7854–7860 Nucleic Acids Research, 2013, Vol. 41, No. 16 Published online 19 June 2013
doi:10.1093/nar/gkt507

Conservation of mRNA secondary structures may filter out mutations in *Escherichia coli* evolution

Andrey Chursov[1], Dmitrij Frishman[1,2,*] and Alexander Shneider[3,*]

[1]Department of Genome Oriented Bioinformatics, Technische Universität München, Wissenschaftzentrum Weihenstephan, Maximus-von-Imhof-Forum 3, D-85354, Freising, Germany, [2]Helmholtz Center Munich—German Research Center for Environmental Health (GmbH), Institute of Bioinformatics and Systems Biology, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany and [3]Cure Lab, Inc., 43 Rybury Hillway, Needham, MA 02492, USA

Received January 11, 2013; Revised April 29, 2013; Accepted May 1, 2013

ABSTRACT

Recent reports indicate that mutations in viral genomes tend to preserve RNA secondary structure, and those mutations that disrupt secondary structural elements may reduce gene expression levels, thereby serving as a functional knockout. In this article, we explore the conservation of secondary structures of mRNA coding regions, a previously unknown factor in bacterial evolution, by comparing the structural consequences of mutations in essential and nonessential *Escherichia coli* genes accumulated over 40 000 generations in the course of the 'long-term evolution experiment'. We monitored the extent to which mutations influence minimum free energy (MFE) values, assuming that a substantial change in MFE is indicative of structural perturbation. Our principal finding is that purifying selection tends to eliminate those mutations in essential genes that lead to greater changes of MFE values and, therefore, may be more disruptive for the corresponding mRNA secondary structures. This effect implies that synonymous mutations disrupting mRNA secondary structures may directly affect the fitness of the organism. These results demonstrate that the need to maintain intact mRNA structures imposes additional evolutionary constraints on bacterial genomes, which go beyond preservation of structure and function of the encoded proteins.

INTRODUCTION

Increasing experimental (1) and computational (2,3) evidence points to the existence of extensive RNA structures in the coding regions of mRNA molecules. RNA secondary structures have been implicated in regulation of translation initiation, elongation and termination in both prokaryotes and eukaryotes (4,5). In particular, the anti-correlation between translation efficiency and the thermodynamic stability of local secondary structure in the vicinity of the translation initiation site has been thoroughly documented (6). RNA hairpins are thought to be involved in controlling mRNA decay (7), localization (8–10) and interaction with other molecules (11). Overall, the mRNA coding regions appear to be more structured than the untranslated regions (1) and have lower minimum folding free energies. Hence, the mRNA coding regions appear to have more stable structures than codon-randomized sequences (12). Owing to the need to simultaneously preserve both the function and structure of the encoded protein, as well as the structural elements of the RNA molecule itself, mRNA coding regions are subject to dual selection pressure.

Using a mammalian system, we have recently shown that mutations altering secondary structures of influenza mRNAs may serve as a functional knockout of the corresponding genes (13). More recently, Moss *et al.* (14) established a direct connection between mutation patterns in the influenza virus genome and the hydrogen-bonding patterns shaping RNA structures. Thus, preservation of viral RNA structures and elimination of mutations disruptive for RNA structures may be a previously unknown mechanism of viral evolution. In the present article, we put forward the hypothesis that conservation of RNA structures may also play a role in bacterial evolution. To examine this hypothesis, we compared the genomes of parental and progeny *Escherichia coli* clones standing 40 000 generations apart. The 'long-term evolution experiment' (15–18) tracking genetic changes in 12 populations of *E. coli* was started by Richard Lenski in February 1988. All 12 replicate populations have originated from a single cell of the baseline strain, which was an *E. coli* B clone, and have been propagated at 37°C in liquid culture. Every 500 generations, samples for each population were frozen

*To whom correspondence should be addressed. Tel: +49 179 538 2799; Fax: +49 8161 712 186; Email: d.frishman@wzw.tum.de
Correspondence may also be addressed to Alexander Shneider. Tel: +1 609 841 1201; Fax: +1 260 572 3405; Email: ashneider@curelab.com

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from http://nar.oxfordjournals.org/ at GSF Forschungszentrum on October 24, 2013

away at −80°C and retained for sequencing and comparison with their predecessors.

