The Evolution of Genetic Coding

Professor Charles Williams Carter, Jr

Scientia

THE EVOLUTION OF GENETIC CODING

The research of **Professor Charles Carter**, University of North Carolina at Chapel Hill, USA, unravels some of the biggest mysteries of molecular evolution. His research is dedicated to investigating how information flows from genes to proteins found in living organisms via genetic coding. Much of this work centres around the structural origins of genetics and ultimately, how chemistry created biology.



The Origins of Genetics

One of the biggest evolutionary questions is 'how did we come to exist in the forms that we do?' The most significant contribution towards answering this question was the famous discovery of the DNA double helix by Watson and Crick in 1952. Due to their work, we now know that the four nucleotide bases, the building blocks of DNA, always line up in a characteristic manner along the double helix backbone; and that adenine and thymine always bind to each other, as do guanine and cytosine, ensuring that genes are passed correctly from one generation to the next. Genes are blueprints for making proteins. In order to access the information hidden in genes, the nucleotide sequence must be read and transcribed into a new version of the blueprint - messenger RNA, or mRNA.

Once transcribed, the RNA blueprint must be interpreted so that amino acids can be correctly strung together. This translation process depends on a code to assign one of the 20 amino acids to each triplet of bases (codons). Translation ultimately produces a unique, coded sequence of amino acids that can fold into 3D structures called proteins. However, exactly how the genetic code itself first arose and how information is stored and recovered remain unclear. These are the mysteries that Professor Charles Carter and his team at the Department of Biochemistry and Biophysics at the University of North Carolina at Chapel Hill, USA, aim to solve.

Two distinct kinds of information have been embedded in nucleic acids for as long as they have served as genes. First, we need to consider the process by which triplet codons consisting of the four bases are turned into the twenty amino acids that make up all possible proteins. This is facilitated by an adaptor molecule called transfer RNA (tRNA). Enzymes called aminoacyltRNA sythetases (aaRS) recognise tRNA, forming cognate pairs. These cognate pairs recognise each other very specifically and together, form the molecular dictionary. Each amino acid is associated with one or more cognate pairs made up of one aaRS and tRNAs, whose recognition properties allow the synthetase enzyme to activate and attach the correct amino acid to one end of its cognate tRNA.

Cognate pairs are capable of translating the genetic code from nucleic acid to protein because the aaRS specifically recognise cognate amino acids, and



transfer them covalently to their cognate tRNAs. Non-cognate pairs, that is, incompletely matching aaRS-tRNAamino acid combinations, may occur but at such low levels that they rarely participate in the assembly of proteins. The aaRS enzymes form the basis of much of Professor Carter's research. He believes that they may be the most important contributor to the translation of the genetic code to proteins.

The second source of information comes from the collection of messenger RNAs (mRNA) found within a cell. These mRNAs represent a database of the amino acid sequences that are able to successfully fold and function usefully in the cell.

'This discovery opened the door to a virtual flood of new experiments that have transformed the way we think about the origin of life itself.'



Translation of aaRS genes is a paradoxical level-crossing strange loop. mRNA operates at one level, furnishing blueprints for making proteins, including aaRSs. Translated proteins fold into functional 3D structures in accordance with amino acid physical chemistry (green ellipse) taking them to another level. However, folded aaRS proteins actually translate genes by enforcing the genetic coding rules by which they were originally assembled (blue ellipse). Professor Carter's work elucidated many details of the Escher-like stairways by studying the rules connecting the two levels.

Making Sense (or Antisense) of Bidirectional Coding

A major surprise came with the discovery in 1990 that there are two distinct and apparently unrelated superfamilies of aaRS. The Rodin-Ohno hypothesis offered a rationale by suggesting that as DNA has two strands, perhaps each of the two strands acts as an individual coding strand, rather than one coding strand and one template (for copying) strand. The existence of bidirectional coding in this way has previously been considered controversial.

