli* transcript abundances, and the bottom row shows total T7 abundances. Points show individual experiments, while lines connect measurements from the same biological replicate. Each bar represents the mean relative mRNA abundance for a given strain and time point. Transcript abundances are shown relative to the total mRNA content of the sample. Wild-type, recoded, and evolved strains are shown, and there are no detectable differences between the three strains.

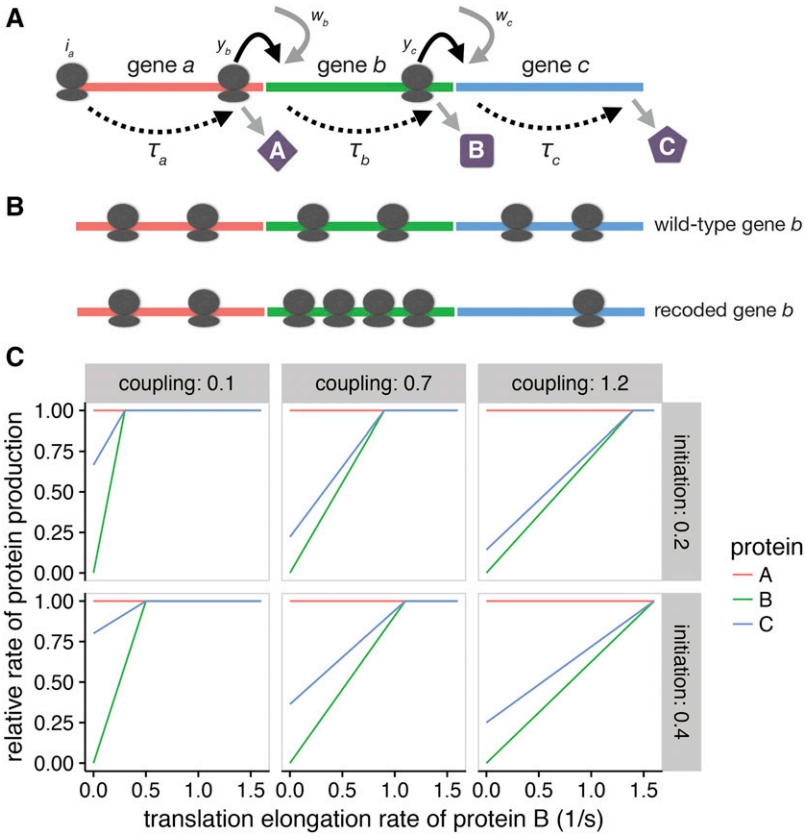

Figure 8 When genes are translationally coupled, the translation rate of upstream genes can affect translation rates of downstream genes. (A) Model of translational coupling. We assume three genes (*a*, *b*, and *c*) are expressed in a polycistronic transcript and are translationally coupled, such that the translation initiation rates of one gene depend partially on the rate of ribosomes reaching the stop codon of the previous gene. Typically, translation initiation is the rate-limiting step in translation. All genes are assumed to be the same length. Translation initiation of gene *a* is given as i_a, and the translation elongation rates of genes *a*, *b*, and *c*, are given by τ_a, τ_b, and τ_c, respectively. The proportion of ribosomes initiating due to upstream translation activity (coupling constant) is given as y_b and y_c for genes *b* and *c*, respectively. The rate of *de novo* initiation is given by w_b and w_c. (B) Recoding gene *b* decreases the rate of translation elongation. Upon recoding gene *b*, we hypothesize that ribosomes will accumulate on the gene *b* transcript and downstream ribosomal densities will decrease. (C) As translation elongation rates increase, the rate of translation of downstream genes also increases. The translation rate of *b* and *c* both increase as the translation elongation rate of gene *b* increases. When the elongation rate exceeds the initiation rate, initiation becomes rate limiting and protein production no longer increases. The left, middle, and right panels show models with coupling constants (*y*, where $y = y_b = y_c$) of 0.1, 0.5, and 1.2, respectively. A larger coupling constant indicates that downstream genes are less likely to initiate translation in the absence of upstream translating ribosomes. As coupling increases, production of protein C becomes more sensitive to translation elongation rates of gene *b*. The rows show different rates of independent translation initiation (*w*, where $w_b = w_c$). The greater the independent initiation rate, the lower the sensitivity of protein-C production to changes in translation elongation of gene *b*.