If our hypothesis is correct, mutations in essential genes that disrupt mRNA secondary structures would lead to insufficient gene expression and, due to the essentiality of those genes, such mutations would be filtered out as lethal. By contrast, selection against mutations disrupting mRNA secondary structures of nonessential genes would be expected to be less pronounced because an altered expression level of nonessential genes would not influence bacterial propagation. Supplementary Figure S1 exemplifies predicted structural perturbations induced by mutations altering minimal free energy of an *E. coli* mRNA.

To demonstrate this effect, one would ideally need to calculate exact secondary structures for both the original and the mutated mRNAs, compare them and make an inference about the changes in the RNA structure caused by mutations. However, a single RNA molecule may fold into more than one conformation (19,20). With increasing sequence length, the number of possible structures that an RNA molecule can adopt with similar (in many cases even the same) values of folding energy increases as well (21), thereby resulting in diminished prediction accuracy. Another well-known complication is that predicting secondary structures with pseudoknots is an NP-hard problem (22), which necessitates using approximations in structure prediction algorithms. Therefore, instead of calculating explicit secondary structure shapes for mRNAs, we pursued an indirect method of assessing whether mutation(s) affect secondary structures by quantifying minimum free energy (MFE) change. While different RNA structures may have exactly the same MFE, different MFE values are guaranteed to correspond to different structures. Despite the fact that a mutation did not change MFE does not mean that the RNA structure remained the same, yet, an opposite situation is reliably conclusive. The more mutations change the MFE, the greater affect on a secondary RNA structure they have.

Using this approach, we investigated how mutations observed in essential and nonessential genes influence the MFE values of mRNA structures. This article presents evidence that mutations in essential genes of *E. coli* that occurred during the 'long-term evolution experiment' changed the MFE of mRNA secondary structures to a lesser extent than mutations in nonessential genes. We emphasize that we focus exclusively on the conservation of secondary structures of mRNA coding regions and do not consider noncoding RNAs. This finding supports our hypothesis that mutations disrupting the mRNA structure of essential genes are filtered out during the course of bacterial evolution.

MATERIALS AND METHODS

Experimental data on evolutionary mutations in *E. coli*

In our analysis, we used data on genetic polymorphisms in *E. coli* accumulated in the course of the 'long-term evolution experiment' (16–18). Specifically, mutations in the 40 000th generation of one of the populations (Ara-1), with the ancestral strain REL606 (GenBank accession number NC_012967.1), were investigated. In this 40 000th clone, 627 single-nucleotide polymorphisms (SNPs) and 26 deletions, insertions and other polymorphisms were detected. Hereinafter, we take into account only SNPs. Ninety-two mutations occurring in intergenic regions as well as six mutations in pseudogenes and one mutation in an insertion sequence element were excluded from consideration. We also ignored one SNP owing to an inconsistency between the mutated nucleic acid, as reported in (18) and the nucleic acid occurring at this position in the complete genome sequence. Two genes with available SNP data were not considered: one with an inconsistency between its nucleotide and amino acid sequences, and another that had one of the reported mutations in its start codon. Our final data set contained 523 mutations involving 485 genes.

Data on essential and nonessential genes of *E. coli*

There is no essentiality data available for the B strain of *E. coli*, but it is closely related to the well-studied *E. coli* K-12 MG1655 (23,24). For this latter strain (GenBank accession number U00096.2) Gerdes *et al*. (25) experimentally identified 620 genes as essential and 3126 genes as dispensable using a genetic footprinting technique. Because of the numerous discrepancies between the gene names, we conducted similarity-based transfer of essentiality data from the MG1655 strain to the REL606 strain, using the bidirectional best hit strategy to identify orthologous genes. Using blastp (26), we aligned all mutated genes from the REL606 genome against all genes from the MG1655 genome and *vice versa*. Genes from the two genomes were considered orthologous if they were the best hits for each other, with amino acid sequence identity >75% and e-value $<10^{-25}$. This procedure enabled us to map 456 out of the 485 mutated genes in REL606 to the MG1655 strain, of which 48 were essential, 348 dispensable and 60 had undefined essentiality according to the MG1655 annotation.

