

Research article

Open Access

Evolution of phage with chemically ambiguous proteomes

Jamie M Bacher^{1,4}, James J Bull^{1,2} and Andrew D Ellington*^{1,3}

Address: ¹Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX, USA 78712, ²Section of Integrative Biology, University of Texas at Austin, Austin, TX, USA 78712, ³Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX, USA 78712 and ⁴The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA, USA 92037

Email: Jamie M Bacher - jbacher@scripps.edu; James J Bull - bull@bull.biosci.utexas.edu; Andrew D Ellington* - andy.ellington@mail.utexas.edu

* Corresponding author

Published: 10 December 2003

Received: 24 June 2002

BMC Evolutionary Biology 2003, 3:24

Accepted: 10 December 2003

This article is available from: <http://www.biomedcentral.com/1471-2148/3/24>

© 2003 Bacher et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The widespread introduction of amino acid substitutions into organismal proteomes has occurred during natural evolution, but has been difficult to achieve by directed evolution. The adaptation of the translation apparatus represents one barrier, but the multiple mutations that may be required throughout a proteome in order to accommodate an alternative amino acid or analogue is an even more daunting problem. The evolution of a small bacteriophage proteome to accommodate an unnatural amino acid analogue can provide insights into the number and type of substitutions that individual proteins will require to retain functionality.

Results: The bacteriophage Q β initially grows poorly in the presence of the amino acid analogue 6-fluorotryptophan. After 25 serial passages, the fitness of the phage on the analogue was substantially increased; there was no loss of fitness when the evolved phage were passaged in the presence of tryptophan. Seven mutations were fixed throughout the phage in two independent lines of descent. None of the mutations changed a tryptophan residue.

Conclusions: A relatively small number of mutations allowed an unnatural amino acid to be functionally incorporated into a highly interdependent set of proteins. These results support the 'ambiguous intermediate' hypothesis for the emergence of divergent genetic codes, in which the adoption of a new genetic code is preceded by the evolution of proteins that can simultaneously accommodate more than one amino acid at a given codon. It may now be possible to direct the evolution of organisms with novel genetic codes using methods that promote ambiguous intermediates.

Background

Organismal proteomes are generally thought of as being chemically distinct, in the sense that a genetic code is maintained by codon:anticodon interactions and the specificities of aminoacyl-tRNA synthetases will almost always lead to the translation of mRNAs into proteins of defined sequence and chemical composition. While alternative codes are known [1], these also yield chemically

distinct proteomes. The evolution of an organism with novel codon:anticodon interactions and aminoacyl-tRNA synthetase specificities may produce proteins whose sequences and compositions differ from those generated by an organism with the 'Universal' code, but still will not produce proteins that have multiple, different amino acids at a given sequence position.

This chemical distinctness of organismal proteomes is maintained by the relatively low rate of amino acid misincorporation that occurs during protein biosynthesis. Many aminoacyl-tRNA synthetases have been found to have at least a thousand-fold preference for their cognate amino acid, after editing (reviewed in [2]). EF-Tu further discriminates between cognate and non-cognate codon:anticodon pairs prior to and after GTP hydrolysis [3]. Because of these mechanisms, the overall error rate for amino acid insertion into proteins is typically at least 3×10^{-3} , and frequently lower [2-4].

Although amino acid misincorporations seldom occur in Nature, chemically ambiguous proteomes can be generated in laboratory settings. Many aminoacyl-tRNA synthetases will efficiently charge tRNA molecules with amino acid analogues [2]. In particular, the ability of the *Bacillus subtilis* tryptophanyl-tRNA synthetase to discriminate against fluorine-substituted analogues of tryptophan has been examined. Discrimination against 4-fluorotryptophan (4fW) was only 6-fold, while discrimination against 6-fluorotryptophan (6fW) was 20-fold [5]. Consistent with this, *Escherichia coli* strains that are transiently grown in the presence of high concentrations of fluorotryptophan analogues will incorporate a mixture of natural and unnatural amino acids throughout their proteomes [6-10]. Similarly, norleucine and norvaline have been shown to be synthesized as side-products of branched chain amino acid biosynthesis [2]. Norleucine is incorporated with alacrity into proteins, replacing up to 20% of methionine residues once methionine has been exhausted during protein overexpression [11,12].

Such chemical ambiguity typically extracts a phenotypic cost. An *E. coli* auxotroph selected to grow continuously on a high proportion of 4fW [6] accumulated 5 (identified) mutations in three genes responsible for tryptophan incorporation (tryptophanyl tRNA synthetase, aromatic amino acid permease, and a transcriptional repressor of aromatic amino acid permease). Nonetheless, the evolved strain grew extremely poorly, and had a doubling time of over a day. *E. coli* mutants selected to grow with cysteine incorporated at a valine codon accumulated mutations in the editing domain of valyl-tRNA synthetase [13]. Increased mischarging led to the substitution of 24% of valines with aminobutyrate. Finally, the yeast *Candida* spp. has been found to ambiguously (albeit inefficiently [14]) translate the leucine codon CUG as serine [15]. This ambiguous tRNA was transferred to *Saccharomyces cerevisiae* on a plasmid, and the dual incorporation of serine and leucine throughout the yeast proteome resulted in a 50% decrease in growth rate [16].

The design or evolution of organisms with novel genetic codes has been undertaken by a number of groups

[6,13,17-19]. One potential route to the evolution of an organism with a novel genetic code is to initially select for the mixed incorporation of natural and unnatural amino acids throughout the proteome [6]. Growth defects that arise from the misincorporation of the amino acid analogue can potentially be ameliorated by the evolution of those proteins whose functions are inhibited by the analogue. Such chemically ambiguous proteomes might then further evolve over time to fully incorporate the analogue. In order to better understand the initial route of adaptation of an organismal proteome to chemical ambiguity, we chose to adapt a simple proteome, that of bacteriophage Q β , to function in the presence of an amino acid analogue.

Results and Discussion

The evolution of phage that could utilize or tolerate an amino acid analogue required the availability of a host that could grow on the analogue. We and others have previously shown that *E. coli* can be grown in high concentrations of fluorotryptophan analogues, with concomitantly high incorporation of the analogues into cellular proteins [6-10]. The replication of Q β phage was therefore examined in an *E. coli* auxotroph grown in the presence of a series of tryptophan analogues. The number of doublings in 20 hours was used as a measure of fitness, and was determined using a standardized assay. While most analogues did not seem to affect phage growth, 6fW significantly depressed Q β fitness, decreasing the number of doublings by ca. 10-fold in a standard assay (Figure 1). This is equivalent to an approximately 180 million-fold smaller increase in titer over 20 hours. We therefore chose to adapt phage to 95% 6fW. As expected based on previous experiments with tryptophan analogues, 6fW was found to be incorporated into cellular proteins at a level of approximately 60%, irrespective of whether a single, isolated protein or the bacterial proteome was analyzed (Table 1).