on the rate of ribosomes moving through the stop codon of an upstream gene. Thus, the relative rate of protein A, B, and C production (assuming that initiation is slower than elongation in each gene) depended on the translation initiation rate of gene *a*. In turn, when the translation elongation rate was less than the initiation rate for gene *b*, translation elongation became rate limiting. Under these elongation rate-limited conditions, the production rates of proteins B and C only increased as the elongation rate of *b* increased, while A production rates remained unaffected (Figure 8C, right panels). Conversely, in a weakly coupled model ($y = 0.1$), even if elongation rates were slower than initiation rates in gene *b*, production of protein C was only weakly affected by recoding of gene *b* (Figure 8C, left panels). In partially coupled models ($y = 0.7$; Figure 8B, middle panels), the production rate of protein C also increased as elongation rates increased until surpassing initiation rates for gene *b*. However, this increase in the rate of C protein production was smaller than the increase observed on a strongly coupled model. Increasing the independent translation initiation rate *w* decreases the dependence of protein-C production on gene *b* elongation rates (Figure 8C, top row to bottom row). This model demonstrated that recoding a gene in a translationally coupled set of genes affects the protein production rates of downstream, but not upstream, genes.

Connecting proteomics to fitness

In previous work, the T7 with a recoded capsid gene had been found to have a fitness of 35.7 doublings/hr, compared to a value of 43.2 for the wild type. Here we consider whether and how the altered proteomics might lead to this fitness reduction. The connection from proteomics to fitness spans two steps: (i) identify the phage life-history components affected by the recoding and evaluate whether that change is compatible with the proteomics, and then (ii) assess whether the magnitude of altered fitness components is compatible with overall fitness.

In the growth conditions used for our assays, fitness is determined by cell density and three phage properties: burst size, lysis time, and adsorption rate (Wang *et al.* 1996; Guyader and Burch 2008; Shao and Wang 2008; Bull 2006; Patwa and Wahl 2009; Bull *et al.* 2011). Burst size refers to the average number of viral particles released from each infected cell. Burst sizes and lysis times were estimated here for the wild-type and recoded phages (there was no expectation that adsorption rate would be affected, which depends on the presence of tail fibers, the product of gene *17*). Although the same cell line was used here as in previous studies (IJ1133), new cell preparations were used, so quantitative agreement with past estimates of burst size and lysis time is not expected, but proportional differences should scale across different cell preparations. No difference in lysis time was observed between strains, but burst size was reduced almost 50% with the recoding (Figure 9).

The proportional reduction in burst size is nearly the same as that for the reduction in capsid protein abundance at 9 min. This reduction in burst size is no doubt caused by the reduction in capsid protein. There is perhaps little basis for arguing that the reductions should match quantitatively, but the agreement between the two numbers poses no dilemma.

Table 2 Parameters for three-gene model of translational coupling

Name	Description
i_a	Translation initiation rate of gene *a*
i_b	Aggregate translation initiation rate of gene *b*
i_c	Aggregate translation initiation rate of gene *c*
y_b	Coupling constant between genes *a* and *b*
y_c	Coupling constant between genes *b* and *c*
w_b	Rate of translation initiation independent of upstream translation rates on gene *b*
w_c	Rate of translation initiation independent of upstream translation rates on gene *c*
a_{in}	Effective initiation rate of ribosomes on gene *a*
$\dot{A}$	Rate of protein-A production
b_{in}	Effective initiation rate of ribosomes on gene *b*
$\dot{B}$	Rate of protein-B production
c_{in}	Effective initiation rate of ribosomes on gene *c*
$\dot{C}$	Rate of protein-C production
τ_a	Translation elongation rate of gene *a*
τ_b	Translation elongation rate of gene *b*
τ_c	Translation elongation rate of gene *c*

The second step in connecting proteomics with fitness is to consider whether a 50% reduction in burst size (with no change in lysis time) is compatible with the observed fitness reduction of 7.5 doublings/hr. The mapping of phage fitness components onto total fitness has been addressed in detail (Bull *et al.* 2011). From numerical trials in that study, a 50% reduction in burst size is compatible with a fitness reduction of the magnitude observed here (table 1, lines L1 and L5, in Bull *et al.* 2011).

