A code within a code: how codons influence mRNA stability

Javier Martinez^{1,2} & Bojan Zagrovic³

The genetic code was deciphered more than 50 years ago, but we are only now becoming aware of a second, hidden code. It is the concept of “codon optimality” that enters the scene of developmental and homeostatic gene expression, linking translation rates, mRNA stability, and tRNA abundance. Both at the biological and methodological levels, work by Giraldez and colleagues in this issue of *The EMBO Journal* paves the way for further analyses of such key regulatory mechanisms.

See also: **AA Bazzini *et al*** (October 2016)

Maternal-to-zygote transition in early embryonic development requires clearance of thousands of maternal mRNAs, illustrating the ability of a cell to carefully regulate gene expression by affecting the stability of its mRNAs. The general mechanisms of how this is achieved, however, remain largely mysterious. Pioneering work from Antonio Giraldez and Alexander Schier characterized the role of microRNAs (miRs), in particular miR-430, in degrading maternal mRNAs during the zygotic program in zebrafish (Giraldez *et al*, 2006). Other groups reported similar observations in *Drosophila* (Bushati *et al*, 2008) and *Xenopus* (Lund *et al*, 2009). Notwithstanding their importance, miRNAs are only responsible for the degradation of a subset of mRNAs.

In a large-scale, comparative analysis, Giraldez and coworkers now show that maternal mRNA clearance in early vertebrate embryogenesis as well as gene expression during homeostasis is directly modulated by

codon composition in a translation-dependent manner and with higher levels of tRNAs decoding optimal codons (Bazzini *et al*, 2016). Simply put, certain well-defined “optimal” codons reproducibly increase—across organisms, tissues, and developmental stages—the mRNA stability and translation efficiency, while other non-optimal codons diminish it (Fig 1A). Although the exact mechanism behind this phenomenon remains to be discovered, and allowing exceptions to this rule (Presnyak *et al*, 2015), it is clear that its overall impact is very strong and is comparable in magnitude, and most probably additive, to that exerted by microRNAs. The work by Bazzini *et al* (2016) complements and extends an exciting line of recent research showing similar effects in bacteria (Boel *et al*, 2016), yeast (Presnyak *et al*, 2015), and zebrafish (Mishima & Tomari, 2016), suggesting that the involved mechanisms may be a fundamental property of living matter, regardless of the level of organization. Furthermore, it gives support to previous ideas of the existence of multiple codes within biological sequences (Trifonov, 2011).

A striking pattern in the data of Bazzini *et al* (2016) is that optimal codons tend to be heavily enriched in purines (A+G), outnumbering the pyrimidines (U+C) by fourfold at the first two codon positions in zebrafish. An opposite situation, albeit somewhat less pronounced, is seen for the non-optimal codons, with pyrimidines taking the lead there (Fig 1A). It is possible that the underlying mechanisms, be it by involving just the ribosome or other specific cellular factors, may depend on this or some related general properties of optimal and

non-optimal codons. An important finding in this context is that the codon optimality becomes apparent only in the presence of active translation, thus linking the fate of mRNA transcripts not only with the ribosome, but also potentially with the cognate proteins encoded by them. Related to this, it has recently been shown that nucleobase-density patterns in mRNA coding sequences are closely related to the nucleobase-affinity patterns of the proteins they encode (Polyansky *et al*, 2013b). For example, pyrimidine-density profiles in mRNAs match their cognate protein's pyrimidine-affinity profiles with a median absolute value of the Pearson correlation coefficient of $|R| = 0.74$. This has been used to propose that cognate mRNAs and proteins, when structurally disordered, may be physicochemically complementary to each other and bind in a co-aligned fashion, generalizing the stereochemical hypothesis of the origin of the genetic code. A corollary of this proposal is that the presence of a given protein would lead to structural disordering of its cognate mRNA and vice versa. In light of the above complementarity hypothesis, it is actually possible that the direct interactions between mRNAs and their cognate proteins may actually be in the background of the observed effect of codon optimality on mRNA stability (Fig 1B). A simple, but by no means the only, mechanism could involve preferential binding of a protein to its cognate mRNA with a preponderance of non-optimal codons and subsequent unfolding and destabilization of the mRNA. Optimized CLIP (cross-linking and immunoprecipitation) protocols might be able to “catch” proteins enriched

