IFSR #183

IFSR #183

# FINE STRUCTURE OF THE GENETIC CODE

Swadesh M. Mahajan

Department of Physics
The University of Texas at Austin
Austin, Texas 78712

April 1985

# FINE STRUCTURE OF THE GENETIC CODE

Swadesh M. Mahajan

Department of Physics
The University of Texas at Austin
Austin, Texas 78712

Simple physical principles force the removal of degeneracy in the genetic code, giving it a fine structure; important biological consequences will probably follow.

The degeneracy of the genetic code, that is, the existence of more than one codon to specify the same amino acid, is very intriguing since it does not follow from any underlying symmetry principle. This purely accidental degeneracy[1] cannot be absolute because only absolute symmetries can lead to absolute degeneracies. I, therefore, predict that the degeneracy in the genetic code must necessarily be removed at an appropriate level of description, and that at that level there must exist observable differences in the functions of the ordinarily 'degenerate' codons.

Thus, the currently understood 'deciphering' of the genetic code is only a leading order understanding; a more complete understanding will emerge only after the 'fine-structure'[2] of the code, brought about by the removal of degeneracy, has been investigated.

The degenerate states of a system are characterized by the same value, x, of a controlling parameter. The removal or breaking of degeneracy implies that the new non-degenerate

states have to be labelled by different values $x_1$, $x_2$ $\cdots$ $x_n$ (for an n-fold degeneracy). Typically, the change $\Delta x \sim (x_n - x_{n-1}) \ll x$, the original value shared by all these states. For example, in the Hydrogen atom, the energy splitting $\Delta E$ due to spin orbit coupling for electronic states labelled by a particular value of the angular momentum is much smaller than the energy E associated with the erstwhile degenerate states.

What is the relevant parameter for the nucleotide triplets coding for the amino acids? One should probably seek the answer in the properties of the charge clouds (moments of charge distribution, for example) associated with these groups of molecules. The members of a particular degenerate group (for example, the codons, UUA, UUG, CUU, CUC, CUA and CUG for Leucine), which we shall call the Primarily Equivalent Codons (P.E. Codons for later use), are clearly indistinguishable at a primary level of cellular interaction: at energy levels $E_C$ and $E_H$ ($E_H < E_C$) associated respectively with the making (or breaking) of covalent and Hydrogen bonds. The energy differences between these triplets, i.e., the energy associated with the breaking of degeneracy, then, must be of the order $E_B \ll E_H$. We could interpret the energy of order $E_B$ as characterizing a new weak bond, the 'Biological Bond'. Thus, the P.E. Codons have equivalent Hydrogen bonding properties but differ in their ability to make the 'Biological Bond'. The manifestations of the fine structure of the genetic code must be sought in this realm of energy ($\sim E_B$).[3]

The biological implications of my proposal are manifold. It will be necessary to view gene coding, and gene expression in

a radically different light. The language of gene coding is hierarchical, and will have to be understood at several levels. The already elucidated version of the genetic code is simply a translation at the primary or the most energetic ($\sim E_H$) level. At this level, a coding change, i.e., a change in the sequence of codons, or the replacement of a codon specifying a particular amino acid by another specifying a different amino acid, is a major change, and manifests itself in the eventual synthesis of a different protein. This is the strong and dominant mode of gene expression, and its purpose is to determine the cell dynamics at the primary or gross level.

By interchanging slightly different P.E. Codons (one or several in two primarily equivalent chains), the gene encodes subtle differences in its information content which will be eventually expressed as subtle differences in the characteristics of the synthesized proteins. Needless to say, these subtle differences could result in important and even crucial differences in the cell-dynamics of two creatures which make the same protein(s) but whose genes are composed of two different but primarily equivalent chains.

The mechanism of expression of subtle genetic differences into functionally different proteins cannot be settled a priori. Some of the likely candidates are:

(1) It can happen through the selective properties of t-RNA which might have different binding efficiencies with different P.E. Codons (different abilities to make the 'Biological Bond'). This will result in different rates of synthesis of the

protein encoded by two primarily equivalent chains. For this mechanism, some evidence exists.[4]

(2) The amino acid molecules retain the memory of their respective codons; the Leucine moelcule coded by UUA is just a bit different from the Leucine molecule coded by CUG. This difference could be in the charge cloud associated with these molecules. This could lead to different functional properties between two protein molecules with exactly the same amino acid sequence. I am not aware of any known experimental evidence for this mechanism. However, I do not believe that the evidence for this possibility has ever been sought.

(3) Two primarily equivalent DNA chains could have different binding strengths with the repressor molecules which prevent transcription. Thus, different quantities or even a different type of an inducer may be needed to allow the transcription to begin. This obviously would change the cell chemistry by changing the rates and kinds of reactions that might occur.

Only experiment could determine the mechanism(s) by which the gene expresses its fine structure. Whatever the mechanism, I believe that our understanding of basic genetics (and hence biology) will remain severely limited until we determine how the gene manages to convert this apparent redundancy (degeneracy) into a powerful tool for coding at a subtler level.

I thank Dr's. Suresh Mahajan, Prashant Valanju, E.C.G. Sudarshan and Satish Arora for their support.

FOOTNOTES

1. It will, of course, be worthwhile to search for plausible explanation for this phenomenal accident; 'wobble' hypothesis[5] is a good example.

2. I have borrowed this phrase from atomic spectroscopy where it is used in an analogous situation.

3. I have used the word "energy" generically to denote a controlling parameter. There may indeed exist a more sensitive signature of the 'charge-distribution' which will distinguish between the P.E. Codons. For example, the spin is a much better attribute than the energy-difference to distinguish between the electron-positron bound state (positronium), and the unbound state of an electron-positron system.

4. T. Ikemura and H. Ozeki in Proceedings of Cold Spring Harbor Symposia on Quantitative Biology, Vol. 47, page 1087 (1982).

5. J.D. Watson, Molecular Biology of the Gene, 3rd Edition, W.A. Benjamin Inc., California, pp. 357-358.