DISCUSSION

In the decade since the first proposals to attenuate viruses by synonymous codon substitutions, it has been established that the method works in many viruses and offers many advantages over earlier methods of attenuation. Yet the mechanism by which silent codon changes attenuate not only remains elusive but seems less clear now than it did at the start. Nor is it clear that a single mechanism underlies the attenuation in different systems. Here, we extended previous work on a bacteriophage system in which the encoding of rare codons in the major capsid gene reduced fitness. Our goal was to refine an understanding of the molecular basis of the attenuation.

The capsid protein (encoded by gene *10*) is the most abundant protein produced during the infection cycle of bacteriophage T7. Deoptimizing 50% of gene *10* codons reduced fitness (Bull *et al.* 2012). In exploring the underlying molecular mechanism by which the recoding has this effect, our primary result is that the protein product of gene *10* is reduced almost 50% by the end of the infection cycle, but protein abundance of genes immediately downstream of gene *10* are also depressed. The differences in protein abundance are not reflected in transcript levels, so it appears that the suppression of protein levels lies in translation. The evidence thus supports a simple interpretation of the fitness impact of recoding the major capsid gene:

1. Capsid protein is expressed at a reduced level, as are a few downstream genes.
2. Burst size is correspondingly reduced ~50% with no change in lysis time, compatible with the observed reduction in total fitness.

Taken together, these two observations indicate that the reduction of capsid protein abundance is a consequence of recoding and that this reduction is a plausible intracellular cause of viral attenuation.

One mechanism we entertained to explain the altered proteomics of the recoded phage is saturation of the ribosomes with gene *10* transcripts. Such a model requires that the production of all T7 proteins would decline late in the infection cycle for the recoded phage. Within the limits of resolution, the proteomics rule out an overall reduction in T7 protein production, indicating that reductions are limited to the recoded *10* and a few downstream genes. This study may provide the first indication that translational effects of the recoding extend beyond the recoded genes. There was also no evidence for aborted gene *10* polypeptides in the recoded strains, as might occur from ribosomes stalled on gene *10* transcripts. An obvious next step is to extend these analyses to ribosome profiling, which would directly indicate whether the recoding does tie up ribosomes on gene *10* (Li *et al.* 2014).

The means by which synonymous codon replacement attenuates may be more straightforward for phage T7 than for eukaryotic viruses.