Initially, two replicate lines of phage were evolved over ten serial passages in tryptophan alone, to help ensure that any mutations that were adaptive for the growth conditions alone would sweep the population in advance. Both lines were split and then further evolved over an additional 15 rounds of selection in W or an additional 25 rounds of selection in 95% 6fW (see also Figure 2). After 25 serial passages the fitness of both replicate lines increased by slightly more than 4-fold on the analogue. The kinetics of fitness improvements were quite different between the replicate lines (Figure 3), indicating that the phage may have taken different evolutionary paths to similar phenotypes. Variant lineages have previously been observed during the natural or directed evolution of other phenotypes, including the evolution of drug-resistant HIV-1 [20], the evolution of ϕ X174 bacteriophage with

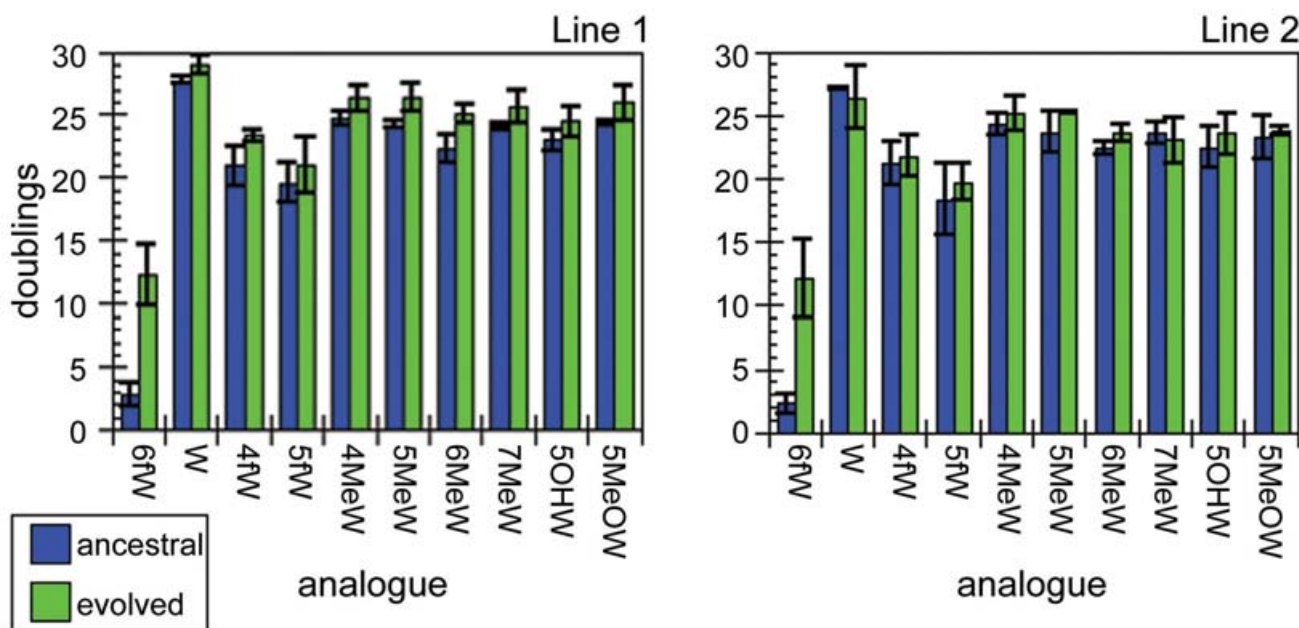


Figure 1
 Fitness of ancestral and selected phage on various tryptophan analogues. Ancestral and round 25 selected phage were tested for fitness on eight additional tryptophan analogues (95% analogue, 5% W) for Line 1 (left) and Line 2 (right). Analogues used were 4-fluorotryptophan (4fW), 5-fluorotryptophan (5fW), 4-methyltryptophan (4MeW), 5-methyltryptophan (5MeW), 6-methyltryptophan (6MeW), 7-methyltryptophan (7MeW), 5-hydroxytryptophan (5OHW) and 5-methoxytryptophan (5MeOW). Data for fitness on 95% 6fW and W are taken from Figures 3 and 5 respectively. Error bars represent standard deviations of at least three replicates.

Table 1: Incorporation of 6fW into proteins.

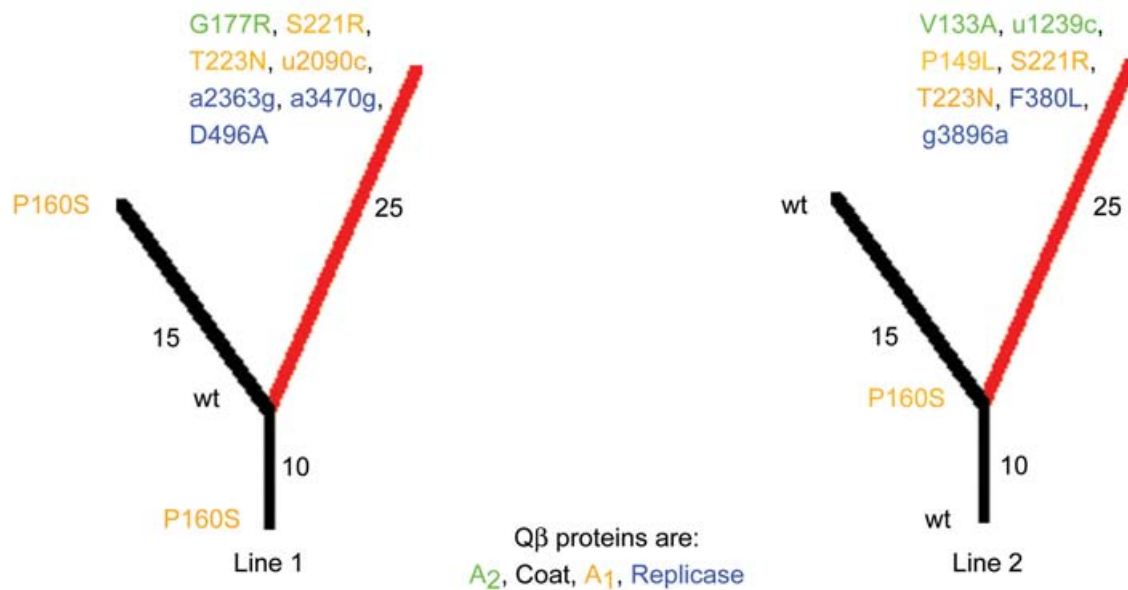
Method of Analysis	%6fW Incorporation
Whole-cell Protein Extract	
- HPLC-ESI	56.5
- HPLC-HPLC	66.7
Purified GFPuv	
- HPLC-ESI	68.6
Average % Incorporation	64.0

altered host ranges and thermal optima [21], and the evolution of ribozymes that could cleave a novel substrate [22]. The increase in fitness appeared to have leveled off after 25 rounds of selection in at least one of the lines (Figure 3), and the selection was therefore stopped and the population further characterized.

