Naturwissenschaften (2012) 99:167–176
DOI 10.1007/s00114-012-0892-6

CONCEPTS & SYNTHESIS

On dating stages in prebiotic chemical evolution

Robert P. Bywater

Received: 2 December 2011 / Revised: 28 January 2012 / Accepted: 30 January 2012 / Published online: 15 February 2012

Abstract The notion that RNA must have had a unique and decisive role in the development of life needs hardly be questioned. However, the chemical complexity and other properties of RNA, such as high solubility in water and vulnerability to degradation, make it improbable that RNA could have had an early presence in the development of life on Earth or on any comparable telluric planet. Rather, the task of origin of life research must surely be to identify those chemical processes which could have taken place on Earth that could accumulate the complexity and rich molecular information content needed to sustain primitive life, and ultimately give rise to RNA. A collection of likely chemical precursors to modern biomolecules is listed here together with calculations of their molecular complexity. These complexity scores are then used to propose an ordering, on a timescale, of when they might have appeared on Earth. These pre-RNA living systems would have flourished during the first ~0.3 Gyrs after the start of the Archaean era (~4.2 Gyr ago). If there ever was an "RNA-world" it could have started after that initial period (~3.9 Gyrs ago), later to be complemented with the appearance of duplex DNA at about ~3.6 Gyrs ago, some time before the earliest known stromatolites (~3.4 Gyr).

Keywords Origin of life · Chemical evolution · Prebiotic chemistry · Lipid bilayer · Lipophilic peptides · Chemoinformatics · Shannon entropy · RNA world

R. P. Bywater (✉)
Magdalen College,
High Street,
Oxford OX1 4AU England, UK
e-mail: romeobravo@adelard.org.uk

Introduction

Despite the wide repertoire of chemical and biological properties of RNA, which make it such an appealing contender for being the first type of molecular species to usher in life onto this planet, there is no explanation for how such a complex chemical species could have arisen in the absence of sophisticated chemical machinery (Luisi et al. 1999). In a previously published origin of life (OoL) model (Bywater and Conde-Frieboes 2005; Bywater 2009) the types of chemistry that led to the emergence of the earliest biomolecular species were discussed. A critical requirement for this to happen would have been the accumulation of organic compounds with increasing complexity—the "bottom up" approach (Luisi et al. 1999), and the development of primitive cells (Luisi et al. 1999; Bywater and Conde-Frieboes 2005; Donoghue and Antcliffe 2010; Deamer and Weber 2010), "protocells", lined by a lipid bilayer that can concentrate and protect these compounds inside the protocell, and even more importantly, within the lipid bilayer itself (Bywater 2009). The generation of ever more complex chemicals required many millions of cycles of synthesis, partial degradation, concentration, selection and re-annealing in combinatorially new ways such that sufficiently diverse species could be produced and reproduced, from which particularly suitable entities survived (Follmann and Brownson 2009). A model based upon repeated ebb and flow of tides on beaches made of finely divided minerals with catalytic properties was proposed (Bywater and Conde-Frieboes 2005). It was later postulated (Bywater 2009) that the very first "information-rich" molecule types were most likely lipophilic peptides which are suitable for being embedded in the protocell membranes. Many of the functions necessary for life could evolve under these conditions leading to the notion that these membranes could have acted as

the "cradle of life". There is a precedent for the notion that oligopeptides can self-replicate (Bujdak and Rode 2004; Fitz et al. 2007; Rode 1999) and the membrane provides the ideal environment for these oligpeptides to self-assemble and begin to evolve some of the key functions that a primitive cells needs.

All of these predated the appearance of RNA by hundreds of millions of years, and in this paper, a dating scheme for the sequential elaboration of ever more complex chemical entities is proposed. This dating scheme is based on calculations of molecular complexity and enthalphy of formation of relevant molecular species, as described in the next section.

Methods

Molecular complexity, C, is here quantified in terms of a scheme devised originally by Bertz (1983) and is defined in terms of a skeletal complexity term C_η and an atom-type diversity term C_E :

$$C = C_\eta + C_E$$

(where lg stands for binary logarithm) and

$$C_\eta = 2\eta \lg \eta - \Sigma_i \eta_i \lg \eta_i$$

with η, a connectivity score for heavy (=non-hydrogen) atoms related inversely to the number of hydrogens on the atom, and η_i, a term designed to reduce the complexity measure based on the number of similar substructures in the molecule and the number of multiple bonds, and

$$C_E = E \lg E - \Sigma_j E_j \lg E_j$$

with E the total number of non-hydrogen atoms and E_j being the number of atoms of type j.

A simpler measure of complexity based on number of chiral centres and number of double bonds is also listed in each table along with the Shannon entropy-based complexity score.

Enthalpy of formation was calculated using a quantum mechanical procedure implemented in the modelling program YASARA (Krieger et al. 2002). The PM3 Hamiltonian (Stewart 1989) was selected for these calculations since it gives more accurate results than other semiempirical methods (Young 2001)).

Prerequisites for life

While there does not seem to be any incontrovertible definition as to what constitutes "life", the following list of essential requirements are deemed to be necessary. (1) Catalytic activity, to accelerate synthetic reactions for needed products and in order to compensate for degradation processes. (2) Energy supply and management. (3) Storage system for synthesized products. (4) Controlled material transport across cell membranes. (5) Communication between cell interior and the outside world. (6) Molecular recognition and selectivity/affinity. (7) Replication of molecular species and ultimately of entire cells.

