REVIEW

Terpenes, hormones and life: isoprene rule revisited

Stephen G Hillier¹ and Richard Lathe²

¹Medical Research Council Centre for Reproductive Health, University of Edinburgh, The Queen’s Medical Research Institute, Edinburgh, UK
²Division of Infection and Pathway Medicine, University of Edinburgh Medical School, Edinburgh, UK

Correspondence should be addressed to S G Hillier or R Lathe: steve.hillier@ed.ac.uk or richard.lathe@ed.ac.uk

Abstract

The year 2019 marks the 80th anniversary of the 1939 Nobel Prize in Chemistry awarded to Leopold Ruzicka (1887–1976) for work on higher terpene molecular structures, including the first chemical synthesis of male sex hormones. Arguably his crowning achievement was the ‘biogenetic isoprene rule’, which helped to unravel the complexities of terpenoid biosynthesis. The rule declares terpenoids to be enzymatically cyclized products of substrate alkene chains containing a characteristic number of linear, head-to-tail condensed, C₅ isoprene units. The number of repeat isoprene units dictates the type of terpene produced (i.e., 2, monoterpenes; 3, sesquiterpenes; 4, diterpenes, etc.). In the case of triterpenes, six C₅ isoprene units combine into C₃₀ squalene, which is cyclized into one of the signature carbon skeletons from which myriad downstream triterpenoid structures are derived, including sterols and steroids. Ruzicka also had a keen interest in the origin of life, but the pivotal role of terpenoids has generally been overshadowed by nucleobases, amino acids, and sugars. To redress the balance, we provide a historical and evolutionary perspective. We address the potential abiotic generation of isoprene, the crucial role that polyprene terpenoids played in early membranes and cellular life, and emphasize that endocrinology from microbes to plants and vertebrates is firmly grounded on Ruzicka’s pivotal insights into the structure and function of terpenes. A harmonizing feature is that all known lifeforms (including bacteria) biosynthesize triterpenoid substances that are essential for cellular membrane formation and function, from which signaling molecules such as steroid hormones and cognate receptors are likely to have evolved.

Introduction

‘The structural similarities of the higher terpenes raise the question as to whether these compounds may have been formed according to a uniform principle in nature’ (Ruzicka 1966)

Terpenes (including sterols, steroids, and related aromatic hydrocarbons) are present in all known life forms where they pivotally impact on individual and population survival (Summons et al. 2006, Nes 2011, Jiang et al. 2016).

Fossilized terpenes have been discovered in geological deposits billions of years old, signifying involvement in the very beginnings of life on Earth (Ourisson & Albrecht 1992, Melendez et al. 2013): involvement so fundamental that as ancient lipids they may represent (along with DNA, RNA, and protein according to the central dogma), a fourth molecular strand of terrestrial life.

The terpenome (the compendium of all known terpenoids) is so vast that it accounts for nearly one-third...
of all compounds currently characterized in the ‘Dictionary of Natural Products’ (http://dnp.chemnetbase.com) (Christianson 2017). Why so many structurally related compounds with such diverse functions should exist is yet to be explained. The clue seemingly lies in their core chemical structure: they are all composed of multiples of a 5-carbon unit called isoprene (2-methyl 1,3-butadiene, isopentene), one of the most common organic chemicals on Earth (Sharkey & Yeh 2001).

2019 marks the eightieth anniversary of the 1939 Nobel Prize in Chemistry awarded to Leopold Ruzicka (Ružička; the French transcription ‘Rougitchka’ may assist in pronunciation), a Croatian national who worked extensively in Switzerland, for research that established the importance of the ‘isoprene rule’ in the elucidation of terpene chemical structures, including the classification of cholesterol and sex steroids as triterpenoids. He also advanced a biochemical extension of the rule that became the basis of present day understanding of terpene biosynthesis, known as the ‘biogenetic isoprene rule’ (Ruzicka 1953, 1959).

The fundamental significance of the biogenetic isoprene rule to biology, immediately evident at the time, continues to become increasingly apparent. In the same way as terpene chemistry is built on a common molecular plan involving isoprene, so too is terpenoid biology.

Ruzicka became increasingly interested in the possible link between the biogenetic isoprene rule and the question of life’s origin (Eschenmoser 1990). The occasion of his 80th Nobel anniversary allows the opportunity to celebrate his seminal contributions and reappraise the connection between terpenoid biochemistry and the existence of life on Earth.

The isoprene rule

‘… the leading question was whether the carbon skeletons of the higher terpenes were also composed of isoprene units. For all the compounds we examined, the answer was positive, and thus the original working hypothesis gradually grew into the isoprene rule’ (Ruzicka 1959)

Isoprene, the prototypic terpene substance, is one of the most copiously produced volatile hydrocarbon chemicals on Earth (McGenity et al. 2018) owing to the global abundance of terpenoid biosynthesis, not vice versa. Approximately 40% of the biogenic volatile organic compounds emitted by plants are in the form of isoprene, and isoprene is the principal hydrocarbon identified in human breath (Gelmont et al. 1981). Cinema audiences exhale more isoprene when watching scenes of suspense, and isoprene levels spike in the air above football fans when goals are scored (Stonner & Williams 2016).