In order to more closely discern similarities or differences in the evolutionary paths taken by the phage, the genomes of populations of ancestral and evolved phage, as well as

genomes of individual variants from the evolved populations, were isolated and sequenced (Figure 2, Figure 4). All phage apparently had a number of sequence differences relative to a previously published sequence of Qβ phage (A558C, A1607G, A2111G, C2944T, T3229C, C3712T, C4019T) [23]. However, the sequences we have determined are consistent with other published sequences of the coat and A1 proteins (Medline accession number M99039) and the replicase protein (accession number X14764). While bacteriophage Qβ can evolve extremely quickly, we believe that our sequences represent the first complete, accurate, and electronically accessible sequence of the bacteriophage genome, and the first detailed examination of the genome of a Qβ quasi-species. Nonetheless, it should be noted that prior to the development of RNA sequencing techniques, Domingo et al. [24] used RNase T1 fingerprinting to demonstrate that bacteriophage Qβ was in fact a quasi-species in which the genome was "a weighted average of a large number of different individual sequences."

Replicate experiments were carried out in parallel on tryptophan media. In these lines, only one mutation, P160S

**Figure 2**

Genotype changes over the course of the selection. A phylogeny of the two lines, along with associated mutations, is shown. Evolution on W is in black, while selection carried out on 6fW is indicated in red; the number of cycles of selection are indicated beside the lines. Mutations are color coordinated as indicated at the bottom of the figure; mutations in green are in A₂, black is the coat protein, orange indicates a mutation in A₁, and green in the replicase; genotypes with no mutations are indicated by 'wt.' Lower case mutations correspond to base changes; upper case indicates protein substitutions.

in the A₁ protein, was fixed; this mutation was not found in the lines evolved on 6fW (Figure 2). Interestingly, proline and serine seem to toggle back and forth at position 160 during the passage of the population. It may also be that these mutations do not alternately sweep the population, but instead vary between high (detectable at the population level) and low (undetectable) frequencies over time. Mutations that cyclically appear have been observed during the evolution of other phage, although usually as a result of iterative passages between different environmental conditions (for example, see [25]). Finally, the dominant genomic sequence of these populations is identical, except for the variable presence of P160S substitutions in A₁. This being the case, we expect that the fitness level of the unselected population, the population after ten rounds of selection on W, and the population after fifteen additional rounds of selection on W to be highly comparable.

Given this control, it is likely that the mutations that were fixed at the population level during growth on 6fW were adaptive. Each evolved population had seven mutations that were either fixed or at high frequency, although only two of these mutations were common to both lines (Figure 2). There were some discrepancies between the muta-

tions found in the population and the mutations identified in individual isolates. Mutations that appeared to be fixed at the population level were found to be missing from either one clone (one instance) or two clones (from different lines, one instance). Conversely, there were two mutations (i.e., t66a and a3309t = I320F in the replicase) that appeared in two clones, but did not appear at the population level.

On average, individual clones from the two populations had 13 mutations (standard deviation of 3.4), 7 fixed mutations and 6 mutations that were unique to a given isolate. Some isolates contained only 2 unique mutations while others had up to 10 unique mutations. A total of more than 50 unique mutations were found. Qβ phage is typically thought of as an error-prone, quasi-species comprised of numerous different variants, and it has been estimated to have mutation rates as high as 6.5 nucleotide substitutions per genome per replicative cycle [26,27]. While our results are also consistent with considering Qβ a quasi-species [24], they may also support the hypothesis that selection is continuing to act on a transient population of variants, especially in line 2, in which fitness may still be increasing. In support of this hypothesis, the final

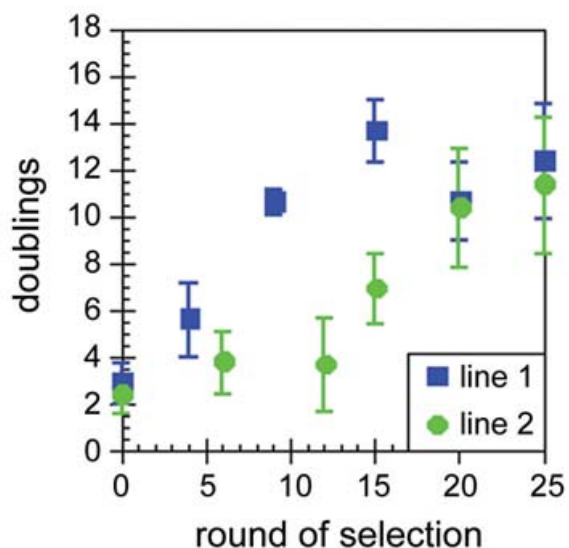


Figure 3

Fitness of selected populations on 95%6fW. Populations were tested for fitness on 95%6fW at various points over the course of the selection. Error bars represent standard deviations of at least three replicates.

population of line 2 was found to have a greater variability when assayed for fitness on W (Figure 1, Figure 5).

Amino acid substitutions were distributed throughout the phage genome (Figure 4, Figure 6a), but the improvement in fitness on 6fW occurred without the isolation of a single mutation in a tryptophan codon, either at the population level or in individual clones. Interestingly, the coat protein contained no tryptophans, and was also found to contain no fixed amino acid substitutions. However the fact that this gene is short and would therefore have accumulated fewer random mutations may also explain this phenomenon. In contrast, the read-through protein A1 contained two fixed amino acid substitutions, S221R and T223N. However, it should be noted that while the S221R substitution was consistently found at the population level it was not found in either clone 1 of Line 1 or clone 3 of Line 2. Because the recombination frequency of Q β phage is known to be low (on the order of 10^{-8} [28]), it is possible that S221R may be a mutation which was accidentally fixed along with another, truly adaptive mutation, and was in the process of being slowly diluted out of the population. Each of the replicate lines also had additional amino acid substitutions that were fixed at the population level. An amino acid substitution (P149L) was

found in the A1 protein in Line 2, there were two different amino acid substitutions (one in each line) in the A2 protein, and two different amino acid substitutions in the replicase. Sequence alignments of the replicase genes from Q β , SP, MS2 and GA, representing the four serotypes of RNA phage, revealed a number of regions of high conservation [29,30]. The F380L substitution in the replicase protein of Line 2 occurred in what was otherwise a phylogenetically conserved residue. Similarly, the amino acid substitutions D250N (clone 3, Line 2) and L290P (clone 3, Line 1) occurred in highly conserved residues. That appearance of mutations in otherwise highly conserved residues strongly suggests that these mutations were adaptive. By comparison, I320F, found in clones 1 and 3 of Line 1, substituted the residue found in Group B single-stranded RNA phage for the residue found in Group A, suggesting that this substitution is functionally conservative [29,30].