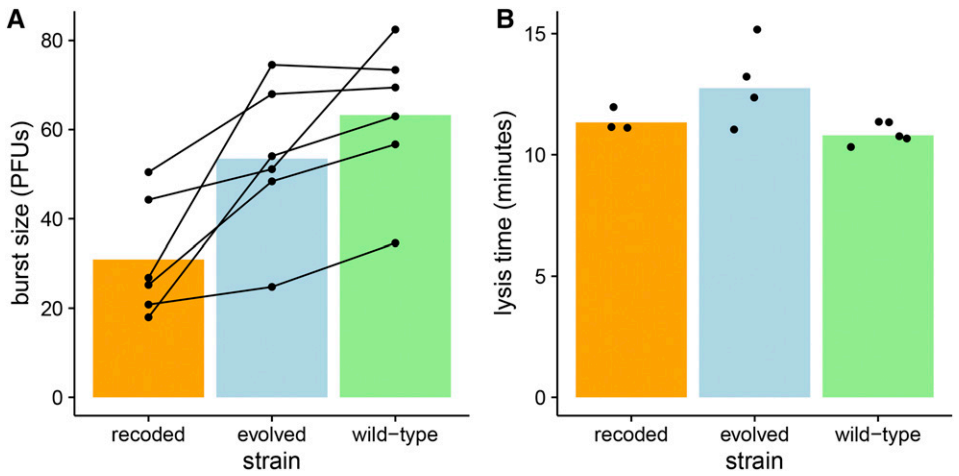

Figure 9 Recoding affects burst size, but not lysis time. Burst size in recoded T7 strain is lower than that of the wild-type and evolved strains, while lysis time is indistinguishable for all three strains. (A) Five replicates each of burst-size measurements are shown for the wild-type, recoded, and evolved T7 strains. Points represent individual measurements in plaque forming units (PFUs), and each set of measurements in a replicate are connected by a line. Bars show the mean burst size for each strain. Burst size of the recoded strain is smaller than that of the evolved ($p = 0.02$, paired *t*-test) and wild-type ($p = 0.002$, paired *t*-test) strains. (B) Lysis time for the three different strains is shown. Points represent individual measurements, and bars represent the mean lysis time for each strain. There were no significant differences between the lysis times of each strain ($p = 0.4$, ANOVA).

Several lines of evidence suggest that the eukaryotic virus attenuation by synonymous codon changes is from the creation of CpG dinucleotides (Burns *et al.* 2009; Atkinson *et al.* 2014; Tulloch *et al.* 2014); indeed, especially powerful evidence to support this interpretation is that evolutionary reversions of attenuated viruses disproportionately reverse CpGs (Burns *et al.* 2006). In contrast, T7 evolutionary reversions did not exhibit any signature suggestive of a dinucleotide basis for attenuation (Bull *et al.* 2012). Nonetheless, we expect that the mechanism of attenuation in T7 will apply across other viruses, even if it is not the only mechanism operating in those viruses.

If our interpretation is correct for the mechanism underlying the effect of the recoding, an evolutionary response to overcome the effect might be duplication of the promoter immediately upstream of gene *10*. Such a duplication would increase the number of *10* transcripts and, in the absence of ribosome saturation, would increase the amount of capsid protein. A duplication of a class-II promoter was observed during adaptation of a different phage (Springman *et al.* 2005), adding credence to the possibility of such an outcome. Yet the mutational origin of a promoter duplication may be highly dependent on surrounding sequences, so perhaps not feasible for all promoters.

We propose that translational coupling may explain why expression of genes downstream of *10* is suppressed by the recoding. Part of that inference is based on a mathematical model of translation. That model is necessarily simplified, however, and there are some obvious improvements needed to increase its realism. First, it assumes a fixed quantity of transcripts, when we know from our RNA-sequencing results that T7 transcripts increase rapidly during infection. Second, the model assumes a per-gene translation elongation rate, but does not model individual codons. A more sophisticated model that includes codon-level detail and the T7 life cycle would be needed to predict the fitness effects of codon deoptimization. Several life-cycle and molecular models of T7 have achieved limited success in predicting the phenotypic effects of genome manipulations (Endy *et al.* 1997; You *et al.* 2002; Kosuri *et al.* 2007; Birch *et al.* 2012), but none enable codon modifications or complex translation mechanisms such as coupling. The proteomics and RNA-sequencing data generated in this study should be useful in future high-resolution modeling studies that scale from the molecular level to that of viral fitness.

Although it is tempting to interpret the slowed translation as a consequence of using rare codons, which then use rare tRNAs, some recoding strategies used in other genomes suggest alternative possibilities. Codon deoptimization of GFP in *E. coli* initially yielded a range of protein expression effects, but these effects were eventually attributed to changes in mRNA secondary structure in the first 28 codons of the GFP sequence (Kudla *et al.* 2009). Codon changes beyond these first 28 codons had a weak effect on protein expression (Kudla *et al.* 2009). In our designs, we explicitly excluded these 5′-end codons from modification. Moreover, some attenuation designs with influenza virus, poliovirus, and arboviruses have achieved attenuation by merely shuffling codons within a gene or genome to create rare codon pairs (Coleman *et al.* 2008; Burns *et al.* 2009; Mueller *et al.* 2010; Nougairede *et al.* 2013; Shen *et al.* 2015). Since the numbers of each codon are not changed in those designs, the mechanism cannot be one of simple tRNA abundance. One possibility is suggested by recent work in yeast and *Salmonella*, whereby the degree of codon clustering is important to rapid translation (Cannarozzi *et al.* 2010; Gamble *et al.* 2016; Chevance and Hughes 2017). We did not attempt to control for codon-pair bias in our recoded T7 constructs. Further experiments will be needed to determine if the *E. coli* translation machinery is sensitive to changes in codon-pair bias.