The centrality of isoprene to terpene chemistry cannot be overestimated (Nes 2011). First obtained from burning rubber by Michael Faraday (Faraday 1826), isopentene was isolated as a distillation product of natural rubber by Williams (1860) who correctly assigned it the empirical formula of C₉H₈ and named it isoprene. Tilden (1884) went on to distil isoprene from turpentine oil and showed that it could be dimerized into ‘dipentene’ (C₁₀H₁₆) by heat or treatment with sulfuric acid, portending modern discussions on the abiogenesis of isoprenoids (see below). Meanwhile, Kekulé coined the name terpen (English ‘terpene’) for the group of hydrocarbons obtained from turpentine oil with C:H ratios of C₃₀H₄₆ of which isoprene (dipentene) was one (Box 1). Remarkably, all this was accomplished before isoprene’s iconic molecular structure was formally established (Ipatiew & Wittorf 1897) (Fig. 1).

When Otto Wallach began his systematic studies of the terpenes, he observed that individual terpene structures contained multiples of a 5-carbon unit that allowed them to be classified according to the number of such units they contained (Wallach 1887). The canonical C₅ unit proved to be isoprene, which Wallach recognized as the core terpenoid structure. That terpenes might be represented as repeating isoprene units became known as the ‘isoprene rule’. In this scenario, C₁₀ monoterpenes contain two head-to-tail linked hemiterpene isoprene units, C₁₅ sesquiterpenes contain three C₅ isoprenes, etc. (Box 1). The isoprene rule was extended to accommodate C₂₀ diterpenes (4 × C₅ = C₂₀), C₃₀ triterpenes (6 × C₅ = C₃₀), and beyond (Ruzicka 1953, Eschenmoser 1990). It is noteworthy that Wallach’s rule was successfully applied a full decade before the molecular structure of isoprene had been fully identified (Ipatiew & Wittorf 1897) or its chemical synthesis unambiguously achieved (Euler 1897).

The isoprene rule languished until the 1920s when it was referenced during the structural determination of cholesterol (Robinson 1932, Wieland 1966) and shaped by Ruzicka into the ‘biogenetic isoprene rule’ (Ruzicka 1953, Ruzicka et al. 1953). Ruzicka’s version of the rule allowed terpenoid structures to be explained or predicted based on accepted reaction mechanisms involving acyclic precursors that were products of isoprene condensation, such as geraniol, farnesol, and geranylgeraniol (Ruzicka 1963). His Nobel lecture vividly illustrates the success of this approach, which led to the classification of cholesterol as a triterpenoid substance several years before any formal demonstration that the carbon atoms in its...
Box 1 Isoprenoids and terpenoids: history and nomenclature

‘Terpene’ refers to a 10-carbon dimer of ‘isoprene’ as well as to the generic class of chemical substances built of polymerized isoprene units. Although the IUPAC recommends that the terms ‘terpenoid’ and ‘isoprenoid’ be used to refer to chemically modified (generally oxygenated or methylated) forms of terpene/isoprenoid polymers, these terms are all used interchangeably in the field. Confusingly, the molecule isoprene is not itself the precursor for many terpenoids. Nevertheless, this nomenclature persists, and farnesol is regarded as a sesquiterpene (based on three isoprene units), whereas squalene, lanosterol, and steroids are triterpenoids (six isoprene units). Because of their sterically constrained structures and multiple chiral forms, these volatile linear and (poly)cyclic terpenoids (‘aromatics’) are well adapted as signaling molecules. Across the kingdom of life the terpene repertoire encompasses signaling molecules – both repellants/antimicrobials and attractants/pheromones – the latter being widely employed in perfumery. Ruzicka was keenly interested; he established that the fragrance of ambergris is based on the triterpene, ambrein (Ruzicka & Lardon 1944, Prelog & Jeger 1980). In the following we examine the origins of the key terms:

Terpenes: named in 1863 by August Kekulé from turpentine/terpine, the aromatic resin of the terebinth tree (Pistacia terebinthus) that grows widely around the Mediterranean – including Croatia, Ruzicka’s birthplace. The major component of terpine is a 10-carbon monoterpane (pinene). Tree resins protect the host tree against invasion by microbes and insects, and thus have potent medicinal properties – essential oils from terpine (also ‘tereminth’) were highly prized and the tree had early religious significance (Barton 1906). A depiction of the terpine leaf was a symbol in the Minoan writing system, and terpine resins were used as a preservative for wine as early as 5400–5000 BCE (McGovern 2003). Tereminthos may be an early Indo-European word denoting/akin to ‘overcomer of the forces of growth/death’ (Beckmann 2012), with roots in ter (= supra) and minth (cf. Greek minth, Latin menta, ‘mint’; Greek methu, ‘wine’, also Welsh medd, English ‘mead’, and ‘menthol’, a monoterpenoid).

