Cell Reports
Physical Science

CellPress
OPEN ACCESS

Review

What it takes to solve the origins of life: An integrated review. Part 2: Theoretical methods and emerging trends

OoLEN (Origin of Life Early-Career Network)[1] Silke Asche,[2,44] Carla Bautista,[3,4,5,6,44] Celia Blanco,[7,44] David Boulesteix,[8,44] Alexandre Champagne-Ruel,[9,10,44] Cole Mathis,[10,11,44,*] Omer Markovitch,[7,12,44] Zhen Peng,[13,14,15,44] Avinash Vicholous Dass,[16,44] Alyssa Adams,[17] Eloi Camprubi,[18] Enrico Sandro Colizzi,[19] Stephanie Colón-Santos,[15] Hannah Dromiack,[20] Valentina Erastova,[21,22] Amanda Garcia,[13] Ghjuvan Grimaud,[23,24]

(Author list continued on next page)

[1]www.oolen.org
[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[3]Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Laval, QC, Canada
[4]Département de Biologie, Faculté des Sciences et de Génie, Université Laval, Laval, QC, Canada
[5]Regroupement québécois de recherche sur la fonction, la structure et l'ingénierie des protéines (PROTEO), Université Laval, Laval, QC, Canada
[6]Centre de Recherche en Données Massives (CRDM), Université Laval, Laval, QC, Canada
[7]Blue Marble Space Institute of Science, Seattle, WA, USA
[8]Laboratoire Génie des Procédés et Matériaux, CentraleSupélec, Gif-sur-Yvette, France
[9]Université de Montréal, Montréal, QC, Canada
[10]Biodesign Institute, Arizona State University, Tempe, AZ, USA
[11]School of Complex Adaptive Systems, Arizona State University, Tempe, AZ, USA
[12]Centro de Química Estrutural, Institute of Molecular Sciences and Department of Chemical Engineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[13]Department of Bacteriology, University of Wisconsin–Madison, Madison, WI, USA
[14]Wisconsin Institute for Discovery, University of Wisconsin–Madison, Madison, WI, USA
[15]Department of Geoscience, University of Wisconsin–Madison, Madison, WI, USA
[16]Origins Institute, Department of Physics and Astronomy, McMaster University, Hamilton, ON, Canada
[17]Cross Labs, Kyoto, Japan

(Affiliations continued on next page)

SUMMARY

The origin(s) of life (OoL), which has puzzled scientists for centuries, remains a major scientific challenge in the 21st century. Understanding the processes relevant to the OoL demands theoretical frameworks that can connect processes across scales, from microscopic dynamics to emergent levels of organization. While experimental studies generate a wealth of data, theoretical and computational approaches provide the structure necessary to interpret and generalize these findings. In Part 1, we examined the most widely used experimental techniques in the field. Here, we focus on the mathematical, physical, and computational techniques used to model phenomena relevant to life's origin(s). We discuss methods ranging from quantum chemistry and molecular dynamics to chemical reaction networks, autocatalysis, and evolutionary modeling, as well as information-theoretic and phylogenetic approaches that link chemical and biological organization. We further highlight emerging trends such as synthetic biology, omics-based methods, and laboratory automation as novel points of contact for theory-experiment integration. Ultimately, we aim to provide an educational tool that can facilitate more post-disciplinary collaborations in OoL research by helping scientists understand what they can do about the problem of life's origins, rather than telling them how to think about it.

INTRODUCTION

The question of how life began on Earth is one of the oldest posed by humankind. Origin(s) of life (OoL) is a multi-disciplinary endeavor searching for *de novo* life. In lieu of this, we wrote this two-part review focusing on tools and techniques used by different disciplines to tackle the question of how life might have started. In Part 1, we focus on experimental techniques, covering spectroscopy, chromatography, mass spectrometry, microscopy, genomic sequencing, and physical and chemical databases.[1]

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Aaron Halpern,[25] Stuart A. Harrison,[25] Seán F. Jordan,[26] Tony Z. Jia,[27,7] Amit Kahana,[28] Artemy Kolchinsky,[29] Odin Moron-Garcia,[30] Ryo Mizuuchi,[31] Jingbo Nan,[32] Yuliia Orlova,[33] Ben K.D. Pearce,[34] Klaus Paschek,[35] Martina Preiner,[36] Silvana Pinna,[37] Eduardo Rodríguez-Román,[38,39] Loraine Schwander,[40] Siddhant Sharma,[41,7] Harrison B. Smith,[27,7] Andrey Vieira,[42] and Joana C. Xavier[43]

[18]School of Integrative Biological and Chemical Sciences, University of Texas Rio Grande Valley, Edinburg, TX, USA
[19]Sainsbury Laboratory, University of Cambridge, Cambridge, UK
[20]Department of Physics, Arizona State University, Tempe, AZ, USA
[21]School of Chemistry, University of Edinburgh, Joseph Black Building, Edinburgh, UK
[22]UK Centre for Astrobiology, School of Physics and Astronomy, University of Edinburgh, Edinburgh, UK
[23]APC Microbiome Ireland, University College Cork, Cork, Ireland
[24]Food Biosciences Department, Teagasc Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland
[25]Department of Genetics, Evolution and Environment, University College London, London, UK
[26]Life Sciences Institute, School of Chemical Sciences, Dublin City University, Dublin, Ireland
[27]Earth-Life Science Institute, Institute of Science Tokyo, Meguro-ku, Tokyo, Japan
[28]School of Chemistry, University of Glasgow, Glasgow, UK
[29]ICREA–Complex Systems Lab, Universitat Pompeu Fabra, Barcelona, Spain
[30]Functional and Evolutionary Ecology Department, Estación Experimental de Zonas Áridas (EEZA-CSIC), Almería, Spain
[31]Department of Electrical Engineering and Bioscience, Faculty of Science and Engineering, Waseda University, Shinjuku, Tokyo, Japan
[32]State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing, China
[33]University of Amsterdam, Swammerdam Institute of Life Sciences, Amsterdam, the Netherlands
[34]Johns Hopkins University, Baltimore, MD, USA
[35]Max Planck Institute for Astronomy, Heidelberg, Germany
[36]Microcosm Earth Center, Max Planck Institute for Terrestrial Microbiology and Philipps-University Marburg, Marburg, Germany
[37]Institut de science et d'ingénierie supramoléculaires (ISIS, Université de Strasbourg & CNRS, UMR 7006), Strasbourg, France
[38]Department of Biology, Emory University, Atlanta, GA, USA
[39]Center for Microbiology and Cell Biology, IVIC, Caracas, Venezuela
[40]Institute of Molecular Evolution, Biology Department, Math.-Nat. Faculty, Heinrich-Heine-Universität, Düsseldorf, Germany
[41]Institute of Science and Technology, Am Campus, Klosterneuburg, Austria
[42]Linnaeuskade 47 2, Amsterdam, the Netherlands
[43]Department of Chemistry, Imperial College London, London, UK
[44]These authors contributed equally
*Correspondence: cole.mathis@asu.edu
https://doi.org/10.1016/j.xcrp.2026.103211

Herein, we focus on theoretical and modeling approaches. These include molecular simulations, chemical thermodynamics, kinetics and networks, and models of (proto)cellular evolution, together with information-theoretic perspectives and molecular phylogenetics (Figure 1). Finally, we conclude by discussing emerging trends that integrate experimental and theoretical work, such as omics studies, laboratory automation, microfluidics, synthetic biology targeting protocells, and evolution and selection experiments. Our goal is to present the methodologies and techniques commonly used in OoL rather than to be an in-depth review. It is our hope that these two parts will facilitate more collaborative work between specialists within the community.

THEORETICAL APPROACHES AND MODELING FRAMEWORKS FOR THE ORIGIN(S) OF LIFE

After thorough experimental investigation and comparison with databases, you are starting to grasp what your advisor's mysterious sample is made of. But composition is only a part of the story. To truly understand a system you need to know not just what it is made of *but what it does*. Predicting and explaining the behavior of systems falls under the realm of modeling. If you want to predict how the sample will respond to external stimuli like heat or light, or if you want to understand the reactions that might have created the sample in the first place, you will need models to guide the way. The questions scientists explore in the context of OoL lead to diverse answers,[2] each requiring different theoretical approaches to capture various aspects of the same phenomena. In some cases, these are first-principles physical approaches such as quantum chemical models and thermodynamic models. In other cases, these approaches are based on the principles of biology, as is the case for molecular phylogenetics. Still in other cases need more abstract models to understand how interactions between individual parts compose a whole or how simple rules can lead to complex outcomes. In this section, we outline various theoretical models, computational approaches, and simulation techniques to address problems in OoL.

Our focus is not on theories of life's origin but rather on methods that offer scientific insights beyond what can be directly accessed through experimental or lab-based approaches. Like the experimental approaches, these methods are diverse, but they do share some commonalities. A summary of different approaches is shown in Table 1. Modeling is typically expressed in the language of mathematics, specifically, linear equations, differential equations, and probability and statistics.

In modern research, mathematical approaches are heavily supported by computation, which allows us to solve differential equations, implement maximum-likelihood estimators, and iterate rule-based systems. While many different models share

Figure 1. Techniques covered in these reviews and their most important relationships

Theoretical and computational techniques are shown in yellow, databases in green, and dark blue indicates emerging trends. Experimental techniques that are covered in Part 1[1] are shown in gray. Gray lines connect related techniques within a given category, pink lines connect experimental and theoretical work, while green lines connect techniques to databases.

common mathematical formalisms, the scientific interpretation and application of these models can vary significantly. For example, quantum mechanics (QM), chemical kinetics, and replicator dynamics all involve solving differential equations, but the meaning of those solutions and the approximations used differ across disciplines. Here, we focus on the domain-specific concepts, interpretation, and application of these models, omitting the formal rigor that can be found in specialized literature.

Molecular modeling and simulations

Molecular modeling focuses on understanding chemical-physical processes at the atomic level.[21,22] In principle, this involves solving the time-independent Schrödinger equation for *N* particles, where each particle is a nucleus or electron in the system, and accounting for the relevant potentials in the problem. The challenge is that solving Schrödinger's equation directly, either analytically or numerically, is impossible for all but the simplest cases. Therefore, we must approximate this equation in various ways and use different approaches to solve these approximations. This generally falls into two main approaches: quantum chemistry and molecular mechanics (MM). In quantum chemistry, we make approximations to the Schrödinger equation itself, retaining some fundamental features of QM. In MM, we approximate the system using classical mechanics, which makes the calculations easier. Figure 2 illustrates the relationship between different methods, system sizes, and the timescales attainable by molecular modeling methodologies. These computational methods often require high-performance computing and can involve long computing times.

The primary limitation in applying these methods is the computational tractability of the problem being addressed. Depending on the problem's nature and scale (i.e., whether dealing with electrons or entire molecules), different methods are suitable. This creates a trade-off between accuracy, resolution, and computation time. Practical constraints determine what types of problems can be modeled using these approaches, based on factors such as the system's size (measured by the number of atoms or molecules) and the complexity of changes occurring (for example, calculating the movement of electrons between orbitals is more computationally demanding than if they remain relatively static). The quantities we typically measure in experiments or expect in thermodynamic considerations correspond to ensemble averages of many microscopic processes. In molecular modeling, obtaining these averages requires generating and exploring many realizations of the system over time. The process of efficiently exploring these configurations—whether by initializing different starting points or dynamically evolving a single trajectory to ensure thorough phase-space coverage—is called sampling. In OoL research, rare events and non-equilibrium processes are particularly important, making enhanced sampling techniques useful as they allow for exploring a larger portion of the configuration space.[23] We explore these topics in more detail in the following sections.

Quantum chemistry

The goal of QM calculations is to predict the arrangement and behavior of electrons in molecules by solving the electronic (time-independent) Schrödinger equation. In OoL studies, where molecules are often complex, approximations are necessary to overcome computational challenges. One fundamental approximation used in many QM calculations is the Born-Oppenheimer approximation. This simplifies the problem by treating the movement of nuclei separately from the movement of electrons, allowing for a numerical solution to the Schrödinger equation to determine the electronic structure and energies of molecules based

Table 1. Systems and formalisms in modeling

Modeling framework	Formalism(s)	Use cases
Molecular modeling	Partial differential equations Schrödinger's equation and approximations	[3–6]
Chemical thermodynamics, kinetics, and networks	systems of linear equations, partial differential equations, chemical master equation, graph theory	[7–11]
Evolutionary modeling	ordinary and partial differential equations, rule base modeling	[12–15]
Information theory	probability theory	[16,17]
Molecular phylogenetics	Bayesian inference, maximum-likelihood methods	[18–20]

on their geometry. Wavefunction-based methods describe electrons as single-particle wavefunctions (orbitals) around a fixed nucleus. Early Hartree-Fock (HF) methods approximate the electronic wavefunction using the simplest combination of molecular orbitals (known as the Slater determinant) and optimize them numerically in a self-consistent field. However, HF methods do not account for electron correlation effects, which are important for chemical bonding.

To overcome this, post-HF methods such as Møller-Plesset perturbation or coupled-cluster theory[24] provide computational routines to account for dynamic electron correlation. Wavefunction-based methods are thus effective for calculating optimal molecular geometries, transition states, and vibrational spectra (see Part 1 Spectroscopy[1]). These methods have been used to interpret extraterrestrial spectra and identify molecular signatures in space,[25] as well as to calculate free energies of molecular interactions, such as the emergence of codons within DNA.[26] However, due to their high computational cost, wavefunction-based methods are generally limited to small molecules. To address this limitation, these methods often rely on sampling the molecular configuration space using lower-cost, lower-accuracy methods, including both quantum- and classical mechanics-based approaches.

Density functional theory (DFT) has gained popularity due to its computational affordability while delivering reliable predictions of molecular geometries and associated ground-state properties of a system. DFT describes the electron system by an electronic density instead of an electronic wavefunction—focusing on approximations that can be made on the electronic Hamiltonian.[27] Typically, the methods are based upon Kohn-Sham theory,[28] which introduces an orbital representation of the electronic density to better evaluate kinetic energy, while offering a different option to approximate the (in)famous exchange, and correlation energy and potential. If this exchange and correlation potential were to be known, Kohn-Sham DFT would be exact. Unfortunately, this is not possible, so the functional form of this potential needs to be approximate, leading to a range of density functional approximation (DFA) methods. DFAs are often classified in families such as local density approximation,[29] generalized gradient approximation,[30] or the very-commonly used hybrid (exact-exchange) functionals. These methods allow reliable prediction of molecular geometries and associated ground-state properties, including spectra, and allow for prediction and identification of the spectroscopic signatures of molecules applicable to their detection[31,32] (see Part 1 Spectroscopy[1]).

Notably, DFT can be used to study reaction mechanisms, predicting the transition states, activation barriers, and pathways taken. For example, developments in the DFT allow us to revisit the hypothesis of amino acid synthesis from alpha-keto acids via catalysis by dinucleotide species.[5] The computational analysis and details of the structures for the intermediates and transition states showed that there was wide scope for interactions between the keto acids and dinucleotide moieties and led to the required proto-metabolic selectivity.[33] Furthermore, quantum chemical calculations help generate hypotheses in the absence of experimental data. For example, they produced a testable mechanism for the formation of formamide.[34]

While finite molecular systems are at the center of abiogenesis research, condensed materials are also of great importance in the OoL setting.[35] Substrates such as mineral surfaces may have a key role in concentrating small molecules, catalyzing reactions toward the increased complexity of biological molecules.[36–38] The study of peptide-bond formation on aluminosilicate surfaces allowed the investigation of this hypothesis computationally, which demonstrated the feasibility for peptide formation on the clay surface.[39] The method itself modeled the substrate as an isolated cluster of atoms, assumed to be representative of the system. This approach represents both the molecule and substrate as a system of orbitals but is not always applicable to materials. For example, crystalline materials or metals will have a continuum of energy levels (bands) that are not centered on nuclei, in which case a planewave approach is adopted.[40]

The applications are often related to identification of potential catalytic surfaces or to provide theoretical evidence for a proposed reaction mechanism on a surface. For instance, a DFT study focused on CO_2 interactions with catalytic minerals such as iron sulfides (e.g., mackinawite: tetragonal FeS) revealed that, when these surfaces are doped with Ni, they exhibit weaker bindings to CO_2.[41] Furthermore, the DFT calculations used alongside ultraviolet-visible (UV-vis, see Part 1 Spectroscopy[1]) experimental studies have been effective in identification of products and intermediates of mineral-assisted formamide conversion to nucleobases, also allowing to identify corresponding features on the experimental spectra.[42] In another example, a QM study, combined with experiments and classical MM simulations (as discussed in "proteomics and transcriptomics"), enables the investigation of larger systems and the inclusion of temperature effects, helping to elucidate interactions between organic molecules and minerals under early Earth conditions.[43]

The mechanism of formation of early organic molecules, such as formamide, has been described by QM models and shown to be possible at extremely low temperatures.[44] However, it would not be fair to assume that the electronic structure of a molecule is not affected by its environment, as not every reaction can happen in a near-vacuum and at absolute zero. The substrate, solvent, temperature, surrounding ions, and proton gradients are all crucial in the study of OoL. From a modeling perspective, a solvent is a many-molecule matrix, coordinating to the solute

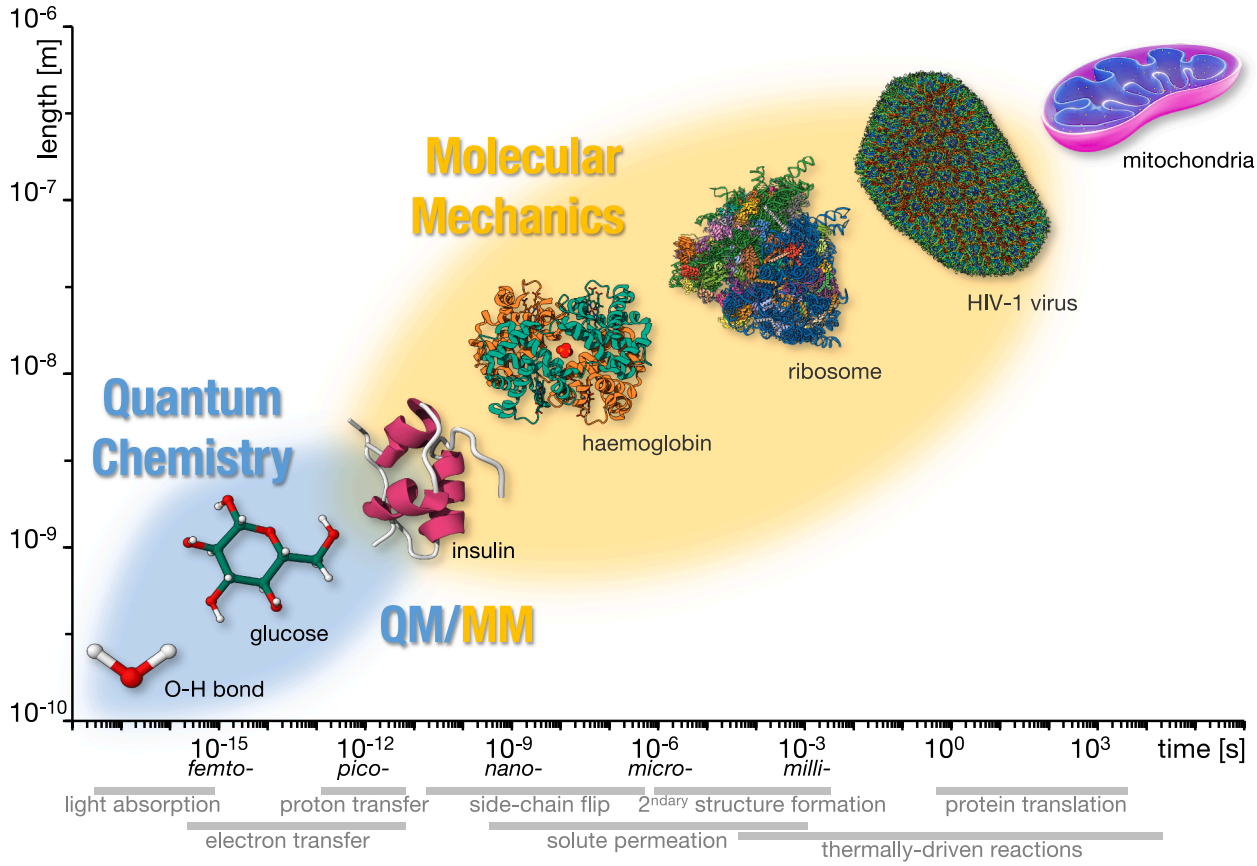

Figure 2. Relationship between timescale and system size in simulations, ranging from gas-phase quantum calculations (small molecules and bonds) to condensed-phase MM (large molecules) and biological systems and processes (viruses, cells)

The shaded areas indicate the approximate limits of quantum chemistry (QM) methods (blue) and MM methods (orange). The overlapping region represents the typical range explored by hybrid QM/MM methods. Visualizations of example systems with dynamics across the specified time and size scales are also shown.

molecule and creating a surrounding electrostatic medium. Due to the sheer number of atoms, this is inherently computationally demanding. The polarizable continuum (PC) model[45] allows the inclusion of dielectric bulk solvent effects, and thereby to drop the explicit solvent molecules. Nevertheless, solvent molecules are coordinating the solute, coupling motions, allowing molecular coordination, damping the motions, and allowing for energy exchange between species.[46] Therefore, these molecules should be included. In this case, the computing effort is dedicated to the region of interest—the molecule—while the necessary but not chemically interesting areas are reduced to a minimal representation.[47] Similarly, QM/MM mixed methods allow the computation of a relatively large system by reducing the surroundings of a molecule of interest to the classical (i.e., electron-free) MM representation. For example, to study a hydrogenation reaction of isocyanic acid on amorphous solid water to form formamide, the QM/MM approach allows screening the binding sites on the surface, calculating the activation energies, and to identify the tunneling mechanism of this reaction occurring at the interstellar temperatures of 103 K.[48] Similarly, a combination of a three-layer QM/MM framework was used to investigate hydrogen-cyanide isomerization on an icy grain. This approach includes a larger number of molecules surrounding the reaction site and it highlights tunneling as a main mechanism at low temperatures.[49]

The temperature of the surrounding environment will result in a molecular motion, affecting properties such as molecular vibrations, anharmonic motions, and diffusion collision of species. Within the Born-Oppenheimer approximation, heavy nuclear motion can be determined through Newton's law, with nuclear forces provided by calculating each time step of the dynamics, following the gradient of the electronic energy obtained with QM calculations. Through the use of molecular dynamics, we can incorporate energy transfer between molecules through collisions, allowing them to overcome the energy barrier necessary for activation of a reaction. To this end, the *ab initio* molecular dynamics (AIMD) nanoreactor approach has been successfully used to study reactivity of aqueous HCN, suggesting that it could be a source of RNA and protein precursors.[3] Obtaining time-average properties requires a long dynamic calculation and makes AIMD an expensive method,[50] typically allowing simulation under a nanosecond timescale.[3] To gather statistically meaningful sampling within the attainable computational resources, AIMD has been combined with machine learning to

gather accurate free-energy profiles for prebiotic chemical reactions.[51]

All of the QM methods discussed above are suitable for the ground-state representation of electronic structure. While ground molecular states are applicable to many chemical scenarios on Earth, these methods cannot be used for the study of light-activated processes, such as ones occurring in interstellar space or atmospheres. These chemical processes are of a particular interest to the formation of proto-biomolecules.[52] In order to study photoexcitation, the extensions to the wavefunction methods are commonly used,[53,54] and, in the world of DFT, linear-response time-dependent DFT is one of the most frequently used approaches to study excited states of a molecule. These methods are orders of magnitude more computationally demanding, which is currently limiting their application to the study of molecules with relevance to OoL. Light has been a key component allowing for the creation of chemical complexity in the interstellar media, as in the study of hydroxylated naphthalene on dust grains discussed above[55] or toward formation of abiotic precursors of the pyrimidine ribonucleotides.[56]

With the developments of additional QM methods and increased availability of computing resources, larger and more complex systems are amenable to these methods. However, there is still a strict limitation of the system sizes and timescales that can be modeled using QM methods, and molecular dynamics enables further analysis at these scales.