These are in turn catered for, in the prebiotic world, in the following way. (1) Catalytic material abounds in the mineral world, and while these catalysts do not possess the efficiency and selectivity of enzymes or ribozymes, they would be sufficient, during the eons of time that we are considering, for catalyzing the synthetic pathways to the earliest prebiotic chemicals (see Tables 1 and 2). It is known that primitive metal ion catalysts were abundant on the early Earth (Nitschke and Russell 2010). In our model, the first key synthetic steps were purported to take place on tidal beaches where there are myriads of catalytic particles (Bywater and Conde-Frieboes 2005). We are of course aware of other hypotheses whereby libraries of potential prebiotic compounds could have been produced by deep-sea vents (Nitschke and Russell 2010; Martin and Russell 2007; Sleep et al. 2011) or delivered from space (Gallori 2011). Both would have access to adequate energy sources and catalytic functions but both also lack key functions such as concentration, cyclical segregation, renewed mixing and combinatorial synthesis that the beach model provides. (2) Energy supplies were abundant—solar, geothermal, cosmic and terrestrial radiation, etc. In particular, the conflict between the thermodynamical imperatives, which require a drive towards increased entropy, and the need for living systems for the accumulation of order and reduced entropy must be considered. Energy must be consumed in order to favour the latter. (3) The importance of lipids has been stressed by several authors (Luisi et al. 1999; Donoghue and Antcliffe 2010; Deamer and Weber 2010) and we have also made a case for this. The protocells referred to above played a key role in all of this. Further, it is not only the space enclosed by the lipid bilayer that is important for storage but, even more importantly, the lipid bilayer itself, especially since many of the early biochemicals would have been lipophilic in nature (see also Tables 1 and 2 and discussion below). (4) The lipid bilayer presents a barrier to material transport, but there are mechanisms of passive transport that could easily operate, both peptide channel based (Bywater 2009) and diffusive (Soh et al. 2010). Active transport would be a much later development. (5) Communications across membranes can be accomplished through material transport (see previous item), but more importantly for the development of future communication systems in a cellular world is the development of signal transduction mechanisms. These are taken care of by assemblies of transmembrane peptides such as have been previously described (Bywater 2009) and studied experimentally (Bak et

Table 1 Complexity metrics for the 20 standard amino acid residue types

Residue type	Number of chiral centres	Number of double bonds in sidechain	Complexity	Enthalpy of formation kJ/mol
G	0	0	36	−528.67
A	1	0	43	−535.39
V	1	0	57	−561.29
S	1	0	63	−710.35
I	2	0	64	−574.97
L	1	0	64	−579.68
C	1	0	68	−495.53
T	2	0	71	−740.66
P	1	0	72	−503.88
K	1	0	77	−148.68
M	1	0	81	−555.13
N	1	1	96	−727.34
Q	1	1	103	−758.05
D	1	1	103	−920.47
E	1	1	110	−956.21
F	1	3	125	−409.92
R	1	1	125	−121.38
H	1	2	129	−424.00
Y	1	3	154	−628.49
W	1	4	190	−357.24

Standard single-letter code abbreviations are used for amino acid residue types. In this and subsequent tables: Complexity refers to measurement of molecular complexity using Shannon information theory and Enthalpy of Formation is calculated in Yasara (Krieger et al. 2002) under the MOPAC MP3 Hamiltonian (see text)

al. 2001). (6) It is in the nature of organic molecules to interact with each other in condensed phases. This is the mechanism behind the elaboration of ever more complex molecular systems through self-assembly. We support the notion that the best way to do this in a really selective manner is when the molecules in question are embedded in lipid, for example the transmembrane peptides referred to above and elsewhere (Killian and Nyholm 2006). It is in this context that molecular

Table 2 Complexity metrics for lipids, ribose, nucleobases and nucleotides

Residue type	Number of chiral centres	Number of double bonds	Complexity	Enthalpy of formation kJ/mol
Heptadecanol	0	0	82	−610.56
C17	0	1	95	−509.45
C16	0	1	121	−803.06
GlyP	1	1	116	−1,622.06
DipalmgPchol	1	3	551	−3,259.88
Ura	0	3	64	−384.47
Cyt	0	3	99	−138.21
Thy	0	3	113	−415.74
Ade	0	4	129	121.73
Gua	0	4	160	−111.02
RibP	4	1	202	−1,975.63
UMP	4	4	391	−2,072.09
CMP	4	4	369	−1,867.72
dTMP	4	4	403	−2,104.77
AMP	4	5	430	−1,545.27
GMP	4	5	464	−1,749.76
Product_1	4	1	160	−729.63
Product_2	4	3	254	1,216.63
Product_3	4	3	382	−2,578.65

C_{17} ($C_{17}H_{35}CO_2-$) and C_{16} ($C_{16}H_{33}C{:}CO-$) structures are typical alkyl chains present in ester and ether lipids, respectively

Product_1, Product_2 and Product_3 are the three proposed nucleotide precursors (Powner et al. 2009).

Gly-P D-glycerol-2-phosphate, *DipalmgPchol* 1,2-dipalmitoyl-sn-glycero-3-phosphocholine, *Ura* uracil, *Cyt* cytosine, *Thy* thymidine, *Ade* adenine, *Gua* guanine, *UMP* uridylate, *CMP* cytidylate, *dTMP* thymidylate, *AMP* adenylate, *GMP* guanidylate, *RibP* ribose phosphate

recognition really has a chance to flourish. (7) Replication is of course an absolute requirement of biological systems and their potential for development and evolution. Obviously, RNA fulfils most of the requirements for this admirably, but as has been stated already, there was no RNA at the earliest stages. Primitive replication mechanisms based on molecular recognition between membrane peptides (Bywater 2009) and evolutionary processes involving inorganic species (Cairns-Smith 2008) have already been described.