Isoprenes: the 5-carbon isoprene molecule was (mis)named in 1860 by British chemist Charles H. Greville Williams as a supposed isomer of the propane/proplyl group of substances, whose name ultimately derives from Greek πίθον ‘fat’. Any chemical substance formally derived from isoprene is an isoprenoid.

Farnesol: the key sesquiterpene is named from the floral essence of the acacia tree (Vachellia farnesiana) that was brought to Europe from the Americas by Cardinal Farnese (1573–1626).

Squalene: the triterpenoid (hexa-isoprene) squalene was first characterized in shark liver oil (Squalidae, the shark family) in 1906 by the Japanese scientist Misumaru Tsujimoto (Popa et al. 2015). Because squalene is lighter than water, cartilaginous fish (which lack a swim bladder) such as sharks reduce their body density with such fats/oils. However, squalene is widespread in plants and animals including humans, where it is secreted by the liver, carried in the blood by LDL and VLDL, and secreted in large quantities from sebaceous glands where it may exert antimicrobial action (Popa et al. 2015).

Lanosterol: a triterpenoid formed by cyclization of squalene, the precursor to steroids, and a component of lanolin (from Latin lāna ‘wool’), the oily water-repellent secretion of sebaceous glands of sheep and other wool-producing animals. Derivatives have antimicrobial properties. Of the primary counting system based on carbon numbers (meth-, eth-, prop-, but-; thereafter classical pent-, hex-, etc.), with the exception of ‘eth’- (from aether, named in 1834/1835 by Justus Liebig and Jacob Berzelius); the others derive from ‘wine’, ‘fat’, and ‘butter’, respectively, with pride of place going to ‘meth’-, denoting wine, tereminth, and thus terpene.

Figure 1
The isoprene unit. (top left) Original depiction of the structure of isoprene (data from Ipatiew & Wittorf (1897)), alongside (top right) a contemporary rendition of the isoprene unit, with numbered carbon atoms, which is (bottom right) the basic building block for polymeric isoprenoid synthesis through linear head-to-tail (C1–C4) condensation of individual isoprene units. (bottom left) Diagram based on a sketch from Ruzicka’s personal notes showing the tetracyclic structure of lanosterol broken (red dashed lines) into constituent 5-carbon isoprene units. The C1–C4 isoprene unit at the boundary of rings B and C is highlighted in blue (data from Eschenmoser (1990)).
as the founder C₅ unit upon which all terpenoid biosynthesis depends (Lynen et al. 1958). The tortuous route to unraveling these key pathways is aptly summarized by Bloch (1992).

All natural triterpenoid substances could now be understood as end-products of a time-honored biogenetic sequence beginning with the formation of IPP – the active C₅ unit in terpenoid biosynthesis. Linear, head-to-tail, coupling of IPP with its isomer, dimethylallyl pyrophosphate (DMAPP), and a secondary molecule of IPP leads to C₁₅ FPP via C₁₀ geranyl pyrophosphate (GPP) (Fig. 2). Reductive, tail-to-tail dimerization of two FPPs leads to formation of C₃₀ squalene (Popjak et al. 1969). Importantly, the biochemical route from IPP to squalene is shared, to a greater or lesser degree, by all domains of life, consistent with the ubiquity of terrestrial terpene biology.

Beyond squalene

‘... a scheme has been developed leading from squalene to the formulas of the basic representatives of all known cyclic triterpene groups ... this result is considered to support the squalene hypothesis of the biogenesis of cyclic triterpenes’ (Eschenmoser et al. 1955)

The isoprene rule morphed into the ‘squalene rule’ when squalene proved to be an essential intermediate in the biosynthesis of cholesterol and the carbon skeletons of all other then known cyclic triterpenes (Eschenmoser et al. 1955). The paper positing cyclization of squalene as the axis of triterpenoid biogenesis (Woodward & Bloch 1953) was published in the same year that Watson and Crick disclosed the structure of DNA (Watson & Crick 1953). Both discoveries had immediate impact and are still absolutely relevant to our understanding of the biochemical basis of life on Earth. Fittingly, all four authors duly received Nobel prizes (Watson and Crick shared the Physiology or Medicine prize with Maurice Wilkins in 1962; Bloch shared the 1964 Physiology or Medicine prize with Feodor Lynen; and Woodward was awarded the full Chemistry prize in 1965).

Squalene cyclization proceeds as an electrophilic reaction cascade catalyzed by phylum-specific terpene cyclase enzymes – whereby the C₃₀ polyalkene chain is effectively rolled into a polycyclic structure comprising up to five interconnected carbon rings and a residual side-chain of variable length. Depending on biological context, squalene cyclization occurs with absolute regio- and stereospecificity to produce the signature multicyclic carbon skeletons from which myriad downstream terpenoid structures are derived (Xu et al. 2004, Nes 2011, Jia & Peters 2017).