That controversy prompted Professor Carter's group to develop new biochemical, bioinformatic, and phylogenetic approaches to test the critical Rodin-Ohno hypothesis that the two unique synthetase classes had evolved from ancestors that, while present on the same gene, were located on opposite sense and antisense strands. They designed and tested a bidirectional gene that produced gene products from both strands (compared to the norm of using just one strand). That gene coded for two specific peptides (amino acid chains which give rise to proteins). When they measured the levels of amino acid activation using enzymatic assays, both gene products accelerated amino acid activation by one million-fold. Professor Carter and his group believe that such a gene could have given rise to the first two enzymes capable of activating amino acids.

The complex process behind this discovery involved help from a colleague Professor Brian Kuhlman, whose computer program, called Rosetta, is able to design protein structures that obey bidirectional coding. This work showed that the peptides from opposite strands of the same gene had entirely different amino acid sequences. It also verified that it was indeed possible to encode two enzymes with very similar activities, and which were much smaller than modern enzymes. Professor Carter described the two enzymes as 'a bone fide molecular Adam and Eve!'

Professor Carter has shown that one ancestral gene could code for two distinct, functional amino acid-coding enzymes, or protozymes. Thus, because two letters represent the simplest possible code, and because that initial code required both tRNA and protein, he suggests that biology did not replace an existing RNA-dependent system but instead arose directly from a very early partnership between RNA and peptides. Bidirectional coding thus limits the theory that life forms originally arose from an RNA-focused world, raising the question: should the focus should be on a peptide-RNA world instead?

Tryptophan and the Chemistry of Amino Acid Side Chains

Professor Carter and his group have worked extensively on the aaRS for one particular amino acid, tryptophan (the precursor of the neurotransmitter, serotonin). Their findings include analyses of how tryptophanyl-tRNA synthetase (the enzyme required for tryptophanyl-tRNA formation) carries out its function.



Professor Carter and Research Group

The ways in which polypeptide chains go on to form very complicated 3D protein structures are due to many variables. Work by the team verified several intricate links between the physical chemistry of amino acid side chains, which are the chemical groups attached to the main backbone of the molecule, and protein folding. The approach used by Professor Carter and his colleague Professor Richard Wolfenden to examine how amino acid side chains behave was multifaceted. First, they observed the polarities of the side chains, by measuring how they distribute between water and cyclohexane, a colourless, flammable liquid. The polarity of a molecule explains how positively or negatively charged it is. They also investigated the size of the side chains by measuring distributions between the vapour phase (a state when the molecule is free from interactions with other molecules) and cyclohexane.

Using these techniques, they also were able to demonstrate that one region of the tRNA crucial for recognition by the synthetases is its acceptor-stem. The code in the tRNA acceptor-stem is related to amino acid size, whereas the anticodon, a distant, alternative part of the molecule is related to amino acid polarity. They then realised that the two descendants of the ancestral bidirectional gene recognised two separate groups of amino acids that can be distinguished by whether they have large or small side chains. Thus, the inherent duality of the two aaRS Classes depends intimately on the physical chemistry of the amino acids that they recognise and is embedded into the tRNA structures.

Strange Loops

The symbolic transformation achieved by genetic coding of aaRS is a paradoxical, level-crossing feedback loop – a 'strange loop'. This is a hierarchy of levels, each of which is somehow linked to another. Remarkably, moving through the levels, one eventually returns to the starting point, i.e., the original level. How this works in genetic coding is illustrated in Figure 1. The feedback mechanism of the strange loop is perhaps best explained by a quote from another of Professor Carter's collaborators, Professor Peter Wills, who explains 'the enforcement of the relationship between genes and amino acids depends on aaRS, which are themselves encoded by genes and made of amino acids.' Professor Wills also notes that this process is essentially computational, and so transcends chemistry. Professors Carter and Wills propose that this feedback cycle was necessary for Nature to search rapidly for an optimal genetic code and for protein sequences that fold into functioning machines. In this sense, it represents an early emergence of 'symbolic meaning'.