MFE values of RNA secondary structures

For each of the 48 essential and 348 nonessential mutated genes, we calculated MFE values of secondary RNA structures for both the original ancestral mRNAs and their 40 000th generation counterparts. We used the RNAfold tool from the Vienna RNA Package with the command line option noLP, which disallowed base pairs that can only occur as helices of length 1 (27).

Generation of randomly mutated mRNAs

For each gene reported in (18) as possessing mutation(s) in the 40 000th generation, we produced an *in silico* family of random counterparts. Synonymous random mutations were introduced into ancestral mRNAs. When compared with the ancestral strain, each computer-generated RNA sequence had the same number of point mutations as the respective 40 000th generation mutant. There are six types of possible nucleotide substitutions: C:G → A:U; A:U → C:G; A:U → U:A; C:G → G:C; C:G → U:A; A:U → G:C. We introduced random mutations in such a

way that the frequency for any given nucleotide substitution type was similar for essential and nonessential gene groups (Supplementary Table S1) and free of transition to transversion bias. For each gene, the ratio of transitions to transversions was calculated, and distributions of these ratios were compared for essential and nonessential genes. According to the Mann–Whitney U test, these distributions do not differ ($P = 0.38$). For the purpose of this work, we did not have to simulate *in silico* relative frequencies of nucleotide substitution observed reported by Wielgoss *et al.* (28). The MFE of secondary RNA structure was calculated for each computer-generated sequence by the RNAfold tool as described above.

The number of computer-generated sequences varied from gene to gene dependent on gene length (Figure 1). If a short gene possesses only one nucleotide substitution *in vitro*, the number of conceivable *in silico* generated sequences having only one nucleotide changed is limited to an exhaustive set of synonymous point mutations (e.g. 516 variants for the gene *yciT* of length 750). For sufficiently long genes (typically >1300 bases), the subset of 1000 sequences with randomly introduced SNPs was used for further analysis.

Statistical test

To find out whether *in vitro* mutations in essential and nonessential genes differ in their affect on MFE and mRNA secondary structures, we applied the following analysis. First, for each gene, we determined the absolute value of the difference between the MFE of the ancestor RNA and the MFE of the *in vitro* mutant, as well as that of each of the computer-generated mutants. Then, we calculated the fraction of computer-generated mutants whose absolute values of MFE differences were lower than the corresponding *in vitro* mutant. Each gene in the data set of essential genes and in the data set of dispensable genes was thus characterized by a percentile value. The Mann–Whitney U test was then used with the null hypothesis (H_0) that the percentile values for essential and nonessential genes are from the same distribution.

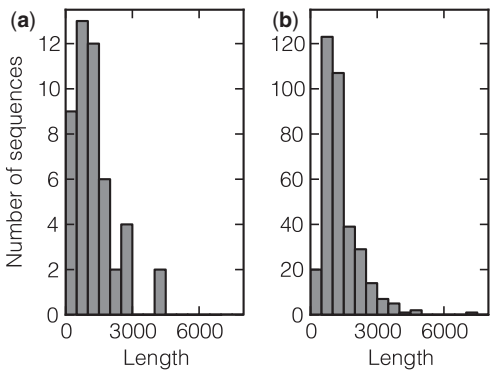


Figure 1. Histograms of gene lengths for essential (**a**) and nonessential (**b**) genes.

Data availability

The lists of defined essential and nonessential genes with the corresponding MFE values are presented in the Supplementary Tables S2 and S3, respectively.

RESULTS

The main scientific questions we addressed in this study are whether purifying selection tends to eliminate mutations that are disruptive for mRNA structures, and whether this effect is more pronounced in essential genes compared with dispensable ones. Our methodology involved a comparison of actual mutations observed in the 40 000th generation of the 'long-term evolution experiment' (18) with a pool of random computer-generated mutations.