Molecular dynamics

MM is another valuable tool for modeling systems in which electronic structural changes are not of primary importance (i.e., no reactivity, excited states or changes in the chemical bonding). In these situations, molecular structures play a crucial role in determining the chemical-physical properties and functionality. The most widely used technique in MM is molecular dynamics (MD). In MD, classical (Newtonian) equations of motion are solved at each time step, resulting in a trajectory of atomic motions over a specified time period. Sufficient sampling is necessary to ensure that the time-averaged properties of the system are representative of the macroscopic thermodynamic ensemble. In MM, molecules are represented by atoms, each depicted as a sphere with specific radius, softness, and charge. These atoms are connected by bonds, represented by springs with specific lengths and stiffness, along with equilibrium angles and dihedrals. These parameters are defined in a force field, which is calibrated based on QM calculations and experimental measurements for specific systems. As a result, various force fields have been developed, many of which are specifically designed for liquid organic or biomolecular systems. It is worth noting that simpler mesoscale simulations, such as coarse-grained MD and dissipative particle dynamics (DPD), can be employed effectively for investigating general chemistry concepts while mitigating computational costs. Coarse-grained MD and DPD simulations are widely accepted as reliable approaches for accurately simulating chemical phenomena.[57] These methods have also been applied in OoL research.[58,59] Furthermore, extension from molecular models to lattice models, powered by parameters derived from the MM simulations, allows modeling slower processes than attainable by MM alone. For instance, this approach is used to study the polymerization of nucleotides, enabled through diffusion and aggregation within a membrane—a large and slow-moving molecular entity.[60]

A notable use of MD in OoL research is to provide a mechanistic explanation for the emergence of selection processes in complex mixtures. This includes selective synthesis of peptides[6,61] and nucleotides[6,62] on mineral surfaces, selective permeation of sugars across membranes,[63] and composition-based selective self-assembly of lipids.[64] A use of reactive force fields may also facilitate covalent reactions to better study their mechanisms[65] or generate chemical reaction networks.[65,66] This theoretical framework has been successfully combined with experimental work,[67] providing detailed predictions that can be explored experimentally regarding the emergence of reproduction and evolution.

Structure prediction techniques have proved useful in identifying conserved motifs and structures within the ribosome[68] and modeling short peptide sequences containing active sites of modern proteins, such as aminoacyl-tRNA synthetases. These techniques enable the confirmation of reasonable 3D structures before synthesizing the sequences for wet-lab experiments.[69,70]

Enhanced sampling techniques

The key to any molecular modeling is ensuring a correct thermodynamic phase-space sampling is achieved. To this end, accelerated sampling techniques that bias or modify the potential energy landscape are employed. Examples of such methods include metadynamics, umbrella sampling, and variationally enhanced sampling.[71–73] These methods can be applied to both QM and MM models. Furthermore, molecular docking allows for steered the assessment of molecular interactions between a larger molecular system and a smaller (docked) molecule by steering the sampling. It provides a useful approach for quickly evaluating interactions with lower computational costs compared to standard equilibrium QM or MD simulations.[74] This speed-up enabled the exploration of problems such as the interaction of all 1,280 combinations of proteinogenic amino acids with all nucleotide triplets.[75] However, its utility may be limited in systems with significant flexibility or when considering the importance of water, ions, charge, and geometry in interactions. An extensive review explores the application of molecular modeling methods to prebiotic chemistry in greater detail.[22]

Chemical thermodynamics, kinetics, and networks

Chemical reactions can be analyzed at a coarser level than the QM and MD scales described in the previous section. In many cases, we have information about the intrinsic properties of molecules we are interested in and what reactions can occur between them. We can use the information about the molecules and the reactions to predict how their concentration will change through time. Broadly, there are two approaches to this: equilibrium chemical thermodynamics, which is primarily interested in the relative abundances of chemical species at equilibrium, and non-equilibrium kinetic calculations, which is interested in the relative rates of different chemical reactions. Many kinetically controlled biochemical systems are modeled using the formalism of chemical reaction networks, a useful approach when many reactions occur simultaneously. Chemical kinetic

describes the rates at which reactions occur in a chemical system and how the concentrations of species change over time, whereas thermodynamic equilibrium only describes properties that are fixed at equilibrium.

In both approaches, we aim to solve a combination of linear systems of equations or coupled differential equations. Both approaches are concerned with chemical reactions that describe how molecules transform. Every reaction takes the following form:

$$\sum_i \nu_i X_i \rightleftarrows \sum_i \gamma_i X_i. \quad \text{(Equation 1)}$$

This expression represents a reversible reaction that converts ν_1 molecules of species X_1, ν_2 molecules of X_2, etc., into γ_1 molecules of species X_1, $\gamma_2$molecules of X_2, and so on. The species consumed by the reaction are called reactants, while the species produced are called products. The stoichiometric coefficients ν_i and γ_i specify the quantity of reactants and products involved in the reaction.

Both equilibrium and non-equilibrium approaches to modeling chemical reactions are common in OoL studies. The equilibrium approach is based on thermodynamic theory and is used in environments that locally reach thermodynamic equilibrium, such as the interiors of asteroids or the lower layers of gas giant planets. The non-equilibrium approach is based on chemical kinetic theory as well as the growing field of non-equilibrium thermodynamics, and it is used in environments that are constantly driven out of equilibrium, such as terrestrial atmospheres, and biological enzymes operating in non-equilibrium living matter. For detailed discussions of these topics, see Kondepudi and Prigogine[76] for chemical thermodynamics basics, Palsson[77] for an introduction to chemical reaction networks and chemical kinetics, and Qian and Ge and Rao and Esposito[78,79] for advanced discussions on non-equilibrium thermodynamics of chemical systems. A thorough review of prebiotic chemical reactions and networks can be found in Ruiz-Mirazo et al.[80] Another general resource about artificial chemistries is Banzhaf and Yamamoto,[81] which includes the contribution of several of these models—such as autocatalytic sets or chemical reaction networks—to OoL research.

Chemical thermodynamics

Thermodynamics explores how energy is transformed and conserved within systems, influencing the behavior and equilibrium of matter. A fundamental concept in this field is thermodynamic equilibrium, where a system achieves a state with the lowest free energy and its macroscopic properties cease to change over time. Closed systems naturally progress toward thermodynamic equilibrium, as do open systems that are coupled to equilibrium environments. However, open systems in non-equilibrium environments can be driven out of equilibrium by free-energy fluxes, for example from temperature gradients, electrochemical potentials, light, and radioactive decay. Exchanges of matter, for instance via the inflow of high-energy chemical fuel or asteroid impacts, can also maintain a system out of equilibrium.

Often the goal of chemical thermodynamics is to predict the concentrations (or related quantities) of different molecules at equilibrium. Doing this correctly requires knowing the standard Gibbs free energy of formation (ΔG_f°) for all the molecules that could exist in the system.[82] The standard Gibbs free energy of formation is the change in free energy of a system when a molecule is assembled from individual elements at standard concentrations,[82] and it depends on the temperature and pressure. Usually this quantity is measured in the lab for individual molecules, for a range of temperatures and pressures, and these values can be used to predict its value in different conditions.[82,83] It is also possible in principle to calculate the standard Gibbs free energy of formation using quantum chemical calculations (described in molecular modeling and simulations), but this is frequently not done.

The equilibrium composition of a system, which is the number of moles of each species, achieves the lowest total Gibbs free energy for the entire system.[82] Equilibrium compositions can be predicted by minimizing the Gibbs free energy using numerical techniques, as performed by various open-source and proprietary software, including Cantera,[84] OpenCalphad,[85] and ChemApp.[86] The inputs for these software tools are the standard Gibbs free energies of formation ΔG_f° for each atom or molecule in the system as a function of temperature and pressure, and the system's initial composition, temperature, and pressure. The output is the equilibrium composition. Reaction yields can be predicted by restricting the model to the species involved in reactions of interest.

At chemical equilibrium, forward and reverse reactions balance each other, resulting in no net conversion of reactants to products or vice versa. This state is also referred to as detailed balance. Outside of equilibrium, the affinity of a chemical reaction indicates whether a reaction is favorable to occur. Reaction affinities are defined as the negative of the Gibbs free-energy changes of the reaction (e.g., $A = -\Delta G_r$). A positive affinity indicates a reaction that can proceed spontaneously, while a negative affinity corresponds to a thermodynamically unfavorable reaction that can only occur by coupling to another thermodynamically favorable process. The simplest method to calculate reaction affinities uses the standard Gibbs free energies of formation of products and reactants,

$$A = -\Delta G_r = \Delta G^\circ_{f,\text{reactants}} - \Delta G^\circ_{f,\text{products}} - RT\ln Q. \quad \text{(Equation 2)}$$

Here, R is the gas constant, T the temperature, and $Q = \prod [X_i]^{\gamma_i - \nu_i}$ is the so-called reaction quotient, which quantifies the contribution from the (possibly non-equilibrium) concentrations of reactants and products (see Equation 1).

Common thermodynamic calculations include computing equilibrium compositions, reaction yields, and chemical reaction affinities (i.e., whether a reaction will occur spontaneously or not). The key data for these calculations are the standard Gibbs free energies of formation for the species involved and initial concentrations. Multiple databases provide these energies, based on laboratory measurements, including GRI-Mech 3.0,[87] CHNOSZ,[88] and JANAF.[89] Gibbs free energies of formation for specific molecules can also be calculated using quantum chemistry methods (e.g., Paschek et al.,[90] and see section molecular modeling and simulation and section quantum chemistry).

The validity of an equilibrium or non-equilibrium approach depends not only on how the timescales of disequilibrium sources compare to those of the parameters being calculated but also on

whether the system is open or closed, as equilibrium assumptions strictly apply only in the infinite time limit and within closed systems. For instance, consider the atmospheres of terrestrial planets such as Earth in contrast to those of gas giants such as Jupiter and Saturn. Earth's atmosphere is transparent to long-wave UV light. The photon energy of UV light breaks molecular bonds keeping Earth's atmosphere out of equilibrium. Ozone (O_3) is produced via disequilibrium UV chemistry; therefore, a purely equilibrium analysis of the Earth's atmosphere would incorrectly predict that no ozone layer is produced, suggesting the need for a non-equilibrium analysis. In contrast, the majority of Jupiter's and Saturn's atmospheres are opaque to UV light. Furthermore, the lower layers of these atmospheres are hot and thermodynamic equilibrium is reached for many molecular species before these gases are transported to cooler regions of the atmosphere and become kinetically inhibited.[91,92]

Typically, chemical systems at lower temperatures take longer to reach thermodynamic equilibrium. To assess whether reactions involving a molecular species (such as HCN in Saturn's atmosphere) have enough time to reach equilibrium, one can look at the time constant for the fastest reactions that produce and destroy that species. For instance, in the deep layers of Saturn's atmosphere where the pressure is 10 kbar and the temperature is 2,000 K, the critical reaction involving HCN is its destruction, represented as $HCN + H_2 \rightarrow CH_2 + NH$. The rate constant for this reaction is $k(T) = 1.08e^{-70{,}456/T}\ M^{-1}s^{-1}$. M is molar, i.e., mol/L. The time constant for this reaction, $t(\text{HCN, 2000 K}) \approx 1/k(2000\text{K})[H_2] \approx 4$ s, which is much shorter than the year-long timescale of convective transport in these regions. Thus, it is likely that the reaction involving HCN can reach equilibrium within these environmental conditions, making a thermodynamic equilibrium approach suitable for analysis.

Thermodynamic calculations have played a role in various OoL studies, analyzing models of nucleobase, ribose, and amino acid synthesis within asteroid interiors,[7] lightning chemistry on primitive Earth,[8] impact-generated chemistry during the Hadean eon,[93] Archean mantle/volcanic outgassing chemistry,[94] and potential ancient metabolisms in hydrothermal systems.[9,95,96] While thermodynamic equilibrium calculations provide useful estimates for the chemistry in these settings, many environments are constantly driven out of equilibrium, making such models less effective. In such scenarios, a non-equilibrium analysis that incorporates kinetic information is often necessary.

Chemical kinetics

Chemical reaction networks (CRNs) provide a general modeling framework for studying the dynamics (sometimes called kinetics) of chemical reactions. A CRN consists of a set of chemical species, representing different types of molecules, and a set of reactions that convert these species into one another. Each reaction can be written in a general form as in Equation 1. A concrete example would be an enzyme C catalyzing the formation of product P from substrate S. This CRN consists of two reactions and four species (enzyme C, substrate S, a bound combination of substrate and enzyme SC, and product P):

$$S + C \rightleftharpoons SC \qquad \text{(Equation 3)}$$

$$SC \rightarrow P + C \qquad \text{(Equation 4)}$$

In CRNs, reactions can be either reversible ($\rightleftharpoons$, e.g., Equation 3) or irreversible ($\rightarrow$, e.g., Equation 4). Reversible reactions occur in both directions, while irreversible reactions proceed only in one direction. Equilibrium reactions are always reversible. It is also important to distinguish elementary and non-elementary reactions. An elementary reaction occurs in a single step, and a non-elementary reaction represents the net effect of a sequence of several elementary reactions. For example, Equations 3 and 4 can be represented by a single non-elementary reaction $S + C \rightarrow P + C$. In fact, this representation of enzymatic catalysis is used ubiquitously in biology, where it is called the Michaelis-Menten scheme.[97,98]

The dynamics of a CRN are typically represented by differential equations that reflect the rates (sometimes called fluxes) at which the different reactions occur. In general, reaction kinetics depend on particular details of the reaction volume, rate constants, temperature, external parameters, etc. The dynamics can be stochastic or deterministic; the deterministic approach is frequently used for large systems where microscopic fluctuations can be ignored. Mass-action kinetics are often used to model well-mixed deterministic systems. For mass-action kinetics, the net flux, J, across the reaction represented by Equation 1 can be written as

$$J = k^+ \prod_i [X_i]^{\nu_i} - k^- \prod_i [X_i]^{\gamma_i}, \qquad \text{(Equation 5)}$$

where k^+ and k^- are the forward and backward rate constants and $[X_i]$ indicates the concentration of species X_i at a given point in time. The concentrations then evolve according to the following differential equation, which is sometimes called the reaction rate equation:

$$\frac{d}{dt}[X_i] = (\gamma_i^r - \nu_i^r)\, J_r. \qquad \text{(Equation 6)}$$

Here, r indexes over different reactions present in the system. For non-elementary reactions, so-called Michael-Menten kinetics and other types of kinetics may be employed.[99,100]

Importantly, for elementary and reversible reactions, the fluxes can be related to the reaction affinities in Equation 2. Specifically, the affinity can be written as $A = RT\ln(J^+/J^-)$, where J^+ and J^- refer to the forward and reverse fluxes, represented by the first and second term in Equation 5. This is an important equation in non-equilibrium chemical thermodynamics since it relates reaction dynamics and thermodynamics.[76] A simple chemical kinetics simulation would solve the differential equation in Equation 6 to obtain the concentrations of all chemical species as a function of time. In simple cases, it is possible to obtain an analytical solution (i.e., exact functional form), while in many cases the solution is numerical (i.e., the time-dependent concentrations are calculated based on the set of rate constants and initial concentrations). In practice, chemical kinetics simulations could solve networks of hundreds to thousands of chemical reactions (i.e., rate equations) simultaneously. In some cases, chemical concentrations in these simulations reach a steady-state solution after a certain amount of time; this is not to be confused with thermodynamic equilibrium, which has no net reaction fluxes ($J_r = 0$ for all reactions r). Many open-source chemical kinetics solvers exist, including Kintecus,[101] Cantera,[84] and

ChemPy.[102] These solvers take as input a collection of chemical reactions with their temperature-dependent rate constants, as well as initial species concentrations, and output chemical abundances under the assumption that the system is well mixed. There exist chemical networks that contain collections of rate constants as a function of temperature calculated from experiments or quantum chemistry simulations or estimated using thermodynamics or similar reactions. Examples include CRAHCN-O,[103–105] KIDA,[106,107] STAND,[108] and UMIST.[109]

More relevant for OoL are non-equilibrium steady states of CRNs. One way to study such non-equilibrium states is using continuous stirred-tank reactors (CSTRs). CSTRs are reaction vessels (simulated or real) in which a constant total concentration is achieved by a continuous inflow of reagents and a continuous outflow while keeping a constant total volume. These conditions specify a non-equilibrium boundary condition and ensure that all reagents and products are diluted out of the system in proportion to their concentration while the total mass in the reactor remains constant. Chemical kinetics analysis has been used in many various OoL studies because almost all modeling approaches that involve tracking molecular concentrations through time involve some kind of kinetic analysis. A notable example is Semenov et al.,[10] where kinetic analysis was performed to characterize a bistable organic reaction network in a CSTR. In this study the authors measured molecular abundances using UV-vis spectroscopy (see Part 1, Spectroscopy[1]), and they modulated the inflow of different chemical species into the CSTR to drive complex dynamics without the use of enzymes. This work showed how the structure of the reaction network, and specifically the presence of an autocatalytic loop within the system, enabled the complex dynamics observed in the experiments. We next discuss the role of entire networks and their structural properties.

CRNs

In the previous sections, we discussed how the compositions or dynamics of CRNs can be studied. But the structure of the reactions and their relationships can also be analyzed without considering how the concentrations change through time or even the identity of the molecules involved.[110] This analysis can be done using CRN formalism.

Chemical reactions can be represented as complexes, which in turn can be represented by vectors. For example, if we consider the reactions discussed above with four species (*S*, *C*, *SC*, *P*), Equation 3 can be represented as a transition between two complexes: *S* + *C* (represented by [1, 1, 0, 0]) and *SC* (represented by [0, 0, 1, 0]).[110] CRNs represented as networks of complexes can be analyzed to determine static properties. Static network properties place limits on dynamical properties of CRNs, for example the deficiency zero theorem places constraints on the number and type of steady solutions a CRN can have, based only on the structure of the network itself.[110]

CRN models and derivatives based on genomic data have been used to identify essential and ancestral metabolic functions in prokaryotes, including tRNA-charging and cofactor metabolism.[11] Network expansion algorithms, which expand a chemical reaction network by iteratively adding all products that can be formed from reagents current in the network, have provided evidence of plausible proto-metabolic networks including predictions about potential phosphate-free[9] or organo-sulfur nitrogen-free metabolism at the OoL.[111]

In many cases, a CRN can be constructed by hand, using *a priori* knowledge of possible reactions and databases of previously cataloged reactions (see Part 1, Databases in OoL[1]). However, in many cases, we do not know the structure of the entire chemical reaction network or we only know some features. In those cases, we would like to be able to directly construct, or estimate, the entire network. This is the goal of automatic reaction network generation (ARNG). ARNG is a rule-based computational method that reconstructs CRNs from a limited set of predefined chemical transformations, called reaction rules, that happen on molecules due to reactions. Molecules are represented either symbolically, as in simplified molecular-input line-entry system (SMILES)[112] or as molecular graphs (bond-electron matrices).[113] Accordingly, the reaction rules are defined either as symbolic or as matrix transformations. Reaction rules typically correspond to reaction types. For example, in the reductive tricarboxylic acid (rTCA) cycle, reductive carboxylation can be represented as one reaction rule capturing atomic transformations between acetyl-CoA and pyruvate and between succinate and oxoglutarate.[114]

Starting with a set of available molecules, reaction rules are algorithmically applied to the reactive sites of the molecules, consequently transforming reactants into products, which in turn can become reactants in the following iterations.[115] Through successive iterations, the pool of available molecules is systematically expanded and represented as a reaction network. The feasibility of a reaction can be accessed by calculating the corresponding Gibbs free energy using either quantum chemical calculations (see molecular modeling and simulations)[116] or group contribution methods.[117,118]

The ARNG methodology has been implemented in software such as Netgen,[119] RMG,[120] AllChemy,[121] and in Arya et al.[122] For example, ARNG was used to suggest that the rTCA hypothesis can be merged with the glyoxylate hypothesis.[123–125] Work presented in Wołos et al.[121] used ARNG to address emergent phenomena in prebiotic reaction networks and trace synthesis of life's building blocks from a set of chemical compounds that were present in the atmosphere of early Earth.

Network autocatalysis

In its traditional definition, autocatalysis refers to any reaction in which a product is also a catalyst, and such a product is called an autocatalyst.[126,127] Network autocatalysis is, by definition, a multistep autocatalytic reaction, but the number of reaction steps can be extensive and the connections between reactions may be complicated, making a reaction network a more suitable representation than a simple net reaction equation. Since all known forms of life convert external nutrients into more of themselves through a vast array of reactions (i.e., life catalyzes the production of life), all forms of life can be viewed as autocatalysts engaged in network autocatalysis. However, present-day forms of life are not the only physicochemical systems capable of performing network autocatalysis. Abiotic reaction systems, such as the formose reaction,[128] the dissolution of copper in nitric acid,[129] and the Belousov-Zhabotinsky reaction,[130] are also examples of network autocatalysis, although their reaction networks are much simpler than a metabolic network.

Table 2. Categorizing models of network autocatalysis

Model	Are catalytic polymers necessary?	Is catalysis of all nodes necessary?	Notes
(*M*, *R*) systems	no	yes	assumes that every component of metabolism has a finite life-time, so the system must be able to 'repair' the components by reactions.
Hypercycles	yes	yes	originally proposed to explain how replication fidelity could be maintained without a long error-correction enzyme.
RAF	no	yes* (spontaneous catalyst can be added in food-set to overcome this)	every reaction needs to be explicitly catalyzed by at least a chemical species in the network. Stoichiometry is not implemented explicitly. Based on Kauffman's CAS model.
COT	no	no	definition of self-maintenance does not guarantee autocatalysis, but it is possible to enforce autocatalysis by assuming environmental openness. Every reaction within a chemical organization must have a positive flux.
Systems of stoichiometrically autocatalytic motifs	no	no	not every reaction within an autocatalytic system must have a positive flux, so it is possible to find minimal stoichiometrically autocatalytic motifs by setting some fluxes to zero. Networks of autocatalytic motifs can be understood as ecological communities.
GARD	no	no	based on catalytic networks. Proposes replication and evolution of compositional information.