In what follows, the emergence of the first six of the above-named functionalities will be discussed first and all the likely molecular species that could have been formed and had a role to play in the emergence of these functionalities will be considered. It is proposed here that as a first approximation, the order in which these organic compounds begin to appear depends on their molecular complexity and to the thermodynamics of their formation. Of course, the formation of any given compound depends on the nature and concentration of precursors, but these in turn will owe their existence to complexity-dependent natural synthetic processes. This ordering suggest a time when they could have appeared on Earth, the dating scheme referred to in the "Introduction". In regard to the seventh of the above-named functionalities, the replication issue, it has been stated (Shapiro 2000) that OoL could have got underway without template-based replication, but it is discussed below anyway.

Molecular complexity

Experiments and theoretical considerations have directed attention towards a collection of candidate chemicals that could have been produced prebiotically, and that these later, in various combinations, could react in ways that generated molecular species of ever increasing complexity, the reactions being powered by solar and geothermal energy and radiation (Bywater and Conde-Frieboes 2005; Bywater 2009; Nitschke and Russell 2010; Martin and Russell 2007; Sleep et al. 2011; Miller 1953; Miller and Urey 1959; Fox 1995; Van der Gulik et al. 2009). This collection of chemicals would have to include some lipids, of sufficient complexity to be able to form the protocells referred to above. In addition, peptides would form that can assemble inside the enveloping membranes of these protocells and exert some of the key functions that will ultimately be required by more advanced living systems. The conditions that were needed to support the synthesis of the necessary organic chemicals were adequate for the production of the simplest peptides and lipids—energy sources and catalysts already referred to—and a feedstock of simple chemicals (Miller 1953; Miller and Urey 1959; Van der Gulik et al. 2009). In order to be able to estimate how early a given compound could have been produced, and to order these on a timescale, it is necessary to quantify this concept of "simplest". In this work, this is done by defining molecular complexity according to a chemoinformatic score of information content based on Shannon entropy (see "Methods"). In the present case, several more complex and more recent compounds have been added.

According to this complexity score, 12 of the now 20 natural (i.e. in the biology of today) amino acid types would have been produced early on, prebiotically. These are shown in Table 1. A cutoff of 100 defines this "minimal set". This may seem arbitrary, but there are several other criteria that have been published elsewhere (Miller 1953; Miller and Urey 1959; Fox 1995; Dryden et al. 2008) that lend support to this selection. In particular, it was stated by Miller (2009): "Just turning on the spark in a basic pre-biotic experiment will yield 11 out of 20 amino acids." These 11 are included in the 12 identified here. (Of course, there could have been other amino acid types that have not survived. This is discussed below in "Conclusions and Comments"). One particularly significant feature of this minimal subset is that the majority of its members are hydrophobic in character, consistent with their preference for lipid-rich media and conducive to a propensity to populate polypeptides that are embedded in lipid. This observation led to the notion of a "transmembrane-peptide-first" model for OoL (Bywater 2009), but this would only earn any credence if the membrane-spanning peptides could exert useful biological functions. This is precisely what they can do (i.e. the fourth, fifth, sixth and even seventh of the above-defined functions). This emphasis on membrane-spanning peptides presupposes a supply of membranes and these are in turn based on lipids which would result from esterification (or etherification) of glycerol by fatty acids of various chain lengths (reckoned here as the number of methylenes), typically up to 17. As shown in Table 2, the chemicals required to make up phospholipids (such as are found in lipid bilayers) are also "simple" by the criteria applied here (complexity score 95 and 121 for typical alkyl chain precursors, 116 for glycerol phosphate). To construct a phospholipid comparable to those in typical "modern" membranes requires the production of relatively complex molecules (score 551) but the lipids lining the earliest protocells would have been much simpler (Luisi et al. 1999).

Fitting complexity scores to a timescale

The classical Miller experiment (Miller 1953; Miller and Urey 1959) suggests that the simplest precursors to biomolecules could have been produced very early on, but it is surmised that more complex molecules will take more time to be generated in the first place and later to accumulate in sufficient quantities and concentrations, and then to evolve into further new species. The hypothesis presented here is that the emergence of these more complex species is dependent on molecular complexity. The minimal set of amino

acids (Table 1) would have been produced very quickly and be present in some abundance. Lipid precursors and pyrimidines and purines (nucleobases) would also have been produced early on (Table 2). But nucleobases on their own have no functions that are useful for replication; they need to be integrated into a polymeric structure. At some point a ribose-phosphate backbone became the favoured construct for this, but it can be seen that the nucleotides needed to form these polynucleotides are now ~4 times more complex (see Table 2) than their parent nucleobases by virtue of the need to append ribose and phosphate onto them. In fact, it is very unlikely that nucleotides were ever produced in this way. While it is easy to sketch plausible synthetic pathways for the peptides, sugars and lipids and even for nucleobases, and indeed, these compounds are produced experimentally in systems designed to mimic the chemistry of the early Earth (Miller 1953; Miller and Urey 1959; Fox 1995), it is not immediately obvious how nucleotides could be correctly assembled. It is unlikely that the nucleoside could be formed from the corresponding nucleobase correctly connected to a ribose moiety without the aid of a sophisticated catalyst. An interesting and highly plausible, synthetic path has been proposed (Powner et al. 2009). This is a very novel idea that nucleotides developed not from nucleobases, but rather, from ribose. The compounds in the proposed synthetic pathway to nucleotides from arabinose aminooxazolines (labelled Product_1 in Table 2) via arabinose anhydronucleoside ("Product_2") to an activated ribonucleotide β-ribocytidine-2′,3′-cyclic phosphate ("Product_3") have complexity scores that increase as one proceeds from a ribose-based structure to a nucleotide-like structure, as one would expect, and these scores are in themselves very much in line with the scores for ribose phosphate itself and nucleotides, respectively (Table 2).