The approach developed here should help elucidate other mechanisms of viral attenuation. For example, the timing of gene expression appears important to fitness: a reciprocal exchange of some middle and late genes had some major fitness effects, and those effects were not recovered on long-term adaptation (Cecchini *et al.* 2013). Ultimately, we envision a future in which an understanding of viral life history at the molecular level enables facile engineering of arbitrary fitness and alternative vaccine designs.

ACKNOWLEDGMENTS

We thank I. J. Molineux for advice on and insight into T7. This work was supported in part by National Institutes of Health grants R01 GM088344, National Science Foundation Cooperative agreement no. DBI-0939454 (BEACON Center), and Army Research Office grant W911NF-12-1-0390. The Texas Advanced Computing Center provided high-performance computing resources. J.J.B. is supported as the University of Texas Miescher Regents Professor.

Author contributions: J.J.B., C.O.W., and D.R.B. conceived and designed the experiments and provided reagents and materials. D.R.B., M.L.P., and B.L.S. performed the experiments. B.R.J., D.R.B., M.L.P., and B.L.S. analyzed the data. B.R.J., C.O.W., and J.J.B. developed the model. J.J.B., C.O.W., D.R.B., M.L.P., B.L.S., and B.R.J. wrote the paper. The authors report no conflicts of interest.

LITERATURE CITED

Aksoy, S., C. L. Squires, and C. Squires, 1984 Translational coupling of the trpB and trpA genes in the *Escherichia coli* tryptophan operon. J. Bacteriol. 157: 363–367.

Atkinson, N. J., J. Witteveldt, D. J. Evans, and P. Simmonds, 2014 The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. Nucleic Acids Res. 42: 4527.

Barrett, T., S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim *et al.*, 2013 NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 41: D991–D995.

Bates, D., M. Mächler, B. Bolker, and S. Walker, 2015 Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67: 1–48.

Birch, E. W., N. A. Ruggero, and M. W. Covert, 2012 Determining host metabolic limitations on viral replication via integrated modeling and experimental perturbation. PLoS Comput. Biol. 8: e1002746.

Bray, N. L., H. Pimentel, P. Melsted, and L. Pachter, 2016 Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34: 525–527.

Bull, J. J., 2006 Optimality models of phage life history and parallels in disease evolution. J. Theor. Biol. 241: 928–938.

Bull, J. J., R. H. Heineman, and C. O. Wilke, 2011 The phenotype-fitness map in experimental evolution of phages. PLoS One 6: e27796.

Bull, J. J., I. J. Molineux, and C. O. Wilke, 2012 Slow fitness recovery in a codon-modified viral genome. Mol. Biol. Evol. 29: 2997–3004.

Burns, C. C., J. Shaw, R. Campagnoli, J. Jorba, A. Vincent *et al.*, 2006 Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J. Virol. 80: 3259–3272.

Burns, C. C., R. Campagnoli, J. Shaw, A. Vincent, J. Jorba *et al.*, 2009 Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. J. Virol. 83: 9957–9969.

Cannarozzi, G., N. N. Schraudolph, M. Faty, P. von Rohr, M. T. Friberg *et al.*, 2010 A role for codon order in translation dynamics. Cell 141: 355–367.

Cecchini, N., M. Schmerer, I. J. Molineux, R. Springman, and J. J. Bull, 2013 Evolutionarily stable attenuation by genome rearrangement in a virus. G3 3: 1389–1397.

Chevance, F. F. V., and K. T. Hughes, 2017 Case for the genetic code as a triplet of triplets. Proc. Natl. Acad. Sci. USA 114: 4745–4750.

