BANBURY CENTER MEETING SUMMARY

Evolution of the translational apparatus and implications for the origin of the Genetic Code

November 13-16, 2016

In 1966 the Cold Spring Harbor Laboratory symposium "The Genetic Code" celebrated the completion of the Genetic Code. Francis Crick noted that the remaining challenge was the Code's origin:

> *"It is almost impossible to discuss the origin of the code without discussing the origin of the actual biochemical mechanisms of protein synthesis. This is very difficult for two reasons: it is complex and many of its details are not yet understood."*
>
> *Francis Crick*

The time has come to discuss the Origin of the Genetic Code and the evolution of the Translational Apparatus.

In November 2016, leading researchers involved in the translational system attended a workshop at the Banbury Center of Cold Spring Harbor Laboratory. The meeting began with a review of the structure, function and evolution of the ribosome.

The Ribosome

The ribosome is composed of a large

subunit (LSU) and a small subunit (SSU). For example, in bacteria the ribosomal RNA of the SSU is 1500 nucleotides long and has 23 ribosomal proteins, while the ribosomal RNA of the LSU is comprised of 2900 nucleotides with 34 ribosomal proteins. Archaeal and eukaryotic ribosomes have similar RNAs and proteins. The evolution of structures of all of these RNAs and proteins can now be studied.

We can also now follow the movements of the LSU and SSU subunits of the ribosome that are coupled to the movement of the messenger RNA, and the movement of the growing peptide in the tunnel of the ribosome. Associated with the movement of the messenger RNA is the movement of the charged tRNAs from the A site to the P site on the ribosome. The activated amino acid on a tRNA in the A site is added to the growing polypeptide attached to a tRNA in the P site. This addition is catalyzed at the Peptidyl Transfer Center (PTC), located on the LSU. The PTC is an RNA structure that catalyzes this exchange and hence is a ribozyme.

Ribosomal RNA

The consensus at this meeting was that the LSU of the ribosomal RNA evolved by a set of sequential RNA additions to the PTC. This model of the evolution of the LSU RNA demonstrates (1) the complex structures of the translational apparatus i.e. the ribosomal RNA began as a set small RNAs, and (2) ligation created larger structures from smaller structures. The LSU RNA grew "like an onion" and generated an evolutionary path from simplicity to complexity. This model was also applied to the evolution of the SSU ribosomal RNA and created a theme for the remainder of the meeting.

Ribosomal Proteins

In order to study the evolution of the ribosomal proteins, it was necessary to study the Universal Ribosomal Proteins (URPs), those proteins, which are homologous in the three cellular domains (Bacteria, Archaea and Eukarya).

In the comparisons of these homologous proteins in Bacteria, Archaea and Eukarya, it was found that these proteins break up into blocks (i.e In the same protein there are blocks of amino acids that go back to a common ancestor), there are blocks that go back to a common ancestor of the Bacteria and there are blocks that go back to a common ancestor of the Archaea and the Eukarya. The simplest explanation for this observation is that the ribosomal proteins were ligated from smaller peptides similar to the growth of the Ribosomal RNAs.

In the URPs, there are two types of ribosomal proteins: globular proteins without extensions and globular proteins with extensions that reach into the RNA of the ribosome.

Concurrent with the evolutionary growth of the ribosomal RNAs were the co-evolution of the ribosomal proteins. The

fully globular proteins are found on the ribosomal surface and are furthest from the PTC. The globular proteins with extensions must be considered separately as the extensions were quite distinct in their amino acid composition and were distinct in their block structure from that of their globular connections The extensions that reach deep into the large ribosomal subunit's interior and interact with the PTC represent some of the very oldest peptides.

In particular there were four ribosomal proteins associated with the LSU, (L2, L3, L4 and L22), whose extensions interacted with the PTC and the associated Tunnel.

There is a gradient from peptides at the PTC, the hairpin in the Tunnel to the globular proteins at the end of the Tunnel and at the surface of LSU The extensions from L2, L3 and L4 were random coils. L22's peptide extension was a beta hairpin that interacted with the RNA of the Tunnel. There are two globular proteins L23 and L29 at the exit of the Tunnel and the protein L23 is an Alpha Beta protein and L29 is a pure Alpha protein.

There is an evolution from the peptide extensions at the PTC and the Tunnel thru to the globular proteins at the surface of the Ribosome that begins with Random coils→Beta Hairpin→ Alpha Beta protein→ Alpha helical protein . This evolution is reflected in the Evolution of the Genetic Code itself.

The Aminoacyl-tRNA synthetases (AARS)

There are twenty aminoacyl tRNA synthetases, one for each amino acid, forming two groups: Class I (Val Leu Ile, Met, Tyr, Trp, Glu, Gln, Cys, Arg), and Class II (Gly Pro Ala, Thr, Ser, Lys, Asp. Asn, Hist, Phe).