This is of course encouraging, but we have to compare polypeptides with RNA. The heavy burden, chemoinformatically speaking, of producing a full coding sequence for a given polypeptide or small protein becomes clear when one considers the complexity scores shown in Tables 3 and 4. Scores for both RNA and its corresponding DNA are given and they are clearly much larger compared to the polypeptide/protein they code for. It has to be remembered that the genetic code for proteins is a triplet code, but in practice, this means that for every amino acid residue there are six nucleobases, since DNA has to exist in a duplex (+/- strand) form. The consequences of this are very clearly seen in the enormously large values for complexity in the DNA-coding regions for $(Ile)_{25}$ (Table 3) and crambin (Table 4), respectively.

These complexity scores are of course related to how proteins are coded for "today" and do not take into account, nor preclude, that there could have been an "RNA world" without the need for DNA duplex structures. Still, RNA as such would be very much more complex than the proteins that it codes for, and therefore be a later arrival than oligopeptides and even some small proteins.

RNA has been held to reign supreme in the OoL research world despite the need to explain the complexity, how it is synthesised in the first place, how it can be protected (during synthesis and certainly afterwards) from degradation and dilution in the oceans. These objections have never been satisfactorily defended.

It remains the case that RNA is, with good reason, credited with many attributes that are essential for the maintenance of life. Principles among these are catalytic activity ("ribozyme") and the replication function. As far as the former is concerned, there is nothing unique about that function in itself and examples of ribozyme activity are not common in the modern world. They operate when they are needed as in ribosomes (Yonath 2005).

The appearance of the oldest living organisms on Earth—cyanobacteria as found in stromatolites—has been estimated to have taken place 3.5 Gyrs ago (Nisbet and Sleep 2001). This is 0.7 Gyrs after the earliest time for anything resembling life to have emerged, at the transition from the Hadean to the Archaean era (Nisbet and Sleep 2001). Double-stranded DNA (complexity ~322,100 for a small protein) must have been in place before that time, so we can at least put that as a lower limit. It is known that the moon is receding, so it is not strictly correct to assume a linear timescale for, e.g. the evolution of complexity. But over a restricted period, if this assumption can be allowed, then it is possible to give approximate times for the appearance of different types of compound. Starting from the simpler

Table 3 Complexity metrics for a 25-meric oligopeptide and its corresponding DNA sequence

Residue type	Number of residues	Number of chiral centres	Number of double bonds	Complexity	Enthalpy of formation kJ/mol
$(Ile)_{25}$	25	25	25	3,824	−1.44E+04
$(AUA)_{25}$	75	300	275	74,458	−1.29E+05
$(ATA:TAT)_{25}$	150	600	550	148,526	−2.74E+05

$(Ile)_{25}$ is a 25-residue-long oligomer of isoleucine chosen to represent a typical membrane-spanning peptide that could easily be synthesised under prebiotic conditions. AUA is one of the three codons for isoleucine, but UAU needs to be included in the calculation to cater for the fact that DNA is double stranded

Table 4 Complexity metrics for a small (46-residue) water-soluble globular protein and its corresponding DNA sequence

Residue type	Number of residues	Number of chiral centres	Number of double bonds	Complexity	Enthalpy of formation kJ/mol
Crambin	46	57	62	6,452	−1.48E+04
mRNA seq for crambin	138	552	616	143,174	−2.50E+05
DNA seq for crambin	276	1,104	1,232	322,127	−5.01E+05

Crambin is a protein (MW 4.72 kD) from the plant kingdom, chosen as a representative of the type of small globular protein that could have emerged at an early stage in OoL

substances (complexity ~100) that were in abundance 4.2 Gyr ago to duplex DNA at 3.5 Gyrs ago, we can, by interpolation, put the time of appearance of phospholipids (complexity ~550) at 1.2 Myr, a transmembrane peptide (~3,800) 8.3 Myr, a small protein (~6,500) 14 Myr and its corresponding RNA (~143,000) at 0.31 Gyr. Thus, the "RNA world" would have started about 3.9 Gyrs ago and then from 3.5 Gyrs the emergence of duplex DNA would start to take place.

Evolution of lipid structures

According to the complexity-based timescale being proposed here, the earliest lipids would have emerged at about the same time as the simplest amino acids, glycerol etc. (see Table 1). The presence of lipids leads to the formation of vesicles which are generally accepted as being critical (Luisi et al. 1999; Deamer and Weber 2010; Budin and Szostak 2011). Ultimately, there would emerge complex lipids similar to "modern" lipids based on derivatives of phosphoglyceryl esters (and ethers) for example. But much simpler lipid-like molecules such as long-chain alcohols (Budin and Szostak 2011) and fatty acids (Luisi et al. 1999; Budin and Szostak 2011) can also form vesicles. It is clear that this represents an important early stage in the steady evolution of ever more complex systems. This in turn exerts constraints on which oligopeptides emerged first. For example, in lipid vesicles made of only alcohols there would not be any way, at least under pH conditions close to neutral, to accommodate helical oligopeptides containing lysine. The complexity of, e.g. heptadecanol would be comparable to that of the very simplest amino acid types (see Table 1). This would have affected the very earliest expressions of peptide evolution and by no means invalidates the argument outlined above.