As an example, for a gene harboring two point mutations, we generated a thousand *in silico* mutants with two mutations each. MFE was calculated for the ancestor mRNA as well as for the mRNA of the 40 000th generation mutant experimentally observed in a Petri dish and for those of the *in silico* mutants. Owing to slight sequence changes, the RNA folding energies of both experimentally recorded and computer-generated mutants will be somewhat different from the MFE of the ancestor's mRNA. We calculated the fraction of *in silico* mutants with a lesser extent of MFE change than the mutant observed *in vitro*. Suppose, for example, that the MFE of the ancestor mRNA was −5 kcal/mol and that the MFE of the *in vitro* mutant differs from it by 2 kcal/mol (it does not matter whether the MFE went down to −7 or went up to −3 kcal/mol). If 700 out of 1000 computer-generated mutants have their MFEs either >-3 or <-7 kcal/mol, it means that for this particular gene a mutant recorded in the *in vitro* experiment changes its MFE to a greater extent than 30% of the randomly mutated sequences.

Suppose that experimentally observed mutations in essential genes lead to bigger MFE changes than only 10% of random mutations, while in the data set of mutations in nonessential genes, MFE changes bigger than those of random mutants are observed in 50% of the cases. This would indicate that the evolutionary constraints acting on mRNA structure in essential genes are stronger that those acting on dispensable genes.

In the *E. coli* genome sampled at 40 000th generation, 523 nucleotide substitutions occurred in 485 genes (11.5% of all *E. coli* genes), of which 48 genes were essential, 348 nonessential and 89 genes either could not be successfully mapped from the REL606 to the MG1655 strain or had unknown essentiality status (Table 1). A great majority of the mutated genes (92.2%) have only one SNP mutation. For the mutated genes, the ratio between the number of essential and nonessential genes is 0.138; while for nonmutated genes, this ratio is 0.206. The latter finding is in agreement with the report by Jordan *et al.* (29) showing that essential genes in bacteria accumulate mutations less frequently than nonessential genes do. The majority of mutations are nonsynonymous (Table 2), with the ratio of synonymous to nonsynonymous

Table 1. The number of all, essential and nonessential genes in which a particular number of SNPs occurred during the 'long-term evolution experiment' between the first and the 40 000th generations

Genes type	Number of mutations per gene			Total
	1	2	3	
All genes[a]	447	36	2	485
Essential genes	43	5	0	48
Nonessential genes	318	29	1	348

[a]Including those with unknown essentiality status.

Table 2. The divisions of synonymous and nonsynonymous mutations

Mutation type	All genes	Essential genes	Nonessential genes
Synonymous	83	4	70
Nonsynonymous	442	49	309

mutations in essential genes (0.082) being somewhat lower than in nonessential genes (0.227). The *P*-value calculated using a binomial test (4 synonymous SNPs out of 53 in essential genes *vs* 70 synonymous SNPs out of 379 in nonessential genes) equals 0.048.

SNPs cause changes in the MFE and structural perturbations in many, though not all, mRNAs (Table 3). Specifically, 27.1% of essential genes do not change the MFE value, while only 14.7% of nonessential genes demonstrate the same MFE values for both original and mutated mRNAs. In general, SNPs in essential genes change the MFE values (median = 0.69 kcal/mol) to a smaller extent than do mutations in nonessential genes (median = 1.10 kcal/mol) (Figure 2). We compared the properties of essential and nonessential genes that could influence MFE calculations, but found no confounding factors (data not shown). Both groups of genes have the same average GC content. While essential genes tend to be somewhat shorter than nonessential ones (Figure 1), neither in essential nor in nonessential genes do the differences in MFE between the native and mutated sequences depend on mRNA length. Additionally, different types of mutations (e.g. C → G) occur in these two data sets equally often. At the same time, mutations observed at the 40 000th generation *in vitro* are more likely to reduce MFE of nonessential genes than the essential ones. MFE value decreased in 56.0% of the nonessential mutants, while only 45.8% of the essential ones demonstrated MFE reduction. A possible interpretation could be that ancestral essential genes were folded in structures that caused the values of their folding energies to be close to the minimum (robust); in contrast, nonessential genes had MFEs more distant from the minimal values. Thus, mutations were less likely to reduce energies of essential genes.

