Review

Structure and evolution of the genetic code viewed from the perspective of the experimentally expanded amino acid repertoire in vivo

N. Budisa, L. Moroder and R. Huber

Proteros Biostructures GmbH, Am Klopferspitz 19, D-82152 Martinsried (Germany), Fax +49 89 7007 6115, e-mail: budisa@proteros.com
Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried (Germany)

Received 30 June 1999; accepted 9 July 1999

Abstract. Much effort has been devoted recently to expanding the amino acid repertoire in protein biosynthesis in vivo. From such experimental work it has emerged that some of the non-canonical amino acids are accepted by the cellular translational machinery while others are not, i.e. we have learned that some determinants must exist and that they can even be anticipated. Here, we propose a conceptual framework by which it should be possible to assess deeper levels of the structure of the genetic code, and based on this experiment to understand its evolution and establishment. First, we propose a standardised repertoire of 20 amino acids as a basic set of conserved building blocks in protein biosynthesis in living cells to be the main criteria for genetic code structure and evolutionary considerations. Second, based on such argumentation, we postulate the structure and evolution of the genetic code in the form of three general statements: (i) the nature of the genetic code is deterministic; (ii) the genetic code is conserved and universal; (iii) the genetic code is the oldest known level of complexity in the evolution of living organisms that is accessible to our direct observation and experimental manipulations. Such statements are discussed as our working hypotheses that are experimentally tested by recent findings in the field of expanded amino acid repertoire in vivo.

Key words. Amino acid repertoire; evolution; genetic code; metabolism; protein folding.

An expanded amino acid repertoire and the genetic code

New terminology
In our first attempts to interpret new experimental findings in the context of the structure and evolution of the genetic code, we found inconsistent terminology for amino acids to be a major stumbling block. In other words, we have been convinced that understanding the genetic code at a deeper level cannot be attained with the current taxonomy of amino acids. Thus, we proposed a new nomenclature which should not be difficult to integrate into the already existing biochemical terminology [1]. In brief, the well-known standard set of 20 amino acids represents canonical amino acids; other amino acids outside this standard set which can be introduced in a codon-dependent manner are non-canonical amino acids. Those amino acids whose introduction is not only codon dependent, but also dependent on context (e.g. selenocysteine or formyl-methionine) are special canonical amino acids. There are also experimental procedures, such as in vitro suppression, that led to context-dependent introduction of
some amino acids with special properties (e.g. cages, sensors) that are named special non-canonical amino acids. Finally, numerous amino acids resulting from secondary metabolism, precursors or post-translational modifications are special biogenic amino acids. This distinction is not a semantic issue, but a biological one. In fact, only such terminological clarification convinced us to propose the concepts about the genetic code where the amino acid repertoire is a central criterion for dissecting its nature, evolution and establishment.

Steps in the flow of genetic information
The flow of genetic information is a complex process. It starts with self-replicating DNA that contains instructions that are converted into biological activity through transcription followed by translation, as