

## NEWS AND COMMENTARY

Whence the genetic code?

# Thawing the 'Frozen Accident'

CW Carter Jr

*Heredity* (2008) 100, 339–340; doi:10.1038/hdy.2008.7;  
published online 13 February 2008

Base-pairing and the double-helical structure of DNA may be the most important discovery humanity has, or ever will, make about itself. Its extraordinary explanatory power derives from the way it stores information for making proteins, which, in turn, depends on the genetic code (Woese, 1965; Crick, 1966). The near-universal translation table assigning codons to amino acids is highly non-random; nearly all possible alternatives are more vulnerable to the effects of both mutational and readout errors (Freeland and Hurst, 1998). The importance of horizontal transfer and primordial sharing of early coding experiments during the optimization of the code have recently been described by Woese (Vestigian *et al.*, 2006).

Missing from this overview is a grammar for attaching meaning to codons and a molecular rationale for the specific choices made by evolution as the codon table froze into its present form. In this issue of *Heredity*, Rodin and Rodin (2008) seek a unified molecular interpretation of the genetic code based largely on two recent papers that propose to fill these gaps (Rodin and Rodin, 2006a, b; Delarue, 2007).

The emergence of genetic code was inseparable from the ancestry of the RNA adaptors and protein catalysts that implement it now. In the absence of catalysts, amino-acid activation is perhaps  $10^7$  times slower than peptide bond formation. The catalytic heavy lifting required to accelerate this rate is divided equally between two unrelated enzyme superfamilies, the class I and class II aminoacyl-transfer RNA synthetases (aaRSs), each activating 10 of the 20 canonical amino acids (Eriani *et al.*, 1990). The two aaRS classes ensure correct aminoacylation of their cognate tRNAs by interacting from opposite sides of the acceptor stem, either with the minor (class I) or major (class II) groove, and hence acylating either the 2' or 3' hydroxyl groups of A76. Both Delarue's and Rodin's papers reorganize the codon table to reflect these contrasting molecular recognition modes. They infer rather different, but perhaps not mutually exclusive rules.

Delarue (2007) argues that the partition of codons according to the aaRS

class distinction facilitated a hierarchical process by which additions to the code reduced codon ambiguity to produce the extant table with just five binary choices. Undifferentiated triplets, NNN, were nonsense codons. Codons were given meaning beginning with the second base (whether it was pyrimidine or purine, then the distinction between U vs C and A vs G) and ending with the third. The NYN triplets could interact with a synthetase, whereas the NRN could not and remained stop codons. Synthetase:NYN-containing tRNA interactions were ambiguous with respect to groove recognition. The identity of the middle base distinguished four codon families, of which NCN was recognized from the major and NUN from the minor groove, while NGN remained ambiguous with respect to groove recognition and NAN remained undefined.

At each step, the ambiguous codon family differentiated to give descendants with opposite groove recognition, while descent of the stop codon family generated a new ambiguous family and retained a stop codon, which therefore was always a feature of the code. These asymmetric division rules provide a unique differentiation order, rendering the exhaustive exploration of the initial assignment of codons plausible, and suggesting that the appearance of the code distilled meaning successively from redundancy by a deterministic elimination of the most frequent errors.

Molecular mechanisms remain obscure in Delarue's differentiation model. On the other hand, the Rodins identify and exploit a dual complementarity evident in synthetase:tRNA recognition to provide molecular detail. tRNA recognition depends on two distinct recognition 'codes,' the anticodon and an 'operational code' (Schimmel *et al.*, 1993) adjacent to the site of aminoacylation at the 3' CCA terminus. Remarkably, tRNAs with complementary anticodons also have statistically significant complementarity in their acceptor stem operational codes (Rodin and Rodin, 2006a, b).

The Rodins propose that this dual complementarity is retained from an earlier stage in the code's development, and thus is a palimpsest of molecular recognition modes in transitional

tRNAs from a more symmetric stage in code evolution, at which triplet reading frames had been established, but only the middle bases of the anticodons had been fixed, perhaps coinciding with the second step of Delarue's differentiation scheme. Furthermore, they propose that the two aaRS classes probably developed major or minor groove recognition at a time when both strands of genes contained coherent messages through sense/antisense coding. They conclude that new codons were recruited in pairs, because translation of both sense and antisense strands would require that meaning be attached to both codons and their anticodons. Evolutionary variation of a duplicated gene encoding primordial synthetases of opposite classes (and tRNA groove recognition) on opposite strands (Pham *et al.*, 2007) could have introduced new codon pairs, as suggested by the inheritance of groove recognition in Delarue's model.