We subsequently compared the absolute values of MFE changes caused in each mRNA by naturally occurring and an equal number of randomly introduced synonymous

Table 3. The number of essential and nonessential genes that decrease, increase or do not change their MFE value on mutation

Gene type	$MFE_{mutant} - MFE_{original}$			Total
	<0	=0	>0	
Essential	22 (45.8%)	13 (27.1%)	13 (27.1%)	48
Nonessential	195 (56.0%)	51 (14.7%)	102 (29.3%)	348

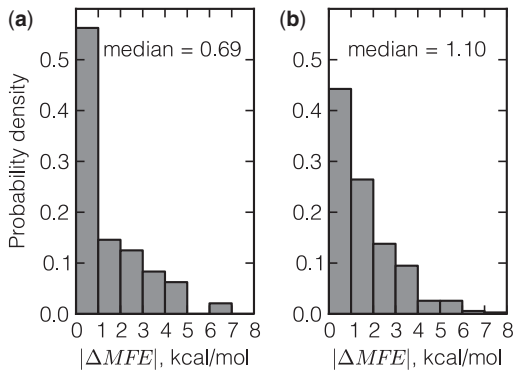


Figure 2. Histograms of absolute changes in MFE values for essential (**a**) and nonessential (**b**) genes.

mutations, thus avoiding those mutations in the sequences, generated *in silico*, that could be eliminated by purifying selection due to their effect on the encoded protein. For each mRNA, we determined the fraction of *in silico* derivatives, which change their MFEs less than the mutant observed *in vitro*. These percentages are much lower in essential *E. coli* genes than in nonessential genes (Table 4), implying that mutations accumulated in essential *E. coli* genes affect MFEs (and hence secondary structure) to a lesser extent than mutations in nonessential genes. This effect is further demonstrated by the fact that the cumulative distribution function corresponding to essential genes elevates considerably faster at the beginning (Figure 3). The difference between values for essential and nonessential genes is statistically significant according to the Mann–Whitney U test ($P = 0.044$). Our results suggest that mRNA secondary structure imposes substantially smaller selective pressure at the mutations taking place in nonessential genes because the median of their effect on MFE is 50.2% of what the random set of mutations would cause. By contrast, the median value of how mutations occurring in essential genes influence MFEs is only 32.6% of what random mutations would do.

DISCUSSION

The importance of mRNA secondary structure for gene expression was demonstrated for many organisms including bacteria (30–33), human (34,35) and *Drosophila* (36). These studies showed that synonymous SNPs altering mRNA folding may result in decreased mRNA stability and may also change expression efficiency. In a recent *in vitro* study, we introduced mutations

Table 4. Summary of MFE changes in mRNA secondary structures of essential and nonessential genes

Gene type	Number of genes	Lower quartile %	Median %	Upper quartile %
Essential	48	9.4	32.6	71.1
Nonessential	348	25.5	50.2	75.0

Lower quartile, median and upper quartile values are presented for the distributions of percentages of computer-generated mutants with randomly introduced mutations that change the MFE less than the naturally occurring mutations.

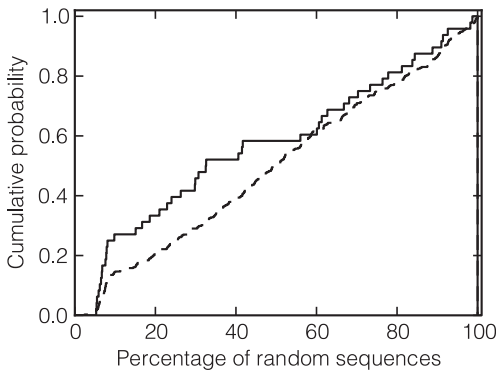


Figure 3. Cumulative distribution functions of the percentages of randomly introduced *in silico* mutations that change the MFE values less than the mutations occurring *in vitro* for essential genes (solid line) and nonessential genes (dashed line). Each curve gives the probability that the MFE change in a particular gene due to an actual mutation will be higher than MFE changes observed in a given percentage of genes with randomly introduced mutations.