Figure 2
Simplified scheme of triterpenoid formation by stepwise condensation of C₅ isoprene units (IPP, DMAPP) via C₁₀ GPP and C₁₅ FPP to form C₃₀ squalene (SQ). Squalene is then cyclized by terpene/squalene cyclase enzymes into the multicyclic triterpenoid (carbon skeletons shown) substrates from which countless downstream secondary metabolites are formed, as discussed in the text. DMAPP, dimethylallyl pyrophosphate (isomer of isoprene); FPP, farnesyl pyrophosphate; GPP, geranyl pyrophosphate; IPP, isopentenyl pyrophosphate.
In bacteria, squalene cyclization is catalyzed by squalene cyclase and does not require prior epoxidation (Summons et al. 2006) (Fig. 3). By contrast, in eukaryotes squalene cyclization catalyzed by oxidosqualene cyclase is preceded by 2,3-epoxidation. This is the first oxygen-dependent step in steroid formation and provides the point at which steroid biosynthesis in metazoans diverges from hopanoid biosynthesis in bacteria.

Ruzicka clearly saw the broader implications of the ‘squalene hypothesis’ in that it logically explained how such a constellation of multicyclic terpenoid structures could be derived from a single precursor molecule (Fig. 4) (Ruzicka 1959).

**Polycyclic terpenoids**

‘After Kekulé, in 1865, first introduced the carbon ring into structural chemistry in his formula for benzene, the six-membered ring maintained its unique position in the taxonomy of organic chemistry for several decades’ (Ruzicka 1966)

The molecular mechanism of squalene cyclization to lanosterol, the primary cholesterol precursor, was established within a decade of Ruzicka’s 1945 Nobel lecture (Woodward & Bloch 1953, Eschenmoser et al. 1955, Eschenmoser & Arigoni 2005) (Fig. 5). Lanosterol is now recognized as the major terpenoid precursor for fungal as well as animal sterols and steroids, whereas cycloartenol gives rise to β-sitosterol and downstream sterol and steroidal metabolites in plants with a photosynthetic lineage (Nes 2011).

Extant animal, plant, and fungal clades have terpenoid signatures corresponding to individual needs for survival within particular ecosystems. Of the estimated >150 multicyclic carbon skeletons known or hypothesized to serve as triterpenoid precursors, only one (lanosterol cation) is the source of cholesterol and true steroids. The rest are the uniquely adapted sources of the colors, scents, poisons, potions, rubbers, and waxes that are, worldwide, the components of the terpenome (Xu et al. 2004, Jiang et al. 2016)

Terpenoids based on the tetracyclic 6-6-6-5 lanostane carbon skeleton form a subsection of the terpenome known as the sterolome (Nes 2011). The sterolome is estimated to comprise at least 1000 natural products derived from lanosterol and related molecules that carry out essential biological functions across all domains of life on Earth. Cholesterol is the parent animal sterol,
while cycloartenol and ergosterol are major plant and fungal equivalents. Chemically, they differ mainly in their side-chain substitution and saturation, as well as in their degree of esterification into glucuronides, glycosides, sulfates, etc. In animals cholesterol is famously converted to bile acids, vitamin D, and true steroids. The major plant sterols are ethylated (β-sitosterol, stigmasterol) or methylated (cycloartenol) at C24 of the side chain, and the balance between ethylation and methylation is specific for individual plant species (Valitova et al. 2016). Plants also metabolize cholesterol into brassinosteroids (plant steroids), glycoalkaloids, cardenolides (poisons), saponins (vegetable soaps), and withanolides (plant defense substances). Present-day insects do not biosynthesize cholesterol de novo but metabolize dietary molecules such as cholesterol into ecdysteroids (moulting hormones). Nematodes convert cholesterol to worm-specific pheromonal sterols (dafachronic acids) (Aguilaniu et al. 2016). Fungal steroids produced from ergosterol include antheriodiol (female sex hormone) and oogoniol (male sex hormone) (Nes 2011).

Thus, although steroids can be regarded as specialized higher terpenoids that fulfill multiple functions vital to metazoan life, the entire sterolome occupies only a miniscule corner of global terpenoid biochemistry.

Special mention must be made of the pentacyclic 6-6-6-6-5 hopanoids. First identified as a terpenoid component of resin extracted from plant genus *Hopea*, they exist in diverse bacteria and some lower plant forms such as algae and lichens, where they function as membrane lipids (Saenz et al. 2015, Belin et al. 2018). The structural and functional similarity of bacterial hopanes to eukaryotic cholestanes renders them of particular interest in interpreting the molecular beginnings of life, not least because geological evidence suggests that hopanoid sterols were (and perhaps still are) among the most abundant natural products on Earth (Ourisson & Albrecht 1992).