These proteins are multi-domain proteins with a catalytic domain containing the active site where the Amino acid is activated by ATP and then attached to the appropriate tRNA. In addition there are an editing and a codon recognition domains.

The two Classes are distinguished by their catalytic domains; Class I has a Rossman Fold and Class II has a Beta Sheet domain. It was proposed that the Class II synthetases evolved first before the Class I synthetases based partially on the simpler Amino Acids such as (Gly, Pro, Thr and Ser) i.e are close to their formation from the reverse citric acid cycle.)

The Catalytic Domain of the Class II synthetases

A detailed analysis of the structure of the active site of these enzymes led to the observation that they were formed by the ligation of three polypeptide hairpins.

The evolution of the Catalytic domains of the Class II synthetases was studied by treeing the amino acid sequences of the active site of the Class II synthetases.

This resulted in the observation that Glycine (GG*) and Proline (CC*)

Synthetases were the first of the Class II Synthetases to form Thus the first amino acids to be encoded in the Genetic Code were Glycine and Proline followed by Alanine.

The structural and functional studies of Glycine and Proline synthetases were then examined and in particular there was a structural analysis of the two distinct Glycyl-tRNA synthetases.

Since there are two distinct Glycyl-tRNA synthetases, the question was posed which came first and could not be resolved.

There were further discussions of the limits of the accuracy of the early code and the importance of the later addition of trans-editing domains that removed incorrectly charged tRNAs.

Finally, there were the further ligation of peptides to the C and N terminals of the catalytic centers of the Glycine, Proline and Alanine synthetases. These peptides interacted with the operational code in the tRNA acceptor arm.

The Thioester World

It was also shown that the catalytic sites of the AARS themselves pass an activated amino acid onto a pantothenic acid to form a thioester. This suggested that the synthetases were involved in the formation of thioesters before they interacted with the tRNAs and hence they evolved in a thioester world.

It has not escaped our attention that it seemed that peptides such as hairpins were present before they were coded for by the translational apparatus. This of course has been a paradox that must be resolved. One answer is that peptides were formed in what has been called the thioester world. Evidence for such a world was given at this meeting by a study of the network of biosynthetic catalysts in which phosphate was removed and the resulting network showed an underlying thioester network of some complexity.

The co- evolution of the tRNA and the Catalytic site of the Class II synthetases.

It is at the catalytic site of the synthetases that the activated amino acid is passed on to the tRNA. At the end of the acceptor arm there are four nucleotides that are not in a helix. They are the discriminator nucleotide and the three nucleotides CCA. It is to the ribose of the A nucleotide to which the amino acid is attached.

The CCA is not coded for and is attached by an enzyme appropriately named the CCA enzyme. This enzyme belongs to a large superfamily of nucleotide transferases including the Poly A polymerase. Like the Class II synthetases, this CCA enzyme and the superfamily to which it belongs all have the same active site due to the ligation of a number of hairpin polypetides.

Like the synthetases and the ribosomal RNAS, the CCA enzyme grew more complex by ligation of peptides. It began small and grew more complex by ligation.

The evolution of the tRNA

Alterations of the CCA enzyme give rise to more complex polynucleotides such as the tetramers UCCA and UGGU.

It is from these tetra-nucleotides that the tRNA evolved again by ligation to form the mini helices and the early messenger RNA. As the Class II synthetases evolved the interaction between the evolving t-RNA the acceptor arm began to interact with the evolving catalytic site and gave rise to the operational code in the acceptor stem of the t-RNA Here we are beginning to see the coevolution of the tRNA and the catalytic module of the Class II synthetases.

It is at this point that the RNA world was born and began to interact with the peptide world.

Summary

This workshop gave us an exciting new perspective on the Origin and Evolution of Genetic code and its underlying relationship to the Origin and Evolution of the translational apparatus.

The proteins of the translational system were derived from smaller peptides by ligation. The source of the original peptides,which came from a thioester world. The ribosomal RNAs and transfer RNAs were evolved from smaller polynucleotides by ligation. In general the RNAs and proteins of the translational apparatus were evolved in a world of small peptides and small polynucleotides.

The evolution of the Genetic Code code can be seen in the evolution of the aminoacyl t-RNA synthetases, especially in their Catalytic Modules. The evolution of t-RNAs is based on ligation of polynucleotide tetramers resulting in an operational code.

This meeting was organized by Hyman Hartman (Massachusetts Institute of Technology) and Temple Smith (Boston University), and was funded by the Cold Spring Harbor Laboratory Corporate Sponsor Program.