The simplest explanation for these results is that the amino acid substitutions in the three Q β proteins somehow compensated for intramolecular disruptions due to the incorporation of 6-fluorotryptophan or for intermolecular disruptions with fluorinated *E. coli* proteins. A number of interactions between phage and host proteins have been described. Interactions between Q β replicase and various *E. coli* proteins are known, including EF-Tu, EF-Ts, ribosomal protein S1, and an RNA-binding protein called Hfq [31-33]. A2 is known to interact with MurA and inhibit cell wall biosynthesis, resulting in cell lysis [34]. Finally, the entry of Q β phage is mediated by the F-pilus. A2 binds to the pilus and uses it for transport of the genome. The read-through protein A1 is also required for this process [30], although its precise function is not yet known [35,36].

The identification of five fixed yet silent substitutions (three in Line 1 and two in Line 2; Figure 6b) was consistent with results from previous directed evolution experiments with Q β and the related RNA phage MS2, which indicated that mutations affecting RNA structure could be as or more important than those affecting proteins. For example, when a hairpin structure that controls the expression levels of the MS2 coat protein was mutated, compensatory mutations were recovered that restored the hairpin [37]. Eight of the selected MS2 operator mutations were silent and retained the wild-type amino acid sequence of the coat protein; only one altered the amino acid sequence [37].

Since it is clear that the secondary structures of RNA phage are under selective pressure, it is formally possible that the amino acid substitutions we observed were not important in and of themselves, but rather were by-products of the evolution of an altered RNA structure. Both the S221R and

Protein	Mutation	Identity	Line 1						Line 2										
			round 0 on W	round 10 on W	round 25 on W	round 25 on 6fW	clone 1	clone 2	clone 3	round 0 on W	round 10 on W	round 25 on W	round 25 on 6fW	clone 1	clone 2	clone 3			
untranslated	A52G A54G																		
A2	T66A	silert P2																	
	T312C	silert V84																	
	T422C	F12-S																	
	T458C	V13JA																	
	C558A	silert R166																	
	T591C	G177R																	
	T681C	silert F207																	
	A740G	N227S																	
	A759G	silert Q233																	
	T780C	silert V240																	
	A1090G	I344V																	
	T1092C	silert I344																	
	C1215T	silert L385																	
	A1217G	H389R																	
T1239C	silert A393																		
coat, A1	A1344G	M1V																	
	T1724C	silert D127																	
	T1742C	I133F																	
A1	C1789T	P148L																	
	C1798T	P152L																	
	C1821T	P166S																	
	C1905T	P188S																	
	A1956G	K205E																	
	T2006G	S22 R																	
	C2011A	T222N																	
	T2051C	silert L236																	
	T2090C	silert D249																	
	T2099C	silert Y252																	
	G2208A	A289T																	
	G2217A	V295M																	
	T2231C	silert F296																	
replicase	A2363G	silert R4																	
	G2451A	A34 T																	
	G2533A	G61 E																	
	T2558C	silert N69																	
	T2774C	silert D141																	
	TG3087-8CA	C248H																	
	G3099A	D256N																	
	T3219C	L296P																	
	G3294A	V318I																	
	A3309T	I320F																	
	A3470G	silert E373																	
	T3489C	F388L																	
	A3517T	S389F																	
	A3653C	silert T434																	
	T3684C	S443P																	
	A3838G	D499A																	
G3896A	silert S515																		
T4019C	G583D																		
G4090A	E588K																		
Totals																			
	Total Unique Mutations	Totals Mutations		W	95%6fW	Line 1		W	95%6fW	Line 2									
		population	clones	population	clones	Total	Total	population	clones	Total	Total								
Silent Mutations:	20	5	15	0	3	8	11	0	2	8	10								
Missense Mutations:	29	8	22	1	4	11	16	1	5	11	17								
Total Mutations:	52	13	39	1	7	20	28	1	7	20	28								

Figure 4

Specific mutations found in Qβ phage selected on W and 6fW. Of the upper portion of the figure, the first column indicates the protein affected by mutations, the second column indicates the specific genetic mutation and the third column indicates the expected protein mutation. Later columns represent specific populations or clones as indicated. Mutations indicated in the second and third columns are present in the phage in question if a particular cell is filled with the color corresponding to the specific protein in question, as indicated in the first column. The lower portion of the figure presents a summary of the mutations found.

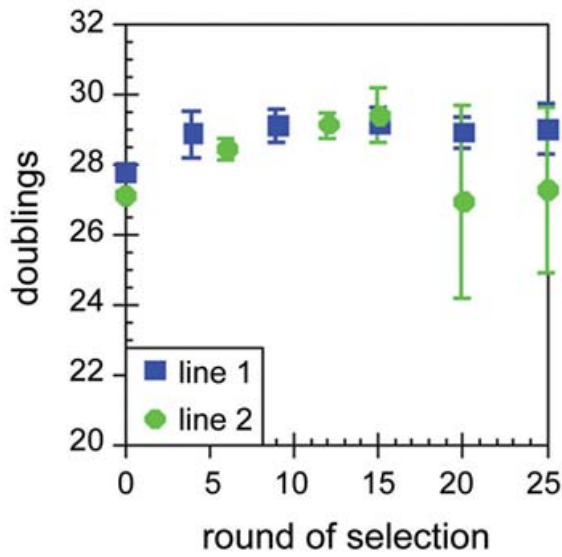


Figure 5
 Fitness of selected populations on W. Populations were tested for fitness on W at various points over the course of the selection. Error bars represent standard deviations of at least three replicates.

T223N substitutions in the Q β A1 protein occur successively in a stem-loop structure [38]. The U2006G (S221R) mutation converts an A:U base-pair into an A:G mismatch, while the C2011A (T223N) mutation converts a G:C base-pair to a G:A mismatch. However, given that both of these mutations would be expected to destabilize the stem structure, it is telling that no non-coding mutations were found that would similarly destabilize this structure. Moreover, if silent mutations were involved in functional alterations of RNA structure then it might be expected that compensatory base-pairing mutations would have been observed. For example, when Q β was selected to grow in a *hfq* host a G:C base pair was found to be mutated to an A:U base pair [31,39]. This covariation destabilized the 3'-terminus of the plus strand and promoted melting of the phage RNA structure, a function ascribed to Hfq. No such compensatory base-pairing mutations were found in our selection. Overall, the simplest explanation for the fixation of amino acid substitutions is that that these substitutions preserved the stability or function of Q β proteins in the presence of a mixture of W and 6fW.

