Molecular fossils probe life's origins

Research into molecular fossils and modern viruses is shedding light on the evolution of archaea, prokaryotes and eukaryotes

Philip Hunter

The fossil record—surviving mineral components or imprints of multicellular life—has provided valuable insights into how animals and plants evolved over millennia, but offers limited scope for discerning the origins of life itself or the separation of organisms into the three domains of eukaryotes, prokaryotes and archaea. The development of new technologies, however, is enabling scientists to analyse molecular fossils, such as the remnants of ancient nucleic acids, sugars, proteins, carbohydrates and lipids, to study the evolution of key metabolic pathways. This knowledge should enable researchers to peer back through time as far as the great oxidation event (GOE) that enabled the emergence of eukaryotic life. In addition to the analysis of the organic molecules themselves, the study of modern genomes to look for ancestral clues might yield knowledge about the protein structures present in the earliest forms of life.

The GOE is thought to have occurred when cyanobacteria released free oxygen into the atmosphere through oxidative photosynthesis. While there is reasonable consensus that this occurred around 2.4 billion years ago, there is still uncertainty over how long it took until sufficient free oxygen accumulated to enable oxidative metabolism and, eventually, the emergence of new life forms. Initially, minerals, including iron, which would have been present in metallic form and plentiful, are thought

to have taken up the oxygen produced by cyanobacteria. Atmospheric oxidation would not have started until the Earth's surface minerals had become saturated, but the estimated time to that point ranges from 100 million years to 1 billion years. Because cyanobacteria are widely believed to be one of the first lifeforms because of their ability to thrive in anoxic conditions, resolution of this question could move scientists closer to establishing the origins of life.

Unlike physical fossils, molecular fossils do not contain material derived directly from the original organism itself, but rather are biomarkers that represent some of its specific chemical composition and provide a 'signature'. Molecular fossils are embedded in rock or sediment and are altered over time by chemical and physical processes. As such, they can only be dated indirectly by analysis of the surrounding rock or sediment. Although direct dating methods are now considered fairly reliable, indirect methods are controversial because they rely on various assumptions, notably that the sample has not been contaminated and has remained fixed relative to its surroundings. "It is assumed the molecules of the microbes present in these rocks are of similar age," said Stefan Schouten, an organic geochemist at the Royal Netherlands Institute for Sea Research, Texel, Netherlands. "Generally this assumption is correct, though in recent sediments offsets of up to 5,000 years have been noted."

The molecules are commonly separated from one another by using gas or high-pressure liquid chromatography and identified by mass spectrometry. The surrounding material is dated, usually by using

The most extensive use of molecular fossils to date has been to search for biomarkers [...] of the [great oxidation event] and the associated emergence of eukaryotic life

well-established radiometric methods, often combined with stratigraphy: the analysis of rock or sediment formation through the accumulation of successive layers, which assumes that a lower layer must be older than the one above it.

Despite the challenges, molecular fossils provide information about the organisms they are derived from and the biosynthetic pathways in operation at the time of their formation. Some of the key biomarkers in old deposits include sesquiterpenes, which indicate that a fossil came from a plant or insect; biphytanes, which point to archaea; hopanes, which suggest bacteria; 2-methylhopanes, which are specifically associated with cyanobacteria; and steranes, which point to eukaryotes. Hopanes, for instance, are derived from hopanoids, which give strength and rigidity to the plasma membranes of bacteria. Sterols fulfil a similar role in eukaryotes and form steranes under the action of sedimentary processes.

The most extensive use of molecular fossils to date has been to search for biomarkers that provide evidence of the GOE and the associated emergence of eukaryotic life. Notable advances have been made, but have raised major controversy over the duration of the GOE. In 1999, Jochen Brocks and colleagues at the University of Sydney, Australia, reported evidence that eukaryotes were present

...molecular fossils provide information about the organisms they are derived from and the biosynthetic pathways in operation at the time of their formation

up to 2.7 billion years ago, which is 1 billion years earlier than had previously been believed [1]. In a paper published in *Science*, the researchers argued that the presence of abundant 2 α -methylhopanes, which are characteristic of cyanobacteria, indicated that oxygenic photosynthesis evolved well before the atmosphere became oxidizing. They also wrote that, "the presence of steranes, particularly cholestane and its 28- to 30-carbon analogues, provides persuasive evidence for the existence of eukaryotes 500 million to 1 billion years before the extant fossil record indicates that the lineage arose."