The Rodins observe that 16 of the 32 codon:anticodon pairs include RY palindromes and hence present potential ambiguity, especially in the context of the anticodon loop, YYNNRRR (Figure 1). Contemporary aaRS invariably use opposite groove recognition for these 16 pairs, suggesting a latent, complementarity-based subcode that avoids confusing palindromic codon:anticodon pairs. This latent code works as follows: if two complementary codons contain YY vs RR at the second and adjacent bases, their aaRS approach the tRNA acceptor stem from the same groove, that is,  $NAR \cdot YUN = \text{minor groove}$ ;  $RGN \cdot NCY = \text{major groove}$ . If adjacent bases alternate—YR or RY—and hence could form a palindrome, then they are recognized from opposite grooves:  $YGN = \text{minor} \cdot NCR = \text{major}$  and  $NAY = \text{minor} \cdot RUN = \text{major}$ . The key link between the aaRS class distinction and error reduction is what the Rodins call the direction of 'spreading,' that is, the manner in which neighboring bases are utilized to enhance specific recognition. Spreading recognition out in opposite directions greatly decreases the risk of incorrect aminoacylation, especially by ribozymal aaRS, for which meeting different nucleotides would have meant going in opposite directions and hence interacting from opposite grooves.

These papers bring two new insights to the codon table. Delarue provides a deterministic mechanism for differentiating an entirely ambiguous triplet code with no information into the present day code specifying 22 amino acids. Key to this notion is that one



**Figure 1** Potential palindromes in the tRNA anticodon loop and the latent, complementarity-based subcode for aminoacyl-tRNA synthetase recognition of tRNA. Four possible configurations of two anticodon bases at the level of purine vs pyrimidine (bold face) are paired together with complementary anticodons and colored (blue  $\Rightarrow$  red) according to increasing vulnerability to misreading. Two successive purines or pyrimidines are unlikely to be confused with their complements. Alternating purine/pyrimidine dinucleotides in the anticodon are quite likely to be confused, as indicated by the yellow dashed lines (adapted from Rodin and Rodin, 2008; Figures 3 and 4 and Table 3).

branch of descent retains no coding information, and is therefore always available for punctuation, that is, the stop codons. Consistent with this differentiation model, the most recently discovered tRNA, which decodes the UAG codon as pyrrolysine in Archaea is the most ambiguous of all, interacting with both class I and class II lysyl tRNA synthetases (Delarue, 2007).

The latent, complementarity-based subcode identified by the Rodins provide a rational link between the aaRS class distinction and acylation error reduction for those sixteen codon:anticodon pairs that would have been the most often confused by ribozymic tRNA synthetase precursors. In

turn, this argues that tRNA acylation was originally carried out by ribozymic synthetases.

If, as the Rodins suggest, the code grew by the (symmetric) inclusion of codon:anticodon pairs, it would imply an early equivalence between the first and third codon bases, and hence would change some evolutionary implications of the wobble hypothesis (Crick, 1966), which attributes the least significance to the third base. Delarue's differentiation model may provide an alternative explanation.

Professor CW Carter Jr is at the Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599-7260, USA.

e-mail: carter@med.unc.edu

Crick FHC (1966). The origin of the genetic code. *J Mol Biol* **38**: 367–379.

Delarue M (2007). An asymmetric underlying rule in the assignment of codons: possible clue to a quick early evolution of the genetic code via successive binary choices. *RNA* **13**: 1–9.

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**: 203–206.

Freeland SJ, Hurst LD (1998). The genetic code is one in a million. *J Mol Evol* **47**: 238–248.

Pham Y, Li L, Kim A, Erdogan O, Weinreb V, Butterfoss GL *et al.* (2007). A minimal TrpRS catalytic domain supports sense/antisense ancestry of class I and II aminoacyl-tRNA synthetases. *Mol Cell* **25**: 851–862.

Rodin SN, Rodin A (2006a). Partitioning of aminoacyl-tRNA synthetases in two classes could have been encoded in a strand-symmetric RNA world. *DNA Cell Biol* **25**: 617–626.

Rodin SN, Rodin A (2006b). Origin of the genetic code: first aminoacyl-tRNA synthetases could replace isofunctional ribozymes when only the second base of codons was established. *DNA Cell Biol* **25**: 365–375.

Rodin SN, Rodin AS (2008). On the origin of the genetic code: signatures of its primordial complementarity in tRNAs and aminoacyl-tRNA synthetases. *Heredity* **100**: 341–355.

Schimmel P, Giegé R, Moras D, Yokoyama S (1993). An operational RNA code for amino acids and possible relationship to genetic code. *Proc Natl Acad Sci USA* **90**: 8763–8768.

Vestigian K, Woese CR, Goldenfeld N (2006). Collective evolution and the genetic code. *Proc Natl Acad Sci USA* **103**: 10696–10701.

Woese CR (1965). On the origin of the genetic code. *Proc Natl Acad Sci USA* **54**: 1546–1552.

#### Editor's suggested reading

Belshaw R, Sensasson D (2006). The rise and falls of introns. *Heredity* **96**: 208–213.

Landry CR *et al.* (2007). Genome clashes in hybrids: insights from gene expression. *Heredity* **99**: 483–493.

Schlötterer C (2007). *Drosophila* unleashed: the 12-genome update of the genetic power horse has been released. *Heredity* **100**: 337–338.