**Ancient organics**

Geologically durable isoprenoids have existed on Earth since the dawn of life. Hydrocarbons assembled from repeating isoprene units are ubiquitous in ancient sediments. Biomarker evidence for eukaryotes comes from terpenoid steranes with diagnostic alkylation patterns in Barney Creek Formation rocks 1.64 billion years (Ga)
before present (Summons et al. 1988, reviewed in Brocks & Summons 2003). Steroid residues have been found in rocks dated to ~2.7 Ga (Summons et al. 2006) and cholesaines have been identified in Ediacaran (protoist) fossils, which are the oldest confirmed macroscopic animals in the rock record (Bobrovskiy et al. 2018). The hopane carbon skeleton is particularly ubiquitous in oil shales and petroleum deposits in sedimentary rocks. Above all else, the presence of hopane residues in geological deposits laid down before the advent of atmospheric oxygen (see below) implies a fundamental contribution of sterols to the earliest stages of the evolution of terrestrial life. However, as discussed in the next section, the origins of prebiotic terpenoids remain enigmatic.

Isoprenoid abiogenesis

‘… the level of perfection achieved by organic chemistry … is enabling biochemistry to penetrate the innermost secrets of life processes on a molecular basis’ [attributed to Ruzicka] (Eschenmoser 1990)

Simple precursor molecules in the primeval soup fueled the origin of life (Haldane 1929, Oparin 1953); these can be generated from a prebiotic milieu containing only H$_2$O, CH$_4$, NH$_3$, and H$_2$ (Miller & Urey 1959). Sugars are polymerization products of HCHO, amino acids are generated by condensation of HCN, HCHO, and NH$_3$, and bases are polymers of HCN/NH$_3$ and/or formamide. Reactive phosphorus moieties were no doubt abundant (Yamagata et al. 1991, Schwartz 2006, Pasek et al. 2013). Thus, precursors to the first three strands of life according to the central dogma (DNA makes RNA makes protein) were available for molecular tinkering.

Much less attention has been paid to the ‘fourth strand’ – lipids – that are essential for the generation of the first micelles and coacervates, the inferred precursors to cellular life (Segré & Lancet 2000). Indeed, the ‘lipid world’ undoubtedly accompanied, or even preceded (Ouirisson & Nakatani 1994, Segré et al. 2001), the emergence of nucleic acid-based life.

Abiotic synthesis of straight-chain hydrocarbons (up to pentane) in Urey–Miller experiments has been confirmed, and extraterrestrial pentacarbon molecules have been widely detected (McGuire 2018). On Titan, the major moon of Saturn, lakes of methane and ethane are present, but these are not the final end-products, and complex photochemical conversions generate diverse hydrocarbons at high altitudes in the Titan stratosphere (Wilson & Atreya 2004, Waite et al. 2007). The terpenoids pristane and phytane have been reported in samples of meteorites falling to Earth (Oro et al. 1966), and Cronin & Pizzarello (1990) found C$_{15}$ to C$_{30}$ branched alkyl-substituted mono-, di-, and tricyclic alkanes in the Murchison meteorite that fell in Australia in 1969 (Cronin & Pizzarello 1990). While it has not been formally possible to exclude terrestrial contamination (reviewed in Sephton 2002), the scene is set for long-chain hydrocarbons, both saturated and unsaturated – although the inferred preponderance of unsaturated C$_5$ isoprene precursors remains unexplained.

Following Ruzicka, isoprene was a likely prebiotic precursor for polymerization into linear polymers (Lazzcano et al. 1983), but the abiotic origin of isoprene remains uncertain. Because isoprene is normally a gas (boiling point=34.1°C) and is intensely insoluble in water, early Miller–Urey-type experiments may have overlooked this important avenue. In his sketches, Ruzicka (reproduced in Eschenmoser 1990, p 9) outlined the formation of 2-butenal (CH$_2$=CH=CH–CHO, a potential isoprene precursor) from elementary components, and no doubt intended to take this further. Ourisson and colleagues (Nakatani et al. 2012) suggested that isoprene could have been formed from isobutene, ethylene, and formaldehyde at high temperatures (with further product being generated by condensation on clays), noting that both isoprene and HCHO are components of volcanic gases (Dong et al. 1994). Other potential routes involve acetic acid and acetylene or deamination and decarboxylation of leucine, an amino acid detected in Miller–Urey experiments in the presence of H$_2$S (Parker et al. 2011). However, none of these routes so far resoundingly explains the inferred preponderance of isoprene moieties in the prebiotic chemosphere. Given the vital importance of isoprenes to the emergence of cellular life, resolving the puzzle of isoprene abiogenesis and its inferred pivotal role remains a priority.

Differential solubility provides an insight. Polyisoprene synthesis requires activation of isoprene by (pyro)phosphorylation (e.g., through the generation of isopentenyl pyrophosphate, IPP, and its isomer, dimethylallyl pyrophosphate, DMAPP). Reactive polyphosphates are generously emitted by volcanic activity (Yamagata et al. 1991) and could react in the gas phase with C=C and C≡C hydrocarbons. Crucially, (unlike isoprene) pyrophosphorylated derivatives of isoprene (i.e., IPP and DMAPP) are relatively soluble in water (7 and 25 g/L, respectively). Dissolution would also have afforded protection from photochemical degradation. This combination of factors could have led to the
accumulation of isoprene (pyro)phosphates in prebiotic oceans, paving the way to later exploitation at the origin of life.