Since self-propagation is a feature shared by life and certain much simpler abiotic systems, there has long been interest in using theoretical models of network autocatalysis to explain the origins of life's attributes, to find candidate processes underlying abiogenesis, and to direct experimental studies on abiogenesis. These models include but are not limited to (*M*, *R*) systems,[131] hypercycles,[12,132] the theory of reflexively autocatalytic food (RAF)-generated sets[133,134] (based on the Collectively Autocatalytic Sets [CASs][135]), and chemical organization theory (COT).[136]

Another type of network-autocatalysis model is the graded autocatalysis replication domain (GARD).[15,137–140] The model demonstrates how life-like properties may emerge in a mutually catalytic network of self-assembling amphiphiles forming assemblies such as micelles and vesicles that can collectively reproduce, store, and propagate compositional information.[15] There has been debate regarding the evolvability of these systems; see Markovitch and Krasnogor[140] for a discussion of this criticism.

These models have different emphases and attributes. To help readers assess which models are suitable for their purposes, we briefly categorize these models in Table 2. Interested readers can also refer to Hordijk and Steel[141] for a review of several models of autocatalytic sets and to Letelier et al.[142] for a broader discussion of the historical context centered on the concept of catalytic closure.

Network-autocatalysis models have multiple applications in OoL research. For example, the theories of autocatalytic sets can help search for collectively propagating RNA systems.[143] One may also use RAF theory and COT to assess the conditions for a pathway to arise from simple nutrients to specific cofactors or more complex molecules.[134,144] In addition, minimal stoichiometrically autocatalytic motifs in databases of chemical reactions can be computationally detected, which makes it easier to design experiments aimed at constructing autocatalytic systems, and analyses of network autocatalysis could suggest possible routes of prebiotic evolution leading from simple systems to more complex ones.[145,146]

Modeling of evolutionary dynamics and (proto)cells

A growing body of work employs differential equations, simulations, and agent-based models to probe the complex dynamics relevant to the OoL that extend beyond traditional chemistry. Various systems lend themselves to this modeling approach, particularly those involving evolutionary dynamics and protocell modeling. Consistent with the previous section, these models are typically encoded as ordinary or partial differential equations,

and the techniques for solving these systems involve different approximations or solution methods. However, an alternative method known as rule based modeling can also be used. Rule-based modeling uses predefined simple rules of interaction between components to understand the collective behavior of the entire system. Agent-based models are the most common example of rule-based systems. The ensuing discussion provides a brief review of prebiotic replicator models, related models incorporating agent-based entities, and computational models of protocells.

Replicator models

Prebiotic evolution encounters challenges similar to those of cellular life, such as maintaining and utilizing information to interact with the environment and other species. Consequently, prebiotic replicator models have both significantly influenced and been influenced by other fields. Replicator models generally assume that a molecular species, such as RNA, can replicate and mutate. Often, these models abstract away the metabolic details of replication to focus specifically on Darwinian evolutionary properties. They typically provide either an explicit selection criterion, such as a fitness function, or an implicit evolutionary pressure, such as cooperation.

A convenient aspect of replicator models is that they decouple the origin of Darwinian evolution from the origin of cellular life, allowing the exploration of replicators and environments that have not yet been experimentally created, such as those set in virtual geological environments like mineral surfaces, ice crystals, and porous rocks.[147] A growing body of research has developed models that display various characteristics of replicators and their environments.[148,149] Below, we briefly outline some key models to demonstrate how they enhance our understanding of prebiotic evolution.

Replicator models are typically formulated in terms of systems of coupled ordinary differential equations. Differential equation models treat populations of replicators as continuous quantities that represent mean concentrations and are therefore not well suited for studying situations where small numbers of replicators are important. However, stochastic solution methods can accommodate small number limits. Similarly rule-based models can often be implemented such that they recapitulate the differential equations when the number of replicators grows.[150] Each equation represents the concentration of a molecular species, which changes according to its replication rate, mutation rate, and ecological interactions with other species. Important applications of this approach are the quasispecies model and hypercycle model.[12,151]

The evolutionary dynamics of replicator models are typically formulated in terms of mutation-selection mechanisms governing an evolving system, which allows for significant complexity. This formulation consists of two parts: a matrix that specifies how each molecular species mutates into others and a growth rate matrix that specifies how fast each molecular species grows in the absence of mutations. The mutation matrix—effectively a network that connects mutationally adjacent molecules—can describe the effect of realistic types of mutations, such as nucleotide substitutions, recombination, or others, while the growth rate matrix can be fitted to experimental data.[152]

These models have shown that replicators with lower replication rates but higher connectivity in the mutational network can outcompete faster-replicating but less-connected species when mutation rates are high[15,153] (the so-called survival-of-the-flattest effect). More recently, these models have clarified the effect of lethal mutations on the survival of the population[154] as well as the effect of selection acting on the phenotype of a species (e.g., its secondary structure).[155] Although these models are often solved numerically, some simplified cases admit analytical solutions.[156]

Replicator models can be classified in terms of their growth kinetics.[157–159] In particular, the concentration of a growing replicator can be modeled with the differential equation $\frac{d}{dt}\,x(t) = c\,x(t)^p$ where c is a constant and p is called the order of a replicator. The most common replicators are called first-order replicators and have $p = 1$, resulting in exponential growth $x(t) = x(0)e^{ct}$. This growth is characteristic of autocatalytic reactions such as $S + X \rightarrow 2X$ or networks of such replicators. Second-order replicators with $p = 2$ are characteristic of autocatalytic reactions $S + 2X \rightarrow 3X$, or networks of such reactions, and lead to so-called hyperbolic growth. Second-order growth occurs when two replicators are necessary to replicate, as in sexual replication or other kinds of mutualistic interactions. On the other hand, so-called parabolic replicators have $p = 1/2$. Such kinetics were originally observed in early experimental work on self-replicating templating RNA molecules.[160,161] An in-depth review and analysis of the theoretical issues can be found elsewhere.[157–159]

It is important to emphasize that the order of autocatalytic growth is not just a mathematical parameter but has important dynamical and evolutionary consequences. In particular, second-order replication leads to a nonlinear differential equation with much richer dynamics than first-order replication. For this reason, second-order autocatalysis plays a central role in several well-known models of bistability, oscillations, and pattern formation. This includes Eigen's hypercycle model of collective self-replication,[12] the Schlögl model of bistability,[162,163] the core dynamics of the Brusselator model of oscillations,[164] and the Gray-Scott model of pattern formation.[165] On the other hand, parabolic and other subexponential replicators favor weaker selection with co-existence at steady state.[166] For a brief review of these topics, see Szathmáry and Szathmáry and Gladkih.[158,166]

Agent-based models

Individual-based models (also called agent-based models) describe populations of individual replicators, where each replicator is assigned specific properties that can determine its replication potential. This approach is closely related to—but still distinct from—replicator models. Upon replication, a copy of the selected individual is made, which may mutate, thus allowing a Darwinian process to take place. These models can be extended more easily than their continuous counterparts to include additional complexity.

One of the most successful extensions consists of introducing a genotype-to-phenotype (GP) map. In the case of RNA, fast minimum-free-energy folding algorithms mapping sequences to secondary structures[167] introduce one such GP map. This allows us to explicitly study RNA evolution as a process whereby

mutations affect the genotype and selection acts on the phenotype (although there are conceptual ambiguities associated with genotypes and phenotypes embedded in the same molecule[168]). The evolutionary dynamics of populations of RNA replicators can be extremely rich, diverse, and life-like—with complex patterns of mutational neutrality,[169] evolvability, and endless potential for innovation.[132,169] Encapsulating a population of RNAs into protocells, each containing a small number of RNAs, has been shown to significantly improve resistance to parasites compared to populations without encapsulation.[13] Group selection at the protocell level can counterbalance template selfishness, as the protocells with a more balanced number of different RNAs (or with fewer parasites) can divide sooner, and even unfit protocells can stochastically re-generate a fit mixture of RNAs (i.e., the stochastic-corrector model).

In individual-based models that include spatial structure, replicators are assumed to occupy nodes on a lattice, representing a porous surface that limits diffusion. Discrete evolutionary rules represent the dynamics of the system in a simplified and computationally efficient manner. Replicators can increase in number by copying themselves into adjacent nodes in the lattice and interact with their immediate neighborhood. Using such models, it was found that local interactions between replicators spatially limit the replication of parasitic templates[148,170] and generate emergent forms of organization between cooperating replicators and their parasites.[171,172] While these models were originally developed as stochastic cellular automata, they are now typically interpreted as spatially extended agent-based models. This enables increased complexity, for instance by equipping the replicators with a GP map as above.[14,173] Replicators can also be modeled to catalyze reactions and collectively generate a metabolism,[174] and chemical variants of replicators, e.g., RNA vs. DNA, can be accounted for.[175]

Alternatively, models based on evolutionary game theory have further simplified replicator dynamics by focusing solely on cooperative phenomena.[176] Game theory has been used to investigate microscale dynamics in the laboratory,[177,178] and analogous lattice-based models have been used to show that cooperation can readily emerge at the microscale.[179]

Whole (proto)cell models

Complementary to replicator models discussed above, other approaches aim to model whole cells to predict phenotype from genotype and the environment, providing a mechanistic understanding of how an entire cell works. Similar to genome-scale metabolic models, these models represent the function of each gene, gene product, and metabolite.[180,181] They also represent multiscale interactions at the cellular level; including the cellular structure, dynamic structure of molecular interactions, and the spatial compartment of the subcellular components.[182] The dynamics of each subsystem is usually described with ordinary differential equations, as a set of linear constraints to be solved, or with stochastic simulations. To date, only two models have been developed: a model for the simple bacterium *Mycoplasma genitalium*,[183] and a model for *Escherichia coli*.[184] More recently, a model of the "minimal biological cell" JCVI-syn3A, a genetically minimal cell with only 493 genes, including 452 protein-coding genes,[185] was developed.[180] This represents the closest attempt to developing a nearly complete, 3D spatially resolved whole-cell kinetic model, describing growth emerging from metabolism and gene expression (including the contribution of integral membrane proteins and lipids, and other cellular components). This model couples ordinary differential equations and stochastic simulations to handle both the kinetics of metabolic networks and the kinetics of genetic processes, respectively, and overall accounts for over 7,000 reactions. The model gives quantitative insight into how the cell balances the demands of metabolism, genetic information, and growth over a cell cycle. The emergent behaviors arising from the simulations provide valuable understanding of the principles of life for minimal cells. Future development of similar models could be used in the context of the OoL.

Information-theoretic approaches

Information theory, the mathematical study of quantification, storage, transfer, and communication, was first formalized by Shannon in 1948.[186] Despite its late application to OoL research, it has gained significant interest in recent decades. Integrating information into OoL studies assumes that all life phenomena are governed by information and therefore must account for the informational capabilities of living cells.[187] In the following sections, we provide a brief review of how the application of concepts related to information, particularly information theory, has contributed to our understanding of the transition from non-life to life. This includes the emergence of precursor mechanisms in physical and abiotic chemical systems, as well as new definitions of the living state.

Information theory introduces the bit as a fundamental quantity. Shannon information is based on the amount of information that is provided to a receiver after they have decoded an unknown message from a sender. A message is modeled as a discrete random process, where $x \in X$ represents a random variable. The probability of an outcome x is denoted as $p(x)$, and the amount of information provided by outcome x is given by $h(x) = log_2 \frac{1}{p(x)}$, measured in bits. Intuitively, $h(x)$ reflects the "amount of surprise" in observing a specific outcome. The entropy, which measures the average surprise across all outcomes in the set X, is given by

$$H(X) = \sum_{x \in X} p(x) \log_2 \frac{1}{p(x)}. \quad \text{(Equation 7)}$$

The Shannon definition has been instrumental in investigating key elements of living systems and the OoL. As evolution represents the change in living systems over time, in informatic terms it is a processing of genetic information across large timescales. This understanding has led authors to propose that natural selection acts as an information acquisition process, and to describe a relationship between both information gained by natural selection and population growth by formalizing the relationship between information and fitness,[188] and has been recently revisited in light of computational learning theory.[189] Similarly, the ways in which living systems acquire information through evolutionary processes have also been investigated using Shannon information.[190]

Information theory has also been applied to the OoL by quantitatively analyzing the probability of the emergence of primitive replicators from an abiotic environment. Calculating the

likelihood of polymer assembly from random processes shows that the likelihood (rate) of monomer formation can be extremely low. However, considering the biased probability distribution in which they are found in functional informative molecules, these likelihoods increase dramatically.[16] These results align with earlier investigations using the Shannon-McMillan-Breiman theorem to derive similar values.[191]

Additional applications of Shannon information in investigating the OoL involve the concept of information complexity. Life is often perceived as more complex than non-life. To make this observation scientifically useful, it needs to be paired with a concrete definition of complexity. Information theory has provided candidate complexity measures that have been used to characterize the differences between non-living and living systems. For instance, information complexity can serve as a proxy for functional complexity,[192] a view that is experimentally supported by *in vitro* work on ribozyme functionality.[193]

Mutual information, a measure derived from Shannon information, can be used to determine how the entropy of one variable relates to the entropy of another variable,[194] and is defined as follows:

$$I(X,Y) = H(X) + H(Y) - H(X,Y). \qquad \text{(Equation 8)}$$

In Equation 8, $H(X)$ and $H(Y)$ are the entropies of two variables X and Y, while $H(X,Y)$ is the entropy of the joint distribution of both, which is calculated as $H(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\, log_2 \frac{1}{p(x,y)}$. This information-theoretic measure of the relationship between variables has been utilized to investigate the mutual dependence between replicators and their environment. Analysis using a model of replicators coupled to their environment through the recycling of resources showed that the transition from non-life to life could coincide with a phase transition measured through the mutual information between them.[17]

Entropy can also be used to compare distributions. The Kullback-Leibler divergence, or relative entropy, can be used to compare two distributions $p(x)$ and $q(x)$.[194] It is defined as

$$D_{KL}(p\|q) = \sum_{x\in X} p(x) log \left(\frac{p(x)}{q(x)}\right). \qquad \text{(Equation 9)}$$

Information accumulation and maintenance in the genome is one example where Kullback-Leibler has been proved useful in investigating the OoL: it has been shown that the amount of information that can be maintained in the genome at a given cost might scale with population size and mutation rate.[195] Other topics, such as fitness and optimal information processing, have also been studied along similar lines.[196]

Related fields study reproducing features of living systems *in silico* or *in vitro* and can also inform OoL research. *In silico* studies investigate problems through an information-theoretic perspective, such as artificial cell design.[197] Other topics of investigation include information storage[198] and advantages of multicellular consortia relative to information storage.[199] Biological networks and organization principles are also relevant to OoL. Examples of these investigations include research focused on information transmission in complex networks typical of living systems,[200] uses of information in deciphering hierarchical organization in biological systems,[201] and Boolean models of biological networks to determine whether information processing is an intrinsic property.[202] Other uses of information theory in research related to the OoL include that of the Shannon's channel capacity theorem to analyze the problem of transmission in translated or decoded codons.[191]

An informational perspective allows for an integration of the different fields of thought associated with OoL research by providing a common mathematical framework enabling cross-disciplinary collaboration. Such a perspective can thus be involved in every step leading from physical systems to biological ones by providing quantifiable measurements of complexity and information dynamics having been identified as a key property of living systems. A common framework is essential to answer such a multi-faceted question as the OoL problem.

Molecular phylogenetics

Molecular phylogenetics studies the hierarchical evolutionary relations, based on shared ancestry, between extant and extinct taxa. Darwin's "Origin of Species" contains the first known phylogenetic tree representing the common origin of many taxa known at that time.[203] Molecular genetics and sequencing technologies (see Part 1, Genomic sequencing[1]) enable the construction of trees using the information embedded in DNA, RNA, and protein polymeric sequences. Many applications in OoL research are common with other OoL problems that center on the evolution of traits such as metabolic pathways and the cell physiology of extinct or unknown organisms. Research focuses in this area include the inference of the last universal common ancestor (LUCA) characteristics based on phylogenetic analyses of genes,[204] the transition from the prebiotic world to biotic worlds,[205,206] and the resolution of taxonomic relationships between bacteria and archaea.[207] Gene and genome trees have shed light on the origin and evolution of novel enzymes and novel metabolic pathways, and their ancestral states have been reconstructed by statistical evolutionary models.[208,209] Phylogenetic methods have been important to establish the ubiquitous nature of lateral gene transfer (LGT) between both ancient and extant organisms.[210,211] Other research has used phylogenetic methods to evaluate specific evolutionary hypotheses regarding the OoL. For example, previous work has demonstrated that the emergence of both LGT and the frozen genetic code likely predated LUCA, explaining how aminoacyl-tRNA synthetases were shared between branches of life.[212,213] Phylogenetic models have also been used to identify the root of the Tree of Life,[214] helping to identify which extant organisms are most likely to share characteristics with LUCA. Similarly, phylogenomics has provided evidence of the origin of eukaryotes within the Archaeal domain, particularly within the Asgard Archaea.[215] This has been recently identified as a key group closely related to the ancestry of eukaryotes.[20,216] Finally, these approaches have reconstructed the evolutionary origin of specific metabolic pathways[18] and the possible metabolic capabilities of early life.[204,217]

Homology and functional gene annotation

In phylogenetics, common ancestry is deduced from trait similarity. However, careful analysis is required to distinguish traits with common ancestry (homologous) from those with close function but different origin (homoplasy) or those sharing ancestry but being an evolutionary novelty (apomorphic). In molecular phylogenetics, homology in DNA, RNA, and protein sequences

is identified from sequence similarity. Homologous sequences that retain similar functions through speciation events are referred to as orthologs. On the other hand, genes that emerge via duplication events, referred to as paralogs, can acquire new or specialized functions in their subsequent, distinct evolutionary path. Paralogs can be an important source of information when building phylogenies, and the distinction may be relevant when using phylogenetic trees for ancestral state reconstruction.[218] Also, sections of the genome can jump horizontally between evolutionarily distant taxa by LGT rather than by vertical inheritance. LGT can invalidate certain assumptions for tree building,[219] distance matrix calculations, and, particularly, "molecular clock" calibration[220,221]

Sequence homology underlies methods for gene and protein functional annotation in the absence of experimental characterization. There are several databases with sequences annotated for functions that are either experimentally known or imputed based on sequence information. Sequence matching algorithms,[222], such as OrthoMCL, inParanoid, or Reciprocal Best Hits, are used to annotate genes or proteins of unknown function. Popular databases are Clusters of Ortholog Sequences, the Kyoto Encyclopedia of Genes and Genomes (KEGG), EggNOG, and OrthoFinder among many others.[223,224] Popular servers for prokaryotic gene annotation are RAST or MEGAN.[225,226] Similar methods are used to annotate protein function in metagenomic assembled genomes in environmental metagenomics studies.[227] Other methodologies permit the reconstruction of metabolic networks from functional annotation by using orthologs linked to metabolic pathways, such as KEGG mapper,[228] COG pathways[229] or modelSEED,[230] and RAST server to search feasible metabolisms in a given environment.[230–232] Ortholog detection algorithms also help to identify LGT in microbial evolution[233,234] and date evolutionary events.[235,236]

Constructing trees

The origin of eukaryotes within the Archaeal domain and the phylogenetic structure of prokaryotes highlight the growing importance of phylogenetic trees in understanding the complex evolutionary relationships of early life forms.[237,238] Phylogenies are built by comparing traits among taxa. Traits are typically morphological, either coded as binary (e.g., presence or absence of an organelle) or numeric (e.g., number of cilia), or functional (e.g., ability to produce and consume a given metabolite or certain amino acid metabolic pathway synthesis).[239] In the case of sequence-based phylogenies, the corresponding traits would be mutations in the DNA, RNA, or protein sequence (i.e., nucleotide and amino acid substitutions, insertions, or deletions). Sets of informative traits are collected in tables and passed to phylogenetic tree-building algorithms that use either the original character matrix or a derived distance matrix. Distance matrices are double entry tables filled with values proportional to the number of sequence changes calculated after sequence alignment. In pairwise sequence alignment, all combinations of sequence pairs are arranged so their homologous sites are adjacent, either using global[240] or local alignments.[241] Molecular phylogenetic reconstruction uses multiple sequence alignments (MSAs) constructed by algorithms such as CLUSTAL-Omega,[242] MAFFT,[243] and MUSCLE.[244] There are recent alignment-free methods usable to calculate distance matrices from molecular data.[245,246] Finally, some alignment methods leverage the greater conservation of protein structure over sequence[247] and integrate available structural information into the alignment procedure.[248]

A second task required for modern phylogenetic inference methods is the selection of a best-fit molecular evolutionary model. These models integrate features of molecular evolution, including the rates of pairwise nucleotide or amino acid substitutions, evolutionary rate heterogeneity across different positions in the MSA, and base frequencies of each nucleotide or amino acid. Model testing can be carried out by maximum-likelihood-based methods, which calculate the likelihood (i.e., probability of the extant sequence data given the model) of different candidate models. Model selection criteria consider the likelihood of candidate models while penalizing overparameterization. Software for model testing include ModelTest-NG[249] and ModelFinder.[250]

Given both an MSA and best-fit evolutionary model, phylogenetic-reconstruction tools infer a tree by resolving branching patterns and branch lengths. Like model testing, tree searches can also be performed by maximum-likelihood-based methods, which calculate the likelihood of candidate trees during the search process (e.g., RAxML,[251] IQ-TREE[252]). For nearly all real-world sequence datasets, possible tree space is massive and computationally infeasible to explore completely. Therefore, heuristic approaches are used to identify the most likely tree. Branch support can be assessed by different metrics, including nonparametric bootstrap (evaluating the frequency of given clade, or cluster of sequences, across different trees reconstructed from resampled sequence data)[253] and likelihood ratio tests, which compare the likelihood of the best tree with that of the next-best tree with the branch in question collapsed.[254] Finally, phylogenetic reconstruction can also be performed through Bayesian inference. Bayesian tools empirically approximate the posterior probability distribution across model parameter space (including both the tree and evolutionary model) by sampling parameter values through Markov chain Monte Carlo-based algorithms. These methods thereby identify the maximum *a posteriori* tree as well as provide a measure of tree uncertainty. Branch support is expressed as the posterior probability of a given clade or the frequency of the clade across sampled trees. We direct the reader to Holder and Lewis[255] for a discussion of different phylogenetic inference methods.