into an influenza gene, particularly into a region of the gene encoding for a functionally important protein domain (13). As a result of the perturbations in the RNA structure caused by these mutations, gene expression was significantly reduced. Mutations altering RNA structures thus had a functional knockout effect. We also demonstrated that mutations disruptive to RNA structure may impair transcription without facilitating mRNA degradation. Thus, a new mechanism of viral evolution was proposed (13). We hypothesized that mutations disruptive to RNA structures would likely be eliminated to preserve the gene regions encoding for functionally important sites of viral proteins. Following this line of thought, the goal of the present study was to examine whether preservation of mRNA structures is implicated in the evolution of bacteria.

An important evolutionary characteristic of bacterial genes is their essentiality for organism survival, which can be experimentally assessed based on absence of growth on knockout. We hypothesized that if some of the mutations causing perturbations in mRNA structures also result in reduction in expression levels of bacterial genes, these mutations are more likely to be eliminated by purifying selection if they take place in essential rather than nonessential genes. Indeed, we found that mRNA secondary structures of essential genes are more conserved than those of nonessential genes in bacteria. Previous work revealed that essential bacterial genes are more evolutionarily conserved than nonessential ones (29,37). It was shown that in the *E. coli* genome paired DNA bases have lower propensities to mutate than unpaired bases (38,39). Based on the comparison of the *E. coli* and *Salmonella typhi* genomes, it was concluded that homologous RNAs of polycistronic genes in both organisms have significantly higher folding potential than randomized sequences, which is a sign that natural selection is acting to preserve RNA secondary structure in the coding regions of polycistronic genes (7). However, to the best of our knowledge, preservation of intact mRNA structures of individual genes has not yet been assessed as a potential constraint on the evolution of bacterial genes.

As the best available proxy for *E. coli* B essential genes we used experimentally determined the essentiality status of genes in the closely related *E. coli* K-12 genome (25). Such homology mapping may not always be accurate, even between similar organisms, owing to possible differences in gene regulation, posttranslational modifications and other cellular processes. An additional factor potentially masking the true magnitude of the effect is that we used changes in MFE as an indirect indication that the secondary structure of the mRNA has changed. However, changes in RNA sequence and the resulting perturbations of its structure may in fact take place without causing MFE changes (Table 3). Owing to these obvious limitations, we believe that our results represent a conservative estimate of the role played by mRNA structure in constraining mutations.

Our results point to the preservation of coding mRNA structures as a previously unappreciated factor influencing bacterial evolution. Until now, selective pressure in coding regions was thought to primarily act against mutations that either impair protein function and stability or affect robustness against mistranslations (40). In particular, selection against mistranslation-induced protein misfolding is currently considered to be the major factor determining the strong dependence of protein evolutionary rate on the level of expression (41). The bulk of this research has thus been devoted to the 'protein half of the equation'—translation, folding and function. In the past few years, attention is being increasingly focused on the noncoding selective pressure in coding regions, which is manifested by the presence of synonymous constraint (42–44). Such noncoding selective pressure may be caused, on one hand, by the presence of various functional elements, such as microRNA binding sites, transcription factor binding sites and splicing enhancers in eukaryotic mRNAs, and, on the other hand, by the formation of RNA structural elements playing a role in mRNA localization, degradation and interactions with other molecules. This article presents the first statistical evidence linking mRNA folding to bacterial evolution. Our principal finding is that purifying selection tends to eliminate those mutations in essential genes that lead to greater changes of MFE values and, therefore, may be more disruptive for the corresponding mRNA secondary structures. This effect is implying that synonymous mutations disrupting mRNA secondary structures may directly affect the fitness of the organism.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3 and Supplementary Figure 1.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Richard Lenski for productive and stimulating discussions, Dr Olivier Tenaillon for illuminating comments on the article and Prof. Andrei Osterman for helpful comments on the essentiality data. The authors gratefully acknowledge the support of the TUM Graduate School's Thematic Graduate Center Regulation and Evolution of Cellular Systems (RECESS) at the Technische Universität München.