The paper was heralded as a breakthrough and highly cited during the following decade. Brocks, however, discovered that some of the sediment samples that his team had used had been contaminated. In 2008, he coauthored a paper with different colleagues that essentially overturned the findings of the 1999 paper [2]. "The most important point is whether these biomarkers in 2.7 [billion year] old rocks are indeed that old," Brocks said. "After many years

of scientific dispute about them, others embraced the earlier findings and published follow up papers apparently vindicating the original 1999 results, but the community has come to the consensus that these hydrocarbons have entered the Archaean rocks at a later point in time" [3].

Simon George, leader of the organic geochemistry group in the Department of Earth and Planetary Sciences at Macquarie University in Sydney, Australia, argues that although the discovery of contamination was a setback for Brocks and others, it does not disprove the validity of all other findings based on samples of an apparently similar age. "Jochen Brocks's [...] inference is that everybody's work is based on contamination. He's certainly proven that some of the samples he worked on were affected by contamination, but it's a bit of a leap to say everyone else's is," George explained. He argued that findings of steranes in ancient samples have been repeated in different geographical locations and by a variety of people at several leading

institutions. "I'd be surprised if everyone was seeing contamination," he said.

Gordon Love, an organic geochemist at the University of California Riverside, CA, USA, is more cautious. He commented that findings based on archaean rocks are often unreliable because the levels of biomarkers are very low, which makes it harder to sift out contaminants. "The pursuit of Archean lipid biomarkers has always been viewed as a very extreme application of molecular organic geochemistry requiring the most sophisticated and sensitive instrumentation to detect any signals at all," he said. "We are talking about trace quantities of biomarkers that wouldn't normally adversely affect ancient biomarker studies or even show up in routine analyses since the absolute yields of these compounds are so low, but which become significant when dealing with highly overmature Archean organic matter [original matter that has been transformed by thermal and chemical processes into oil and gas]." Nevertheless, Love noted that the conclusion that eukaryotes evolved over 2.5 billion years ago might still be correct.

"We cannot say that the absence of steranes shows that eukaryotes had not evolved. The most appropriate conclusion, in my strong opinion, is that that organic matter found in Archean rocks has been so thermally transformed that we have no way of knowing whether eukaryotic biomarkers were ever present as original lipid constituents."

Answers might ultimately come from another promising line of research into archaean evolution that relies on the analysis of molecular fossils obtained from 'fluid inclusions' within sediment rocks. These are small microscopic bubbles of liquid and gas—typically 0.1–1.0 mm in diameter—trapped within crystals. Because they have been trapped since their formation, they are almost guaranteed to be free from contamination. The problem so far is that they have had to be analysed in bulk to provide enough fluid for separation and mass spectrometry. This requirement makes the work less reliable the further back you go because there is not enough fluid in single samples to date accurately anything that is older than 2.4 billion years.

George and colleagues are now applying the well-known technique of 'time-of-flight' mass spectrometry to analyse individual fluid inclusions without having to extract them. This technique was developed more than 60 years ago and has long been used in inorganic chemistry. It works by directing ion beams at a sample to identify molecules by their mass-to-charge ratio. "You use beams of ions to drill down into rock and then when you get to the required depth you put the analysis beam on," George explained. The principle is that the velocity of an ion depends on its mass-to-charge ratio, which can be calculated by measuring the time that it takes for the particle to reach a detector, thus identifying the particle. "It's a way of assessing very small samples," George said; adding that a lot more work on instrument development and the interpretation of results will be needed before we can say which biomarkers are present in fluid inclusion zones. "We think there are steranes in there, but it is hard for us to prove it," George said. "We can definitely see hydrocarbons in there and be sure it is really old material, but because we can't do gas chromatography separation on these we can't be sure [...] If we are able to prove fluid inclusions are holding this larger chemical record of life, it becomes a very important tool for

understanding life's origins, because we can go back as far as 3.2 billion years."

Love said that analysis of oil-bearing fluid inclusions, as practised by Simon George's group, is a promising approach because the effects of extreme thermal maturity on organic molecules are often not as acute for migrated petroleum fluids trapped in sandstones under high pressure as they are for rock bitumens found in the parent rocks. But he cautioned that the migration of these fluid inclusions away from the parent rock where they originated introduces some uncertainty over dating. "There will always be some degree of ambiguity concerning the age and stratigraphic position of the parent source rock that actually generated the oil trapped in the inclusion," Love explained. "At the same time, I look forward to seeing what they will generate from new, cleanly drilled Archean cores using the fluid inclusion approach."

Schouten likewise anticipates the results of this research and suggested that such work could finally settle the question of whether eukaryotes diverged from archaea and prokaryotes during the archaean period. "It is the work of Simon George on fluid inclusions which makes me think that we cannot fully dismiss that possibility yet," he said.