Molecular clocks

Emile Zuckerkandl and Linus Pauling first suggested the molecular clock hypothesis: if the mutation rate in nucleotide and protein sequences correlates linearly with time and remains relatively constant under neutral evolution, this relationship can be used to date evolutionary divergences.[256] The original strict molecular clock method measures lineage- or sequence-specific parameters (e.g., the substitution rate per year). This can be computed as, for example, a linear model between genetic distances and time divergence.[257] Later models control for branch specific substitution rate (i.e., "multi-rate" and "relaxed" clocks), and advanced versions estimate the parameters by means of Markov chain Monte Carlo-based Bayesian parametric statistics, maximum-likelihood methods, and others.[258] When possible, models are time calibrated by including nodes dated

from fossil records.[259] For example, fossil lipids can be used to calibrate molecular clocks early in the evolution of microbes [524]. In some cases, calibrations can be based on geological or climate data [518], and, for rapidly evolving taxa with poor fossil records, calibrations can be based on estimated molecular substitution rates or sampling dates [519]. Recently, it has been shown that LGT can be used to date phylogenetic events such as bacterial radiation[260] and methanogen evolution.[261] Molecular clock models and phylogenomics can help date major events in the evolution of microbial taxa, including the origin of Archaea [521] and LUCA [522,523].

Ancestral reconstruction

Ancestral sequence reconstruction (ASR) was introduced by Pauling and Zuckerkandl[262] as an application for molecular phylogenetics, molecular clocks, and orthology relations.[263] ASR is a method for inferring the sequence content of ancestral proteins or genes corresponding to internal nodes of a phylogenetic tree. Typically, a researcher is interested not only in the ancestral sequence itself but of its phenotypic outcome (e.g., the biochemical or biophysical properties of a particular ancestral protein). With modern gene synthesis services, it is practical to synthesize the encoding gene of an ancestral protein and "resurrect" it in the laboratory by expression in a modern host organism, followed by phenotypic characterization.[264,265]

ASR relies on many of the same computational methods that underlie phylogenetic inference. Earliest ASR studies relied on maximum parsimony methods,[266,267] which seek to minimize the number of substitutions in a phylogeny.[268] ASR is more commonly performed today by probabilistic methods such as maximum likelihood (e.g., PAML[269]) or Bayesian inference (e.g., MrBayes[270]).

Regardless of the method, uncertainty in ASR is a key consideration for downstream analysis. Even for relatively small genes or proteins, the probability of the reconstructed ancestral sequence will typically be very low. For a 100-amino-acid protein, even if the reconstructed residue at every site has a probability of 0.9, the probability of the entire sequence (the joint probability of all residues in that sequence) = $0.9^{100} \approx 3\times10^{-5}$. Thus, much work has been focused on understanding the significant sources of ASR inaccuracy, including uncertainty associated with multiple sequence alignment, evolutionary model parameters, and phylogenetic tree topology.[271–273] Nevertheless, previous experiments have demonstrated that phenotypic properties of interest can be robust to ancestral sequence uncertainty. Therefore, in these cases, the reconstructed sequence need not be accurate to draw scientific conclusions at the level of phenotype. A useful strategy is to integrate phenotypic characterization across a range of plausible ancestral sequences or to incorporate a "worst-case" alternate ancestor, constructed by replacing any ambiguously reconstructed residues with their second most probable residue.[272,274]

Because ancient proteins do not, except in relatively recent and rare cases, leave direct fossil records, ASR is a powerful tool to infer the properties of early-evolved proteins and metabolic processes that have been central to the development of the biosphere. For example, ASR has been used in both *in silico* and experimental studies to investigate the temperature stabilities of ancient proteins dating as far back as LUCA,[275,276] the specificity and functionality of early translation machinery,[277,278] the development of the genetic code,[279,280] and the evolution of key metabolic processes and biogeochemical cycles.[41,281,282]

For OoL studies, ancestral reconstruction also need not only be applied to molecular sequence information. Similar statistical approaches can be leveraged to directly reconstruct ancestral phenotypic traits, given a matrix of trait information associated with extant proteins or taxa rather than a multiple sequence alignment. This approach, referred to more broadly as ancestral-state reconstruction, has been used to infer the minimal gene-set of LUCA,[283] the cell shape and taxonomic affinity of the last bacterial common ancestor,[206] and the ecological attributes of earliest photosynthetic organisms.[284,285] Future work in this area may incorporate recently developed methods for inferring features of the complex cellular networks early in life's history, including protein interactions.[286]

EMERGING TRENDS

As the questions related to the OoL have evolved, they have incorporated more heterogeneous sources of data and more sophisticated techniques. For example, the rise of omics approaches enabled the systematic investigation of biochemistry and has facilitated new approaches to systems chemistry. Similarly advances in *in vitro* selection experiments pioneered in evolutionary biology are being applied to chemical systems to understand selection in proto-biological systems. New technologies have provided OoL scientists with new tools, leading to new approaches, as is the case with automation of laboratory experiments. A common trend in these tools is the feedback or interaction between experimental workflows and large datasets or modeling approaches that provide the opportunity to test old hypotheses while continuing to explore new theoretical models (Figure 3). We consider these types of tools, which allow for tighter integration of computational workflows, and experimental approaches to be particularly important for the future of the OoL field. Therefore, we address here the comprehensive range of concepts and methodologies that, while not techniques themselves, are emerging trends in integrating experimental and theoretical work (Figure 3), enhancing the understanding and simultaneous application of different knowledge from various fields.

Omics

Omics refers to a variety of biological disciplines (metagenomics, metabolomics, proteomics, transcriptomics, etc.) whose overall objective is to describe and quantify cellular biological molecules indicating their composition, structure, dynamics, and function.[287] The emergence of omics techniques has impacted many questions and fields in biology, including OoL, for instance in the study of LUCA.[2,288] The diversification of omics approaches enables the exploration and understanding of different perspectives within various fields: metagenomics focuses on nucleotide sequences, proteomics and transcriptomics analyze proteins and RNA sequences, while metabolomics investigates metabolites.[289–292] However, most of the omics approaches were initially developed for biological research, posing

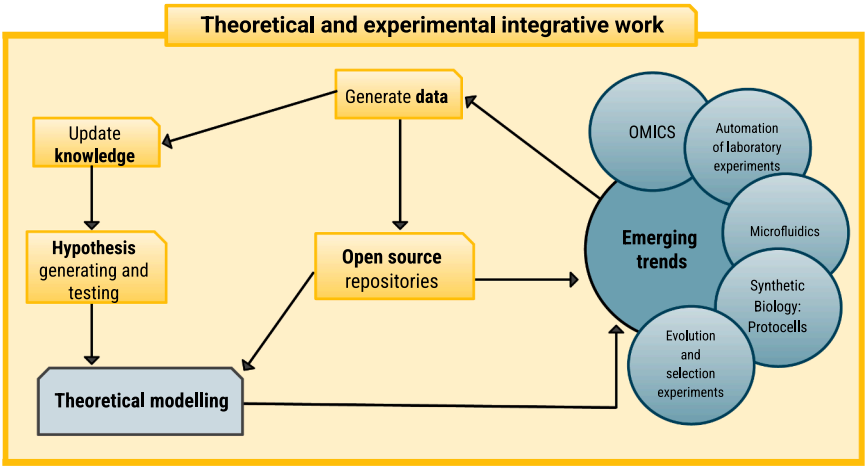

Figure 3. Integrating theoretical and experimental approaches

Interplay between experimental workflows, modeling approaches and large datasets enables the testing of hypotheses while fostering the development of new theoretical models.

challenges when applied to abiotic systems. Nevertheless, significant efforts have been made to adapt these approaches. For instance, techniques such as high-resolution mass spectrometry (MS) (Part 1, mass spectrometry) and nuclear magnetic resonance (NMR) spectroscopy (Part 1, Spectroscopy) have been employed to detect and quantify small molecules and inorganic compounds in prebiotic environments.[1] Consequently, bioinformatics tools originally designed for analyzing biological data have been modified to process data from abiotic systems. Similarly, the ongoing adaptation of databases to incorporate such information poses a significant challenge in the integration of complex data on abiotic systems.

Metagenomics

One of the most widely used techniques to study the structure and function of nucleotide sequences in environmental samples is metagenomics.[293] It generally consists of analyzing heterogeneous microbial communities to obtain their genomic composition (i.e., both in terms of taxonomic and functional composition). Functional metagenomics[294,295] involves cloning these previously identified DNA fragments from the environment, expressing them as genes in a model organism, and screening for enzymatic activity. Metagenomics is particularly relevant when studying extremophiles to identify genes with important and shared functions that are present within the community and to determine their adaptive pathways in the face of extreme conditions, which may be similar to those in early life.[296] Metagenomic analyses are especially useful due to the fact that these microorganisms do not have to be cultivated under laboratory conditions,[297] which is critical since the extreme conditions in which these microorganisms are found are almost impossible to replicate in the laboratory.[58,59] Indeed, the optimum salinity range for cultivable and isolated microorganisms is between 0% and 35%.[57]

Proteomics and transcriptomics

While it is widely recognized that proteins play a crucial role in contemporary life,[298] their involvement in the origins of life is less certain and remains a topic of ongoing research. Studying proteomics can therefore shed light on early biological evolution and offer valuable insights into how life may have originated. Similarly, transcription is an essential process in biology, and understanding how modern cells transcribe DNA into RNA in a variety of environmental conditions may shed light on how such processes emerged in the first place. There are two key omics fields related to these topics: proteomics and transcriptomics. Proteomics refers to the study of the structure and function of the complete set of proteins expressed in an organism, also known as the proteome.[299] On the other hand, transcriptomics focuses on the comprehensive set of coding and non-coding RNA molecules.

Over the past decade, the evolution of various techniques has contributed to the advancement of proteomics. The general principle involves characterizing the proteins encoded by a genome and their expression.[300] MS is the most widely used technique in this field (described in Part 1, mass spectrometry[1]), which enables the high-throughput profiling of proteins through electrophoresis or isotopic labeling.[301] While MS can elucidate the identity of many compounds in complex mixtures,[301] its broad applicability is hindered by the lack of integration between databases designed for extant biological purposes and those for abiotic systems, which are currently nonexistent.[302,303] This absence poses a significant challenge in realizing the full potential of modern omics techniques in studying proto-proteins/peptides.[303]

Transcriptomics aims to quantitatively assign reads to transcripts within a given genome. Early work in transcriptomics involved the use of DNA microarrays.[67] However, the most common contemporary technique for characterizing the transcriptome is RNA sequencing (RNA-seq). Its greatest advantage over previous methods is the ability to detect both known and unknown genes without the need for prior knowledge.[67] Single-cell transcriptomics has further enabled single-cell RNA-seq (scRNA-seq), allowing for enhanced analysis at the level of individual cells. For more information on techniques, we recommend reviewing Lowe et al.[67] These techniques have proved to be invaluable in the study of OoL, particularly in relation to the RNA world or protein-first hypothesis.[68,298] Proteomics and transcriptomics studies provide deeper insights into early life and evolution than genetic analysis alone, allowing for the investigation of fundamental components of the cell, such as the reconstruction of eukaryotic chromatin evolution[69] or the evolution of histones.[70] Furthermore, the study of structural domains of proteins holds significant potential in OoL research, such as in understanding the early origins of viruses.[71]

Metabolomics

The identification of metabolites, which are low-molecular-weight organic compounds produced by living cells, is of significant interest in OoL studies, particularly when considering the

complexity of prokaryotic and eukaryotic biology. Metabolomics studies employ (high-resolution) MS analyzers and chromatography separation to identify metabolites and metabolomic pathway changes. The metabolome of a microbial community is not only influenced by its genetic capacity, which can be studied using sequencing techniques (described in Part 1, genomic sequencing) but also varies based on environmental conditions such as temperature, pressure, pH, salt diversity and concentration, and irradiation exposure. Metabolomics studies focus on various research objectives: (1) investigating metabolites in environmental samples found in meteorites, stromatolites, fossilized matrices or hydrothermal vents on a smaller scale[72]; (2) studying metabolomic pathways of existing microorganisms with primitive traits and theoretical pathways of potential primary metabolomes to understand early Earth conditions (e.g., without O_2, with sulfur intake, in hot environments) on a medium scale[73]; and (3) comprehending and reconstructing the metabolome of microorganisms within a population and their interaction with the surrounding community on the largest scale.[74] In fact, several hypotheses about primitive cells and contemporary extremophiles arise from the interactions (e.g., symbiosis) between individuals.[75]

Metabolomic studies employ a wide range of analytical techniques to detect and identify biochemical compounds (see Part 1[1]). These techniques include combinations of analytical methods such as (1) high-performance liquid MS (HLPC-MS),[304] (2) gas chromatography-mass spectrometry,[305] (3) tandem mass spectrometry,[306] and (4) NMR spectroscopy.[307] Sample analyses can be enhanced with extraction methods, isotopic labeling and screening, capillary electrophoresis, and microfluidics to improve the detection and identification of various organic compounds, ranging from light to complex compounds.[308] These sample pre-treatments save time, aid in the better separation of organic or enantiomeric compounds, and enable the extraction of a wide and quantitative range of organics from large to small sample quantities, which subsequently become detectable.[309]

Among the techniques, HPLC-MS is the most commonly used and convenient method, allowing for the analysis of polar and non-polar or organic and aqueous phases independently or consecutively. Moreover, the high sensitivity of MS provides a high level of certainty in detection (ranging from nmol to fmol concentrations) and enables additional isotopic analyses on the same dataset. These studies, in combination with complementary microscopy and statistical analyses, help establish connections between molecular compounds and biological activities influenced by the community, abiotic stress factors, and the genome. MS-based metabolomics is used to complement metabolic pathways that may not be detected through genome annotations,[310] demonstrate the function of theoretical pathways and metabolites in cells or biofilms,[311,312] and discover novel metabolic pathways.[311]

Automation of laboratory experiments

Automation, by which we mean platforms that execute (often repetitive) experimental laboratory tasks, with minimal or no human intervention, can help overcome experimental limitations and generate vast amounts of data. The first automated peptide synthesizer was published in 1963,[313] and now most laboratories have a range of tools that can automate specific tasks, usually procedures of sequential repeating actions. This can make syntheses and experiments much more reliable than executed manually. Automation is employed more often in biological research than in chemistry, although its most common usage is for molecular procedures such as real-time PCR experiments and next-generation sequencing (see Part 1, genomic sequencing; high-throughput sequencing[1]).[314] The advent of liquid-handling robots enabled efficient automation of more complex protocols,[315,316] from peptide synthesis to enzyme-linked immunosorbent assays, and often utilizing microfluidics capacities (see microfluidics).[317] Population-level biological methodologies are sometimes automated as well, such as in culturing of microorganisms, community control, and high-throughput genetic modifications.[318] Arguably, automation technologies are even more crucial for synthetic biology, where reliable models require large amounts of high-quality data that can only be generated by automation, thoroughly testing a large diversity of systems with high reproducibility.[319] Lastly, while much of the detailed automated procedures are applicable to chemistry research, fully automated chemistry experiments remain rare. A few groups have started developing more advanced automated projects, but those are focused on drug discovery[320] or organic synthesis.[321]

Many classical OoL experiments, such as the Miller-Urey experiment,[322] relied on environmental cycling to drive chemical reactions, and work on wet-dry cycles[323,324] has further emphasized the role of naturally occurring processes in prebiotic chemistry. However, these approaches differ from laboratory automation, which requires programmable control over experimental conditions, which minimizes the need for human intervention. Such fully automated systems have not been widely applied to exploratory experiments in the OoL field despite novel hypotheses about the long-term evolution of chemical systems under automated control.[324–327]

One of the first explicitly automated platforms in the field was a system for peptide coupling,[328] which demonstrated the synthesis of glycine oligopeptides in high yield. The same laboratory developed a microfluidic platform for the study of osmotically driven droplet growth and population behavior.[329] Another group reported a system that enabled automatically controlled wet-dry cycles under anaerobic conditions.[330] A completely different example is the development of a 3D printed stirring device, which further leverages the already established automated functions of an autosampler and enables dynamic combinatorial library experiments with direct injection into the analytical system.[331] While these examples are steps of increasing levels of automation, they still require human intervention at various stages of the experimental process. In a new approach, network-driven exploration has been combined with an automated experimental platform with a closed-loop analytical system.[332] In this workflow, cyclic reactions were carried out over several weeks with an algorithm making decisions in real time based on analytical feedback. The use of automation can ease workload on repetitive tasks such as pipetting or sampling and can increase reproducibility of an experiment through improved documentation.

Automating laboratory work can be daunting for researchers unfamiliar with programming, and the high initial cost of robotic platforms makes them riskier than established manual protocols. However, as technology advances, automation is becoming increasingly accessible and cost-effective.[333]

Microfluidics

Microfluidics refers to the control and manipulation of fluids constrained to a small scale (often μL). It has emerged as a powerful technique for studying complex geochemical scenarios, particularly those involving heterogeneous phases and out-of-equilibrium dynamics. A strict control of experimental variables can be achieved using this technique, yielding robust and reproducible results.[334,335] One key advantage of microfluidics is its ability to maintain a laminar flow regime—unlike turbulent flow—which allows for the establishment of stable, non-equilibrium conditions. This enables controlled interactions between solutions of differing compositions without immediate mixing and dissipation of associated free energy. For example, Möller et al.[336] demonstrated how microfluidics can generate steep and stable pH gradients over μm distances. Microfluidics is not only useful for controlling the fluid dynamics of an experiment, but modifications in the chip's architecture allow for precise and *in situ* measurements of reaction parameters such as voltage, temperature, or pH.[337,338]

Although handling heterogeneous phases in microfluidic systems presents challenges—such as the risk of fouling or blocking—there are cases where microfluidic platforms uniquely facilitate heterogeneous catalysis under controlled conditions. For instance, Hudson et al. successfully precipitated minerals at the flow interface, using them as solid catalysts to promote CO_2 reduction into organics. Such experiments, requiring both heterogeneous catalysis and sustained out-of-equilibrium conditions, highlight the distinct capabilities of microfluidics in OoL studies.[338] Another advantage of microfluidics is its capacity to generate large volumes of data efficiently. For example, microfluidic platforms have been used to analyze vast ensembles of prebiotic compartments and facilitate evolutionary and selection experiments.[339]

While microfluidic experiments in heterogeneous chemistry are often slower due to their nature, automation and high-throughput screening approaches are increasingly being integrated into microfluidic platforms, enabling extensive data collection.[339] Liquid samples from microfluidic chips can be analyzed using various analytical chemistry techniques, including MS and chromatography (see Part 1[1]), although there are challenges with scaling this approach. However, on-line spectroscopy techniques can be integrated if the chip design permits optical access. Some microfluidic setups also allow for post-experimental retrieval of solid catalysts or reaction products for further characterization. These capabilities expand the scope of microfluidics beyond direct imaging and contribute to its utility in studying complex chemical systems.

A significant challenge in prebiotic chemistry is the necessity for a natural purification mechanism that enhances the concentration of reactants or products. Various substance-specific mechanisms have been proposed, such as the crystallization of nucleotide precursors or selective adsorption and enrichment of RNA. However, a broadly applicable natural mechanism has yet to be identified. Recent studies have explored how heat flows through thin geo-compartments could influence prebiotic reactions. These water-filled fractures, presumed to be widespread on early Earth, are subject to thermophoretic movement, leading to the accumulation of organics in colder regions. Experiments using geologically inspired microfluidic heat flow chambers, combined with numerical simulations, have demonstrated localized enrichment of over 50 OoL-relevant substances, including glycine dimerization reactions driven by trimetaphosphate. These enrichments have been shown to be effective across a wide range of pH and solvent conditions, scaling exponentially with temperature differences and network size. Such findings suggest that this mechanism could have played a crucial role in amplifying reactant concentrations to drive prebiotic chemistry.[340]

Synthetic biology: Protocells

One way to understand how non-living molecules could give rise to life is to reconstitute a living cell from scratch, a central goal of bottom-up synthetic biology. Beyond being a topic of independent research,[341,342] protocells have emerged as invaluable experimental and theoretical tools for exploring the fascinating stage between non-living matter and cellular life. Typically, biological molecules, such as nucleic acids, peptides, and proteins, can be encapsulated in synthetic compartments to build model protocells that exhibit life-like properties.[343,344] By using well-defined (purified) molecules, it is possible to clearly observe how dynamical properties emerge without the many unknown factors present in complex extant cells. The simplicity of protocells also allows a direct comparison with computational and theoretical studies.[345] Notably, biomolecules have been studied individually and in chemical networks perhaps more extensively than any other type of complex chemistry, and they are the only type of chemistry directly associated with life. As such, synthetic biology can help identify the minimal requirements for the emergence of life-like behaviors.

There are several common methods for generating protocellular models, especially lipid vesicles and liquid-liquid phase-separated droplets,[346,347] among others.[147,348] Their formation often relies on the self-assembly of lipids or biopolymers, associated with the encapsulation of desired molecules. In the case of phospholipid-based vesicles, the simplest preparation method is to add an aqueous solution onto a dried lipid film deposited on a glass surface to swell the film, leading to spontaneous vesicle formation, although the control of vesicle structures and the efficiency of macromolecular encapsulation are limited. Another method is to transfer water-in-oil droplets (with a phospholipid monolayer) through an oil-water interface (which also has a phospholipid monolayer) to form vesicles with phospholipid bilayers. Vesicle structures and molecular encapsulation can be well controlled during the preparation of droplets. Microfluidics (see microfluidics) is also used to generate vesicles with precisely controlled sizes. Protocell structures and molecules to encapsulate are chosen depending on the phenomena under investigation, and the level of complexity. For example, combining short RNA with fatty acid vesicles could allow the exploration of possible mechanisms of ancient cellular

self-reproduction,[349] whereas integration of an artificial genome and proteins with phospholipid vesicles allows us to probe more complex cellular behaviors, which may have been displayed by LUCA.[344] In the latter case, most central biological functions, such as genome replication, protein translation, metabolism, and energy generation have already been partially reconstituted. Non-biological cell-like compartments (e.g., water-in-oil emulsion) are also used to investigate the role of cellular structures in a more conceptual fashion, such as in molecular evolution (see evolution and selection experiments).[350,351] Although most studies in this field focus on membranous compartments, recent progress in reconstituting key biological reactions, including ribozyme catalysis,[352] genome replication, and protein translation in liquid-liquid phase-separated droplets,[353,354] also highlights the versatility of membrane-free compartments as protocell structures.[355]

Evolution and selection experiments

Evolution generally requires the following three steps: (1) replication, in which copying of a molecule to another one which inherits its trait; (2) mutation, in which imperfect replication leads to diversification of traits, caused by replication errors or environmental perturbation; (3) selection, in which a replicator with a particularly advantageous trait propagates more than others in a population. Similar to any living system, previously constructed molecular systems that can undergo these processes generally have replicable nucleic acids,[356] such as RNA, as a carrier of genetic information that determines their traits. In this case, the mutation is equivalent to the change of nucleic acid compositions. A major topic of study is whether and how evolution can occur without genetic materials, and novel experimental paradigms are facilitating empirical work on the topic. Specifically, the emergence of accessible laboratory automations and microfluidic platforms (see automation of laboratory experiments) are enabling long large-scale chemical experiments to search for signatures of evolution and selection in complex chemical systems.[332,339,357]

Once a molecular system that replicates and mutates is available, one might witness natural selection and evolution. A typical evolution experiment is performed through serial dilution or under continuous flow, with the supplement of nutrients (substances required for replication). During the experiment, repetitive replication generates mutated offspring with different traits, and dilution essentially causes the death of some and makes room for the progenies to reproduce within finite resources and space. If a replicator that can replicate faster emerged, it would dominate the population through successive replication and dilution (i.e., evolution occurs).