FUNDING

DFG International Research Training Group 'Regulation and Evolution of Cellular Systems' [GRK 1563]; Russian Foundation for Basic Research [RFBR 09-04-92742]. Funding for open access charge: DFG International Research Training Group 'Regulation and Evolution of Cellular Systems' [GRK 1563].

Conflict of interest statement. None declared.

REFERENCES

1. Wan,Y., Kertesz,M., Spitale,R.C., Segal,E. and Chang,H.Y. (2011) Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.*, **12**, 641–655.
2. Meyer,I.M. and Miklós,I. (2005) Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. *Nucleic Acids Res.*, **33**, 6338–6348.
3. Findeiß,S., Engelhardt,J., Prohaska,S.J. and Stadler,P.F. (2011) Protein-coding structured RNAs: a computational survey of conserved RNA secondary structures overlapping coding regions in drosophilids. *Biochimie*, **93**, 2019–2023.
4. Gray,N.K. and Hentze,M.W. (1994) Regulation of protein synthesis by mRNA structure. *Mol. Biol. Rep.*, **19**, 195–200.
5. Kozak,M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13–37.
6. Gu,W., Zhou,T. and Wilke,C.O. (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Comput. Biol.*, **6**, e1000664.
7. Katz,L. and Burge,C.B. (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.*, **13**, 2042–2051.
8. Gonzalez,I., Buonomo,S.B., Nasmyth,K. and von Ahsen,U. (1999) ASH1 mRNA localization in yeast involves multiple secondary structural elements and Ash1 protein translation. *Curr. Biol.*, **9**, 337–340.
9. Chartrand,P., Meng,X.H., Singer,R.H. and Long,R.M. (1999) Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle *in vivo*. *Curr. Biol.*, **9**, 333–336.
10. Olivier,C., Poirier,G., Gendron,P., Boisgontier,A., Major,F. and Chartrand,P. (2005) Identification of a conserved RNA motif essential for She2p recognition and mRNA localization to the yeast bud. *Mol. Cell Biol.*, **25**, 4752–4766.
11. Ankö,M.L. and Neugebauer,K.M. (2012) RNA-protein interactions *in vivo*: global gets specific. *Trends Biochem. Sci.*, **37**, 255–262.
12. Seffens,W. and Digby,D. (1999) mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Res.*, **27**, 1578–1584.
13. Ilyinskii,P.O., Schmidt,T., Lukashev,D., Meriin,A.B., Thoidis,G., Frishman,D. and Shneider,A.M. (2009) Importance of mRNA secondary structural elements for the expression of influenza virus genes. *Omics*, **13**, 421–430.
14. Moss,W.N., Priore,S.F. and Turner,D.H. (2011) Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA*, **17**, 991–1011.
15. Lenski,R.E., Rose,M.R., Simpson,S.C. and Tadler,S.C. (1991) long-term experimental evolution in *Escherichia coli* .I. Adaptation and divergence during 2,000 generations. *Am. Nat.*, **138**, 1315–1341.
16. Cooper,T.F., Rozen,D.E. and Lenski,R.E. (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli. Proc. Natl Acad. Sci. USA*, **100**, 1072–1077.
17. Blount,Z.D., Borland,C.Z. and Lenski,R.E. (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli. Proc. Natl Acad. Sci. USA*, **105**, 7899–7906.
18. Barrick,J.E., Yu,D.S., Yoon,S.H., Jeong,H., Oh,T.K., Schneider,D., Lenski,R.E. and Kim,J.F. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, **461**, 1243–1247.
19. Schultes,E.A. and Bartel,D.P. (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science*, **289**, 448–452.
20. Höbartner,C. and Micura,R. (2003) Bistable secondary structures of small RNAs and their structural probing by comparative imino proton NMR spectroscopy. *J. Mol. Biol.*, **325**, 421–431.
21. Tinoco,I. Jr and Bustamante,C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271–281.
22. Lyngsø,R.B. (2004) Complexity of pseudoknot prediction in simple models. *Lect. Notes Comput. Sc.*, **3142**, 919–931.
23. Jeong,H., Barbe,V., Lee,C.H., Vallenet,D., Yu,D.S., Choi,S.H., Couloux,A., Lee,S.W., Yoon,S.H., Cattolico,L. *et al.* (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.*, **394**, 644–652.
24. Studier,F.W., Daegelen,P., Lenski,R.E., Maslov,S. and Kim,J.F. (2009) Understanding the differences between genome sequences of *Escherichia coli* B Strains REL606 and BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *J. Mol. Biol.*, **394**, 653–680.
25. Gerdes,S.Y., Scholle,M.D., Campbell,J.W., Balázsi,G., Ravasz,E., Daugherty,M.D., Somera,A.L., Kyrpides,N.C., Anderson,I., Gelfand,M.S. *et al.* (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.*, **185**, 5673–5684.
26. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
27. Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,L.S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–188.
28. Wielgoss,S., Barrick,J.E., Tenaillon,O., Wiser,M.J., Dittmar,W.J., Cruveiller,S., Chane-Woon-Ming,B., Médigue,C., Lenski,R.E. and Schneider,D. (2013) Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl Acad. Sci. USA*, **110**, 222–227.
29. Jordan,I.K., Rogozin,I.B., Wolf,Y.I. and Koonin,E.V. (2002) Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res.*, **12**, 962–968.
30. Kudla,G., Murray,A.W., Tollervey,D. and Plotkin,J.B. (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, **324**, 255–258.
31. Hall,M.N., Gabay,J., Débarbouillé,M. and Schwartz,M. (1982) A role for mRNA secondary structure in the control of translation initiation. *Nature*, **295**, 616–618.
32. Qing,G., Xia,B. and Inouye,M. (2003) Enhancement of translation initiation by A/T-rich sequences downstream of the initiation codon in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.*, **6**, 133–144.