1 IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Vienna, Austria

2 Max F. Perutz Laboratories (MFPL), Department of Medical Biochemistry, Medical University of Vienna, Vienna, Austria

3 Max F. Perutz Laboratories (MFPL), Department of Structural and Computational Biology, University of Vienna, Vienna, Austria. E-mail: bojan.zagrovic@univie.ac.at

DOI 10.15252/embj.201695283 | Published online 25 August 2016

Figure 1. Codon optimality and its correlates.

(A) A second genetic code, based on codon optimality, determines mRNA fate in maternal-to-zygotic transition in different organisms, as well as gene expression during homeostasis across human tissues. Codon optimality correlates with gene expression levels, tRNA levels, mRNA half-life, translation efficiency, and poly(A) tail length. (B) Hypothesis: A potential regulatory event relies on the translating polypeptide recognizing its own mRNA according to the preponderance of non-optimal codons, and leading to the destabilization of the mRNA, probably associated with a slower translation rate and lower overall translation efficiency. In yellow, the growing polypeptide chain; in dark blue, translation efficiency (full line, high; dashed line, low). AG rich are optimal codons; CU rich are non-optimal codons.

in non-optimal codons when binding their coding mRNAs, ultimately depending on the strength of those interactions.

An exciting aspect of the work by Bazzini *et al* (2016) and others in this context (Presnyak *et al*, 2015; Boel *et al*, 2016; Mishima & Tomari, 2016) is a demonstration of a new form of coupling between mRNAs and their cognate proteins to be added to those studied before (Watts *et al*, 2009; Pechmann & Frydman, 2013; Polyansky *et al*, 2013a). Namely, codon optimality and its effect on mRNA stability also directly influence the amino acid repertoire encoded

by these mRNAs with concordant levels of mature tRNAs. Conversely, the cellular requirements for certain types of protein sequences also, through the codon optimality code, directly affect the stability of mRNA messages that code for them.

The worlds of mRNAs and proteins, coupled through the genetic code, affect each other in numerous, subtle ways, making the case for them actually being two parts of the same whole. In other words, the essence of the biological function may not reside in mRNAs or proteins alone, but rather in the complex, dynamic relationship between the two existing

on multiple levels. It will be exciting to probe this fundamental question further.

References

- Bazzini AA, del Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, Yao J, Khokha MK, Giraldez AJ (2016) Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J* 35: 2087–2103
- Boel G, Letso R, Neely H, Price WN, Wong KH, Su M, Luff JD, Valecha M, Everett JK, Acton TB, Xiao R, Montelione GT, Aalberts DP, Hunt JF (2016) Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature* 529: 358–363
- Bushati N, Stark A, Brennecke J, Cohen SM (2008) Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in *Drosophila*. *Curr Biol* 18: 501–506
- Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312: 75–79
- Lund E, Liu M, Hartley RS, Sheets MD, Dahlberg JE (2009) Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos. *RNA* 15: 2351–2363
- Mishima Y, Tomari Y (2016) Codon usage and 3' UTR length determine maternal mRNA stability in Zebrafish. *Mol Cell* 61: 874–885
- Pechmann S, Frydman J (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat Struct Mol Biol* 20: 237–243
- Polyansky AA, Hlevnjak M, Zagrovic B (2013a) Analogue encoding of physicochemical properties of proteins in their cognate messenger RNAs. *Nat Commun* 4: 2784
- Polyansky AA, Hlevnjak M, Zagrovic B (2013b) Proteome-wide analysis reveals clues of complementary interactions between mRNAs and their cognate proteins as the physicochemical foundation of the genetic code. *RNA Biol* 10: 1248–1254
- Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, Collier J (2015) Codon optimality is a major determinant of mRNA stability. *Cell* 160: 1111–1124
- Trifonov EN (2011) Thirty years of multiple sequence codes. *Genomics Proteomics Bioinformatics* 9: 1–6
- Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460: 711–716