Given (pyro)phosphate derivatives of isoprene, the generation of long-chain isoprenoids is chemically unchallenging, but the cyclization of squalene to lanosterol and related molecules also represents an enigma because of the multiple different boat/chair configurations of the ring systems, whereas only a specific subset (or small number of subsets) is seen in extant molecules of this class. In a cyclic dry-down and dilute scenario (Lathe 2005), it is possible that clays could have fostered specific configurations (Ouirsson & Nakatani 1994), and templated cyclization mediated by product (noting that lanosterol readily forms crystals; Liu & Sawant 2002) is a further possibility.

Life and the great oxygenation event (GOE)

'Somewhere in between (bio)chemical and biological evolution we must assume a point where life was created' [attributed to Ruzicka] (Eschenmoser 1990)

At the origin of life the planetary atmosphere contained little if any oxygen. Any proto-lifeform must have been independent of oxygen, with its lipid biochemistry adjusted accordingly. By contrast, the synthesis of ‘modern’ long-chain fatty acids (LCFAs) and sterols (the major components of cellular membranes in all organisms except Archaea) requires oxygen for synthesis. This has important implications for the biosynthesis of prebiotic/co-biotic polymers.

Free \( \text{O}_2 \) only became available with the advent of oxygenc photosynthesis by primitive cyanobacteria-like organisms, leading to the onset of the GOE, where \( \text{O}_2 \) was an incidental (and toxic) byproduct of sequestration of carbon from \( \text{CO}_2 \) (Fig. 6). The timing of this transition has been accurately dated, based on isotope ratios and paleomagnetic studies, to between 2.46 and 2.43 Ga (Lyons et al. 2014, Gumsley et al. 2017).

Low levels of free \( \text{O}_2 \) may have been available a little earlier (e.g., at 2.5–2.95 Ga; Anbar et al. 2007, Planavsky et al. 2014), but solubility is a further factor constraining \( \text{O}_2 \) availability. Following the formation of the Earth–Moon system at around 5.5 Ga, the first rains fell to give oceans with a temperature of \( \sim 100^\circ \text{C} \), and the mean temperature declined roughly linearly from that time until today (Sleep 2010, Garcia et al. 2017, but see Pope et al. 2012) pointing to surprisingly high temperatures of \( \sim 90^\circ \text{C} \) at the origin of life (ca. 3.9 Ga, or possibly a little earlier; Battistuzzi et al. 2004; Fig. 6), where \( \text{O}_2 \) was many-fold less soluble in water, further accentuating oxygen limitation.

Thus, life is presumed to have emerged, in the near-total absence of oxygen, through extreme thermophiles most closely related to the Archaea (Gribaldo & Brochier-Armanet 2006, Eme et al. 2017). Although the exact relationship between the Archaea, Bacteria, and Eukaryota remains contentious, for simplicity we retain the Archaea/Bacteria distinction and follow the idea that Archaea preceded both Bacteria and Eukaryota, with the

![Figure 6: Terpenoids and the diversification of life forms. S indicates a period of possible symbiosis between Archaea and Bacteria that may have preceded the generation of the first eukaryote generated by incorporation of an alphaproteobacterium within an archaeal host cell. Ga, billion years; GOE, great oxidation event.](https://joe.bioscientifica.com/doi/10.1530/JOE-19-0084)
unusual membrane composition of Archaea (see below) representing an obstacle to alternative phylogenies (Eme et al. 2017).

Centrally, polyisoprenes can be synthesized in the absence of oxygen. It has been argued that terpenes (and not LCFAs) were the essential components of primitive membranes (Ourisson 1989, Ourisson & Nakatani 1994, Nakatani et al. 2012, 2014), as well as those of present-day Archaea (Langworthy et al. 1982). Indeed, polypropyl phosphates spontaneously form membrane vesicles (Nakatani et al. 2014). That these molecules are the functional equivalents of LCFAs was elegantly demonstrated by the engineering of E. coli whose membranes contain up to 30% of archaeal lipids – these bacteria are viable, and under some conditions the hybrid membrane even confers a growth advantage (Caforio et al. 2018).

As with present-day membrane lipids, these early polyisoprene units are presumed to have been linked to glycerol. Glycerol phosphates may have been abundant at the origin of life (Pasek et al. 2013); indeed, the triose glycerol may even have preceded the pentoses ribose/deoxyribose in nucleic acids (Joyce et al. 1987). Importantly, however, the glycerol linkage of membrane terpenoids differs significantly from that of present-day LCFAs. Membrane polyisoprene units in early life (as in present-day Archaea) are inferred to have been linked to glycerol via ether bonds [R–O–R] (Langworthy et al. 1982, de Rosa et al. 1986, Sprott 1992), contrasting with the ester bonds [R–(C=O)–R] that typify modern membrane LCFAs diacylglycerols, thereby further reducing the requirement for oxygen (Fig. 7).

