



Published in final edited form as:

Heredity (Edinb). 2008 April ; 100(4): 339–340. doi:10.1038/hdy.2008.7.

Thawing the “Frozen Accident”

C. W. Carter Jr.

Professor of Biochemistry and Biophysics at the University of North Carolina, Chapel Hill, 27599-7260

C. W. Carter: carter@med.unc.edu

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 (Crick, 1966, Woese, 1965). 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), Rodin and Rodin seek a unified molecular interpretation of the genetic code based largely on two recent papers that propose to fill these gaps (Delarue, 2007, Rodin and Rodin, 2006).

The emergence of the genetic code was inseparable from the ancestry of the RNA adaptors and protein catalysts that implement it now. Uncatalyzed 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 II aminoacyl-tRNA synthetases, aaRS, each activating ten of the twenty canonical amino acids (Eriani, et al., 1990). The two aaRS classes ensure correct aminoacylation of their cognate transfer RNAs (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 the Delarue and Rodin papers reorganize the codon table to reflect these contrasting molecular recognition modes. They infer rather different, but perhaps not mutually exclusive rules.

Delarue (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 (pyrimidine vs purine, then U vs. C, and A vs. G), beginning with the second base and ending with the third. Differentiation between NRN and NYN triplets enabled the latter to interact with a synthetase while the former could not, and remained stop codons. Synthetase:NYN-containing tRNA interactions were ambiguous with respect to groove recognition. Defining the middle base generated 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 Delaure's differentiation model. 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, 2006)

The Rodins propose that this dual complementarity 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 differentiated, perhaps coinciding with the second differentiation step of Delarue's differentiation scheme, which defined identities for the four middle codon bases. Further, 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 via 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. Duplicating a 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 presenting potential ambiguity, especially in the context of the anticodon loop, YYNNRR (Figure 1). Contemporary aaRS invariably use opposite groove recognition for these 16 pairs, suggesting a latent, complementarity-based sub-code 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: NAR•YUN == minor groove; RGN•NCY == major groove. If adjacent bases alternate - YR or RY - and hence could form a palindrome, then they are recognized from opposite grooves: YGN==minor•NCR==major and NAY==minor•RUN==major. The key link between the aaRS class distinction and error reduction is what the Rodins call the direction of "spreading", i.e. 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 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 twenty-two amino acids. Key to this notion is that one branch of descent retains no coding information, and is therefore always available for punctuation, i.e., 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 a class I and class II lysyl tRNA synthetases (Delarue, 2007).

The latent, complementarity based sub-code provides a compelling, 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 to the wobble hypothesis.

References

- Crick FHC. *Journal of Molecular Biology*. 1966; 38:367–379. [PubMed: 4887876]
Delarue M. *RNA*. 2007; 13:1–9. [PubMed: 17123956]
Eriani G, et al. *Nature*. 1990; 347:203–206. [PubMed: 2203971]
Freeland SJ, Hurst LD. *Journal of Molecular Evolution*. 1998; 47:238–248. [PubMed: 9732450]
Pham Y, et al. 2007; 25:851–862.
Rodin SN, Rodin A. *DNA and Cell Biology*. 2006; 25:617–626. [PubMed: 17132092]
Rodin SN, Rodin A. *DNA and Cell Biology*. 2006; 25:365–375. [PubMed: 16792507]
Rodin SN, Rodin AS. *Heredity*. 2008 In Press.
Schimmel P, et al. *Proceedings of the National Academy of Sciences, USA*. 1993; 90:8763–8768.
Vestigian K, Woese CR, Goldenfeld N. *Proceedings of the National Academy of Sciences USA*. 2006; 103:10696–10701.
Woese CR. *Proceedings of the National Academy of Sciences USA*. 1965; 54:1546–1552.



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 => 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, 4, and Table 3).