BIOCHEMISTRY

Hidden code in the protein code

Apparently redundant codons may not be redundant after all.

The instructions for building proteins are given in triplets of the four nucleotides. This yields 64 codons, each dictating either the addition of a particular amino acid or the end of protein synthesis. With only 20 amino acids to choose from, several different codons code for the same protein residue, but recent research shows that so-called synonymous codons may provide subtly different information to a cell's gene-reading machinery.

Researchers led by Anna Kashina at the University of Pennsylvania found that different codons for the same amino acid may affect how quickly mRNA transcripts are translated, and that this pace can influence post-translational modifications (Zhang *et al.*, 2010). Despite being highly homologous, the mammalian cytoskeletal proteins beta- and gamma-actin contain notably different post-translational modifications: though both proteins are actually post-translationally arginylated, only arginylated beta-actin persists in the cell. This difference is essential for each protein's function.

To investigate whether synonymous codons might have a role in how arginylated forms persist, Kashina and colleagues swapped the synonymous codons between the genes for beta- and gamma-actin and found that the patterns of post-translational modification switched as well. Next, they examined translation rates for the wild-type forms of each protein and found that gamma-actin accumulated more slowly. Computational analysis suggested that differences between the folded mRNA structures might cause differences in translation speed. When the researchers added an antibiotic that slowed down translation rates, accumulation of arginylated actin slowed dramatically. Subsequent work indicated that N-arginylated proteins may, if translated slowly, be subjected to ubiquitination, a post-translational modification that targets proteins for destruction.

Thus, these apparently synonymous codons can help explain why some arginylated proteins but not others accumulate in cells. "One of the bigger implications of our work is that post-translational modifications are actually encoded in the mRNA," says Kashina. "Coding sequence can define a protein's translation rate, metabolic fate and post-translational regulation."

Researchers led by Eran Segal at the Weizmann Institute of Science took a computational approach to find evidence for additional codes in the protein code (Itzkovitz *et al.*, 2010). They defined all possible 6-base-pair or 7-base-pair sequences in real genomes. They then tallied up the frequencies of occurrence of the defined short sequences in randomly generated coding sequences and in actual genomes coding for the same proteins. They looked for enrichment or depletions across phyla and found pronounced differences between real and randomized coding regions, particularly for bacteria. The technique identified previously known patterns, such as enrichment for microRNA target sites in eukaryotic genomes as well as for codons contributing to the three-dimensional shape of mRNA molecules, and also suggested that whereas some 'overlapping codes' have been evolutionarily conserved, others have been enriched in different phyla.

Segal and colleagues have made their randomization software available and believe that it could be useful to other scientists who also want to search for additional information: perhaps, for example, identifying patterns in specific families of genes and assessing whether these patterns may represent binding sites of yet-uncharacterized transcription factors or RNA-binding proteins.

A more practical application, Segal says, is in synthetic biology: "Designing the 'optimal' synthetic gene could become a fascinating combinatorial problem that could integrate the constraints that evolution had converged upon and that we believe our study reveals." Kashina also believes that scientists will eventually be able to pick the right codons to engineer desired post-translational modifications. "The idea that nucleotides encode more than the protein primary structure is one of those very simple and very fundamental mechanisms that have been out there on the surface all the time," says Kashina. The code within a code surely has more secrets to reveal.

Monya Baker

RESEARCH PAPERS

Itzkovitz, S. *et al.* Overlapping codes within protein-coding sequences. *Genome Res.* advance online publication (14 September 2010).

Zhang, F. *et al.* Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. *Science* **329**, 1534–1537 (2010).