

Restricting the amino acid usage of a resurrected ancestral protein

Satoshi Akanuma^{1*}, Masami Shimada¹, Yoshiki Nakajima¹, Shin-ichi Yokobori¹, Akihiko Yamagishi¹

¹Department of Applied Bioscience, Tokyo University of Pharmacy and Life Sciences

It is common among life on Earth that proteins are composed of 20 or nearly 20 kinds of amino acids. However, it is uncertain that, from the standpoint of astrobiology, proteins are always composed of the 20 kinds of amino acids. Even for the evolution of proteins on Earth, it has been argued that primordial proteins that existed before the last universal common ancestor involved less than 20 kinds of amino acids (Crick, 1968). Given simpler protein synthesis system, the primitive proteins, which might have comprised a reduced set of amino acids, must have had a sufficiently adequate structure for functional interactions and catalysis. To address this issue experimentally, we used the protein simplification engineering (Akanuma et al., 2002) to examine whether a protein composed of less than 20 types of amino acids can retain its stable structure and biological function. To this end, we first resurrected several ancestral proteins and then restricted the amino acid usage of a resurrected protein to a reduced amino acid set.

As the model protein, we chose a house keeping enzyme, nucleoside diphosphate kinase (NDK). Because extant genes are evolutionary descendants of ancient genes, ancestral sequences of a particular protein can be inferred by comparing extant homologous protein sequences. Along this line, we first inferred several ancestral amino acid sequences of NDK by phylogenetic analysis of the extant homologous sequences. The inferred sequences were then genetically reconstructed. One of the reconstructed protein, Arc1, is very thermally stable (unfolding temperature = 113°C) and shows catalytic efficiency similar to those of the modern NDKs.

It is currently impossible to infer amino acid sequences that existed before the last universal common ancestor. Therefore, using Arc1 as the starting molecule, we reconstructed a reduced amino acid set variant, Arc1-s2, in which Met, Gln, Lys, Tyr and Asn were replaced by other amino acids. Because cysteine is absent from Arc1, Arc1-s2 consists of only 14 amino acid letters. Arc1-s2 retains thermal stability similar to that of Arc1; whereas, no detectable level of catalytic activity was observed for Arc1-s2. Therefore, the fourteen amino acid types are sufficient to encode a thermally stable protein but more amino acid types would be required for its function. To reevaluate the individual contributions of the 20 amino acid types to the stability and activity of Arc1, we reconstructed 19 Arc1 variants in which one of the 19 amino acid types was all replaced by other amino acids. The result will be also present in the meeting.

(1) Crick, *JMB* 38, 367 (1968); (2) Akanuma et al., *PNAS* 99, 13549 (2002)