Enthalpy of formation

The likelihood that any organic compound would be synthesised abiotically is most appropriately measured by the above complexity scores, but thermodynamics obviously plays a role in determining the overall stability of a given compound. Any consideration of the kinetics of formation of these compounds would of course require knowledge of the free energy of the synthetic reaction, but here we are considering very long time periods in which quasi-steady states can be established. Stability is in these circumstances adequately represented by the enthalpy of formation. Accordingly, calculations were performed (see "Methods") on all the compounds studied here to determine their enthalpy of formation. The results are listed in the rightmost column Tables 1, 2, 3 and 4. Compounds whose overall synthesis from their constituent elements is endothermic (negative) are shown in red. Normally, one would expect compounds such as these to be produced endothermically. As can be seen from the tables, oxygen-rich compounds tend to require an input of relatively more energy while nitrogen-rich compounds show the opposite behaviour, with adenine as a particularly striking example (black lettering indicating exothermic synthesis). Already at the time of the Miller experiment (Miller 1953; Miller and Urey 1959), large amounts of adenine were produced and this tempted many researchers to conclude that RNA could be quickly assembled from nucleotides derived from this nucleobase, as well as from the three other nucleobases, even though the latter were considerably less abundant. As we have seen, this route to nucleotides is extremely unlikely, at least prebiotically, and a much more plausible scheme has been proposed (Powner et al. 2009). An alternative scheme (Prieur 2001) which does originate from nucleobases and involves boric acid catalysis deserves to be mentioned, however, and this issue is still undergoing further examination (Benner et al. 2010). Yet another alternative route has been proposed (Costanzo et al. 2007) whereby formamide together with phosphate-rich minerals is reported to yield a complete set of nucleic bases and is purported to provide one more missing link in the identification of a single unifying chemical frame for the self-organization of nucleic polymers in prebiotic conditions. This is clearly an important alternative to the scheme of Powner et al. (2009).

As with molecular complexity, a ranking order seems to emerge from a consideration of the enthalpy of formation data (Tables 1, 2, 3 and 4) and this only serves to support the contention that there was a distinct ordering of the times in which different molecular species could have appeared. The ordering derived from complexity calculations are considered

to be of primary significance, it is just interesting that enthalpy data supports this ordering in a general way (much as did the "ease of synthesizability", reported earlier (Bywater 2009)).

Further, while enthalpy of formation is a guide as to what can be expected to form through abiotic chemistry, the importance of kinetics cannot be ignored. Throughout all of the developments prior to and after OoL, catalysis must have been a major factor. The importance of mineral surfaces as catalysts to assist the appearance and accumulation of relevant species cannot be overstressed and indeed has been referred to extensively in earlier literature (Nitschke and Russell 2010). While the earliest manifestations of catalysis took place in a mineral setting, the species which were produced in those early processes and survived would have to take on the burden of performing catalysis themselves (e.g. in the lipid bilayer and within the vesicle itself). The notion of vesicular or micellar catalysis has already been referred to. The incorporation of phosphate into organic species like glycerol phosphates, phospholipids and, much later, nucleotides obviously made a great difference as did the incorporation of diverse metal catalysts into peptides where the amino acid side chains serve to chelate these metal ions. Similarly, "proto-enzymes" in the form of mineral clusters, which of course predate coded enzyme structures, have been proposed elsewhere (Nitschke and Russell 2010). The catalysts needed to assemble highly complex chemicals such as RNA would have to be much more sophisticated and would have to include a molecular recognition function in addition to the catalytic function. It is often repeated in "RNA first" circles that RNA itself is catalytically active. But the catalytic repertoire of RNA is very restricted and does not include oxidation/reduction possibilities that are the hallmark of metal-ion catalysis.

Other authors (Benner et al. 2010; Kim et al. 2011) have similarly considered the energetics of formation of various classes chemicals that could have been candidates for involvement in OoL and prebiotic chemistry, although not in the quantitative way presented here.

Other factors that might affect the order of appearance of prebiotic chemicals

While the focus in this work is on molecular complexity, supported by thermodynamic data, it is important to mention in passing that other factors could have played a role, the most important of which would be environmental. These could include changes in the diurnal periodicity as the Moon receded from Earth, weathering of volcanic rocks to produce new minerals, changes in volcanic activity and activity of undersea vents (Nitschke and Russell 2010; Martin and Russell 2007; Sleep et al. 2011), impact of large meteors (Kvenvolden et al. 1970; Herd et al. 2011). Water has been singled out as a critical prerequisite for life (Bywater and Conde-Frieboes 2005; Cui 2010) and even held to be the only solvent in the primitive soup (Cui 2010). Indeed, it would have been the only solvent, but reactions can proceed in the absence of water, for example in the dry state or as highly concentrated tars.

Molecular recognition and replication

Finally we now return to the replication issue and ask the question: what kinds of molecular species were involved in the first replication processes? This question has already been satisfactorily answered in previous publications (Bywater 2009; Rode 1999; Bujdak and Rode 2004; Lee et al. 1996; Fitz et al. 2007), so what follows is some new material which supports the earlier arguments.

Intrahelix interactions have been described in detail (Bywater et al. 2001) in both membrane-spanning peptides and globular proteins. In the membrane, intrahelical bonds can arise when a single hydrogen-bonding amino acid residue turns up in the central (lipophilic) region of the peptide, a common (and most likely, given the complexity scale of Table 1) one would be asparagine (Bywater et al. 2001; Choma et al. 2000). The side chain of asparagine contains both donor and acceptor groups, so helices that are furnished with this residue type would self-associate. An interaction involving charged residues could be envisaged but single charges would not be stable in the lipophilic domain unless there was some kind of compensating charge of the opposite sign. A helix with a positive charge at one extremity and a negative charge at the other would self-associate, forming a symmetrical dimer. In the membrane, hydrophobic interactions would not be important, but geometrical compatibility such as the ability of bulky groups (methionine, phenylalanine, tyrosine, tryptophan) to occupy "empty" regions caused by the presence of, e.g. glycine could be important. Again, a helix with a protuberance (methionine, phenylalanine, tyrosine, tryptophan) at one end and a cavity (glycine) at the other would self-associate symmetrically. This is the "knobs-into-holes" model of Crick (1953) which was proposed and shown to be important in coiled-coils. But it could equally be important here, and indeed, tendencies towards supercoiling have already been shown to play a role in helix membrane bundles (Bywater et al. 2001).