Additional experiments revealed that the adaptive mutations allowed the phage to better tolerate a mixture of

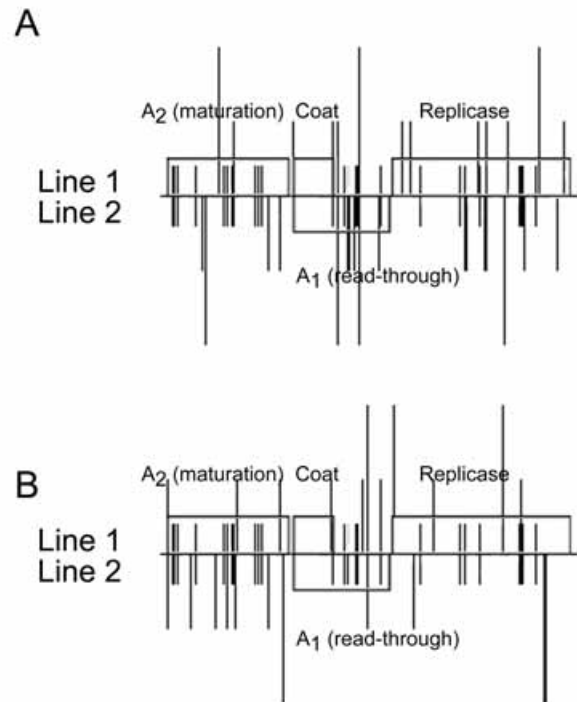


Figure 6
 Distribution of mutations in the Q β genome. (A) Missense and (B) Silent mutations are mapped onto the genome of Q β . Short bars represent locations of codons for tryptophan, mid-sized bars represent mutations found only in clones and full-height bars represent mutations found in populations. Mutations extending upwards show mutations found in Line 1, while mutations extending downwards from the genome represent mutations found in Line 2.

tryptophan and 6-fluorotryptophan, without loss of fitness on the wild-type amino acid (Figure 5). Moreover, fitness remained the same or improved slightly when evolved phage were assayed on eight other tryptophan analogues (Figure 1). The retention of fitness under multiple growth conditions was not a foregone conclusion. For example, when ϕ X174 phage were adapted to grow on *Salmonella typhimurium*, they lost the ability to infect *E. coli* C [25]. *E. coli* adapted to grow on glucose elicited no growth improvement on maltose [40,41]. While the same bacteria evolved to grow at 37°C lost fitness at temperatures further from optimal, they gained fitness at nearby temperatures [42]. One likely explanation for the lack of a trade-off during growth on the unnatural amino acid is that the natural amino acid was still present, and thus any given tryptophan codon would have had to accommodate both compounds at some point in the evolutionary history of the phage. This may also explain why there was no

loss of fitness on a number of other tryptophan analogues.

The most important aspect of these results, though, is that they reveal that it is unlikely that the original diminution in phage fitness and subsequent evolutionary recovery were a consequence of the diminished growth rate of the host on the unnatural amino acid. Strain C600p, a strain of *E. coli* closely related to the host strain used here, has been shown to grow robustly in 95%6fW, but approximately half as well in 95%4fW [6]. In contrast, Q β phage grew poorly on hosts grown in 95%6fW, but grew as well in hosts grown in 95%4fW as hosts grown in pure tryptophan (Figure 1). Thus, it is the effects of the amino acid on the phage itself that seem to be functionally important, as opposed to any indirect effects due to changes in host fitness.

Overall, these results have implications for the origins of alternate genetic codes. Several competing hypotheses for codon reassignment have been proposed (reviewed in [43]). The first of these hypotheses, the 'disappearing intermediate' hypothesis [44-46], posits that certain codons were eliminated by genetic drift throughout genomes that evolved skewed GC or AT contents. Following codon loss, relevant tRNA adaptors became functionless and were deleted. At some later point in evolution sequence composition changed again, and a different tRNA adaptor duplicated, mutated at the anticodon position, and recaptured the codon which had previously disappeared. A variant of this hypothesis suggests that evolutionary pressure on a number of genotypic characteristics, including genome size and organization as well as composition, may have influenced codon reassignment [47].

Alternatively, in the 'ambiguous intermediate' hypothesis [48-50] a duplicated and mutated tRNA could recognize a normally non-cognate codon and insert its amino acid in competition with the cognate amino acid. Propagation of organisms with ambiguous proteomes could occur if the non-cognate amino acid were either close to selectively neutral or provided a net selective advantage that overcame any deficits in the function of individual proteins. The further evolution of those proteins whose functions were compromised by amino acid substitutions would eventually repair any minor decreases in fitness. Following the adaptation of individual proteins, a discrete but altered genetic code could be re-established.

In our system, incorporation of the amino acid analogue was beneficial relative to growth in the presence of low or no tryptophan, yet still caused a decrease in phage fitness. This is analogous to the finding that a yeast tRNA that ambiguously encoded serine and leucine allowed growth

in diverse environments, yet also led to a 50% decrease in growth rate [14-16]. Thus, the requirements for an experimental test of the ambiguous intermediate hypothesis were established. The fact that fitness deficits in the phage Q β proteome were overcome by amino acid substitutions unrelated to the ambiguous amino acid itself strongly supports the 'ambiguous intermediate' hypothesis. The evolved phage are a plausible, experimental example of the penultimate step in amino acid substitution under the ambiguous intermediate model. By way of comparison, a failure to isolate phage with increased fitness on 6fW or the widespread elimination of tryptophan codons would have indicated that codon ambiguity was not an acceptable evolutionary path. Of course, the substitution of an even more chemically dissimilar amino acid might have generated an intractable barrier to evolution.

Conclusion

As in the natural selection of an ambiguous intermediate, evolutionary engineering of an unnatural organism should occur in stages. First, the incorporation of an unnatural amino acid into a proteome, and second the adaptation of the proteome to the unnatural amino acid. Previous experiments have focused largely on the first stage. Taken together, our experiments now suggest that while amino acid ambiguity is poorly tolerated initially, a secondary, proteomic adaptation to ambiguity is possible. Of course, the number of proteins in the phage proteome is of course small relative to larger, organismal proteomes. In this regard, our results with Q β phage can be seen as either discouraging or encouraging. From one vantage, the fact that three of the four Q β proteins accumulated substitutions in order to increase the fitness of the phage may imply that literally thousands of independent mutations may be required to isolate organisms that can fully utilize unnatural amino acids. Alternatively, only a few proteins critical for growth may need to adapt to chemical ambiguity, and the highly interdependent phage proteins may therefore all have been under selection pressure. This latter interpretation is most in keeping with the single example of an organism that has been evolved to have an altered genetic code. Starting with a *B. subtilis* auxotroph, Wong evolved a strain that could not only fully substitute 4fW throughout its proteome, but actually preferred 4fW for growth [19]. While the number and type of genomic mutations responsible for this phenotype are not known, the strain was generated via only four sequential rounds of mutation and selection. The most parsimonious hypothesis for these results, that only a few key proteins in the bacteria were mutated, is consonant with our observation that a relatively small number of mutations were required to adapt the Q β phage proteome for chemical ambiguity. Irrespective of whether critical targets were spread throughout an organismal proteome or concentrated in the highly interdependent phage proteome, these targets

Table 2: Oligonucleotides used for reverse transcription, PCR amplification and sequencing of the Q β genome.