**Sterol–hopanoid homology: membrane stabilization**

‘It has taken all my life’s work to convince myself that life is chemistry; and now you come along and tell us it is physics’ [Ruzicka’s reaction to Manfred Eigen’s theory of evolution based on mathematical principles, as recounted by Eschenmoser] (Eschenmoser 1990)

Terpenoid-based membranes are compatible with life, but terpene-only membranes tend to be unstable. Cholesterol, by contrast, is a molecule beautifully crafted by evolution to stabilize the membrane bilayer. However, cholesterol synthesis requires multiple oxygen-dependent events (Bae et al. 1999) (Fig. 3), and cholesterol synthesis could therefore not have taken place before GOE (Summons et al. 2006; discussed in Galea & Brown 2009). Instead, in the absence of oxygen, squalene and other long-chain terpenes may have contributed to membrane structure – squalene can partially rigidify membranes (Spanova et al. 2012, Gilmore et al. 2013) and may have played a similar role in the earliest life forms.

Cyclization of squalene generates lanosterol and related molecules (Fig. 5), such as bacterial hopanoids, that surpass squalene in their ability to rigidify membranes. All these molecules contain the 4-dimethyl motif of bacterial hopanoids and, given their hydrophobic nature and rigid planar structure, readily intercalate into lipid membranes where they exert significant stabilizing effects as functional equivalents of cholesterol (Rohmer et al. 1979, reviewed by Ourisson et al. 1987).

Although squalene cyclization in eukaryotes can be precipitated by the addition of a single oxygen atom (Fig. 3), oxygen-independent enzymatic cyclization of squalene could have taken place in bacteria via the squalene-hopene cyclase enzyme (Reinert et al. 2004; discussed in Poralla 2004), the antecedent to prokaryotic and eukaryotic sterol synthases (Pearson et al. 2003). Thus, early bacterial and proto-eukaryotic membranes are likely to have contained, in addition to long-chain polypropene lipids, sterols related to lanosterol/hopanoids, whereas the emergence of modern-day cholesterols (and steroids) awaited the advent of plentiful atmospheric oxygen.
Terpenoids and the origins of steroid signaling

‘The ability of so many living organisms to biosynthesize all these compounds with always the same configuration, out of hundreds of alternatives, must surely be one of the most remarkable chemical achievements of Nature’ (Ruzicka 1956)

There is consensus that eukaryotes evolved via intracellular symbiosis (see above) probably between an Archaea (the host) and a bacterium (the endosymbiote), although Archaea and Bacteria may have lived in close association for a protracted period – only later culminating in formal endosymbiosis (reviewed by Gribaldo & Brochier-Armanet 2006, Martin et al. 2015) (Fig. 6). Given that bacteria are likely to have been capable of synthesizing lanosterol (a better membrane-stabilizer than squalene), whereas the Archaea lack the necessary enzymes (Desmond & Gribaldo 2009), this might have provided another driving force for symbiosis leading to the generation of eukaryotes (the first geological fossils of eukaryotes date to somewhere between 1.6 Ga and 2.1–2.3 Ga (Butterfield 2015)). In this conceptual scenario, one partner might have provided a soluble precursor (e.g., farnesyl pyrophosphate, FPP) and received lanosterol or similar molecule in return. This conjecture – that a triterpene was the currency of exchange – has implications for the evolution of nuclear receptors (NRs) and steroid signaling.

In higher eukaryotes, derivatives of terpenoid sterols and steroids orchestrate multiple aspects of growth, development, and reproduction by targeting intracellular NRs, but the origin of steroid signaling remains open to debate. The ancestral ligand for the first NRs may have been a terpenoid or long-chain fatty acid (Moore 1990, Bridgham et al. 2010). Studies of the sponge Amphimedon queenslandia that contains only two NR polypeptides revealed that both NR1 and NR2 bind long-chain fatty acids (Bridgham et al. 2010), but terpenoids were not investigated.

Were farnesoids the forerunners to steroids? In multiple species the sesquiterpene farnesol and its derivatives are the key signaling molecules, not steroids. The key quorum-sensing molecules in the yeast Candida albicans (Polke et al. 2018) and the juvenile hormones of insects and crustacea (Qu et al. 2018) are all farnesoids. In plants the cyclic farnesoid, abscisic acid, is a crucial phytohormone (Cutler et al. 2010), and the same molecule has also been implicated in signaling in vertebrates (Bassaganya-Riera et al. 2011).

