the Proteobacteria, with 28 predicted ECF σ factors.

Supplemental web pages
Access to additional web pages containing supplemental material related to this article can be obtained via the following URL: http://www.cbs.dtu.dk/services/GenomeAtlas/suppl/GenUp018/

Acknowledgements
This work was supported by a grant from the Danish Center for Scientific Computing.

Kristoffer Kil, Tim T. Binnewies, Thomas Sichertz-Pontén, Hanni Willenbrock, Peter F. Hallin, Trudy M. Wassenaar† and David W. Ussery

Center for Biological Sequence Analysis, BioCentrum-DTU, Building 208, The Technical University of Denmark, DK-2800 Kgs Lyngby, Denmark

†Present address: Molecular Microbiology and Genomics Consultants, Zotzenheim, Germany.

Correspondence: David W. Ussery (dave@cbs.dtu.dk)


DOI 10.1099/mic.0.28339-0

Intragenic position of UUA codons in streptomyces

Streptomyces have huge linear genomes (> 8 Mbp), extreme GC content (around 70 mol%), and their life cycles involve vegetative growth, a phase with formation of aerial hyphae and a sporulation stage. In addition, they produce a range of secondary metabolites, including many of the antibiotics used in clinics today, and for this reason they have immense practical importance. Production of antibiotics is typically closely linked to differentiation. One of the key genes involved in the differentiation switch is theldA, which encodes the tRNA recognizing the very rare UUA (leucine) codons (Lawlor et al., 1987). It has been shown that the production of this tRNA is subject to temporal regulation as it becomes abundant in old cultures (Leskiw et al., 1993). Conversely, genes containing UUA codons are often linked to antibiotic production or other aspects of differentiation (see reviews by Leskiw et al., 1991a; Chater, 1993). As a consequence, heterologous expression of proteins, which often takes place in liquid cultures under vegetative growth, can be problematic if the foreign gene contains UUA codons (Leskiw et al., 1991b; Ueda et al., 1993). Codon usage in the streptomyces is therefore an interesting phenomenon that deserves full attention.

Wright & Bibb (1992) investigated a limited number of streptomyces genes and concluded that codon usage largely reflected mutational bias. A full genomic analysis based on codon usage in Streptomyces coelicolor and Streptomyces avermitilis (both are fully sequenced and their genomes are available through GenBank) was recently published in Microbiology by Wu et al. (2005). The paper used a measure of synonymous codon usage bias called the codon adaptation index (Sharp & Li, 1987) to predict highly expressed genes. Technically, this involves identification of genes using codons that are particularly abundant in highly expressed genes such as ribosomal genes. Thus, UUA codons and other single codons had no particular focus in that study. In fact, UUA has to my knowledge not received particular focus in bioinformatic studies of the streptomyces genomes so far. The aim of this Comment therefore is to apply a genomic view on UUA, thereby supplementing the work by Wu et al. and adding to the general knowledge about codon usage in these important bacteria.

I hypothesize that since transcripts for genes containing UUA codons are produced throughout the life cycle, initiation of translation on such mRNAs is futile, and on this account the fitness among different mutants in a population could depend on the loss of energy associated with this futile
synonymous codons ending in A. UUA early in genes can therefore not be ascribed to a general preponderance of all synonymous codons ending in A. The leftward displacement is highly significant (#P < 0.0001; Wilcoxon signed rank test). ○, UUA; ○, UUG; □, CUA; □, CUC; △, CUG; △, CUU. (b) Cumulative histogram for all synonymous codons ending in A (○), C (□), G ( ■) or U (●). Note that there is a considerable difference between the curve for UUA and NNA. The preponderance of UUA early in genes can therefore not be ascribed to a general preponderance of synonymous codons ending in A.