The first *in vitro* evolution was reported in 1967 by Spiegelman's group, using an RNA with the supplement of a purified RNA polymerase to facilitate replication.[358] The RNA gradually improved its replication efficiency in the course of evolution, but it became shorter and shorter. Various evolutionary phenomena, including diversification, complexification, and niche partitioning, have been observed for more complex replicators, such as an RNA that replicates using its self-encoded RNA polymerase[359] and a catalytic RNA (ribozyme) that replicates using reverse-transcription and transcription enzymes.[360] A system may need to be encapsulated in a compartment to link the encoded information with its replication, akin to a contemporary cell. A similar experiment can also be conducted without a protein enzyme. For example, a ribozyme that makes a copy of it by ligating RNA fragments can essentially evolve if pre-mutated fragments are supplied.[360,361] Although its evolutionary potential is not high due to the limited variety of mutations, it will increase if an RNA polymerase ribozyme capable of sustained replication is engineered in the future. Using synthetic chemical replicators, it has been demonstrated that replicators can spontaneously increase in complexity under out-of-equilibrium conditions.[362]

In addition to the Darwinian evolution experiments described above, the evolutionary principle can be applied to screen DNA, RNA, peptides, and protein molecules with desired functions. Some techniques use biological organisms such as bacteria and viruses, whereas others complete the entire experiment *in vitro*.[363,364] The latter, known as *in vitro* selection, has been used to identify functional RNA and peptides that may be relevant to the OoL because of its ability to screen a large number of molecules and compatibility with non-canonical environments. A typical *in vitro* selection experiment first prepares a pool of randomized DNA sequences with constant regions for PCR amplification, followed by *in vitro* transcription to synthesize RNA depending on what to select. For screening nucleic acids, they are subjected to a specific selection process (e.g., ligation or binding with a substrate), and selected molecules are collected (e.g., by collecting the substrate attached to the molecules of interest). Compartmentalizing each molecule also facilitates the selection of complex activities, such as *trans*-reaction and multiple turnovers.[365] For peptides and proteins, each molecule is first cell-free translated from an RNA library in a compartment or via attachment to RNA that encodes each protein, usually through conjugation with a linker (mRNA display) or without releasing from the ribosome (ribosome display).[366] These processes ensure the linkage of protein activity (phenotype) with its genotype (RNA sequence); however, this is not necessarily required for the selection of nucleic acids, as they can possess catalytic activities by themselves. After selection, sequences are amplified by reverse-transcription PCR to obtain a new pool. Through multiple cycles of these procedures, molecules with target functions gradually become enriched from even a single copy in the population, similar to evolutionary processes. At any point, a pool can be sequenced, as described in Part 1, genomic sequencing, and subjected to biochemical characterization.[1] For OoL research, these techniques have been used to find ribozymes that catalyze biological reactions, including ligation,[365,367] polymerization,[368] and aminoacylation,[368,369] as well as peptides and proteins with various functions.[368–371]

Evolution and selection experiments can also be conducted on living organisms. These experimental evolution approaches often include extreme conditions as a selection source and model organisms such as bacteria (*E. coli*[372]) or yeast (*Saccharomyces cerevisiae*[373]), among others. The study of adaptive pathways in extreme conditions and the extensive genomic knowledge of these model species facilitate research on the OoL and genome evolution,[374] including DNA repeats[375] or genomic plasticity[376] in prokaryotes. Similarly, new computational and *in silico* approaches are combined with *in vivo* experiments to study the origin of adaptations.[377]

CONCLUSIONS

Formulating a scientific explanation for the spontaneous animation of inanimate matter, along with experimental demonstrations of this process, would be a landmark achievement in the history of science. If accomplished in the coming decades, it will rely, at least in part, on the experimental and theoretical methodologies outlined here. However, it is unlikely that such a significant achievement will be realized within the confines of a single academic lab or research group. Solving the OoL will require extensive coordination and collaboration among diverse teams. A necessary precondition for the success of these collaborations is the establishment of community standards for the exchange and validation of data and software.

The OoL community must identify and adopt rigorous standards for sharing and distributing scientific data. This is an area where OoL science can benefit from following the lead of other scientific communities by adopting proven standards. For instance, in fields such as geosciences, the findable, accessible, interoperable, and reusable (FAIR) data standard is already becoming the norm.[378] The FAIR principles are general enough to be applicable to all fields in OoL research. The challenge for early-career scientists is to implement these principles in a way that aligns with their own disciplinary backgrounds and the broader OoL community.

Similarly, software and scripting are critical for producing and understanding scientific results, particularly as data sources and analysis pipelines become more sophisticated and complex. To properly document and distribute analysis code, the OoL community should look to other disciplines for resources, such as the Turing Project, which originated in data science but is applicable to quantitative sciences in general.[379]

As experimental procedures become more advanced, and their results yield increasingly life-like phenomena, it is essential that these results are clearly documented and reproducible. Developing universal standards for experimental procedures will be an effective tool for documenting and sharing these processes.[380] The OoL community should decide which tools and standards are useful for our science. The necessary tools to improve the rigor and quality of our science already exist; we now need to adopt them.

If scientists working on OoL aim to transform the field into a more cohesive and productive scientific community, effective communication of their advances is essential. We hope this review provides a starting point for those technical discussions. By focusing on empirical and theoretical results within the context of their technical validity, rather than conceptual narratives, we aim to accelerate discovery and conceptual synthesis in the OoL community. While this review cannot cover every technical topic in exhaustive detail, we hope it serves as a roadmap for future OoL scientists to begin their journey of understanding the diverse techniques and frameworks in the field.

ACKNOWLEDGMENTS

This work is a collaborative effort of the titled authors as part of the Origin of Life Early Career Network (OoLEN). We chose to add OoLEN as the first author to give a better representation of this team effort, rather than listing any single author as the first author. We hope such a thing can be adopted by others. We indicate that authors 2–9 (S.A., C.B., C. Blanco, D.B., A.C.-R., C.M., O.M., Z.P., and A.V.D.) have made a more distinct contribution. All authors are listed alphabetically by their last names. We would like to acknowledge all current and past members of OoLEN for their contributions to our community. In particular, we would like to acknowledge Evrim Fer, who helped with molecular phylogenetics. We would like to thank the anonymous referees for reviewing Parts 1 and 2 of this manuscript; this work was significantly improved through their feedback. S.A. acknowledges support from NASA through the postdoctoral Program at GSFC. C. Bautista acknowledges support from "la Caixa" Foundation (ID 100010434) under agreement (LCF/BQ/AA16/11580051) and by the Fonds de recherche du Québec Nature et technologies (FRQNT) (#274987). C. Blanco acknowledges support from NASA under award 80NSSC21K0595. D.B. acknowledges support from Centre national d'études spatiales (CNES) and postdoctoral support from LGPM-CentralSupélec and NASA under award 80NSSC23K1477. E. Camprubi acknowledges support from UT System for a STARs award. A.C.-R. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada (grant number RGPIN/05278–2018), the Fonds de recherche Nature et Technologies of Québec (grant number 314488), and the Fondation J. Armand Bombardier Excellence Scholarship. A.C.-R.'s research was supported by an appointment to the NASA Postdoctoral Program from the NASA Astrobiology Program administered by Oak Ridge Associated Universities under contract with NASA. S.F.J. acknowledges support from "la Caixa" Foundation (ID 100010434) and from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement no. 847648 (the fellowship code is LCF/BQ/PI21/11830015). T.Z.J. acknowledges support from Japan Society for the Promotion of Science (JSPS) grants-in-aid 18K14354 and 21K14746, a Tokyo Institute of Technology Yoshinori Ohsumi Fund for Fundamental Research, the Mizuho Foundation for the Promotion of Sciences, and by the Temporary Assistant Program by the DE&I Section of Science Tokyo. A.K. acknowledges support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant agreement no. 101068029. C.M. acknowledges support from NASA through the postdoctoral Fellowship Program. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NASA. O.M. acknowledges support from The John Templeton Foundation (#62828) and the Foundation for Science and Technology (2023.05971.CEECIND). B.K.D.P. acknowledges support from the NSERC Banting Postdoctoral Fellowship. K.P. acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster) and is a fellow of the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD).

AUTHOR CONTRIBUTIONS

A.A. proofread "information-theoretic approaches" (part 2). S.A. wrote an initial draft of "chromatography and hyphenated techniques" (part 1), "mass spectrometry" (part 1), and "automation of laboratory experiments" (part 2); edited the entire manuscript (parts 1 and 2); made an initial draft of Figure 2 (part 1) and edited Figure 1 (part 1); and helped organize the writing effort. C. Bautista wrote an initial draft of "omics" (part 2), "metagenomics" (part 2), and "proteomics and transcriptomics" (part 2); edited "genomic sequencing" (part 1), "biochemical and biological databases" (part 1), "emerging trends" (part 2), "metabolomics" (part 2), "automation of laboratory experiments" (part 2), and "evolution and selection experiments" (part 2); edited the entire manuscript (parts 1 and 2); made an initial draft of Figure 3 (part 2); and edited Figures 1 and 3 (part 1) and Figure 3 (part 2). C. Blanco edited "genomic sequencing" (part 1) and "databases in OoL studies" (part 1); and edited the entire manuscript (parts 1 and 2). D.B. edited "experimental techniques for studying the OoL" (part 1), "spectroscopy" (part 1), "chromatography and hyphenated techniques" (part 1), "mass spectrometry" (part 1), and "metabolomics" (part 2); and made an initial draft and edited Figure 1

and the tables in part 1. E. Camprubi wrote and edited "electron microscopy" (part 1), "quantum chemistry" (part 2) and "microfluidics" (part 2). E. Colizzi wrote an initial draft of "replicator models" (part 2). A.C.-R. wrote an initial draft of "information-theoretic approaches" (part 2), edited "chemical thermodynamics, kinetics, and networks" (part 2), "replicator models" (part 2) and "agent-based models" (part 2); and edited the entire manuscript (parts 1 and 2) S.C.-S. wrote an initial draft of the manuscript. A.V.D. wrote and edited "mass spectrometry" (part 1), "Raman spectroscopy" (part 1), "nuclear magnetic resonance spectroscopy" (part 1), "X-ray diffraction" (part 1), and "a case study" (part 1). H.D. edited "information-theoretic approaches" (part 2). V.E. wrote an initial draft of "molecular modelling and simulations" (part 2); edited "introduction" (part 1), "spectroscopy" (part 1) and "mass spectrometry" (part 1); and made an initial draft of Figure 2 (part 2). A.G. wrote an initial draft of "molecular phylogenetics" (part 2). G.G. wrote an initial draft of "whole (proto)cell models" (part 2); and edited "chemical thermodynamics, kinetics, and networks" (part 2) and "information-theoretic approaches" (part 2). A.H. edited "molecular modelling and simulations" (part 2). S.A.H. edited "UV-vis spectroscopy" (part 1). S.F.J. wrote and edited "experimental techniques for studying the OoL" (part 1). T.Z.J. wrote an initial draft of "microscopy techniques" (part 1), "light and fluorescence microscopy" (part 1), "confocal microscopy and optical coherence tomography" (part 1), and edited "genomic sequencing" (part 1), "Sanger sequencing" (part 1) and "high-throughput sequencing" (part 1). A.K. wrote an initial draft of "automation of laboratory experiments" (part 2); edited "molecular modelling and simulations" (part 2), "information-theoretic approaches" (part 2), and "emerging trends" (part 2). A.K. wrote an initial draft of "chemical kinetics" (part 2) and "chemical reaction networks" (part 2); edited "replicator models" (part 2), "molecular modelling and simulations" (part 2), and "information-theoretic approaches" (part 2); and proofread and edited the entire manuscript (parts 1 and 2). C.M. (cole.mathis.ool@gmail.com) coordinated the writing process, organized the first draft, edited the abstract and the introduction for parts 1 and 2, "chemical thermodynamics, kinetics, and networks" (part 2), "modelling of evolutionary dynamics and (proto)cells" (part 2), and "information-theoretic approaches" (part 2); edited the entire manuscript (parts 1 and 2), and handled the submission. O.M.-G. edited "molecular phylogenetics" (part 2). O.M. wrote and edited the abstract (parts 1 and 2), "theoretical approaches and modeling frameworks for the OoL" (part 2), "chemical thermodynamics, kinetics, and networks" (part 2), "chemical kinetics calculations" (part 2), "chemical reaction networks" (part 2), and "information-theoretic approaches" (part 2); and edited the entire manuscript (parts 1 and 2). R.M. wrote an initial draft of "synthetic biology: protocells" (part 2) and "evolution and selection experiments" (part 2), edited section "genomic sequencing" (part 1) and "replicator models" (part 2). J.N. wrote an initial draft of "electron microscopy" (part 1). Y.O. wrote an initial draft of "chemical reaction networks" (part 2). B.K.D.P. wrote and edited "chemical thermodynamics, kinetics, and networks" (part 2), "chemical kinetics calculations" (part 2), and "chemical reaction networks" (part 2). K.P. wrote an initial draft of "chemical thermodynamics" (part 2). M.P. wrote an initial draft of "electron microscopy" (part 1) and edited "experimental techniques for studying the OoL" (part 1). S.P. wrote an initial draft of "spectroscopy" (part 1), "mass spectrometry" (part 1). Z.P. wrote the initial draft of "databases in OoL" (part 1) and "network autocatalysis" (part 2) and edited the entire manuscript (parts 1 and 2). E.R.-R. edited "molecular phylogenetics" (part 2). L.S. edited "spectroscopy" (part 1), "mass spectrometry" (part 1), "microscopy techniques" (part 1), "databases in OoL studies" (part 1), "molecular phylogenetics" (part 2), and "automation of laboratory experiments" (part 2). S.S. edited "Raman spectroscopy" (part 1), "physical and chemical databases" (part 1), and "biochemical and biological databases" (part 1). A.V. wrote an initial draft of the manuscript. J.C.X. helped organize the first draft; contributed to "introduction" (part 1), "experimental techniques for studying the OoL" (part 1), "databases in OoL" (part 1), and "chemical thermodynamics, kinetics, and networks" (part 2); and edited the entire manuscript (parts 1 and 2).

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Asche, S., Bautista, C., Boulesteix, D., Champagne-Ruel, A., Mathis, C., Markovitch, O., Peng, Z., Adams, A., Dass, A.V., et al.; OoLEN (Origin of Life Early-career Network) (2026). What it takes to solve the Origin(s) of Life: An integrated review. Part 1: Experimental methods and data repositories. Cell Rep. Phys. Sci. *7*, 103212. https://doi.org/10.1016/j.xcrp.2026.103212.
2. Preiner, M., Asche, S., Becker, S., Betts, H.C., Boniface, A., Camprubi, E., Chandru, K., Erastova, V., Garg, S.G., Khawaja, N., et al. (2020). The Future of Origin of Life Research: Bridging Decades-Old Divisions. Life *10*, 20.
3. Wang, L.-P., Titov, A., McGibbon, R., Liu, F., Pande, V.S., and Martínez, T.J. (2014). Discovering chemistry with an ab initio nanoreactor. Nat. Chem. *6*, 1044–1048.
4. Rimola, A., Sodupe, M., and Ugliengo, P. (2019). Role of Mineral Surfaces in Prebiotic Chemical Evolution. In Silico Quantum Mechanical Studies. Life *9*, 10. https://doi.org/10.3390/life9010010.
5. Yaman, T., and Harvey, J.N. (2021). Computational Analysis of a Prebiotic Amino Acid Synthesis with Reference to Extant Codon-Amino Acid Relationships. Life *11*, 1343. https://doi.org/10.3390/life11121343.
6. Erastova, V., Degiacomi, M.T., G Fraser, D., and Greenwell, H.C. (2017). Mineral surface chemistry control for origin of prebiotic peptides. Nat. Commun. *8*, 2033.
7. Pearce, B.K.D., and Pudritz, R.E. (2016). Meteorites and the RNA World: A Thermodynamic Model of Nucleobase Synthesis within Planetesimals. Astrobiology *16*, 853–872.
8. Pearce, B.K.D., Molaverdikhani, K., Pudritz, R.E., Henning, T., and Cerrillo, K.E. (2022). Toward RNA life on early Earth: From atmospheric HCN to biomolecule production in warm little ponds. Astrophys. J. *932*, 9.
9. Goldford, J.E., Hartman, H., Smith, T.F., and Segrè, D. (2017). Remnants of an Ancient Metabolism without Phosphate. Cell *168*, 1126–1134.e9.
10. Semenov, S.N., Kraft, L.J., Ainla, A., Zhao, M., Baghbanzadeh, M., Campbell, V.E., Kang, K., Fox, J.M., and Whitesides, G.M. (2016). Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. Nature *537*, 656–660.
11. Xavier, J.C., Patil, K.R., and Rocha, I. (2018). Metabolic models and gene essentiality data reveal essential and conserved metabolism in prokaryotes. PLoS Comput. Biol. *14*, e1006556.
12. Eigen, M., and Schuster, P. (2012). The Hypercycle: A Principle of Natural Self-Organization (Springer Science & Business Media).
13. Szathmáry, E., and Demeter, L. (1987). Group selection of early replicators and the origin of life. J. Theor. Biol. *128*, 463–486.
14. Colizzi, E.S., and Hogeweg, P. (2014). Evolution of functional diversification within quasispecies. Genome Biol. Evol. *6*, 1990–2007.
15. Markovitch, O., and Lancet, D. (2012). Excess mutual catalysis is required for effective evolvability. Artif. Life *18*, 243–266.
16. Adami, C., and Labar, T. (2017). From entropy to information: Biased typewriters and the origin of life. In From Matter to Life, S.I. Walker, P.C.W. Davies, and G.F.R. Ellis, eds. (Cambridge: Cambridge University Press), pp. 130–154.
17. Mathis, C., Bhattacharya, T., and Walker, S.I. (2017). The Emergence of Life as a First-Order Phase Transition. Astrobiology *17*, 266–276.
18. Caetano-Anollés, G., Kim, H.S., and Mittenthal, J.E. (2007). The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. Proc. Natl. Acad. Sci. USA *104*, 9358–9363.
19. Alvarez-Carreño, C., Penev, P.I., Petrov, A.S., and Williams, L.D. (2021). Fold Evolution before LUCA: Common Ancestry of SH3 Domains and OB Domains. Mol. Biol. Evol. *38*, 5134–5143.
20. Williams, T.A., Cox, C.J., Foster, P.G., Szöllősi, G.J., and Embley, T.M. (2020). Phylogenomics provides robust support for a two-domains tree of life. Nat. Ecol. Evol. *4*, 138–147.

