Generation of overlapping open reading frames

Genome sequencing programs have brought us to the position where we have identified many open reading frames (ORFs), beginning with start codons and ending with stop codons; but how do we decide whether an ORF is coding? The classification of ORFs as coding sequences is even more problematic for overlapping frames.

As discussed recently in *Trends in Genetics*[1], mutation pressure introducing guanine and cytosine instead of adenine and thymine into silent third positions of codons can generate ORFs inside existing (reference) reading frames. This type of 'baby ORF' overlaps in such a way that the first position of codons in the baby ORF overlaps the third position of the codons in the reference frame. Thus, baby ORFs can be generated in one phase on the same strand (phase 3; Fig. 1) and one phase on the antisense strand of DNA (phase 2'; Fig. 1). However, some data suggest that many overlapping frames do not have this phase relationship.

For example, if we look at ORFs longer than 100 codons in the yeast chromosome II (Ref. 2) phase 1 and search for ORFs of at least the same length on the antisense strand, the following results are obtained: 3 ORFs in phase 2'; 18 ORFs in phase 3'; and no ORFs in phase 1'. Baby ORFs can be generated by mutation pressure only in phase 2'. Therefore, mutation pressure in the third position of the reference ORF is not the best explanation for the existence of overlapping ORFs.

Reference phase	(1) 1 2 3 1 2 3 1 2 3
Sense phases	(2) – 1 2 3 1 2 3 1 2
	(3) – – 1 2 3 1 2 3 1
Antisense phases	(1') 3 2 1 3 2 1 3 2 1
	(2') – 3 2 1 3 2 1 3 2
	(3') – – 3 2 1 3 2 1 3

FIGURE 1. The phase relationships between the reference phase (1) and the other sense and antisense phases (as discussed in the text).

An alternative explanation for the overlapping reading frames in phase 3' is based on the observation that amber and ochre stop codons have two-base palindromes. Thus, amber and ochre codons on one strand will tend to generate amber and ochre codons on the antisense strand in the related phase. Conversely, the elimination of these stop codons in one phase by selection for a reading frame lowers the probability of occurrence of these codons in the related phase on the opposite strand.

To test this hypothesis, we used a computer to construct an artificial long DNA molecule by 'splicing' all the ORFs of yeast chromosome II, phase 1, and next deleting all the stop codons except the last one in this phase. In this way, we obtained a 'chromosome' with one ORF of 86198 codons in phase 1. Deletion of the 2061 amber and ochre stops in phase 1 of the spliced molecule simultaneously resulted in: the elimination of 883 amber and ochre stops in phase 3', the generation of 650 stops in phase 2 and 549 stops in phase 2' (almost exactly as expected), and the sequestering of almost half of phase 3' into ORFs longer than 70 codons.

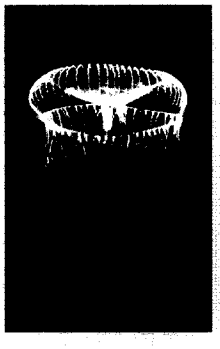
This relationship between phases is observed for other sequences and even stochastic DNA sequences. Thus, this relationship appears to be an intrinsic feature of the genetic code. It is interesting that the overlapping ORFs conserve degeneracy of the third positions of codons and that overlaps that would reduce degeneracy seem to be forbidden. If the generation of long ORFs was a critical step in the evolution of coding sequences, then one could expect that the generation of ORFs inside other ORFs was also exploited during evolution. After duplication of the overlapping ORFs, both sequences could exist and function independently. Therefore, it should be possible to find the examples of sequence homology between one gene and the antisense strand of a second gene. This finding would confirm the hypothesis that the genetic code possesses properties enhancing the generation of long ORFs.

STANISLAW CEBRAT
Institute of Microbiology, Wroclaw University, ul. Przybyszewskiego 63/77, 51-148 Wroclaw, Poland.

MIROSLAW R. DUDEK
mdudek@mirek.ift.uni.wroc.pl
Institute of Theoretical Physics, Wroclaw University, pl. Maxa Borna 9, 50-205 Wroclaw, Poland.

References

1 Boldogköi, Z., Murvai, J. and Fodor, I. (1995) *Trends Genet.* 11, 125–126
2 Feldmann, H. *et al.* (1994) *EMBO J.* 13, 5795–5809



Credit where credit is due

It has come to our attention that the photograph of *Aequoria victoria* that was used on the front cover and for the poster in the August issue of *TIG* was produced by Claudia Mills at the Friday Harbor Laboratories, Washington. The image was used without the knowledge or permission of Dr Mills. We apologize to Dr Mills for these omissions.