Intriguingly, the terpenoid FPP (but not farnesol) can activate multiple present-day steroid NRs, and it was suggested that this might reflect a common structural feature that was present in an ancestral NR (Das et al. 2007) – with the inference that FPP was the primeval ligand for this class of receptors. In support, activation of diverse NRs by FPP has been confirmed (Goyanka et al. 2010), and FPP is to this day an important human metabolite that regulates oxidative stress, in part by acting through the glucocorticoid receptor (Pastar et al. 2016).

There are, moreover, intriguing sequence and structural similarities between extant squalene synthases (SQSs) – that assemble two molecules of FPP into a single squalene molecule – and the ligand-binding domains of NRs (R Lathe & SG Hillier unpublished observations). Perhaps SQS adopted an early signaling role in response to binding of its substrate, FPP? – pointing to terpenoid forerunners of NR signaling in the bacterial symbionts of the earliest eukaryotes. Further research to address this intriguing possibility is certainly warranted, and if confirmed this would cast light on the later emergence of steroid signaling in higher eukaryotes.

Conclusions

‘Attempts may be made to interpret the isoprene rule, not only as a working hypothesis in the laboratory, but also as a structural principle employed by nature’ (Ruzicka 1966)

The isoprene rule was successfully used to elucidate the chemical structures of some of the most important natural substances on Earth. Ruzicka’s special contribution was to recognize the power of the rule to explain and predict complex polycyclic chemical structures. Crucially, he saw how the isoprene rule might provide a unifying principle for resolving the multiple mysteries of triterpenoid biochemistry. In the same way as Kekulé peered into the fire and saw the licking flames shape a six-membered benzene ring, Ruzicka’s musings on aromatic chemistry led to the multimembered alicyclic picture of the terpenoid world.

Ruzicka operated in an era when organic chemistry was informing the emerging discipline of bio(logical) chemistry. He is quoted by his longstanding colleague Albert Eschenmoser (Eschenmoser 1990) as saying ‘To understand biochemistry you need to know at least as much organic chemistry as for organic chemistry itself’. The extrapolation from acid-catalyzed polyene cyclization
Isoprene rule revisited


(encountered in his early days as a perfume chemist) to enzymatically catalyzed squalene cyclization in natural triterpenoid biogenesis illustrates this truism.

He was particularly fascinated by the chemical origins of life on Earth and saw a potential role for isoprenoid chemistry in its understanding. His prescience is underscored by the evidence, which continues to grow, that unicellular and metazoan cell membranes share structural and functional similarities based on terpenoid biochemistry, pointing to lipids as a fourth strand in the evolution of life alongside nucleobases, sugars, and proteins.

Ruzicka’s stereochemical vision was famously matched by his prowess as a synthetic chemist. Having classified cholesterol as a triterpenoid he entertained the possibility that sex hormones ‘of the oestrane and androstan type’ might be molecules (later known as steroids) in which the side-chain of cholesterol had been split off. He formally proved this to be the case, and thereby achieved the first synthesis of a sex hormone: androsterone. This in turn led to his synthesis of the principal male sex steroid testosterone, for which he shared the 1939 Nobel Prize in Chemistry with Adolf Butenandt.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding
This work did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

Author contribution statement
Both co-authors contributed equally to the manuscript.

Acknowledgments
The authors gratefully acknowledge Professors Albert Eschenmoser (ETH Zurich Laboratorium für Organische Chemie), Yoichi Nakatani (Université de Strasbourg), and J Ian Mason (University of Edinburgh) for their helpful comments during preparation of the manuscript.

References


Amur BH, Rilling H & Bloch K 1957 The enzymatic conversion of mevalonic acid to squalene. Journal of the American Chemical Society 79 2646–2647. (https://doi.org/10.1021/ja01567a077)


© 2019 Society for Endocrinology
Published by Bioscientifica Ltd.
Printed in Great Britain

https://doi.org/10.1530/JOE-19-0084

Downloaded from Bioscientifica.com at 08/04/2019 07:08:23AM
via free access

© 2019 Society for Endocrinology
Published by Bioscientifica Ltd.
Printed in Great Britain

https://doi.org/10.1530/JOE-19-0084

Downloaded from Bioscientifica.com at 08/04/2019 07:08:23AM
via free access
Isoprene rule revisited

S-rich spark discharge experiment.

Robinson R 1932 Constitution of cholesterol.


Acc. Chem. Res. 41, 113-123.

Experientia 49, 451-457.


PNAS 108, 5526-5531.

J. Phys. Chem. 102, 2542-2463.


Biochimie 89, 1203-1208.


Arch. Biochem. Biophys. 489, 120-121.

Arch. Biochem. Biophys. 225, 90-94.


Nature 486, 1200-1203.


Cell 141, 225-236.

Am. J. Bot. 94, 1275-1285.


Nature 486, 1200-1203.


Nature 486, 1200-1203.


Nature 486, 1200-1203.


Nature 486, 1200-1203.


Nature 486, 1200-1203.


Nature 486, 1200-1203.


Received in final form 18 April 2019
Accepted 3 May 2019
Accepted Preprint published online 3 May 2019