21. Szabo, A., and Ostlund, N.S. (2012). Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory (Courier Corporation).
22. Pérez-Villa, A., Pietrucci, F., and Saitta, A.M. (2020). Prebiotic chemistry and origins of life research with atomistic computer simulations. Phys. Life Rev. *34–35*, 105–135.
23. Hénin, J., Lelièvre, T., Shirts, M.R., Valsson, O., and Delemotte, L. (2022). Enhanced sampling methods for molecular dynamics simulations. Preprint at arXiv. http://arxiv.org/abs/2202.04164.
24. Bartlett, R.J., and Musiał, M. (2007). Coupled-cluster theory in quantum chemistry. Rev. Mod. Phys. *79*, 291–352. https://doi.org/10.1103/revmodphys.79.291.
25. Barone, V., and Puzzarini, C. (2022). Toward accurate formation routes of complex organic molecules in the interstellar medium: The paradigmatic cases of acrylonitrile and cyanomethanimine. Front. Astron. Space Sci. *8*, 814384. https://doi.org/10.3389/fspas.2021.814384.
26. Moghadam, S.A., Klobukowski, M., and Tuszynski, J.A. (2020). A search for the physical basis of the genetic code. Biosystems *195*, 104148.
27. Teale, A.M., Helgaker, T., Savin, A., Adamo, C., Aradi, B., Arbuznikov, A.V., Ayers, P.W., Baerends, E.J., Barone, V., Calaminici, P., et al. (2022). DFT exchange: sharing perspectives on the workhorse of quantum chemistry and materials science. Phys. Chem. Chem. Phys. *24*, 28700–28781.
28. Hohenberg, P., and Kohn, W. (1964). Inhomogeneous Electron Gas. Phys. Rev. *136*, B864–B871. https://doi.org/10.1103/physrev.136.b864.
29. Vosko, S.H., Wilk, L., and Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Can. J. Phys. *58*, 1200–1211. https://doi.org/10.1139/p80-159.
30. Perdew, J.P., Burke, K., and Ernzerhof, M. (1996). Generalized Gradient Approximation Made Simple. Phys. Rev. Lett. *77*, 3865–3868. https://doi.org/10.1103/physrevlett.77.3865.
31. Ceselin, G., Salta, Z., Bloino, J., Tasinato, N., and Barone, V. (2022). Accurate Quantum Chemical Spectroscopic Characterization of Glycolic Acid: A Route Toward its Astrophysical Detection. J. Phys. Chem. A *126*, 2373–2387.
32. Fornaro, T., Burini, D., Biczysko, M., and Barone, V. (2015). Hydrogen-bonding effects on infrared spectra from anharmonic computations: uracil-water complexes and uracil dimers. J. Phys. Chem. A *119*, 4224–4236.
33. Copley, S.D., Smith, E., and Morowitz, H.J. (2005). A mechanism for the association of amino acids with their codons and the origin of the genetic code. Proc. Natl. Acad. Sci. USA *102*, 4442–4447.
34. Rimola, A., Skouteris, D., Balucani, N., Ceccarelli, C., Enrique-Romero, J., Taquet, V., and Ugliengo, P. (2018). Can formamide be formed on interstellar ice? An atomistic perspective. ACS Earth Space Chem. *2*, 720–734.
35. Lambert, J.-F. (2008). Adsorption and polymerization of amino acids on mineral surfaces: a review. Orig. Life Evol. Biosph. *38*, 211–242.
36. Surman, A.J., Rodriguez-Garcia, M., Abul-Haija, Y.M., Cooper, G.J.T., Gromski, P.S., Turk-MacLeod, R., Mullin, M., Mathis, C., Walker, S.I., and Cronin, L. (2019). Environmental control programs the emergence of distinct functional ensembles from unconstrained chemical reactions. Proc. Natl. Acad. Sci. USA *116*, 5387–5392.
37. Doran, D., Abul-Haija, Y.M., and Cronin, L. (2019). Emergence of Function and Selection from Recursively Programmed Polymerisation Reactions in Mineral Environments. Angew. Chem. Int. Ed. Engl. *58*, 11253–11256.
38. Ferris, J.P. (2006). Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. Philos. Trans. R. Soc. Lond. B Biol. Sci. *361*, 1777–1786. [discussion: 1786].
39. Rimola, A., Sodupe, M., and Ugliengo, P. (2007). Aluminosilicate surfaces as promoters for peptide bond formation: an assessment of Bernal's hypothesis by ab initio methods. J. Am. Chem. Soc. *129*, 8333–8344.
40. Lejaeghere, K., Bihlmayer, G., Björkman, T., Blaha, P., Blügel, S., Blum, V., Caliste, D., Castelli, I.E., Clark, S.J., Dal Corso, A., et al. (2016). Reproducibility in density functional theory calculations of solids. Science *351*, aad3000.
41. Oliver, T., Sánchez-Baracaldo, P., Larkum, A.W., Rutherford, A.W., and Cardona, T. (2021). Time-resolved comparative molecular evolution of oxygenic photosynthesis. Biochim. Biophys. Acta. Bioenerg. *1862*, 148400.
42. Signorile, M., Pantaleone, S., Balucani, N., Bonino, F., Martra, G., and Ugliengo, P. (2020). Monitoring the Reactivity of Formamide on Amorphous SiO by In-Situ UV-Raman Spectroscopy and DFT Modeling. Molecules *25*, 2274. https://doi.org/10.3390/molecules25102274.
43. Grégoire, B., Erastova, V., Geatches, D.L., Clark, S.J., Greenwell, H.C., and Fraser, D.G. (2016). Insights into the behaviour of biomolecules on the early Earth: The concentration of aspartate by layered double hydroxide minerals. Geochim. Cosmochim. Acta *176*, 239–258.
44. Sanchez, R., Ferris, J., and Orgel, L.E. (1966). Conditions for purine synthesis: did prebiotic synthesis occur at low temperatures? Science *153*, 72–73.
45. Tomasi, J., Bonaccorsi, R., Cammi, R., and del Valle, F.J.O. (1991). Theoretical chemistry in solution. Some results and perspectives of the continuum methods and in particular of the polarizable continuum model. J. Mol. Struct.: THEOCHEM *234*, 401–424. https://doi.org/10.1016/0166-1280(91)89026-w.
46. Orr-Ewing, A.J. (2017). Taking the plunge: chemical reaction dynamics in liquids. Chem. Soc. Rev. *46*, 7597–7614.
47. Puzzarini, C., and Barone, V. (2011). Extending the molecular size in accurate quantum-chemical calculations: the equilibrium structure and spectroscopic properties of uracil. Phys. Chem. Chem. Phys. *13*, 7189–7197.
48. Song, L., and Kästner, J. (2016). Formation of the prebiotic molecule NHCHO on astronomical amorphous solid water surfaces: accurate tunneling rate calculations. Phys. Chem. Chem. Phys. *18*, 29278–29285.
49. Baiano, C., Lupi, J., Barone, V., and Tasinato, N. (2022). Gliding on Ice in Search of Accurate and Cost-Effective Computational Methods for Astrochemistry on Grains: The Puzzling Case of the HCN Isomerization. J. Chem. Theory Comput. *18*, 3111–3121.
50. Hassanali, A.A., Cuny, J., Verdolino, V., and Parrinello, M. (2014). Aqueous solutions: state of the art in ab initio molecular dynamics. Philos. Trans. A Math. Phys. Eng. Sci. *372*, 20120482.
51. Devergne, T., Magrino, T., Pietrucci, F., and Saitta, A.M. (2022). Combining Machine Learning Approaches and Accurate Enhanced Sampling Methods for Prebiotic Chemical Reactions in Solution. J. Chem. Theory Comput. *18*, 5410–5421.
52. Rapf, R.J., Perkins, R.J., Dooley, M.R., Kroll, J.A., Carpenter, B.K., and Vaida, V. (2018). Environmental Processing of Lipids Driven by Aqueous Photochemistry of α-Keto Acids. ACS Cent. Sci. *4*, 624–630.
53. Hättig C. Structure Optimizations for Excited States with Correlated Second-Order Methods: CC2 and ADC(2). Response Theory and Molecular Properties (A Tribute to Jan Linderberg and Poul Jørgensen). 2005. pp. 37–60. doi:10.1016/s0065-3276(05)50003-0.
54. Finley, J., Malmqvist, P.Å., Roos, B.O., and Serrano-Andrés, L. (1998). The multi-state CASPT2 method. Chem. Phys. Lett. *288*, 299–306. https://doi.org/10.1016/s0009-2614(98)00252-8.
55. Potenti, S., Manini, P., Fornaro, T., Poggiali, G., Crescenzi, O., Napolitano, A., Brucato, J.R., Barone, V., and d'Ischia, M. (2018). Solid state photochemistry of hydroxylated naphthalenes on minerals: Probing polycyclic aromatic hydrocarbon transformation pathways under astrochemically-relevant conditions. ACS Earth Space Chem. *2*, 977–1000.
56. Szabla, R., Tuna, D., Góra, R.W., Šponer, J., Sobolewski, A.L., and Domcke, W. (2013). Photochemistry of 2-aminooxazole, a hypothetical prebiotic precursor of RNA nucleotides. J. Phys. Chem. Lett. *4*, 2785–2788.

57. Merino, N., Aronson, H.S., Bojanova, D.P., Feyhl-Buska, J., Wong, M.L., Zhang, S., and Giovannelli, D. (2019). Living at the Extremes: Extremophiles and the Limits of Life in a Planetary Context. Front. Microbiol. *10*, 780.

58. Vartoukian, S.R., and Palmer, R.M. (2010). Strategies for culture of "uncculturable"bacteria. FEMS microbiology. https://academic.oup.com/femsle/article-abstract/309/1/1/460001.

59. John, J., Siva, V., Kumari, R., Arya, A., and Kumar, A. (2020). Unveiling Cultivable and Uncultivable Halophilic Bacteria Inhabiting Marakkanam Saltpan, India and Their Potential for Biotechnological Applications. Geomicrobiol. J. *37*, 691–701.

60. Saha, R., Poduval, P., Baratam, K., Nagesh, J., and Srivastava, A. (2024). Membrane Catalyzed Formation of Nucleotide Clusters and Their Role in the Origins of Life: Insights from Molecular Simulations and Lattice Modeling. J. Phys. Chem. B *128*, 3121–3132.

61. Stewart, S.V., and Erastova, V. (2024). Understanding the Role of Layered Minerals in the Emergence and Preservation of Proto-Proteins and Detection of Traces of Early Life. Acc. Chem. Res. *57*, 2453–2463.

62. Mathew, D.C., and Luthey-Schulten, Z. (2010). Influence of montmorillonite on nucleotide oligomerization reactions: a molecular dynamics study. Orig. Life Evol. Biosph. *40*, 303–317.

63. Wei, C., and Pohorille, A. (2009). Permeation of membranes by ribose and its diastereomers. J. Am. Chem. Soc. *131*, 10237–10245.

64. Kahana, A., Lancet, D., and Palmai, Z. (2022). Micellar Composition Affects Lipid Accretion Kinetics in Molecular Dynamics Simulations: Support for Lipid Network Reproduction. Life *12*, 955. https://doi.org/10.3390/life12070955.

65. Senftle, T.P., Hong, S., Islam, M.M., Kylasa, S.B., Zheng, Y., Shin, Y.K., Junkermeier, C., Engel-Herbert, R., Janik, M.J., Aktulga, H.M., et al. (2016). The ReaxFF reactive force-field: development, applications and future directions. npj Comput. Mater. *2*, 15011. https://doi.org/10.1038/npjcompumats.2015.11.

66. Zeng, J., Cao, L., Chin, C.-H., Ren, H., Zhang, J.Z.H., and Zhu, T. (2020). ReacNetGenerator: an automatic reaction network generator for reactive molecular dynamics simulations. Phys. Chem. Chem. Phys. *22*, 683–691.

67. Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. PLoS Comput. Biol. *13*, e1005457.

68. Jheeta, S., Chatzitheodoridis, E., Devine, K., and Block, J. (2021). The Way forward for the Origin of Life: Prions and Prion-Like Molecules First Hypothesis. Life *11*, 872. https://doi.org/10.3390/life11090872.

69. Grau-Bové, X., Navarrete, C., Chiva, C., Pribasnig, T., Antó, M., Torruella, G., Galindo, L.J., Lang, B.F., Moreira, D., López-Garcia, P., et al. (2022). A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution. Nat. Ecol. Evol. *6*, 1007–1023.

70. Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nat. Struct. Biol. *10*, 882–891.

71. Nasir, A., Kim, K.M., and Caetano-Anollés, G. (2017). Phylogenetic Tracings of Proteome Size Support the Gradual Accretion of Protein Structural Domains and the Early Origin of Viruses from Primordial Cells. Front. Microbiol. *8*, 1178.

72. Lutz, R.A., and Kennish, M.J. (1993). Ecology of deep-sea hydrothermal vent communities: A review. Rev. Geophys. *31*, 211–242.

73. Braakman, R., and Smith, E. (2013). The compositional and evolutionary logic of metabolism. Phys. Biol. *10*, 011001.

74. Crits-Christoph, A., Robinson, C.K., Barnum, T., Fricke, W.F., Davila, A.F., Jedynak, B., McKay, C.P., and Diruggiero, J. (2013). Colonization patterns of soil microbial communities in the Atacama Desert. Microbiome *1*, 28.

75. Douglas, A.E. (2014). Symbiosis as a general principle in eukaryotic evolution. Cold Spring Harb. Perspect. Biol. *6*, a016113. https://doi.org/10.1101/cshperspect.a016113.

76. Kondepudi, D.K., and Prigogine, I. (2014). Modern thermodynamics : from heat engines to dissipative structures, Second edition (Wiley), pp. 3–41.

77. Palsson, B.Ø. (2011). Systems Biology: Simulation of Dynamic Network States (Cambridge University Press).

78. Qian, H., and Ge, H. (2021). Stochastic Chemical Reaction Systems in Biology (Springer Nature).

79. Rao, R., and Esposito, M. (2016). Nonequilibrium Thermodynamics of Chemical Reaction Networks: Wisdom from Stochastic Thermodynamics. Phys. Rev. X *6*, 041064.

80. Ruiz-Mirazo, K., Briones, C., and de la Escosura, A. (2014). Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. Chem. Rev. *114*, 285–366.

81. Banzhaf, W., and Yamamoto, L. (2024). Artificial Chemistries (London, England: MIT Press).

82. Anderson, G. (2017). Thermodynamics of Natural Systems: Theory and Applications in Geochemistry and Environmental Science (Cambridge University Press).

83. Helgeson, H.C., and Kirkham, D.H. (1974). Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures; I, Summary of the thermodynamic/electrostatic properties of the solvent. Am. J. Sci. *274*, 1089–1198.

84. Goodwin, D.G., Moffat, H.K., Schoegl, I., Speth, R.L., and Weber, B.W. (2022). Cantera: An Object-Oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes (Zenodo). https://doi.org/10.5281/ZENODO.6387882.

85. Sundman, B., Kattner, U.R., Palumbo, M., and Fries, S.G. (2015). OpenCalphad - a free thermodynamic software. Integr. Mater. Manuf. Innov. *4*, 1–15.

86. Petersen, S., and Hack, K. (2007). The thermochemistry library ChemApp and its applications. Int J Mat Res . *98*, 935–945.

87. Website. Available: http://www.me.berkeley.edu/gri_mech/

88. Dick, J.M. (2019). CHNOSZ: Thermodynamic Calculations and Diagrams for Geochemistry. Front. Earth Sci. *7*, 180. https://doi.org/10.3389/feart.2019.00180.

89. Stull DR, Prophet H. JANAF thermochemical tables, second edition. 1971. doi:10.6028/nbs.nsrds.37.

90. Paschek, K., Kohler, K., Pearce, B.K.D., Lange, K., Henning, T.K., Trapp, O., Pudritz, R.E., and Semenov, D.A. (2022). Possible Ribose Synthesis in Carbonaceous Planetesimals. Life *12*, 404. https://doi.org/10.3390/life12030404.

91. Fegley, B.J., and Prinn, R.G. (1985). Equilibrium and nonequilibrium chemistry of Saturn's atmosphere - Implications for the observability of PH3, N2, CO, and GeH4. Astrophys. J. *299*, 1067.

92. Fegley, B., Jr., and Lodders, K. (1994). Chemical models of the deep atmospheres of Jupiter and Saturn. Icarus *110*, 117–154.

93. Zahnle, K.J., Lupu, R., Catling, D.C., and Wogan, N. (2020). Creation and evolution of impact-generated reduced atmospheres of early Earth. Planet. Sci. J. *1*, 11.

94. Wogan, N., Krissansen-Totton, J., and Catling, D.C. (2020). Abundant atmospheric methane from volcanism on terrestrial planets is unlikely and strengthens the case for methane as a biosignature. Planet. Sci. J. *1*, 58.

95. Chandru, K., Gilbert, A., Butch, C., Aono, M., and Cleaves, H.J. (2016). The Abiotic Chemistry of Thiolated Acetate Derivatives and the Origin of Life. Sci. Rep. *6*, 29883.

96. Wimmer, J.L.E., Xavier, J.C., Vieira, A.D.N., Pereira, D.P.H., Leidner, J., Sousa, F.L., Kleinermanns, K., Preiner, M., and Martin, W.F. (2021). Energy at Origins: Favorable Thermodynamics of Biosynthetic Reactions in the Last Universal Common Ancestor (LUCA). Front. Microbiol. *12*, 793664.

97. Lehninger, A.L., Nelson, D.L., and Cox, M.M. (2005). Lehninger Principles of Biochemistry (Macmillan).

98. Cornish-Bowden, A. (2013). Fundamentals of Enzyme Kinetics (John Wiley & Sons).

99. Keil, F.J. (2014), G.B. Marin and G.S. Yablonsky, eds. Kinetics of Chemical Reactions: Decoding Complexity (Wiley), pp. 910–911.

100. Avanzini, F., Penocchio, E., Falasco, G., and Esposito, M. (2021). Nonequilibrium thermodynamics of non-ideal chemical reaction networks. J. Chem. Phys. *154*, 094114.

101. Ianni, J.C. (2003). A comparison of the Bader-Deuflhard and the Cash-Karp Runge-Kutta integrators for the GRI-MECH 3.0 model based on the chemical kinetics code Kintecus. Computational Fluid and Solid Mechanics *2003*, 1368–1372. https://doi.org/10.1016/b978-008044046-0.50335-3.

102. Dahlgren, B. (2018). ChemPy: A package useful for chemistry written in Python. J. Open Source Softw. *3*, 565.

103. Pearce, B.K.D., Ayers, P.W., and Pudritz, R.E. (2019). A Consistent Reduced Network for HCN Chemistry in Early Earth and Titan Atmospheres: Quantum Calculations of Reaction Rate Coefficients. J. Phys. Chem. A *123*, 1861–1873.

104. Pearce, B.K.D., Molaverdikhani, K., Pudritz, R.E., Henning, T., and Hébrard, E. (2020). HCN production in titan's atmosphere: Coupling quantum chemistry and disequilibrium atmospheric modeling. Astrophys. J. *901*, 110.

105. Pearce, B.K.D., Ayers, P.W., and Pudritz, R.E. (2020). CRAHCN-O: A Consistent Reduced Atmospheric Hybrid Chemical Network Oxygen Extension for Hydrogen Cyanide and Formaldehyde Chemistry in CO-N-HO-CH-and H-Dominated Atmospheres. J. Phys. Chem. A *124*, 8594–8606.

106. Hébrard, E., Dobrijevic, M., Loison, J.C., Bergeat, A., and Hickson, K.M. (2012). Neutral production of hydrogen isocyanide (HNC) and hydrogen cyanide (HCN) in Titan's upper atmosphere. Astron. Astrophys. *541*, A21.

107. Venot, O., Hébrard, E., Agúndez, M., Decin, L., and Bounaceur, R. (2015). New chemical scheme for studying carbon-rich exoplanet atmospheres. Astron. Astrophys. *577*, A33.

108. Rimmer, P.B., and Helling, C. (2016). A chemical kinetics network for lightning and life in planetary atmospheres. Astrophys. J. Suppl. Ser. *224*, 9.

109. McElroy, D., Walsh, C., Markwick, A.J., Cordiner, M.A., Smith, K., and Millar, T.J. (2013). The UMIST database for astrochemistry 2012. Astron. Astrophys. *550*, A36.

110. Feinberg, M. (2019). Foundations of Chemical Reaction Network Theory. 1st Ed. (Cham, Switzerland: Springer Nature).

111. Goldford, J.E., Hartman, H., Marsland, R., and Segrè, D. (2019). Environmental boundary conditions for the origin of life converge to an organo-sulfur metabolism. Nat. Ecol. Evol. *3*, 1715–1724. https://doi.org/10.1038/s41559-019-1018-8.

112. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. *28*, 31–36.

113. Ugi, I., Bauer, J., Brandt, J., Friedrich, J., Gasteiger, J., Jochum, C., and Schubert, W. (1979). New applications of computers in chemistry. Angew. Chem. Int. Ed. Engl. *18*, 111–123.

114. Zubarev, D.Y., Rappoport, D., and Aspuru-Guzik, A. (2015). Uncertainty of prebiotic scenarios: the case of the non-enzymatic reverse tricarboxylic acid cycle. Sci. Rep. *5*, 8009.

115. Sharma, S., Arya, A., Cruz, R., and Cleaves Ii, H.J. (2021). Automated Exploration of Prebiotic Chemical Reaction Space: Progress and Perspectives. Life *11*, 1140. https://doi.org/10.3390/life11111140.

116. Stewart, J. JP. Stewart Computational Chemistry. http://openmopac.net/. 2007 [cited 31 Mar 2022]. Available: https://ci.nii.ac.jp/naid/10028164469/

117. Mavrovouniotis, M., Stephanopoulos, G., and Stephanopoulos, G. (1992). Synthesis of biochemical production routes. Comput. Chem. Eng. *16*, 605–619.

118. Mavrovouniotis, M.L. (1991). Estimation of standard Gibbs energy changes of biotransformations. J. Biol. Chem. *266*, 14440–14445.

119. Broadbelt, L.J., Stark, S.M., and Klein, M.T. (1994). Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates. Ind. Eng. Chem. Res. *33*, 790–799.

120. Magoon, G.R., and Green, W.H. (2013). Design and implementation of a next-generation software interface for on-the-fly quantum and force field calculations in automated reaction mechanism generation. Comput. Chem. Eng. *52*, 35–45. https://doi.org/10.1016/j.compchemeng.2012.11.009.

121. Wołos, A., Roszak, R., Żądło-Dobrowolska, A., Beker, W., Mikulak-Klucznik, B., Spólnik, G., Dygas, M., Szymkuć, S., and Grzybowski, B.A. (2020). Synthetic connectivity, emergence, and self-regeneration in the network of prebiotic chemistry. Science *369*, eaaw1955. https://doi.org/10.1126/science.aaw1955.

122. Arya, A., Ray, J., Sharma, S., Cruz Simbron, R., Lozano, A., Smith, H.B., Andersen, J.L., Chen, H., Meringer, M., and Cleaves, H.J., 2nd. (2022). An open source computational workflow for the discovery of autocatalytic networks in abiotic reactions. Chem. Sci. *13*, 4838–4853.

123. Guzman, M.I., and Martin, S.T. (2010). Photo-production of lactate from glyoxylate: how minerals can facilitate energy storage in a prebiotic world. Chem. Commun. *46*, 2265–2267.

124. Butch, C., Cope, E.D., Pollet, P., Gelbaum, L., Krishnamurthy, R., and Liotta, C.L. (2013). Production of tartrates by cyanide-mediated dimerization of glyoxylate: a potential abiotic pathway to the citric acid cycle. J. Am. Chem. Soc. *135*, 13440–13445.

125. Orgel, L.E. (2008). The implausibility of metabolic cycles on the prebiotic Earth. PLoS Biol. *6*, e18.

126. The International Union of Pure, (iupac) AC. IUPAC - autocatalytic reaction (A00525). [cited 30 Jun 2022]. doi:10.1351/goldbook.A00525.

127. Bissette, A.J., and Fletcher, S.P. (2013). Mechanisms of autocatalysis. Angew. Chem. Int. Ed. Engl. *52*, 12800–12826.

128. Breslow, R. (1959). On the mechanism of the formose reaction. Tetrahedron Lett. *1*, 22–26. https://doi.org/10.1016/s0040-4039(01)99487-0.

129. Saleh, R.M., Abd El Kader, J.M., El Hosary, A.A., and Shams El Din, A.M. (1975). A thermometric study of the dissolution of copper in acid solutions. J. Electroanal. Chem. Interfacial Electrochem. *62*, 297–310. https://doi.org/10.1016/0022-0728(75)80047-7.

130. Ueno, N., and Goto, H. (2015). A Preliminary Study on Belousov-Zhabotinsky (BZ) Reaction for Consideration of Basic Chemical Reaction in Origin of Life. International Letters of Chemistry. Physics and Astronomy *46*, 23–25. https://doi.org/10.18052/www.scipress.com/ilcpa.46.23.

131. Rosen, R. (1958). A relational theory of biological systems. Bull. Math. Biophys. *20*, 245–260. https://doi.org/10.1007/bf02478302.

132. Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften *58*, 465–523.

133. Hordijk, W., and Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. J. Theor. Biol. *227*, 451–461.

134. Xavier, J.C., Hordijk, W., Kauffman, S., Steel, M., and Martin, W.F. (2020). Autocatalytic chemical networks at the origin of metabolism. Proc. Biol. Sci. *287*, 20192377.

135. Kauffman, S.A. (1986). Autocatalytic sets of proteins. J. Theor. Biol. *119*, 1–24. https://doi.org/10.1016/s0022-5193(86)80047-9.

136. Dittrich, P., and Speroni Di Fenizio, P. (2007). Chemical Organisation Theory. Bull. Math. Biol. *69*, 1199–1231. https://doi.org/10.1007/s11538-006-9130-8.

137. Segré, D., Ben-Eli, D., Deamer, D.W., and Lancet, D. (2001). The lipid world. Orig. Life Evol. Biosph. *31*, 119–145.

138. Lancet, D., Zidovetzki, R., and Markovitch, O. (2018). Systems protobiology: origin of life in lipid catalytic networks. J. R. Soc. Interface *15*, 20180159. https://doi.org/10.1098/rsif.2018.0159.

139. Kahana, A., and Lancet, D. (2021). Self-reproducing catalytic micelles as nanoscopic protocell precursors. Nat. Rev. Chem *5*, 870–878.

140. Markovitch, O., and Krasnogor, N. (2018). Predicting species emergence in simulated complex pre-biotic networks. PLoS One *13*, e0192871. https://doi.org/10.1371/journal.pone.0192871.

141. Hordijk, W., and Steel, M. (2018). Autocatalytic Networks at the Basis of Life's Origin and Organization. Life *8*, 62. https://doi.org/10.3390/life8040062.

142. Letelier, J.-C., Cárdenas, M.L., and Cornish-Bowden, A. (2011). From L'Homme Machine to metabolic closure: steps towards understanding life. J. Theor. Biol. *286*, 100–113.

