Algebraic Evolution of the Genetic Code

Takashi Kawamura
kawamura@dmb.med.keio.ac.jp

Department of Molecular Biology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

Keywords: genetic code, symmetry breaking, freezing

1 Introduction

The evolution of the genetic code has been speculated by many authors, but all of them lack determinative proofs. Jukes [4] inferred the evolution from mitochondrial codes, which are different from the universal codes in some respects. He postulated an archetypal code containing 14 amino acids, from which the universal code could evolve by gene duplication followed by mutational changes. This process could have included 28 amino acids in the code, but it has only 20 ones due to the freezing, which is defined as some codons that could have coded for different amino acids code for the same one.

Hornos et al. [3] devised an algebraic model for the evolution of the genetic code. They looked for Lie groups that have a 64-dimensional irreducible representation, which is called a codon representation. Especially they studied Sp(6), the symplectic group of degree 6. If the codon representation of Sp(6) is restricted to its subgroup Sp(4)*SU(2), then it is decomposed into 6 irreducible representations, which is called symmetry breaking. Each representation corresponds to one of the primordial amino acids and the termination codon. Decomposing the representation according to the chain of 4 subgroups, they obtained 27 subspaces assigned to the amino acids and the termination codon. As there are only 20 amino acids in the code, 6 spaces are redundant, which is freezing.

Forger and Sachse [1] explored the codon representations of Lie superalgebras. If the codon representation of the orthosymplectic algebra osp(5|2) is restricted to its subalgebra sp(2)+so(5), then it is decomposed into 3 irreducible representations. Continuing the decomposition according to the chain of subalgebras sp(2)+so(5), sl(2)+sl(2)+sl(2), sl(2)+sl(2), sl(2)+o(5) and o(5)+o(5), we obtain 3, 10, 14, 18 and 26 subspaces, respectively, corresponding to amino acids and the termination codon. But they did not explain the mechanism of freezing, and did not assign amino acids to the subspaces.

We assign the amino acids to the 26 subspaces of [1], and try to explain the mechanism of freezing.

2 Method and Results

The orthosymplectic algebra osp(5|2) has a 64-dimensional irreducible representation. If the representation is restricted to sp(2)+so(5), it is decomposed into 3 irreducible representations, which are 32, 20 and 12-dimensional respectively. When we continue this process according to the chain of subalgebras sp(2)+so(5), sl(2)+sl(2)+sl(2), sl(2)+sl(2), sl(2)+o(5) and o(5)+o(5), we obtain 3, 10, 14, 18 and 26 subspaces, respectively, corresponding to amino acids and the termination codon. As the universal code contains only 20 amino acids, 5 amino acids are redundant due to freezing.

We assign the amino acids to the 26 subspaces based on many biological findings. We think that the genetic code has evolved in 6 steps. Each step corresponded to the 5 subalgebras of osp(5|2), and included 2, 9, 13, 17, 21 and 20 amino acids respectively. The first step included only glycine and serine. Proline, leucine, valine, ornithine, asparagines, aspartic acid and glutamine were added in the second step. The third step involved alanine, phenylalanine, tryptophan and tyrosine in addition to the 9 amino acids in the second step. Isoleucine, methionine, cysteine, and glutamic acid were added in the fourth step. The fifth step included threonine, arginine, lysine and histidine in addition to the 17 amino acids in the fifth step. And ornithine has completely been replaced by arginine in the present code.

The mechanism of freezing is the use of phenylalanine, methionine, tryptophan and arginine instead of 4 new amino acids and the intrusion of arginine in place of ornithine.

3 Discussion
In the first step the 32-dimensional subspace corresponds to the termination codon, the 20-dimensional one to glycine, and the 12-dimensional one to serine. The genetic code was assumed to have originated with the coupling of glycine to the anticodon CC mediated by a copper-montmorillonite [2]. Weber et al. [5] observed that serine could be formed from glycine in the primitive ocean in some conditions.

In the second step the 32-dimensional subspace is decomposed into 4 subspaces, to which proline, leucine, valine and the termination codon are assigned. Hartman [2] stated that the polymerization of proline followed that of glycine when it was coupled to the anticodon GG.

The subspace corresponding to glycine is decomposed into 4 subspaces, to which ornithine, glycine asparagine and aspartic acid are assigned. Jukes [4] proposed that arginine was preceded by ornithine in the early genetic code. Hartman [2] observed that alternative hydrophobic-hydrophilic polypeptide, alanine and ornithine, was coded for by the alternative CG copolymer.

The third step involved 13 amino acids. The subspace corresponding to proline is decomposed into an octet and a quartet, to which proline and alanine are assigned respectively. The subspace corresponding to leucine is decomposed into a nonet and a triplet, to which leucine and phenylalanine are assigned respectively.

The fourth step included 17 amino acids. The nonet corresponding to leucine is decomposed into a sextet and a triplet, to which leucine and isoleucine are assigned respectively. The triplet coding for phenylalanine broke into a doublet and a singlet, which code for phenylalanine and methionine respectively.

The fifth step included 21 amino acids. The sextet coding for leucine broke into a quartet and a doublet, to which leucine and phenylalanine are assigned respectively. The sextet should be frozen, but the doublet is considered to have coded for phenylalanine.

The triplet coding for isoleucine is decomposed into a doublet and a singlet, to which isoleucine and methionine are assigned respectively. The triplet should be frozen, but AUA in the mitochondrial genetic code corresponds to methionine. So the singlet is thought to have coded for methionine.

The triplet corresponding to the termination codon is decomposed into a doublet and a singlet, to which the termination codon and tryptophan are assigned. The triplet must be frozen, but UGA in the mitochondrial genetic code corresponds to tryptophan. Hence the singlet is considered to have coded for tryptophan.

The sextet coding for ornithine broke into a quartet and a doublet, which code for ornithine and arginine respectively. The sextet corresponding to serine is decomposed into a quartet and a doublet, which code for serine and arginine respectively.

The sixth step is the present genetic code. The singlet AUA corresponding to methionine has come to pair with the anticodon *CAU of the isoleucine tRNA, where *C has been modified so that the anticodon pairs with AUA but not with AUG.

The singlet UGA coding for tryptophan has come to correspond to the termination codon as a result of the mutation of the anticodon UCA to CCA leaving UGA without an anticodon [4]. The quartet corresponding to ornithine has come to code for arginine due to more intrusion of arginine.

Acknowledgments
The author is grateful to Professor Nobuyoshi Shimizu, Dr. Shuichi Asakawa and Mr. Susumu Mitsuyama for encouraging him and giving him valuable comments.

References