Coleman, J. R., D. Papamichail, S. Skiena, B. Futcher, E. Wimmer *et al.*, 2008 Virus attenuation by genome-scale changes in codon pair bias. Science 320: 1784–1787.

Dunn, J. J., and F. W. Studier, 1983 Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. 166: 477–535.
Endy, D., D. Kong, and J. Yin, 1997 Intracellular kinetics of a growing virus: a genetically structured simulation for bacteriophage T7. Biotechnol. Bioeng. 55: 375–389.
Fredrick, K., and M. Ibba, 2010 How the sequence of a gene can tune its translation. Cell 141: 227–229.
Gamble, C. E., C. E. Brule, K. M. Dean, S. Fields, and E. J. Grayhack, 2016 Adjacent codons act in concert to modulate translation efficiency in yeast. Cell 166: 679–690.
Guyader, S., and C. L. Burch, 2008 Optimal foraging predicts the ecology but not the evolution of host specialization in bacteriophages. PLoS One 3: e1946.
Heineman, R. H., and J. J. Bull, 2007 Testing optimality with experimental evolution: lysis time in a bacteriophage. Evolution 61: 1695–1709.
Hellmuth, K., G. Rex, B. Surin, R. Zinck, and J. E. G. McCarthy, 1991 Translational coupling varying in efficiency between different pairs of genes in the central region of the atp operon of *Escherichia coli*. Mol. Microbiol. 5: 813–824.
Houser, J. R., C. Barnhart, D. R. Boutz, S. M. Carroll, A. Dasgupta *et al.*, 2015 Controlled measurement and comparative analysis of cellular components in *E. coli* reveals broad regulatory changes in response to glucose starvation. PLoS Comput. Biol. 11: e1004400.
Kosuri, S., J. R. Kelly, and D. Endy, 2007 TABASCO: a single molecule, base-pair resolved gene expression simulator. BMC Bioinformatics 8: 480.
Kudla, G., A. W. Murray, D. Tollervey, and J. B. Plotkin, 2009 Coding-sequence determinants of gene expression in *Escherichia coli*. Science 324: 255–258.
Lesage, P., C. Chiaruttini, M. Graffe, J. Dondon, M. Milet *et al.*, 1992 Messenger RNA secondary structure and translational coupling in the *Escherichia coli* operon encoding translation initiation factor IF3 and the ribosomal proteins, L35 and L20. J. Mol. Biol. 228: 366–386.
Li, G. W., D. Burkhardt, C. Gross, and J. S. Weissman, 2014 Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell 157: 624–635.
Molineux, I. J., 2006 The T7 group, pp. 277–301 in *The Bacteriophages*, edited by Calendar, R., and S. T. Abedon. Oxford University Press, New York.
Mueller, S., D. Papamichail, J. R. Coleman, S. Skiena, and E. Wimmer, 2006 Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J. Virol. 80: 9687–9696.
Mueller, S., J. R. Coleman, D. Papamichail, C. B. Ward, A. Nimnual *et al.*, 2010 Live attenuated influenza virus vaccines by computer-aided rational design. Nat. Biotechnol. 28: 723–726.
Nougairede, A., L. De Fabritus, F. Aubry, E. A. Gould, E. C. Holmes *et al.*, 2013 Random codon re-encoding induces stable reduction of replicative fitness of chikungunya virus in primate and mosquito cells. PLoS Pathog. 9: 1–18.
Oppenheim, D. S., and C. Yanofsky, 1980 Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics 95: 785–795.
Patwa, Z., and L. Wahl, 2009 The impact of host-cell dynamics on the fixation probability for lytic viruses. J. Theor. Biol. 259: 799–810.
Plotkin, J. B., and G. Kudla, 2010 Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12: 32–42.
Qu, X., J. D. Wen, L. Lancaster, H. F. Noller, C. Bustamante *et al.*, 2011 The ribosome uses two active mechanisms to unwind messenger RNA during translation. Nature 475: 118–121.
R Core Team, 2014 *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
Raveh, A., M. Margaliot, E. D. Sontag, and T. Tuller, 2016 A model for competition for ribosomes in the cell. J. R. Soc. Interface 13: e1002127.