143. Vaidya, N., Manapat, M.L., Chen, I.A., Xulvi-Brunet, R., Hayden, E.J., and Lehman, N. (2012). Spontaneous network formation among cooperative RNA replicators. Nature *491*, 72–77.

144. Hordijk, W., Steel, M., and Dittrich, P. (2018). Autocatalytic sets and chemical organizations: modeling self-sustaining reaction networks at the origin of life. New J. Phys. *20*, 015011.

145. Xavier, J.C., and Kauffman, S. (2022). Small-molecule autocatalytic networks are universal metabolic fossils. Philos. Trans. A Math. Phys. Eng. Sci. *380*, 20210244.

146. Peng, Z., Linderoth, J., and Baum, D.A. (2022). The hierarchical organization of autocatalytic reaction networks and its relevance to the origin of life. PLoS Comput. Biol. *18*, e1010498.

147. Mizuuchi, R., and Ichihashi, N. (2021). Primitive Compartmentalization for the Sustainable Replication of Genetic Molecules. Life *11*, 191. https://doi.org/10.3390/life11030191.

148. Takeuchi, N., and Hogeweg, P. (2012). Evolutionary dynamics of RNA-like replicator systems: A bioinformatic approach to the origin of life. Phys. Life Rev. *9*, 219–263. https://doi.org/10.1016/j.plrev.2012.06.001.

149. Szilágyi, A., Zachar, I., Scheuring, I., Kun, Á., Könnyű, B., and Czárán, T. (2017). Ecology and Evolution in the RNA World Dynamics and Stability of Prebiotic Replicator Systems. Life *7*, 48. https://doi.org/10.3390/life7040048.

150. Colizzi, E.S., and Hogeweg, P. (2016). High cost enhances cooperation through the interplay between evolution and self-organisation. BMC Evol. Biol. *16*, 31.

151. Bull, J.J., Meyers, L.A., and Lachmann, M. (2005). Quasispecies made simple. PLoS Comput. Biol. *1*, e61.

152. Kun, Á., Santos, M., and Szathmáry, E. (2005). Real ribozymes suggest a relaxed error threshold. Nat. Genet. *37*, 1008–1011. https://doi.org/10.1038/ng1621.

153. Schuster, P., and Swetina, J. (1988). Stationary mutant distributions and evolutionary optimization. Bull. Math. Biol. *50*, 635–660.

154. Takeuchi, N., and Hogeweg, P. (2007). Error-threshold exists in fitness landscapes with lethal mutants. BMC Evol. Biol. *7*, 15. author reply 15.

155. Takeuchi, N., Poorthuis, P.H., and Hogeweg, P. (2005). Phenotypic error threshold; additivity and epistasis in RNA evolution. BMC Evol. Biol. *5*, 9.

156. Saakian, D.B., and Hu, C.-K. (2006). Exact solution of the Eigen model with general fitness functions and degradation rates. Proc. Natl. Acad. Sci. USA *103*, 4935–4939.

157. von Kiedrowski, G. (1993). Minimal Replicator Theory I: Parabolic Versus Exponential Growth. In Bioorganic Chemistry Frontiers, H. Dugas and F.P. Schmidtchen, eds. (Springer Berlin / Heidelberg), pp. 113–146.

158. Szathmáry, E. (1991). Simple growth laws and selection consequences. Trends Ecol. Evol. *6*, 366–370.

159. Paul, N., and Joyce, G.F. (2004). Minimal self-replicating systems. Curr. Opin. Chem. Biol. *8*, 634–639.

160. von Kiedrowski, G., Wlotzka, B., Helbing, J., Matzen, M., and Jordan, S. (04/1991). Parabolic Growth of a Self-Replicating Hexadeoxynucleotide Bearing a 3'-5'-Phosphoamidate Linkage. Angew. Chem. Int. Ed. Engl. *30*, 423–426.

161. Zielinski, W.S., and Orgel, L.E. (1987). Autocatalytic synthesis of a tetranucleotide analogue. Nature *327*, 346–347.

162. Schlögl, F. (1972). Chemical reaction models for non-equilibrium phase transitions. Z. Phys. *253*, 147–161.

163. Vellela, M., and Qian, H. (2009). Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. J. R. Soc. Interface *6*, 925–940.

164. Lavenda, B., Nicolis, G., and Herschkowitz-Kaufman, M. (1971). Chemical instabilities and relaxation oscillations. J. Theor. Biol. *32*, 283–292.

165. Pearson, J.E. (1993). Complex patterns in a simple system. Science *261*, 189–192.

166. Szathmáry, E., and Gladkih, I. (1989). Sub-exponential growth and coexistence of non-enzymatically replicating templates. J. Theor. Biol. *138*, 55–58.

167. Lorenz, R., Bernhart, S.H., Siederdissen, CH zu, Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithm Mol. Biol. *6*, 26. https://doi.org/10.1186/1748-7188-6-26.

168. Ivica, N.A., Obermayer, B., Campbell, G.W., Rajamani, S., Gerland, U., and Chen, I.A. (2013). The paradox of dual roles in the RNA world: resolving the conflict between stable folding and templating ability. J. Mol. Evol. *77*, 55–63.

169. Huynen, M.A., Stadler, P.F., and Fontana, W. (1996). Smoothness within ruggedness: the role of neutrality in adaptation. Proc. Natl. Acad. Sci. USA *93*, 397–401.

170. Boerlijst, M.C., and Hogeweg, P. (1991). Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. Phys. Nonlinear Phenom. *48*, 17–28. https://doi.org/10.1016/0167-2789(91)90049-f.

171. Colizzi, E.S., and Hogeweg, P. (2016). Parasites Sustain and Enhance RNA-Like Replicators through Spatial Self-Organisation. PLoS Comput. Biol. *12*, e1004902.

172. von der Dunk, S.H.A., Colizzi, E.S., Hogeweg, P., and Hogeweg, P. (2017). Evolutionary Conflict Leads to Innovation: Symmetry Breaking in a Spatial Model of RNA-Like Replicators. Life *7*, 43. https://doi.org/10.3390/life7040043.

173. Takeuchi, N., and Hogeweg, P. (2008). Evolution of complexity in RNA-like replicator systems. Biol. Direct *3*, 11. https://doi.org/10.1186/1745-6150-3-11.

174. Czárán, T., Könnyű, B., and Szathmáry, E. (2015). Metabolically Coupled Replicator Systems: Overview of an RNA-world model concept of prebiotic evolution on mineral surfaces. J. Theor. Biol. *381*, 39–54.

175. Takeuchi, N., Hogeweg, P., and Koonin, E.V. (2011). On the origin of DNA genomes: evolution of the division of labor between template and catalyst in model replicator systems. PLoS Comput. Biol. *7*, e1002024.

176. Nghe, P., Hordijk, W., Kauffman, S.A., Walker, S.I., Schmidt, F.J., Kemble, H., Yeates, J.A.M., and Lehman, N. (2015). Prebiotic network evolution: six key parameters. Mol. Biosyst. *11*, 3206–3217. https://doi.org/10.1039/c5mb00593k.

177. Bohl, K., Hummert, S., Werner, S., Basanta, D., Deutsch, A., Schuster, S., Theissen, G., and Schroeter, A. (2014). Evolutionary game theory: molecules as players. Mol. Biosyst. *10*, 3066–3074. https://doi.org/10.1039/c3mb70601j.

178. Turner, P.E., and Chao, L. (1999). Prisoner's dilemma in an RNA virus. Nature *398*, 441–443. https://doi.org/10.1038/18913.

179. Champagne-Ruel, A., and Charbonneau, P. (2022). A Mutation Threshold for Cooperative Takeover. Life *12*, 254. https://doi.org/10.3390/life12020254.

180. Thornburg, Z.R., Bianchi, D.M., Brier, T.A., Gilbert, B.R., Earnest, E.E., Melo, M.C.R., Safronova, N., Sáenz, J.P., Cook, A.T., Wise, K.S., et al. (2022). Fundamental behaviors emerge from simulations of a living minimal cell. Cell *185*, 345–360.e28. https://doi.org/10.1016/j.cell.2021.12.025.

181. Goldberg, A.P., Szigeti, B., Chew, Y.H., Sekar, J.A., Roth, Y.D., and Karr, J.R. (2018). Emerging whole-cell modeling principles and methods. Curr. Opin. Biotechnol. *51*, 97–102.

182. Marucci, L., Barberis, M., Karr, J., Ray, O., Race, P.R., de Souza Andrade, M., Grierson, C., Hoffmann, S.A., Landon, S., Rech, E., et al. (2020). Computer-Aided Whole-Cell Design: Taking a Holistic Approach by Integrating Synthetic With Systems Biology. Front. Bioeng. Biotechnol. *8*, 942.

183. Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from genotype. Cell *150*, 389–401.

184. Macklin, D.N., Ahn-Horst, T.A., Choi, H., Ruggero, N.A., Carrera, J., Mason, J.C., Sun, G., Agmon, E., DeFelice, M.M., Maayan, I., et al. (2020). Simultaneous cross-evaluation of heterogeneous datasets via mechanistic simulation. Science *369*, eaav3751. https://doi.org/10.1126/science.aav3751.

185. Breuer, M., Earnest, E.E., Merryman, C., Wise, K.S., Sun, L., Lynott, M.R., Hutchison, C.A., Smith, H.O., Lapek, J.D., Gonzalez, D.J., et al. (2019). Essential metabolism for a minimal cell. eLife *8*, e36842. https://doi.org/10.7554/elife.36842.

186. Shannon, C.E. (1948). A Mathematical Theory of Communication. Bell Syst. Tech. J. *27*, 379–423.

187. Davies, P. (2001). The Origin of Life II: How Did It Begin? Sci. Prog. *84*, 17–29.

188. Kimura, M. (1961). Natural Selection as the Process of Accumulating Genetic Information in Adaptive Evolution. Genet. Res. *2*, 127–140.

189. McGee, R.S., Kosterlitz, O., Kaznatcheev, A., Kerr, B., and Bergstrom, C.T. (2022). The Cost of Information Acquisition by Natural Selection. Preprint at bioRxiv. https://doi.org/10.1101/2022.07.02.498577.

190. Schneider, T.D. (2000). Evolution of Biological Information. Nucleic Acids Res. *28*, 2794–2799.

191. Yockey, H.P. (2005). Information Theory, Evolution, and the Origin of Life, 2nd ed. edition (New York: Cambridge University Press).

192. Adami, C. (2023). Evolution of Biological Information: How Evolution Creates Complexity, from Viruses to Brains (Princeton University Press).

193. Szostak, J.W. (2003). Functional Information: Molecular Messages. Nature *423*, 689.

194. Stone, J.V. (2022). Information Theory: A Tutorial Introduction, 2nd. Edition (Packt Publishing, Limited).

195. Hledík, M., Barton, N., and Tkačik, G. (2022). Accumulation and Maintenance of Information in Evolution. Proc. Natl. Acad. Sci. *119*, e2123152119.

196. Donaldson-Matasci, M.C., Bergstrom, C.T., and Lachmann, M. (2010). The fitness value of information. Oikos *119*, 219–230.

197. Solé, R.V., Munteanu, A., Rodriguez-Caso, C., and Macía, J. (2007). Synthetic Protocell Biology: From Reproduction to Computation. Philos. Trans. R. Soc. Lond. B Biol. Sci. *362*, 1727–1739.

198. Barker, T.S., Pierobon, M., and Thomas, P.J. (2022). Subjective Information and Survival in a Simulated Biological System. Entropy *24*, 639.

199. Urrios, A., Macia, J., Manzoni, R., Conde, N., Bonforti, A., de Nadal, E., Posas, F., and Solé, R. (2016). A Synthetic Multicellular Memory Device. ACS Synth. Biol. *5*, 862–873.

200. Tasnim, F., Freitas, N., and Wolpert, D.H. (2023). The Fundamental Thermodynamic Costs of Communication. Preprint at arXiv. https://doi.org/10.48550/arXiv.2302.04320.

201. Walker, S.I., Callahan, B.J., Arya, G., Barry, J.D., Bhattacharya, T., Grigoryev, S., Pellegrini, M., Rippe, K., and Rosenberg, S.M. (2013). Evolutionary Dynamics and Information Hierarchies in Biological Systems. Ann. N. Y. Acad. Sci. *1305*, 1–17.

202. Kim H, Davies P, Walker SI. New Scaling Relation for Information Transfer in Biological Networks. 2015. Available: http://arxiv.org/abs/1508.04174

203. Gregory, T.R. (2008). Understanding evolutionary trees. Evolution *1*, 121–137.

204. Weiss, M.C., Preiner, M., Xavier, J.C., Zimorski, V., and Martin, W.F. (2018). The last universal common ancestor between ancient Earth chemistry and the onset of genetics. PLoS Genet. *14*, e1007518.

205. Wimmer, J.L.E., Vieira, A.d.N., Xavier, J.C., Kleinermanns, K., Martin, W.F., and Preiner, M. (2021). The Autotrophic Core: An Ancient Network of 404 Reactions Converts H2, CO2, and NH3 into Amino Acids, Bases, and Cofactors. Microorganisms *9*, 458. https://doi.org/10.3390/microorganisms9020458.

206. Xavier, J.C., Gerhards, R.E., Wimmer, J.L.E., Brueckner, J., Tria, F.D.K., and Martin, W.F. (2021). The metabolic network of the last bacterial common ancestor. Commun. Biol. *4*, 413.

207. Zhu, Q., Mai, U., Pfeiffer, W., Janssen, S., Asnicar, F., Sanders, J.G., Belda-Ferre, P., Al-Ghalith, G.A., Kopylova, E., McDonald, D., et al. (2019). Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. Nat. Commun. *10*, 5477.

208. Joy, J.B., Liang, R.H., McCloskey, R.M., Nguyen, T., and Poon, A.F.Y. (2016). Ancestral Reconstruction. PLoS Comput. Biol. *12*, e1004763.

209. Baidouri, F.E., Venditti, C., Suzuki, S., Meade, A., and Humphries, S. (2021). Phenotypic reconstruction of the last universal common ancestor reveals a complex cell. Preprint at bioRxiv. https://doi.org/10.1101/2020.08.20.260398.

210. Zhaxybayeva, O., and Doolittle, W.F. (2011). Lateral gene transfer. Curr. Biol. *21*, R242–R246.

211. Daubin, V., and Szöllősi, G.J. (2016). Horizontal Gene Transfer and the History of Life. Cold Spring Harb. Perspect. Biol. *8*, a018036.

212. Fournier, G.P., Andam, C.P., and Gogarten, J.P. (2015). Ancient horizontal gene transfer and the last common ancestors. BMC Evol. Biol. *15*, 70.

213. Fournier, G.P., Andam, C.P., Alm, E.J., and Gogarten, J.P. (2011). Molecular evolution of aminoacyl tRNA synthetase proteins in the early history of life. Orig. Life Evol. Biosph. *41*, 621–632.

214. Lake, J.A., Servin, J.A., Herbold, C.W., and Skophammer, R.G. (2008). Evidence for a new root of the tree of life. Syst. Biol. *57*, 835–843.

215. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353–358.

216. Harish, A., and Morrison, D. (2020). The deep(er) roots of Eukaryotes and Akaryotes. F1000Res. *9*, 112.

217. Di Giulio, M. (2019). A qualitative criterion for identifying the root of the tree of life. J. Theor. Biol. *464*, 126–131.

218. Koonin, E.V. (2005). Orthologs, paralogs, and evolutionary genomics. Annu. Rev. Genet. *39*, 309–338.

219. Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. Proc. Biol. Sci. *277*, 819–827.

220. Syvanen, M. (2012). Evolutionary implications of horizontal gene transfer. Annu. Rev. Genet. *46*, 341–358.

221. Syvanen, M. (1987). Molecular clocks and evolutionary relationships: possible distortions due to horizontal gene flow. J. Mol. Evol. *26*, 16–23.

222. Altenhoff, A.M., Boeckmann, B., Capella-Gutierrez, S., Dalquen, D.A., DeLuca, T., Forslund, K., Huerta-Cepas, J., Linard, B., Pereira, C., Pryszcz, L.P., et al. (2016). Standardized benchmarking in the quest for orthologs. Nat. Methods *13*, 425–430.

223. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238.

224. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314.

225. Bağcı, C., Patz, S., and Huson, D.H. (2021). DIAMOND MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences. Curr. Protoc. *1*. https://doi.org/10.1002/cpz1.59.

226. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res. *17*, 377–386.

227. Mitra, S., Stärk, M., and Huson, D.H. (2011). Analysis of 16S rRNA environmental sequences using MEGAN. BMC Genom. *12*, S17. Suppl 3.

228. Kanehisa, M., Sato, Y., and Kawashima, M. (2022). KEGG mapping tools for uncovering hidden features in biological data. Protein Sci. *31*, 47–53.

229. Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D., and Koonin, E.V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. *49*, D274–D281.

230. Henry, C.S., DeJongh, M., Best, A.A., Frybarger, P.M., Linsay, B., and Stevens, R.L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat. Biotechnol. *28*, 977–982.

231. Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., et al. (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genom. *9*, 75.

232. Smith, H.B., Drew, A., Malloy, J.F., and Walker, S.I. (2021). Seeding Biochemistry on Other Worlds: Enceladus as a Case Study. Astrobiology *21*, 177–190.

233. Novichkov, P.S., Omelchenko, M.V., Gelfand, M.S., Mironov, A.A., Wolf, Y.I., and Koonin, E.V. (2004). Genome-wide molecularclock and horizontal gene transfer in bacterial evolution. J. Bacteriol. *186*, 6575–6585.

234. Ravenhall, M., Škunca, N., Lassalle, F., and Dessimoz, C. (2015). Inferring horizontal gene transfer. PLoS Comput. Biol. *11*. 100e1004095.

235. Davín, A.A., Tannier, E., Williams, T.A., Boussau, B., Daubin, V., and Szöllősi, G.J. (2018). Gene transfers can date the tree of life. Nat. Ecol. Evol. *2*, 904–909.

236. Szöllősi, G.J., Boussau, B., Abby, S.S., Tannier, E., and Daubin, V. (2012). Phylogenetic modeling of lateral gene transferreconstructs the pattern and relative timing of speciations. Proc. Natl. Acad. Sci. USA *109*, 17513–17518.

237. Pace, N.R., Sapp, J., and Goldenfeld, N. (2012). Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. Proc. Natl. Acad. Sci. USA *109*, 1011–1018.

238. Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA *74*, 5088–5090.

239. Kirschning, A. (2022). On the Evolutionary History of the Twenty Encoded Amino Acids. Chemistry *28*, e202201419.

240. Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. *48*, 443–453.

241. Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. J. Mol. Biol. *147*, 195–197.

242. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

243. Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

244. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf. *5*, 113.

245. Zielezinski, A., Girgis, H.Z., Bernard, G., Leimeister, C.-A., Tang, K., Dencker, T., Lau, A.K., Röhling, S., Choi, J.J., Waterman, M.S., et al. (2019). Benchmarking of alignment-free sequence comparison methods. Genome Biol. *20*, 144.

246. Haubold, B. (2014). Alignment-free phylogenetics and population genetics. Brief. Bioinform. *15*, 407–418.

247. Holm, L., and Sander, C. (1996). Mapping the protein universe. Science *273*, 595–603.

248. Garriga, E., Di Tommaso, P., Magis, C., Erb, I., Mansouri, L., Baltzis, A., Floden, E., and Notredame, C. (2021). Multiple Sequence Alignment Computation Using the T-Coffee Regressive Algorithm Implementation. Methods Mol. Biol. *2231*, 89–97.

249. Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. *37*, 291–294.

250. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

251. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453–4455.

252. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534.

253. Felsenstein, J. (1985). CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. Evolution *39*, 783–791.

254. Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. *22*, 521–565.

255. Holder, M., and Lewis, P.O. (2003). Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. *4*, 275–284.

256. Zuckerkandl, E. (1962). Molecular disease, evolution, and genetic heterogeneity. Horiz. Biochem. Biophys., 189–225.

257. Bromham, L., and Penny, D. (2003). The modern molecular clock. Nat. Rev. Genet. *4*, 216–224.

258. Ho, S.Y.W., and Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. Mol. Ecol. *23*, 5947–5965.

259. Donoghue, P.C.J., and Benton, M.J. (2007). Rocks and clocks: calibrating the Tree of Life using fossils and molecules. Trends Ecol. Evol. *22*, 424–431.

260. Davín, A.A., Tannier, E., Williams, T.A., Boussau, B., Daubin, V., and Szöllősi, G.J. (2018). Gene transfers can date the tree of life. Nat. Ecol. Evol. *2*, 904–909.

261. Wolfe, J.M., and Fournier, G.P. (2018). Horizontal gene transfer constrains the timing of methanogen evolution. Nat. Ecol. Evol. *2*, 897–903.

262. Pauling, L., Zuckerkandl, E., Henriksen, T., and Lövstad, R. (1963). Chemical paleogenetics. Molecular "restoration studies" of extinct forms of life. Acta Chem. Scand. *17*, 9–16.

263. Selberg, A.G.A., Gaucher, E.A., and Liberles, D.A. (2021). Ancestral Sequence Reconstruction: From Chemical Paleogenetics to Maximum Likelihood Algorithms and Beyond. J. Mol. Evol. *89*, 157–164.

264. Garcia, A.K., and Kaçar, B. (2019). How to resurrect ancestral proteins as proxies for ancient biogeochemistry. Free Radic. Biol. Med. *140*, 260–269.

265. Benner, S.A., Sassi, S.O., and Gaucher, E.A. (2007). Molecular paleoscience: systems biology from the past. Adv. Enzymol. Relat. Areas Mol. Biol. *75*, 1–132. xi.

266. Stackhouse, J., Presnell, S.R., McGeehan, G.M., Nambiar, K.P., and Benner, S.A. (1990). The ribonuclease from an extinct bovid ruminant. FEBS Lett. *262*, 104–106.

267. Malcolm, B.A., Wilson, K.P., Matthews, B.W., Kirsch, J.F., and Wilson, A.C. (1990). Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. Nature *345*, 86–89.

268. Reconstructing ancestral character states under Wagner parsimony (1987). Math. Biosci. *87*, 199–229.

269. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

270. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. *61*, 539–542.

271. Hanson-Smith, V., Kolaczkowski, B., and Thornton, J.W. (2010). Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. Mol. Biol. Evol. *27*, 1988–1999.

272. Eick, G.N., Bridgham, J.T., Anderson, D.P., Harms, M.J., and Thornton, J.W. (2017). Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. Mol. Biol. Evol. *34*, 247–261.

273. Vialle, R.A., Tamuri, A.U., and Goldman, N. (2018). Alignment Modulates Ancestral Sequence Reconstruction Accuracy. Mol. Biol. Evol. *35*, 1783–1797.

274. Liberles, D.A. (2007). Ancestral Sequence Reconstruction (Oxford University Press on Demand).

275. Akanuma, S., Nakajima, Y., Yokobori, S.-I., Kimura, M., Nemoto, N., Mase, T., Miyazono, K.i., Tanokura, M., and Yamagishi, A. (2013). Experimental evidence for the thermophilicity of ancestral life. Proc. Natl. Acad. Sci. USA *110*, 11067–11072.