01010

011

01 01

0110

0110

01

010

Next Steps in Understanding the Origin of Life

Recent work by the group has established the basis of the mechanisms by which aaRS and tRNAs recognise each other (the underlying mechanisms of the blue Escher-like stairways in Figure 1) and how this duality was exploited to form the cognate pairs necessary to translate the genetic code.

Crucial for supporting their conclusions, was to understand how and why tRNAs recognised specifically by each aaRS Class obeyed the same duality. A major achievement for the team came from examining the crystal structures of aaRStRNA cognate pairs. This revealed that the main distinction between the two different types of cognate pairs (small or large side chain) was that the aaRS specific for large side chain amino acids bind to a different groove in the tRNA, in this case requiring formation of a hairpin loop. Only Class I aaRS can induce and recognise that hairpin only in Class I cognate tRNAs, whereas Class II aaRS prefer to bind the undistorted acceptor stems of Class II cognate tRNAs. Thus, architectural features of the two aaRS Classes dictate their interactions with their cognate tRNAs. The pattern of acceptor-stem bases that form the hairpin became hidden as the amino acid alphabet grew to its modern form. Importantly, this suggests that aspects of current day tRNAs may form a palimpsest, a molecule with a slightly altered structure, but which still bears resemblance to its original form as it would have been many, many years ago.

The methods involved in reaching this conclusion included both experimental and computational approaches, which together with other analyses such as protein engineering and phylogenetic analysis, advanced the toolset that will be required for future research in this field.

Professor Carter's team continue to discover new aspects of the genetic code that are intimately associated with the structures of the two synthetase classes. They have also provided new understanding about how the physical chemistry of amino acids drives 3D protein folding. This cutting-edge research, along with the exciting work that the group have planned for the future, provides new avenues to discovering how the very first biological life forms emerged from chemical reactions. This may ultimately challenge the status quo that RNA alone was behind the start of life.

WWW,SCIENTIA.GLOBAL



Meet the researcher

Professor Charles Williams Carter, Jr

Department of Biochemistry School of Medicine University of North Carolina Chapel Hill, NC USA

Professor Charles Carter completed his PhD at the University of California, San Diego, in 1972. He then completed post-doctoral work at the Medical Research Laboratory of Molecular Biology, Cambridge, UK. His doctoral research led to his appointment as Professor of Biochemistry and Biophysics at the University of North Carolina at Chapel Hill, USA. In recognition of his work focusing on macromolecular structure and function, Professor Carter received the Fulbright International Fellowship award in 2010 for work at the Pasteur Institute in Paris and was Elected Fellow in the American Association for the Advancement of Science in 2013 for his work on molecular evolution. Professor Carter was President of the American Crystallographic Association in 2002 and served as their representative to the American Institute of Physics (AIP) until 2016. He also sits on the Publishing Partners Advisory Committee for AIP Publishing. He remains a long-term supporter of the Biophysical Society, the American Society of Biological Chemists, and the Society for Molecular Biology and Evolution.

CONTACT

101

an all

E: carter@med.unc.edu T: (919) 966-3263 W: http://carterlab.web.unc.edu





KEY COLLABORATORS

Richard Wolfenden, University of North Carolina Chapel Hill, USA Peter R. Wills, Auckland University, New Zealand

FUNDING

National Institute of Health Ffame/John Templeton Foundation

FURTHER READING

CW Carter Jr, Wills PR, Hierarchical groove discrimination by Class I and II aminoacyl-tRNA synthetases reveals a palimpsest of the operational RNA code in the tRNA acceptor-stem bases, Nucleic Acids Research, 2018, 46, 9667–9683.

CW Carter, P Wills, Interdependence, reflexivity, fidelity, impedance matching, and the evolution of genetic coding, Molecular Biology and Evolution, 2017, 35, 269–286.

CW Carter, Coding of Class I and II aminoacyl-tRNA synthetases. Advances in Experimental Medicine and Biology, Protein Reviews, 2016, 18, 103–148.

CW Carter, An alternative to the RNA world, Natural History, 2016, 125, 28–33.

WWW.SCIENTIA.GLOBAL