Oligo name	Position on phage genome	Sequence (5' --> 3')
Qb5'2	1–40	GGGGACCCCCTTTAGGGGGTCACCTCACACAGCAGTACTT
Seq1A	385–411	GCGAACCTTGTCATTAATGACTTTCTG
Seq1revA	510–486	CATAGTGCCATACTTGGCATTAAAC
Seq2A	826–852	GCCGTTGCTTACTTTAAACTGAAAGGG
Qb.rtA3	948–928	GGACACAGGCCGGAAGGGAACC
Seq3A	1295–1315	TTTTATTAACCCAACGCGTAA
Qb.rtA4	1427–1401	ATTTACCCACGCGGATTGAGGACCAG
Seq4A	1730–1752	GAACCCAGCGTATTGAACACTGC
Qb.rtA5	1948–1922	ACATCAAATTCGCGAGGCTGGAGTTCA
AmpB	1998–2020	CGGCTTAGTTATACCACGTTCCG
Qb.rtA	2190–2171	CGCTAAGAGAGCAATAAGCA
Seq1B	2374–2392	GTAACCTCTCTCAGCGCACAA
Seq1revB	2495–2470	AGAGTTAAATGGCGACTGACCATAGG
Seq2B	2820–2841	GGTGCTACAACAACGAATAACC
Qb.rtB4	2947–2925	CGAAGATGTGTAAAGCTATGGTC
Seq3B	3280–3303	CTGACGGTATGTGTTGTACCTACG
Qb.rtB3	3385–3364	TCACAAACGGAACGAGCGAGAG
Seq4B	3738–3760	ATACCTGATGGTTACGGTATGG
Qb.rtB2	3852–3828	GGTCCCTTGATGGTCCGTAATCAC
3'Qb-pTlink	4217–4201	B-AAAAAAAAAAAAAAAAAATGGGAGGAGAGAGGGCA ^a

^aThe primer 3'Qb-pTlink has a 5' biotin and poly-A tract. While this oligo was used for amplification and sequencing, these features were not required for either process.

evolved in response to the change in the genetic code. The evolution of phage with chemically ambiguous proteomes now provides a springboard to the evolution of phage with novel genetic codes, and a means to quantify the relative evolutionary costs of such changes.

Methods

Strains, reagents, and media

Q β bacteriophage was a gift of D. R. Mills (Health Science Center at Brooklyn, State University of New York). *E. coli* C600F (thi-1 thr-1 leuB6 lacY1 tonA21 supE44 mcrA Δ trpE F'kanR) was a gift of I. J. Molineux (University of Texas at Austin). The strain C600F(DE3) was constructed using the DE3 lysogenization kit (Novagen, Madison, WI). Media were as described [6]; 6fW and other tryptophan analogues were from Sigma (St. Louis, MO). Oligonucleotides used in these experiments are listed in Table 2 and were obtained from IDT (Coralville, IA).

Selection on W

Each round of selection consisted of plating approximately 1000 plaque-forming units with C600F on M9B1TLW + Kn. Once plaques were visible, top agar was scraped off of the plate and incubated with 5 ml of PBS at 37°C for 15'. Chloroform (5 ml) was added and the solution was mixed thoroughly with a vortex. Organic and aqueous phases were separated by centrifugation at 6000 rpm for 15'. The aqueous phase was filtered through a 0.22 μ m filter (Pall Gelman Laboratories, Ann Arbor, MI)

to a fresh tube and two aliquots (1 ml each) were taken for storage. Phage were titered on LB + Kn, and diluted appropriately such that approximately 1000 plaque-forming units were used for subsequent rounds of selection. This process was repeated for twenty-five cycles.

Selection on 95% 6fW

Phage from round 10 of the selection on M9B1TLW + Kn plates were further selected for 25 rounds on M9B1TL95%6fW + Kn plates. Since plaques were never visible, each round of selection was carried out for a standard 20 hours. PBS (2 ml) was used to recover the phage, and 100 μ l of solution was used in the subsequent rounds of selection. Phage were titered with C600F on LB + Kn. The phage solution was not extracted with chloroform.

Fitness assays

Host bacteria for fitness assays were grown up in LB + Kn to a concentration of approximately 10⁸ colony-forming units/ml. The culture was spun down and resuspended in 1/100 volume of 20% glycerol, aliquoted, and stored at -80°C. Aliquots were thawed as needed and grown for 1 hour in 100 volumes M9B1TLW + Kn before plating. This procedure served to standardize the physiological state of cells used for assays.

After 1 hour of growth, bacteria were plated with ca. 1000 phage from the population being assayed. Plates varied in terms of what amino acids or analogues were added, but

were always M9B1TL + Kn. After 20 hours of growth at 37°C, the top agar was scraped away, and phage were eluted in PBS (2 ml), spun down at 6000 rpm for 15', and phage were titered on LB + Kn in parallel with phage stocks used to initiate the fitness assay. Plaques were counted and fitness was expressed as the number of doublings in a 20 hour period according to the equation $\log_2(\# \text{ of phage at end of assay}) - \log_2(\# \text{ phage at start of assay})$.

Sequencing of phage genomes

Phage RNA genomes were purified essentially as previously described [51]. In brief, 100 µl of a solution of phage, either directly from the selection or from a population grown on C600F in LB, was extracted with phenol:chloroform:0.1% SDS, chloroform extracted, ethanol precipitated, resuspended in 50 µl of water, and passed through a Centri-Sep column (Princeton Separations, Adelphia, NJ) to remove unincorporated small molecules.

Purified phage genome was used for reverse transcription (Superscript II RT kit, Invitrogen, Carlsbad, CA). In short, 10 µl of phage RNA, 9 µg of random hexamers and water to a total of 21 µl was heated to 70°C for 3', placed on ice, and the remainder of the reaction was assembled according to the manufacturer's instructions. Reverse transcription reactions were incubated for 1 hour at 42°C. A portion of this reaction (4 µl) was used to seed polymerase chain reactions (100 µl). Different reactions contained different primers to amplify different portions of the phage genome. PCR products were gel-purified (QIAquick Gel Extraction Kit, Qiagen, Valencia, CA) prior to sequencing. The complete sequence of the wild-type phage genome has been deposited in GenBank (accession number AY099114). The primers used for the amplification of the genome limited our ability to identify sequence changes to nucleotides 40–4200 of the phage genome.

In some instances, phage were first grown on C600F in LB + Kn prior to reverse transcription and sequencing. In order to ensure that growth on LB did not drastically affect the distribution of phage genotypes, a 1.6 Kb region of the phage genome was also sequenced from non-LB-grown phage stocks. The sequences were found to be identical to those from LB-grown phage stocks.