276. Garcia, A.K., Schopf, J.W., Yokobori, S.-I., Akanuma, S., and Yamagishi, A. (2017). Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean. Proc. Natl. Acad. Sci. USA *114*, 4619–4624.

277. Fer, E., McGrath, K.M., Guy, L., Hockenberry, A.J., and Kaçar, B. (2022). Early divergence of translation initiation and elongation factors. Protein Sci. *31*, e4393.

278. De Tarafder, A., Parajuli, N.P., Majumdar, S., Kaçar, B., and Sanyal, S. (2021). Kinetic Analysis Suggests Evolution of Ribosome Specificity in Modern Elongation Factor-Tus from "Generalist" Ancestors. Mol. Biol. Evol. *38*, 3436–3444.

279. Fournier, G.P., and Alm, E.J. (2015). Ancestral Reconstruction of a Pre-LUCA Aminoacyl-tRNA Synthetase Ancestor Supports the Late Addition of Trp to the Genetic Code. J. Mol. Evol. *80*, 171–185. https://doi.org/10.1007/s00239-015-9672-1.

280. Furukawa, R., Yokobori, S.-I., Sato, R., Kumagawa, T., Nakagawa, M., Katoh, K., and Yamagishi, A. (2022). Amino Acid Specificity of Ancestral Aminoacyl-tRNA Synthetase Prior to the Last Universal Common Ancestor Commonote commonote. J. Mol. Evol. *90*, 73–94.

281. Garcia, A.K., Harris, D.F., Rivier, A.J., Carruthers, B.M., Pinochet-Barros, A., Seefeldt, L.C., and Kaçar, B. (2023). Nitrogenase resurrection and the evolution of a singular enzymatic mechanism. eLife *12*, e85003. https://doi.org/10.7554/eLife.85003.

282. Sephus, C.D., Fer, E., Garcia, A.K., Adam, Z.R., Schwieterman, E.W., and Kacar, B. (2022). Earliest Photic Zone Niches Probed by Ancestral Microbial Rhodopsins. Mol. Biol. Evol. *39*, msac100. https://doi.org/10.1093/molbev/msac100.

283. Ouzounis, C.A., Kunin, V., Darzentas, N., and Goldovsky, L. (2006). A minimal estimate for the gene content of the last universal common ancestor—exobiology from a terrestrial perspective. Res. Microbiol. *157*, 57–68.

284. Blank, C.E., and Sánchez-Baracaldo, P. (2010). Timing of morphological and ecological innovations in the cyanobacteria–a key to understanding the rise in atmospheric oxygen. Geobiology *8*, 1–23.

285. Sánchez-Baracaldo, P., Raven, J.A., Pisani, D., and Knoll, A.H. (2017). Early photosynthetic eukaryotes inhabited low-salinity habitats. Proc. Natl. Acad. Sci. USA *114*, E7737–E7745.

286. Liebeskind, B.J., Aldrich, R.W., and Marcotte, E.M. (2019). Ancestral reconstruction of protein interaction networks. PLoS Comput. Biol. *15*, e1007396.

287. Vailati-Riboni, M., Palombo, V., and Loor, J.J. (2017). What Are Omics Sciences? In Periparturient Diseases of Dairy Cows: A Systems Biology Approach, B.N. Ametaj, ed. (Cham: Springer International Publishing), pp. 1–7.

288. Woese, C. (1998). The universal ancestor. Proc. Natl. Acad. Sci. USA *95*, 6854–6859.

289. López-García, P., and Moreira, D. (2006). Selective forces for the origin of the eukaryotic nucleus. Bioessays *28*, 525–533.

290. Lombard, J., López-García, P., and Moreira, D. (2012). The early evolution of lipid membranes and the three domains of life. Nat. Rev. Microbiol. *10*, 507–515.

291. Myllykallio, H., Lopez, P., López-García, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H., and Forterre, P. (2000). Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. Science *288*, 2212–2215.

292. Moreira, D., and Lopez-Garcia, P. (1998). Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J. Mol. Evol. *47*, 517–530.

293. Wooley, J.C., Godzik, A., and Friedberg, I. (2010). A primer on metagenomics. PLoS Comput. Biol. *6*, e1000667.

294. Lam, K.N., Cheng, J., Engel, K., Neufeld, J.D., and Charles, T.C. (2015). Current and future resources for functional metagenomics. Front. Microbiol. *6*, 1196.

295. Mirete, S., Morgante, V., and González-Pastor, J.E. (2016). Functional metagenomics of extreme environments. Curr. Opin. Biotechnol. *38*, 143–149.

296. Burns, B.P., Anitori, R., Butterworth, P., Henneberger, R., Goh, F., Allen, M.A., Ibañez-Peral, R., Bergquist, P.L., Walter, M.R., and Neilan, B.A. (2009). Modern analogues and the early history of microbial life. Precambr. Res. *173*, 10–18.

297. Cowan, D.A., Ramond, J.-B., Makhalanyane, T.P., and De Maayer, P. (2015). Metagenomics of extreme environments. Curr. Opin. Microbiol. *25*, 97–102.

298. Dill, K.A., and Agozzino, L. (2021). Driving forces in the origins of life. Open Biol. *11*, 200324.

299. Blackstock, W.P., and Weir, M.P. (1999). Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol. *17*, 121–127.

300. Aslam, B., Basit, M., Nisar, M.A., Khurshid, M., and Rasool, M.H. (2017). Proteomics: Technologies and Their Applications. J. Chromatogr. Sci. *55*, 182–196.

301. Gross, J.H. (2004). Mass Spectrometry (Springer International Publishing).

302. Scheubert, K., Hufsky, F., and Böcker, S. (2013). Computational mass spectrometry for small molecules. J. Cheminform. *5*, 12.

303. Doran, D., Clarke, E., Keenan, G., Carrick, E., Mathis, C., and Cronin, L. (2021). Exploring the sequence space of unknown oligomers and polymers. Cell Rep. Phys. Sci. *2*, 100685.

304. Schrimpe-Rutledge, A.C., Codreanu, S.G., Sherrod, S.D., and McLean, J.A. (2016). Untargeted Metabolomics Strategies-Challenges and Emerging Directions. J. Am. Soc. Mass Spectrom. *27*, 1897–1905.

305. Drechsel, D., Dettmer, K., and Engewald, W. (2003). Studies of thermally assisted hydrolysis and methylation-GC-MS of fatty acids and triglycerides using different reagents and injection systems. Chromatographia *57*, S283–S289.

306. Hodge, K., Have, S.T., Hutton, L., and Lamond, A.I. (2013). Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. J. Proteomics *88*, 92–103.

307. Silverstein, R.M., Webster, F.X., Kiemle, D.J., and Bryce, D.L. (2014). Spectrometric Identification of Organic Compounds (John Wiley & Sons).

308. Vuckovic, D., and Pawliszyn, J. (2011). Systematic evaluation of solid-phase microextraction coatings for untargeted metabolomic profiling of biological fluids by liquid chromatography-mass spectrometry. Anal. Chem. *83*, 1944–1954.

309. He, Y., Buch, A., Morisson, M., Szopa, C., Freissinet, C., Williams, A., Millan, M., Guzman, M., Navarro-Gonzalez, R., Bonnet, J.Y., et al. (2019). Application of TMAH thermochemolysis to the detection of nucleobases: Application to the MOMA and SAM space experiment. Talanta *204*, 802–811.

310. Tang, Y.J., Yi, S., Zhuang, W.-Q., Zinder, S.H., Keasling, J.D., and Alvarez-Cohen, L. (2009). Investigation of carbon metabolism in "Dehalococcoides ethenogenes" strain 195 by use of isotopomer and transcriptomic analyses. J. Bacteriol. *191*, 5224–5231.

311. Fürch, T., Preusse, M., Tomasch, J., Zech, H., Wagner-Döbler, I., Rabus, R., and Wittmann, C. (2009). Metabolic fluxes in the central carbon metabolism of Dinoroseobacter shibae and Phaeobacter gallaeciensis, two members of the marine Roseobacter clade. BMC Microbiol. *9*, 209.

312. Peyraud, R., Kiefer, P., Christen, P., Massou, S., Portais, J.-C., and Vorholt, J.A. (2009). Demonstration of the ethylmalonyl-CoA pathway by using 13C metabolomics. Proc. Natl. Acad. Sci. USA *106*, 4846–4851.

313. Merrifield, R.B. (1963). Solid phase peptide synthesis. I. the synthesis of a tetrapeptide. J. Am. Chem. Soc. *85*, 2149–2154.

314. Mardis, E.R. (2013). Next-generation sequencing platforms. Annu. Rev. Anal. Chem. *6*, 287–303.

315. Tegally, H., San, J.E., Giandhari, J., and de Oliveira, T. (2020). Unlocking the efficiency of genomics laboratories with robotic liquid-handling. BMC Genom. *21*, 729.

316. Kong, F., Yuan, L., Zheng, Y.F., and Chen, W. (2012). Automatic liquid handling for life science: a critical review of the current state of the art. J. Lab. Autom. *17*, 169–185.

317. Kothamachu, V.B., Zaini, S., and Muffatto, F. (2020). Role of Digital Microfluidics in Enabling Access to Laboratory Automation and Making Biology Programmable. SLAS Technol. *25*, 411–426.

318. Wong, B.G., Mancuso, C.P., Kiriakov, S., Bashor, C.J., and Khalil, A.S. (2018). Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nat. Biotechnol. *36*, 614–623.

319. Carbonell, P., Radivojevic, T., and García Martín, H. (2019). Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation. ACS Synth. Biol. *8*, 1474–1477.

320. Schneider, G. (2018). Automating drug discovery. Nat. Rev. Drug Discov. *17*, 97–113.

321. Sanderson, K. (2019). Automation: Chemistry shoots for the Moon. Nature *568*, 577–579.

322. Criado-Reyes, J., Bizzarri, B.M., García-Ruiz, J.M., Saladino, R., and Di Mauro, E. (2021). The role of borosilicate glass in Miller-Urey experiment. Sci. Rep. *11*, 21009.

323. Rajamani, S., Vlassov, A., Benner, S., Coombs, A., Olasagasti, F., and Deamer, D. (2008). Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Orig. Life Evol. Biosph. *38*, 57–74.

324. Damer, B., and Deamer, D. (2020). The Hot Spring Hypothesis for an Origin of Life. Astrobiology *20*, 429–452.

325. University of, G., and Cronin, L. (2014). Hybrid Chemo-Robotic Systems for Embodied Chemical Evolution. Artificial Life 14: Proceedings of the Fourteenth International Conference on the Synthesis and Simulation of Living Systems (The MIT Press), pp. 3–5.

326. Robinson, W.E., Daines, E., van Duppen, P., de Jong, T., and Huck, W.T.S. (2022). Environmental conditions drive self-organization of reaction pathways in a prebiotic reaction network. Nat. Chem. *14*, 623–631.

327. Guttenberg, N., Virgo, N., Chandru, K., Scharf, C., and Mamajanov, I. (2017). Bulk measurements of messy chemistries are needed for a theory of the origins of life. Philos. Trans. A Math. Phys. Eng. Sci. *375*, 20160347. https://doi.org/10.1098/rsta.2016.0347.

328. Rodriguez-Garcia, M., Surman, A.J., Cooper, G.J.T., Suárez-Marina, I., Hosni, Z., Lee, M.P., and Cronin, L. (2015). Formation of oligopeptides in high yield under simple programmable conditions. Nat. Commun. *6*, 8385.

329. Doran, D., Rodriguez-Garcia, M., Turk-MacLeod, R., Cooper, G.J.T., and Cronin, L. (2017). A recursive microfluidic platform to explore the emergence of chemical evolution. Beilstein J. Org. Chem. *13*, 1702–1709.

330. Fox, S., Pleyer, H.L., and Strasdeit, H. (2019). An automated apparatus for the simulation of prebiotic wet–dry cycles under strictly anaerobic conditions. Int. J. Astrobiol. *18*, 60–72.

331. Markovitch, O., Ottelé, J., Veldman, O., and Otto, S. (2020). Automated device for continuous stirring while sampling in liquid chromatography systems. Commun. Chem. *3*, 180. https://doi.org/10.1038/s42004-020-00427-5.

332. Asche, S., Cooper, G.J.T., Keenan, G., Mathis, C., and Cronin, L. (2021). A robotic prebiotic chemist probes long term reactions of complexifying mixtures. Nat. Commun. *12*, 3547.

333. Christensen, M., Yunker, L.P.E., Shiri, P., Zepel, T., Prieto, P.L., Grunert, S., Bork, F., and Hein, J.E. (2021). Automation isn't automatic. Chem. Sci. *12*, 15473–15490.

334. Gale, B.K., Jafek, A.R., Lambert, C.J., Goenner, B.L., Moghimifam, H., Nze, U.C., and Kamarapu, S.K. (2018). A review of current methods in microfluidic device fabrication and future commercialization prospects. Inventions *3*, 60.

335. Pattanayak, P., Singh, S.K., Gulati, M., Vishwas, S., Kapoor, B., Chellappan, D.K., Anand, K., Gupta, G., Jha, N.K., Gupta, P.K., et al. (2021). Microfluidic chips: recent advances, critical strategies in design, applications and future perspectives. Microfluid. Nanofluidics *25*, 99.

336. Möller, F.M., Kriegel, F., Kieß, M., Sojo, V., and Braun, D. (2017). Steep pH Gradients and Directed Colloid Transport in a Microfluidic Alkaline Hydrothermal Pore. Angew. Chem. Int. Ed. Engl. *56*, 2340–2344.

337. Mo, Y., Lu, Z., Rughoobur, G., Patil, P., Gershenfeld, N., Akinwande, A.I., Buchwald, S.L., and Jensen, K.F. (2020). Microfluidic electrochemistry for single-electron transfer redox-neutral reactions. Science *368*, 1352–1357.

338. Hudson, R., de Graaf, R., Strandoo Rodin, M., Ohno, A., Lane, N., McGlynn, S.E., Yamada, Y.M.A., Nakamura, R., Barge, L.M., Braun, D., and Sojo, V. (2020). CO2 reduction driven by a pH gradient. Proc. Natl. Acad. Sci. USA *117*, 22873–22879.

339. Lu, H., Blokhuis, A., Turk-MacLeod, R., Karuppusamy, J., Franconi, A., Woronoff, G., Jeancolas, C., Abrishamkar, A., Loire, E., Ferrage, F., et al. (2024). Small-molecule autocatalysis drives compartment growth, competition and reproduction. Nat. Chem. *16*, 70–78.

340. Matreux, T., Aikkila, P., Scheu, B., Braun, D., and Mast, C.B. (2024). Heat flows enrich prebiotic building blocks and enhance their reactivity. Nature *628*, 110–116.

341. Kurihara, K., Okura, Y., Matsuo, M., Toyota, T., Suzuki, K., and Sugawara, T. (2015). A recursive vesicle-based model protocell with a primitive model cell cycle. Nat. Commun. *6*, 8352.

342. Sugawara, T., Kurihara, K., and Suzuki, K. (2013). Constructive Approach Towards Protocells. In Engineering of Chemical Complexity (WORLD SCIENTIFIC), pp. 359–374.

343. Schwille, P., Spatz, J., Landfester, K., Bodenschatz, E., Herminghaus, S., Sourjik, V., Erb, T.J., Bastiaens, P., Lipowsky, R., Hyman, A., et al. (2018). MaxSynBio: Avenues Towards Creating Cells from the Bottom Up. Angew. Chem. Int. Ed. Engl. *57*, 13382–13392.

344. Gaut, N.J., and Adamala, K.P. (2021). Reconstituting Natural Cell Elements in Synthetic Cells. Adv. Biol. *5*, e2000188.

345. Stano, P., and Mavelli, F. (2015). Protocells Models in Origin of Life and Synthetic Biology. Life *5*, 1700–1702. https://doi.org/10.3390/life5041700.

346. Walde, P., Cosentino, K., Engel, H., and Stano, P. (2010). Giant Vesicles: Preparations and Applications. Chembiochem *11*, 848–865. https://doi.org/10.1002/cbic.201000010.

347. Sarkar, S., Das, S., Dagar, S., Joshi, M.P., Mungi, C.V., Sawant, A.A., Patki, G.M., and Rajamani, S. (2020). Prebiological Membranes and Their

Role in the Emergence of Early Cellular Life. J. Membr. Biol. *253*, 589–608.

348. Gözen, I., Köksal, E.S., Põldsalu, I., Xue, L., Spustova, K., Pedrueza-Villalmanzo, E., Ryskulov, R., Meng, F., and Jesorka, A. (2022). Protocells: Milestones and Recent Advances. Small *18*, e2106624.
349. Joyce, G.F., and Szostak, J.W. (2018). Protocells and RNA Self-Replication. Cold Spring Harb. Perspect. Biol. *10*, a034801. https://doi.org/10.1101/cshperspect.a034801.
350. Matsumura, S., Kun, Á., Ryckelynck, M., Coldren, F., Szilágyi, A., Jossinet, F., Rick, C., Nghe, P., Szathmáry, E., and Griffiths, A.D. (2016). Transient compartmentalization of RNA replicators prevents extinction due to parasites. Science *354*, 1293–1296.
351. Mizuuchi, R., and Ichihashi, N. (2018). Sustainable replication and coevolution of cooperative RNAs in an artificial cell-like system. Nat. Ecol. Evol. *2*, 1654–1660.
352. Pressman, A.D., Liu, Z., Janzen, E., Blanco, C., Müller, U.F., Joyce, G.F., Pascal, R., and Chen, I.A. (2019). Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. J. Am. Chem. Soc. *141*, 6213–6223.
353. Drobot, B., Iglesias-Artola, J.M., Le Vay, K., Mayr, V., Kar, M., Kreysing, M., Mutschler, H., and Tang, T.Y.D. (2018). Compartmentalised RNA catalysis in membrane-free coacervate protocells. Nat. Commun. *9*, 3643.
354. Mizuuchi, R., and Ichihashi, N. (2020). Translation-coupled RNA replication and parasitic replicators in membrane-free compartments. Chem. Commun. *56*, 13453–13456.
355. Ji, Y., Mu, W., Wu, H., and Qiao, Y. (2021). Directing Transition of Synthetic Protocell Models via Physicochemical Cues-Triggered Interfacial Dynamic Covalent Chemistry. Adv. Sci. *8*, e2101187.
356. Ichihashi, N. (2019). What can we learn from the construction of in vitro replication systems? Ann. N. Y. Acad. Sci. *1447*, 144–156.
357. Foote, S., Sinhadc, P., Mathis, C., and Walker, S.I. (2023). False Positives and the Challenge of Testing the Alien Hypothesis. Astrobiology *23*, 1189–1201.
358. Mills, D.R., Peterson, R.L., and Spiegelman, S. (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. Proc. Natl. Acad. Sci. USA *58*, 217–224.
359. Mizuuchi, R., Furubayashi, T., and Ichihashi, N. (2022). Evolutionary transition from a single RNA replicator to a multiple replicator network. Nat. Commun. *13*, 1460.
360. Voytek, S.B., and Joyce, G.F. (2009). Niche partitioning in the coevolution of 2 distinct RNA enzymes. Proc. Natl. Acad. Sci. USA *106*, 7780–7785.
361. Lincoln, T.A., and Joyce, G.F. (2009). Self-sustained replication of an RNA enzyme. Science *323*, 1229–1232.
362. Yang, S., Schaeffer, G., Mattia, E., Markovitch, O., Liu, K., Hussain, A.S., Ottelé, J., Sood, A., and Otto, S. (2021). Chemical Fueling Enables Molecular Complexification of Self-Replicators. Angew. Chem. Int. Ed. Engl. *60*, 11344–11349.
363. Lipovsek, D., and Plückthun, A. (2004). In-vitro protein evolution by ribosome display and mRNA display. J. Immunol. Methods *290*, 51–67.
364. Wilson, D.S., and Szostak, J.W. (1999). In vitro selection of functional nucleic acids. Annu. Rev. Biochem. *68*, 611–647.
365. Griffiths, A.D., and Tawfik, D.S. (2006). Miniaturising the laboratory in emulsion droplets. Trends Biotechnol. *24*, 395–402.
366. Blanco, C., Verbanic, S., Seelig, B., and Chen, I.A. (2020). High throughput sequencing of in vitro selections of mRNA-displayed peptides: data analysis and applications. Phys. Chem. Chem. Phys. *22*, 6492–6506.
367. Bartel, D.P., and Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences. Science *261*, 1411–1418.
368. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., and Bartel, D.P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science *292*, 1319–1325.
369. Lee, N., Bessho, Y., Wei, K., Szostak, J.W., and Suga, H. (2000). Ribozyme-catalyzed tRNA aminoacylation. Nat. Struct. Biol. *7*, 28–33.
370. Giacobelli, V.G., Fujishima, K., Lepšík, M., Tretyachenko, V., Kadavá, T., Makarov, M., Bednárová, L., Novák, P., and Hlouchová, K. (2022). In Vitro Evolution Reveals Noncationic Protein-RNA Interaction Mediated by Metal Ions. Mol. Biol. Evol. *39*, msac032. https://doi.org/10.1093/molbev/msac032.
371. Seelig, B., and Szostak, J.W. (2007). Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature *448*, 828–831.
372. Bruckbauer, S.T., Trimarco, J.D., Martin, J., Bushnell, B., Senn, K.A., Schackwitz, W., Lipzen, A., Blow, M., Wood, E.A., Culberson, W.S., et al. (2019). Experimental Evolution of Extreme Resistance to Ionizing Radiation in Escherichia coli after 50 Cycles of Selection. J. Bacteriol. *201*, e00784-18. https://doi.org/10.1128/JB.00784-18.
373. Bautista, C., Marsit, S., and Landry, C.R. (2021). Interspecific hybrids show a reduced adaptive potential under DNA damaging conditions. Evol. Appl. *14*, 758–769.
374. Blum, P., Rudrappa, D., Singh, R., McCarthy, S., and Pavlik, B. (2016). Experimental Microbial Evolution of Extremophiles. In Biotechnology of Extremophiles: Advances and Challenges, P.H. Rampelotto, ed. (Cham: Springer International Publishing), pp. 619–636.
375. Ussery, D.W., Binnewies, T.T., Gouveia-Oliveira, R., Jarmer, H., and Hallin, P.F. (2004). Genome update: DNA repeats in bacterial genomes. Microbiology *150*, 3519–3521.
376. Romero, D., and Palacios, R. (1997). Gene amplification and genomic plasticity in prokaryotes. Annu. Rev. Genet. *31*, 91–111.
377. Ibarra, R.U., Edwards, J.S., and Palsson, B.O. (2002). Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature *420*, 186–189.
378. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data *3*, 160018.
379. The Turing Way Community (2021). The Turing Way: A Handbook for Reproducible, Ethical and Collaborative Research (Zenodo). https://doi.org/10.5281/ZENODO.5671094.
380. Mehr, S.H.M., Craven, M., Leonov, A.I., Keenan, G., and Cronin, L. (2020). A universal system for digitization and automatic execution of the chemical synthesis literature. Science *370*, 101–108.