The importance of a peptide world that precedes an RNA world has already been stressed (Bywater 2009; Rode 1999; Bujdak and Rode 2004; Lee et al. 1996; Fitz et al. 2007). Peptide–peptide recognition not only fulfils an important selection and sorting mechanism for the peptides themselves within the confines of their protocell but the process can continue after the protocell has been split into two or more "daughter protocells". New synthesis, recognition and selection steps can continue to happen during many billions of cycles of these protocell division events, and this represents the basis of peptide-based replication.

Conclusions and Comments

In this paper, the emergence of complex chemicals required to allow and sustain life is seen as a process of chemical evolution (Follmann and Brownson 2009) in which increasing molecular complexity develops over time. We arrive at the following conclusions:

1. The candidate chemicals needed to give rise to living systems as well as late arrivals such as RNA and DNA and proteins have been ranked according to molecular complexity as defined in Methods.
2. This scoring system suggests a way to rank them according to the time at which they might have appeared in a terrestrial setting. The simplest chemicals could have been produced by undersea volcanic activity or delivered from space, but more complex chemicals would have required the synthetic cycles and combinatorial chemistry that a tidal beach provides.
3. The results of this analysis suggest that the simplest (defined above) amino acid types, some primitive lipids (including, at some point glycerol and glycerol-phosphate), pyrimidines and purines could all have been produced at a very early stage starting at −4.2 Gyr ago. This is, of course, a conjecture, but it accords well with suggestions made elsewhere.
4. This list is similar to that of the organic compounds found on the Murchison meteorite (Kvenvolden et al. 1970). This does not mean that such compounds were necessarily "planted" on Earth by earlier meteorites (the Murchison meteorite being a contemporary event). But it suggests that such chemicals may have been in circulation for some time. There is some discrepancy between the composition of the meteorites and the ranking presented here for the amino acids. Aspartate and glutamate are found on the Murchison meteorite (Kvenvolden et al. 1970). These two residue types are only just beyond the "100 mark" (which anyway had an arbitrary character, as already stated) and they would certainly also have been present on early Earth. Their presence would, especially in the case of apsartate, considerably enhance the functional, including catalytic, repertoire of the early oligopeptides. It is more noteworthy that 12 or so "non-protein" amino acid types were found on the Murchison meteorite (Kvenvolden et al. 1970), which prompts the question as to why they do not seem to have been utilised in the emerging biology on Earth. Pyrimidines and purines were found in small amounts on the Murchison meteorite. But as pointed out earlier, these nucleobases are not valid starting points for the synthesis of nucleic acids. No phospholipids or steroids, for example, were found on the Murchison meteorite but there were some necessary lipid precursors such as glycerol and fatty acids. In order to make long-chain fatty acids, some process involving repeated cycles of catalyst-aided chain lengthening is required. This would have to be a terrestrial process involving water and tides. It is hard to see how this could happen on a meteor or asteroid.
5. Nucleobases were by all accounts "early chemicals" but their conversion into the corresponding nucleotides is an unlikely event (very difficult chemistry). Nevertheless, we do not want to disregard a proposal (Sowerby and Petersen 2002) that nucleobases in adsorbed phases could act as templates for some kind of primitive replication process. This is an intriguing proposal but at some point the bases would need to be covalently linked in a linear polymer, not just adsorbed in crystalline arrays. More convincing alternatives have been presented, as discussed elsewhere in this paper.
6. The lipids produced on Earth would be complex and abundant enough to allow formation of the vesicles that are deemed to be critical for life (Luisi et al. 1999; Bywater and Conde-Frieboes 2005; Deamer and Weber 2010; Budin and Szostak 2011).
7. The vesicle membrane is the ideal setting for the assembly of the first information-rich chemicals (Bywater 2009). Apart from being a protecting and concentrating environment, there are catalytic processes associated with vesicles (Luisi et al. 1999). Further support for the notion that lipids and membranes are critical come from the recognition of the importance of phosphates by authors who sought a prebiotic chemistry that leads to both ribonucleotides and lipids in a single system (Powner and Sutherland 2011).
8. It has been observed (Benner et al. 2010) that mixtures of organic chemicals if left to their own devices under conditions of high energy input will form tars that may be hard to resolve. We have proposed two sorting mechanisms for organic compounds generally: chromatography on tidal beaches (Bywater and Conde-Frieboes 2005) and

Table 5 Comparison of complexity for a simple fatty acid and its thioanalogs

Residue type	Number of chiral centres	Number of double bonds	Complexity	Enthalpy of formation kJ/mol
C16	0	1	121	−803.06
C16-thio	0	1	119	−548.50
C16-thiono	0	1	119	−502.36
C16-dithio	0	1	121	−262.96

C_{16}-thio, C_{16}-thiono and C_{16}-dithio are the monothio, thiono- and dithio analogs of C_{16} respectively (see Table 2)

incorporation into ordered lipid bilayer structures (Bywater 2009). Thirdly, for oligopeptides, there exists, as already mentioned, a sophisticated sorting and selection system (Choma et al. 2000).