Determination of amino acid incorporation ratios

Global amino acid incorporation ratios were determined from 100 ml overnight cultures of C600F grown on M9B1TLW + Kn or M9B1TL95%6fW + Kn. The bacteria were spun down and lysed in 200 µl B-PER II (Pierce, Beverly, MA). Half of this volume was passed through a Centri-Sep column. The eluant was dried down and

hydrolyzed overnight in 5.4 M HCl, 10% thioglycolic acid at 110°C under vacuum. Hydrolysates were again dried down, and then resuspended in 50 µl of water. These hydrolysates were analysed by HPLC-ESI at the Mass Spectrometry Facility at the University of Texas at Austin. Hydrolysates were also analyzed by HPLC. Samples (20 µl) were injected onto a C-18 column and eluted with 50 mM NH₄OAc, pH 5.0 in a 3% to 1% MeOH gradient. Peaks were collected and lyophilized, followed by reinjection on the same column and developed with 0.1 M NaH₂PO₄, pH 2.5, 10% MeOH. Identities of peaks that absorb at 280 nm were confirmed by determining the elution times of standards.

Amino acid incorporation ratios in a single protein, green fluorescent protein, were also determined. The gene for the highly fluorescent protein GFPuv [52] (Clontech, Palo Alto, CA) was PCR-amplified with Vent DNA polymerase (NEB, Beverly, MA) with primers CFPA (5'-CACCACG-GCCACTGTGGCCATGAGTAAAGGAGAAGAAGACTT-3') and CFPB (5'-GGCCATCGGGGCCCTATTTGTATAGT-TCATCCATGCC-3'). The GFPuv gene was cloned into the plasmid pET100/D/topo (Invitrogen) and transformed into TOP10 cells. The resultant plasmid pET100GFPuv was purified (QIAprep Miniprep Spin Kit, Qiagen) and used to transform C600F(DE3). Overnight M9B1TLW + Kn + Ap cultures of C600F(DE3)+pET100GFPuv were diluted 1:100 into 400 mL M9B1TLW + Kn + Ap. At mid-log phase, cultures were split and spun down. Pellets were resuspended in 200 mL of either M9B1TLW + Kn + Ap or M9B1TL95%6fW + Kn + Ap, each with 1 mM IPTG and grown for an additional 3 hours. Cultures were again spun down, lysed in 3 mL B-PER, and purified on 3 mL Ni-NTA columns at room temperature as recommended by the manufacturer (Novagen). Purified GFPuv was hydrolyzed as described above and analyzed by HPLC-ESI.

Abbreviations

4-fluorotryptophan, 4fW. 6-fluorotryptophan, 6fW.

Authors' contributions

J.M.B. jointly conceived of the experiment, performed all of the experiments, and analyzed the data; J.J.B. assisted with experimental design and data interpretation, and A.D.E. jointly conceived of the experiment, and participated in experimental design and data interpretation.

All authors read and approved the final manuscript.

Acknowledgements

This work was supported by grants from the NASA Astrobiology Institute, NCC2-1055 to A. D. E. and NIH GM 57756 to J. J. B. J. M. B. was supported as a Harrington Dissertation Fellow. We thank L. Jayant and D. R. Mills for providing their version of the sequence of the Qβ genome.