33. Griswold,K.E., Mahmood,N.A., Iverson,B.L. and Georgiou,G. (2003) Effects of codon usage versus putative 5′-mRNA structure on the expression of *Fusarium solani* cutinase in the *Escherichia coli* cytoplasm. *Protein Expr. Purif.*, **27**, 134–142.
34. Nackley,A.G., Shabalina,S.A., Tchivileva,I.E., Satterfield,K., Korchynskyi,O., Makarov,S.S., Maixner,W. and Diatchenko,L. (2006) Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science*, **314**, 1930–1933.
35. Duan,J., Wainwright,M.S., Comeron,J.M., Saitou,N., Sanders,A.R., Gelernter,J. and Gejman,P.V. (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.*, **12**, 205–216.
36. Carlini,D.B., Chen,Y. and Stephan,W. (2001) The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr. *Genetics*, **159**, 623–633.
37. Wilson,A.C., Carlson,S.S. and White,T.J. (1977) Biochemical evolution. *Annu. Rev. Biochem.*, **46**, 573–639.
38. Wright,B.E., Reschke,D.K., Schmidt,K.H., Reimers,J.M. and Knight,W. (2003) Predicting mutation frequencies in stem-loop structures of derepressed genes: implications for evolution. *Mol. Microbiol.*, **48**, 429–441.
39. Hoede,C., Denamur,E. and Tenaillon,O. (2006) Selection acts on DNA secondary structures to decrease transcriptional mutagenesis. *PLoS Genet.*, **2**, e176.
40. Pál,C., Papp,B. and Lercher,M.J. (2006) An integrated view of protein evolution. *Nat. Rev. Genet.*, **7**, 337–348.
41. Drummond,D.A. and Wilke,C.O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, **134**, 341–352.
42. Chen,H. and Blanchette,M. (2007) Detecting non-coding selective pressure in coding regions. *BMC Evol. Biol.*, **7(Suppl.1)**, S9.
43. Lin,M.F., Kheradpour,P., Washietl,S., Parker,B.J., Pedersen,J.S. and Kellis,M. (2011) Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. *Genome Res.*, **21**, 1916–1928.
44. Chamary,J.V. and Hurst,L.D. (2005) Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol.*, **6**, R75.