Reuveni, S., I. Meilijson, M. Kupiec, E. Ruppin, and T. Tuller, 2011 Genome-scale analysis of translation elongation with a ribosome flow model. PLoS Comput. Biol. 7: e1002127.
Rex, G., B. Surin, G. Besse, B. Schneppe, and J. E. McCarthy, 1994 The mechanism of translational coupling in *Escherichia coli*. Higher order structure in the atpHA mRNA acts as a conformational switch regulating the access of de novo initiating ribosomes. J. Biol. Chem. 269: 18118–18127.
Schümperli, D., K. McKenney, D. A. Sobieski, and M. Rosenberg, 1982 Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. Cell 30: 865–871.
Scott, M., C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, 2010 Interdependence of cell growth and gene expression: origins and consequences. Science 330: 1099–1102.
Shah, P., and M. A. Gilchrist, 2011 Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. Proc. Natl. Acad. Sci. USA 108: 10231–10236.
Shao, Y., and N. Wang, 2008 Bacteriophage adsorption rate and optimal lysis time. Genetics 180: 471–482.
Shen, S. H., C. B. Stauft, O. Gorbatsevych, Y. Song, C. B. Ward *et al.*, 2015 Large-scale recoding of an arbovirus genome to rebalance its insect *vs.* mammalian preference. Proc. Natl. Acad. Sci. USA 112: 4749–4754.
Silva, J. C., M. V. Gorenstein, G. Z. Li, J. P. C. Vissers, and S. J. Geromanos, 2006 Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol. Cell. Proteomics 5: 144–156.
Spanjaard, R. A., and J. van Duin, 1989 Translational reinitiation in the presence and absence of a Shine and Dalgarno sequence. Nucleic Acids Res. 17: 5501–5508.
Springman, R., M. Badgett, I. Molineux, and J. Bull, 2005 Gene order constrains adaptation in bacteriophage T7. Virology 341: 141–152.
Tian, T., and H. M. Salis, 2015 A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. Nucleic Acids Res. 43: 7137–7151.
Torgov, M. Y., D. M. Janzen, and M. K. Reddy, 1998 Efficiency and frequency of translational coupling between the bacteriophage T4 clamp loader genes. J. Bacteriol. 180: 4339–4343.
Tuller, T., A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan *et al.*, 2010 An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141: 344–354.
Tulloch, F., N. J. Atkinson, D. J. Evans, M. D. Ryan, and P. Simmonds, 2014 RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife 3: e04531.
Vind, J., M. A. Sørensen, M. D. Rasmussen, and S. Pedersen, 1993 Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. Expression from reporter genes does not always reflect functional mRNA levels. J. Mol. Biol. 231: 678–688.
Vizcaíno, J., A. Csordas, N. del Toro, J. A. Dianes, J. Griss *et al.*, 2016 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 44: D447–D456.
Wang, N., D. E. Dykhuizen, and L. B. Slobodkin, 1996 The evolution of phage lysis timing. Evol. Ecol. 10: 545–558.
Wickham, H. 2009 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
Wimmer, E., S. Mueller, T. M. Tumpey, and J. K. Taubenberger, 2009 Synthetic viruses: a new opportunity to understand and prevent viral disease. Nat. Biotechnol. 27: 1163–1172.
Xu, Y., P. Ma, P. Shah, A. Rokas, Y. Liu *et al.*, 2013 Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. Nature 495: 116–120.
You, L., P. F. Suthers, and J. Yin, 2002 Effects of *Escherichia coli* physiology on growth of phage T7 in vivo and in silico. J. Bacteriol. 184: 1888–1894.
Zhou, M., J. Guo, J. Cha, M. Chae, S. Chen *et al.*, 2013 Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature 495: 111–115.
Zhou, Z., Y. Dang, M. Zhou, L. Li, Ch. Yu *et al.*, 2016 Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proc. Natl. Acad. Sci. USA 113: E6117–E6125.
Zur, H., and T. Tuller, 2016 Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution. Nucleic Acids Res. 44: 9031–9049.

Communicating editor: B. J. Andrews