9. Following on from the simpler substances (complexity ~100) referred to in "Conclusions and Comments" and in "Prerequisites for life" section above, the appearance of phospholipids (complexity ~550) is calculated to have occurred at ~1.2 Myr, transmembrane peptides (~3,800) at ~8.3 Myr, small proteins (~6,500) at ~14 Myr and its corresponding RNA (~143,000) at ~0.31 Gyr. The "RNA world" would have started about 3.9 Gyrs ago and then from 3.5 Gyrs the emergence of duplex DNA would start to take place. Again, this is a conjecture and furthermore, it assumes a linear timescale. This is of course a very rough approximation, not only on geophysical grounds, but, as more complex chemicals accumulate, opportunities for rapid changes in these processes can take place as a result of catalysis and self-assembly. However, given these *caveats*, the proposed timings seem reasonable.
10. A replication mechanism involving membrane-embedded oligopeptides could have arisen early, certainly well before the −3.9-Gyr period. Peptides could easily be synthesised under prebiotic conditions (Rode 1999; Bujdak and Rode 2004) and their recognition and selection properties form the basis of a primordial system of life (Rode 1999; Bujdak and Rode 2004).
11. We direct attention towards a proposal for a plausible pathway to RNA from simple precursors, which bypasses the purported dead-end represented by the appearance of nucleobases. The four precursors referred to by these authors as having different temporal appearances (Powner et al. 2009): cyanamide, cyanoacetylene, glycolaldehyde and glyceraldehyde could all have been generated in large amounts within the time-scale referred to here. The necessary "conditions of heating and progressive dehydration followed by cooling, rehydration and ultraviolet irradiation" are amply catered for by our tidal beach model. It will be recalled from earlier remarks under Enthalpy of Formation that there are alternatives to the scheme of Powner et al. (2009), in particular, a scheme based on catalysis by borate (Prieur 2001) and yet another (Costanzo et al. 2007) based on interaction between formamide and phosphate-rich minerals.
12. It is acknowledged that the early Earth was not as oxygen rich as it later became, and that a sulfur-rich environment could have supported early primitive microorganisms (~3.5 Gyr ago) that have been deposited as microfossils in Western Australia (Wacey et al. 2011). Preliminary complexity calculations for the chemicals required to support these life forms show that they would fit well into the timescale outlined above. For example, serine and cysteine have almost the same complexity scores, 63 and 68, respectively. For other thio or thiono analogs, the complexity is about the same as for the corresponding oxygen analogs while the enthalpies of formation are a good deal lower (Table 5). Particular note is taken that these authors propose (Wacey et al. 2011) that a tidal beach was necessary for the formation of these microorganisms. We have earlier alluded to the importance of beaches (Bywater and Conde-Frieboes 2005), which were at least as important as "warm little ponds" (Follmann and Brownson 2009). The warm little ponds are exactly what are left over after a tide has ebbed. These ponds were certainly important, although they would contribute more to the rate of chemical evolution by becoming concentrated by drying out at regular intervals and then being replenished by fresh water and new chemicals.

Acknowledgements This work was commenced during the tenure of a Visiting Fellowship at Magdalen College, Oxford. The author wishes to express his gratitude to the College for the excellent facilities and ambiance under which this work was conducted. Elmar Krieger and Gert Vriend are thanked for kindly supplying the YASARA program.

Conflicts of interest There are no competing interests,

References

Bak M, Bywater RP, Hohwy M, Thomsen JK, Adelhorst K, Jakobsen HJ, Sørensen OW, Nielsen NC (2001) Conformation of alamethicin in oriented phospholipid bilayers determined by (15)N solid-state nuclear magnetic resonance. Biophys J 81:1684–1698. doi:10.1016/S0006-3495(01)75822-5

Benner SA, Kim HJ, Kim MJ, Ricardo A (2010) Planetary organic chemistry and the origins of biomolecules. Cold Spring Harb Perspect Biol 2:a003467. doi:10.1101/cshperspect.a003467

Bertz SH (1983) On the complexity of graphs and molecules. Bull Math Biol 45:849–855. doi:10.1007/BF02460054

Budin I, Szostak JW (2011) Physical effects underlying the transition from primitive to modern cell membranes. Proc Nat Acad Sci USA 108:5249–5254. doi:10.1073/pnas.1100498108

Bujdak J, Rode BM (2004) On the mechanisms of oligopeptide reactions in solution and clay dispersion. J Pept Sci 10:731–737

Bywater RP (2009) Membrane-spanning peptides and the origin of life. J Theor Biol 261:407–413. doi:10.1016/j.jtbi.2009.08.001

Bywater RP, Conde-Frieboes K (2005) Did life begin on the beach? Astrobiology 5:568–574. doi:10.1089/ast.2005.5.568

Bywater RP, Thomas D, Vriend G (2001) A sequence and structural study of transmembrane helices. J Comput Aided Mol Des 15:533–552. doi:10.1023/A:1011197908960

Cairns-Smith AG (2008) Chemistry and the missing era of evolution. Chemistry 14:3830–3839. doi:10.1002/chem.200701215

Choma C, Gratkowski H, Lear JD, DeGrado WF (2000) Asparagine-mediated self-association of a model transmembrane helix. Nat Struct Mol Biol 7:161–166. doi:10.1038/72440

Costanzo G, Saladino R, Crestini C, Ciciriello F, Di Mauro E (2007) Nucleoside phosphorylation by phosphate minerals. J Biol Chem 282:16729–16735. doi:10.1074/jbc.M611346200

Crick FHC (1953) The packing of α-helices: simple coiled-coils. Acta Cryst 6:689–697. doi:10.1107/S0365110X53001952

Cui S (2010) The possible roles of water in the prebiotic chemical evolution of DNA. Phys Chem Chem Phys 12:10147–10153. doi:10.1039/C002414G