translation activity. Selection should therefore promote UUA codons at the beginning of genes (proximal to the start codon), rather than at the end (proximal to the stop codon), simply because the energy waste is proportional to the number of codons that will be translated until the problematic UUA codon is encountered. Thus, the purpose of this paper was to test if UUA codons in Streptomyces actually do occur more often at the beginning than at the end of genes. This is not straightforward to test since genes lengths differ (codon number X may be closer to the start codon in gene A but not in gene B, if gene B is shorter than gene A). The intragenic position of a codon can be normalized to a relative position (RP) between 0 and 1 using \( \text{RP} = (i-1)/(L-1) \), where \( L \) is the length (measured in codons) and \( i \) is the codon number \( i \in \{1, 2, 3, \ldots, L\} \) of a sequence in which the start and stop codons are removed. Thus, when \( \text{RP} = 0 \) we are dealing with the codon right after the start codon, and when \( \text{RP} = 1 \) we are looking at the codon immediately upstream of the stop codon. Notably, a value of 0·5 signifies exactly the middle, so we could in principle use 0·5 as our null value; the hypothesis states that UUA should occur nearer the start codon than to the stop codon, so \( H_0: \text{median } \text{RP} \geq 0.5 \) could be tested. This is, however, potentially confounded by other variations in codon usage and/or amino acid usage. The aforementioned null hypothesis would only apply if the median RP of the other leucine codons were 0·5, and if there were no overall variations in composition along the length of genes. Therefore, the median RP for UUA has to be viewed in light of the RP for the other leucine codons. In Fig. 1(a) I have shown a cumulative histogram of all six leucine codons collected in bin widths of 0·01 units of intragenic relative position for \( S.\ coeli\)color. In this plot we can clearly see how unusual the usage of UUA is compared to the usage of the other leucine codons. Generally, the UUA histogram is displaced leftwards compared to the other leucine codon histograms; about 75% of the UUA codons are found closer to the start codon than to the stop codon (relative position < 0·5) while the other synonyms are much closer to a straight line with unit slope and zero intercept. In Fig. 1(b) I have plotted a similar plot for all A,-C,-G- or T-ending synonymous codons. Although there is a slight preponderance of A-ending codons (as well as T-ending codons) at the beginning of genes, this phenomenon alone cannot account for the extreme values for UUA observed in Fig. 1(a). For \( S.\ avermitilis \) the same phenomenon is observed (data not illustrated). Figs 1(a) and (b) thus show that UUA is predominant at the beginning of genes and even more so than the other leucine codons. All in all, the observations in this study are therefore fully in line with the evolutionary considerations laid out above.

However, it needs to the emphasized that the data do not prove the model as such. In perspective, it should be mentioned that Chen & Inouye (1990, 1994) studied the usage of rare codons in \( E\). col\i and found that rare arginine codons were predominantly found at the beginning of genes. Expression experiments showed that the yield was inversely related to the number of these rare arginine codons and that the limitations imposed by rare codons could be overcome by supplementation with the relevant tRNA (argU). There are no reports on temporal regulation of tRNA availability in \( E.\ col\i \), so we cannot readily conclude that the usage of UUA codons in streptomycetes is exactly the same phenomenon as the usage of rare arginine codons in \( E.\ col\i \). Nevertheless, it is well known that codon usage in \( E.\ col\i \) generally reflects ‘translational selection’, in that not all tRNAs are expressed at equal levels, and codon usage in highly expressed genes reflects this (Ikemura, 1981). In view of this, I would hypothesize that codon usage in streptomycetes also reflects translational selection, but in these interesting species it has been taken to the very extreme.

Plots of the kind shown in Figs 1(a) and (b) could be valuable for the characterization of codon usage in a wider range of organisms, so hopefully this Comment will inspire further genomic studies of rare codons as a relevant contrast of optimal or abundant codons.

Anders Fuglsang

Danish University of Pharmaceutical Sciences, 2 Universitetsparken, DK-2100 Copenhagen O, Denmark, and Norwegian Medicines Agency, Sven Oftedals Vei 8, N-0950, Oslo, Norway

Correspondence: Anders Fuglsang (anfu@dfuni.dk)


DOI 10.1099/mic.0.28352-0