Modern genomes and proteins are also fossils in their own way, littered with evidence of genes that were useful to ancestral organisms but that have been abandoned or adapted in today's species. The comparative analysis of genomes and surface proteins, for example, could help scientists understand how viruses evolved before the three domains of life split. Viruses are believed to have been around at the time, having coevolved with early life, but their presence is impossible to establish directly because they do not even leave molecular fossils. As such, viral evolution can only be studied through observation of their comparative structures and genetic sequences across life's three domains. Only recently have virologists been able to look at viral origins and correlate them with the emergence of prokaryotes, eukaryotes and archaea.

Sarah Butcher, from the Institute of Biotechnology at the University of Helsinki, Finland, led an international team that made a significant breakthrough in March 2013, based on an archaeal virus found in a salt pan [4]. "The major insight in the paper was to show that the molecular architecture

Modern genomes and proteins are also fossils in their own way, littered with evidence of genes that were useful to ancestral organisms...

of an archaeal virus is conserved both with the most common bacterial viruses and a wide range of eukaryotic viruses in the herpes family," Butcher said. The study combined genomic analysis with electron microscopy and computerized image reconstruction to determine that the major coat protein of the isolated archaeal *Haloarcula sinaiensis* tailed virus 1 (HSTV-1) has an almost identical structure to that of the bacterial virus Hong Kong 97 (HK97), which is one of the so called head-tailed dsDNA bacteriophages. This similarity had been predicted, but the study provided the first physical proof, backed up by equally compelling genomic evidence, according to Butcher. The analysis revealed that hallmark proteins found in dsDNA bacteriophages, such as terminases and portals, are present in the HSTV-1 genome. Furthermore, the genomes themselves had common structural features.

Earlier work had already identified the same fold in herpes viruses, which infect a wide variety of animals, including humans [5]. As Butcher pointed out, the same lineage identified first in the HK97 bacteriophage has now been found in viruses from all three domains of life. Viruses have evolved unique mechanisms for infecting cells dependent on their hosts, but the capsid proteins seem to share common structural features. "The basic argument is that capsid proteins with the same fold share common ancestry, even when, as is often the case, they no longer share any detectable sequence similarity," commented Roger Hendrix, a viral evolution specialist at the University of Pittsburgh, PA, USA. He explained that the idea is that the adaptation to different host environments led to fundamental sequence changes in ancestral viruses as new functions were acquired, but that there was no corresponding selective pressure to alter the fold of the capsid. "The simple interpretation of this is that there were viruses with some resemblance to modern viruses before cells started to divide into the three current domains and some of the then-existing viruses stuck with and coevolved with each of the three emerging cellular domains," he said.

He cautioned, however, that another explanation for the common fold could be that at some stage in evolution, after the three cellular domains of life split, a virus jumped across the domains, spreading the common fold. This is not likely to have occurred recently because the cells of the different domains have diverged so far that such a viral jump would be difficult, though it could have occurred only shortly after life split into three domains and cannot be ruled out.

A third possible explanation for the common fold would be that it coevolved in parallel in each of the three domains, but this is less likely to have happened, according to Hendrix. "Co-evolution is a formal possibility but [...] unlikely since I think

it would imply that some protein folds are 'ideal' and selected in certain viruses of all three domains but not others," he said. "There are too many successful ways to make a viral capsid for this to make sense, at least to me."

At any rate, the evidence does indicate that diverse viruses were around at the time of the last common ancestor of the three domains of life and contributed to their evolution through their ability to mutate quickly and donate genes to host genomes. As such, the work being done to untangle viral evolution by studying modern viruses is very much related to the advances being made with molecular fossils to probe our way back towards the origins of life.

CONFLICT OF INTEREST

The author declares that he has no conflict of interest.

REFERENCES

1. Brocks JJ, Logan GA, Buick R, Summons RE (1999) *Science* **285**: 1033–1036
2. Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR (2008) *Nature* **455**: 1101–1104
3. Brocks JJ (2011) *Geochim Cosmochim Acta* **75**: 3196–3213
4. Pietilä MK, Laurinmäki P, Russell DA, Ko CC, Jacobs-Sera D, Hendrix RW, Bamford DH, Butcher SJ (2013) *Proc Natl Acad Sci USA* **110**: 10604–10609
5. Veesler D, Johnson JE (2012) *Annu Rev Biophys* **41**: 473–496

Philip Hunter is a freelance journalist in London, UK

EMBO reports (2013) **14**, 964–967; published online 15 October 2013; doi:10.1038/embor.2013.162