Deamer DW, Weber AL (2010) Bioenergetics and life's origins. Cold Spring Harb Perspect Biol 2010; Feb. 2 : a004929. doi:10.1101/cshperspect.a004929

Donoghue PCJ, Antcliffe JB (2010) Origins of multicellularity. Nature 466:41–42. doi:10.1038/466041a

Dryden DT, Thomson AR, White JH (2008) How much protein sequence space has been explored by life on earth? J R Soc Interface 5:953–956. doi:10.1098/rsif.2008.0085

Fitz D, Reiner H, Rode BM (2007) Chemical evolution toward the origin of life. Pure Appl Chem 79:2101–2117. doi:10.1351/pac200779122101

Follmann H, Brownson C (2009) Darwin's warm little pond revisited: from molecules to the origin of life. Naturwissenschaften 96:1265–1292. doi:10.1007/s00114-009-0602-1

Fox SW (1995) Thermal synthesis of amino acids and the origin of life. Geochim Cosmochim Acta 59:1213–1214

Gallori E (2011) Astrochemistry and the origin of genetic material. Rend Lincei 22:113–118. doi:10.1007/s12210-011-0118-4

Herd CDK, Blinova A, Simkus DN, Huang YS, Tarozo R, Alexander CM, O'D GF, Nittler LR, Cody GD, Fogel ML, Kebukawa Y, Kilcoyne ALD, Hilts RW, Slater GF, Glavin DP, Dworkin JP, Callahan MP, Elsila JE, De Gregorio BT, Stroud RM (2011) Origin and evolution of prebiotic organic matter as inferred from the Tagish Lake meteorite. Science 332:1304–1307. doi:10.1126/science.1203290

Killian JA, Nyholm TKM (2006) Peptide in lipid bilayers, the power of simple models. Curr Opin Struc Biol 16:473–479. doi:10.1016/j.sbi.2006.06.007

Kim HJ, Ricardo A, Illangkoon HI, Kim MJ, Carrigan MA, Frye F, Benner SA (2011) Synthesis of carbohydrates in mineral-guided prebiotic cycles. J Am Chem Soc 133:9457–9468, Epub 2011 May 27

Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. Proteins 47:393–402. doi:10.1002/prot.10104

Kvenvolden KA, Lawless J, Pering K, Peterson E, Flores JH, Ponnamperuma C, Kaplan IR, Moore C (1970) Evidence for extraterrestrial amino-acids and hydrocarbons in the Murchison meteorite. Nature 228:923–926. doi:10.1038/228923a0

Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR (1996) A self-replicating peptide. Nature 382:525–528. doi:10.1038/382525a0

Luisi PL, Walde PW, Oberholzer T (1999) Lipid vesicles as possible intermediates in the origin of life. Curr Opin Colloid In 4:33–39. doi:10.1016/S1359-0294(99)00012-6

Martin W, Russell MJ (2007) On the origin of biochemistry at an alkaline hydrothermal vent. Philos Trans R Soc B362:1887–1926. doi:10.1098/rstb.2006.1881

Miller SL (1953) Production of amino acids under possible primitive earth conditions. Science 117:528–529

Miller SL (2009) Exobiology: an interview with Stanley L. Miller. http://www.accessexcellence.org/WN/NM/miller.php. Accessed 20 August 2009.

Miller SL, Urey HC (1959) Organic compound synthesis on the primitive Earth. Science 130:245–251. doi:10.1126/science.130.3370.245

Nisbet EG, Sleep NH (2001) The habitat and nature of early life. Nature 409:1083–1091. doi:10.1038/35059210

Nitschke W, Russell MJ (2010) Just like the Universe the emergence of life had high enthalpy and low entropy beginnings. J Cosmol 10:3200–3216

Powner MW, Sutherland JD (2011) Prebiotic chemistry: a new *modus operandi*. Philos Trans R Soc 366:2870–2877. doi:10.1098/rstb.2011.0134

Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459:239–242. doi:10.1038/nature08013

Prieur BE (2001) Etude de l'activite prebiotique de l'acide borique. C R Acad Sci Paris Chimie 4:1–4

Rode BM (1999) Peptides and the origin of life. Peptides 20:773–786. doi:10.1016/S0196-9781(99)00062-5

Shapiro R (2000) A replicator was not involved in the origin of life. IUBMB Life 49:173–176. doi:10.1080/152165400306160

Sleep NH, Bird DK, Pope EC (2011) Serpentinite and the dawn of life. Philos Trans R Soc B366:2857–2869. doi:10.1098/rstb.2011.0219

Soh S, Byrska M, Kandere-Grzybowska K, Grzybowski AB (2010) Reaction-diffusion systems in intracellular molecular transport and control. Angew Chem Int Edit 49:4170–4198. doi:10.1002/anie.200905513

Sowerby SJ, Petersen GB (2002) Life before RNA. Astrobiology 2:231–239. doi:10.1089/153110702762027826

Stewart JJP (1989) Optimization of parameters for semiempirical methods. J Comput Chem 10:209–264. doi:10.1002/jcc.540100208

Van der Gulik P, Massar S, Gilis D, Buhrman H, Rooman M (2009) The first peptides: the evolutionary transition between prebiotic amino acids and early proteins. J Theor Biol 261:531–539. doi:10.1016/j.jtbi.2009.09.004

Wacey D, Kilburn MR, Saunders M, Cliff J, Brasier MD (2011) Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. Nature Geosci 4:698–702. doi:10.1038/ngeo1238

Yonath A (2005) Ribosomal crystallography: peptide bond formation, chaperone assistance and antibiotics activity. Mol Cell 20:1–16

Young DC (2001) Computational chemistry: a practical guide for applying techniques to real world problems. New York USA.