References

1. Fox TD: **Natural variation in the genetic code.** *Annu Rev Genet* 1987, **21**:67-91.
2. Jakubowski H, Goldman E: **Editing of errors in selection of amino acids for protein synthesis.** *Microbiol Rev* 1992, **56**:412-429.
3. Rodnina MV, Wintermeyer W: **Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms.** *Annu Rev Biochem* 2001, **70**:415-435.
4. Radman M, Matic I, Taddei F: **Evolution of evolvability.** *Ann N Y Acad Sci* 1999, **870**:146-155.
5. Xu ZJ, Love ML, Ma LY, Blum M, Bronskill PM, Bernstein J, Grey AA, Hofmann T, Camerman N, Wong JT: **Tryptophanyl-tRNA synthetase from *Bacillus subtilis*. Characterization and role of hydrophobicity in substrate recognition.** *J Biol Chem* 1989, **264**:4304-4311.
6. Bacher JM, Ellington AD: **Selection and Characterization of *Escherichia coli* Variants Capable of Growth on an Otherwise Toxic Tryptophan Analogue.** *J Bacteriol* 2001, **183**:5414-5425.
7. Browne DR, Kenyon GL, Hegeman GD: **Incorporation of monofluorotryptophans into protein during the growth of *Escherichia coli*.** *Biochem Biophys Res Commun* 1970, **39**:13-19.
8. Zhang QS, Shen L, Wang ED, Wang YL: **Biosynthesis and characterization of 4-fluorotryptophan-labeled *Escherichia coli* arginyl-tRNA synthetase.** *J Protein Chem* 1999, **18**:187-192.
9. Pratt EA, Ho C: **Incorporation of fluorotryptophans into proteins of *Escherichia coli*.** *Biochemistry* 1975, **14**:3035-3040.
10. Parsons JF, Xiao G, Gilliland GL, Armstrong RN: **Enzymes harboring unnatural amino acids: mechanistic and structural analysis of the enhanced catalytic activity of a glutathione transferase containing 5-fluorotryptophan.** *Biochemistry* 1998, **37**:6286-6294.
11. Randhawa ZI, Witkowska HE, Cone J, Wilkins JA, Hughes P, Yamanishi K, Yasuda S, Masui Y, Arthur P, Kletke C, Bitsch F, Shackleton CHL: **Incorporation of norleucine at methionine positions in recombinant human macrophage colony stimulating factor (M-CSF, 4-153) expressed in *Escherichia coli*: structural analysis.** *Biochemistry* 1994, **33**:4352-4362.
12. Tsai LB, Lu HS, Kenney WC, Curless CC, Klein ML, Lai PH, Fenton DM, Altrock BW, Mann MB: **Control of misincorporation of de novo synthesized norleucine into recombinant interleukin-2 in *E. coli*.** *Biochem Biophys Res Commun* 1988, **156**:733-739.
13. Döring V, Mootz HD, Nangle LA, Hendrickson TL, de Crécy-Lagard V, Schimmel P, Marlière P: **Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway.** *Science* 2001, **292**:501-504.
14. Perreau VM, Keith G, Holmes WM, Przykorska A, Santos MA, Tuite MF: **The *Candida albicans* CUG-decoding ser-tRNA has an atypical anticodon stem-loop structure.** *J Mol Biol* 1999, **293**:1039-1053.
15. Santos MA, Ueda T, Watanabe K, Tuite MF: **The non-standard genetic code of *Candida* spp.: an evolving genetic code or a novel mechanism for adaptation?** *Mol Microbiol* 1997, **26**:423-431.
16. Santos MA, Cheesman C, Costa V, Moradas-Ferreira P, Tuite MF: **Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida* spp.** *Mol Microbiol* 1999, **31**:937-947.
17. Mehl RA, Anderson JC, Santoro SW, Wang L, Martin AB, King DS, Horn DM, Schultz PG: **Generation of a bacterium with a 21 amino acid genetic code.** *J Am Chem Soc* 2003, **125**:935-939.
18. Wang L, Brock A, Herberich B, Schultz PG: **Expanding the genetic code of *Escherichia coli*.** *Science* 2001, **292**:498-500.
19. Wong JT: **Membership mutation of the genetic code: loss of fitness by tryptophan.** *Proc Natl Acad Sci U S A* 1983, **80**:6303-6306.
20. Iglesias-Ussel MD, Casado C, Yuste E, Olivares I, Lopez-Galindez C: **In vitro analysis of human immunodeficiency virus type 1 resistance to nevirapine and fitness determination of resistant variants.** *J Gen Virol* 2002, **83**:93-101.
21. Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ: **Different trajectories of parallel evolution during viral adaptation.** *Science* 1999, **285**:422-424.
22. Hanczyc MM, Dorit RL: **Replicability and recurrence in the experimental evolution of a group I ribozyme.** *Mol Biol Evol* 2000, **17**:1050-1060.
23. Mekler P: **Determination of nucleotide sequence of the bacteriophage QB genome: organization and evolution of an RNA virus.** *University of Zurich*; 1981.
24. Domingo E, Sabo D, Taniguchi T, Weissmann C: **Nucleotide sequence heterogeneity of an RNA phage population.** *Cell* 1978, **13**:735-744.
25. Crill WD, Wichman HA, Bull JJ: **Evolutionary reversals during viral adaptation to alternating hosts.** *Genetics* 2000, **154**:27-37.
26. Drake JW: **Rates of spontaneous mutation among RNA viruses.** *Proc Natl Acad Sci U S A* 1993, **90**:4171-4175.
27. Drake JW, Charlesworth B, Charlesworth D, Crow JF: **Rates of spontaneous mutation.** *Genetics* 1998, **148**:1667-1686.
28. Palasingam K, Shaklee PN: **Reversion of Q beta RNA phage mutants by homologous RNA recombination.** *J Virol* 1992, **66**:2435-2442.
29. Mills DR, Priano C, DiMauro P, Binderow BD: **Q beta replicase: mapping the functional domains of an RNA-dependent RNA polymerase.** *J Mol Biol* 1989, **205**:751-764.
30. van Duin Jan: **The single-stranded RNA bacteriophages.** *The Viruses: The Bacteriophages* Edited by: Calendar Richard. New York, Plenum Press; 1988:2 v..
31. Schuppli D, Miranda G, Tsui HC, Winkler ME, Sogo JM, Weber H: **Altered 3'-terminal RNA structure in phage Qbeta adapted to host factor-less *Escherichia coli*.** *Proc Natl Acad Sci U S A* 1997, **94**:10239-10242.
32. Schumacher MA, Pearson RF, Møller T, Valentin-Hansen P, Brennan RG: **Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein.** *Embo J* 2002, **21**:3546-3556.
33. Møller T, Franch T, Højrup P, Keene DR, Bächinger HP, Brennan RG, Valentin-Hansen P: **Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction.** *Mol Cell* 2002, **9**:23-30.
34. Bernhardt TG, Wang IN, Struck DK, Young R: **A protein antibiotic in the phage Qbeta virion: diversity in lysis targets.** *Science* 2001, **292**:2326-2329.
35. Kozlovskaya TM, Cielens I, Dreilinn D, Dislers A, Baumanis V, Ose V, Pumpens P: **Recombinant RNA phage Q beta capsid particles synthesized and self-assembled in *Escherichia coli*.** *Gene* 1993, **137**:133-137.
36. Arora R, Priano C, Jacobson AB, Mills DR: **cis-acting elements within an RNA coliphage genome: fold as you please, but fold you must!!** *J Mol Biol* 1996, **258**:433-446.
37. Olsthoorn RC, Lics N, van Duin J: **Leeway and constraints in the forced evolution of a regulatory RNA helix.** *Embo J* 1994, **13**:2660-2668.
38. Skripkin EA, Jacobson AB: **A two-dimensional model at the nucleotide level for the central hairpin of coliphage Q beta RNA.** *J Mol Biol* 1993, **233**:245-260.
39. Schuppli D, Georgijevic J, Weber H: **Synergism of mutations in bacteriophage Qbeta RNA affecting host factor dependence of Qbeta replicase.** *J Mol Biol* 2000, **295**:149-154.
40. Travisano M, Mongold JA, Bennett AF, Lenski RE: **Experimental tests of the roles of adaptation, chance, and history in evolution.** *Science* 1995, **267**:87-90.
41. Travisano M: **Long-term experimental evolution in *Escherichia coli*. VI. Environmental constraints on adaptation and divergence.** *Genetics* 1997, **146**:471-479.
42. Cooper VS, Bennett AF, Lenski RE: **Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment.** *Evolution Int J Org Evolution* 2001, **55**:889-896.
43. Knight RD, Freeland SJ, Landweber LF: **Rewiring the keyboard: evolvability of the genetic code.** *Nat Rev Genet* 2001, **2**:49-58.
44. Osawa S, Jukes TH: **Codon reassignment (codon capture) in evolution.** *J Mol Evol* 1989, **28**:271-278.
45. Osawa S, Jukes TH: **Evolution of the genetic code as affected by anticodon content.** *Trends Genet* 1988, **4**:191-198.
46. Jukes TH, Osawa S: **Evolutionary changes in the genetic code.** *Comp Biochem Physiol B* 1993, **106**:489-494.
47. Andersson SG, Kurland CG: **Genomic evolution drives the evolution of the translation system.** *Biochem Cell Biol* 1995, **73**:775-787.
48. Yarus M, Schultz DW: **Further comments on codon reassignment. Response.** *J Mol Evol* 1997, **45**:3-6.
49. Schultz DW, Yarus M: **On malleability in the genetic code.** *J Mol Evol* 1996, **42**:597-601.

50. Schultz DW, Yarus M: **Transfer RNA mutation and the malleability of the genetic code.** *J Mol Biol* 1994, **235**:1377-1380.
51. Beekwilder MJ, Nieuwenhuizen R, van Duin J: **Secondary structure model for the last two domains of single-stranded RNA phage Q beta.** *J Mol Biol* 1995, **247**:903-917.
52. Cramer A, Whitehorn EA, Tate E, Stemmer WP: **Improved green fluorescent protein by molecular evolution using DNA shuffling.** *Nat Biotechnol* 1996, **14**:315-319.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

