Cell Reports
Physical Science

CellPress
OPEN ACCESS

Review

What it takes to solve the origin of life: An integrated review. Part 1–Experimental methods and data repositories

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] Aaron Halpern,[25]

(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, Québec, QC, Canada
[4]Département de Biologie, Faculté des Sciences et de Génie, Université Laval, Québec, QC, Canada
[5]Regroupement québécois de recherche sur la fonction, la Structure et l'ingénierie des protéines (PROTEO), Université Laval, Québec, QC, Canada
[6]Centre de Recherche en Données Massives (CRDM), Université Laval, Québec, 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

(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. Research on OoL spans many disciplines, including chemistry, physics, biology, planetary sciences, computer science, and mathematics. The sheer number of different scientific perspectives relevant to the problem has resulted in the coexistence of diverse tools, techniques, data, and software in OoL studies. This has made communication between the disciplines relevant to the OoL extremely difficult because the interpretation of data, analyses, or standards of evidence varies dramatically. Here, we hope to bridge this wide field of study by providing common ground via the consolidation of techniques rather than positing a unifying view on how life emerges. In part 1 of this review, we cover common experimental techniques that have been used significantly in OoL studies in recent years, while in part 2, we review theoretical, computational, and integrative methods. Here, we discuss the use of spectroscopy, spectrometry, chromatography, microscopy, and sequencing methods for characterizing diverse materials. We further discuss the role of data repositories in facilitating the analysis and dissemination of experimental data. This review provides a baseline expectation and understanding of the analytical aspects of origins' research. 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. For millennia, the seemingly ethereal nature of living beings was attributed to supernatural forces that imbued inanimate matter with unearthly properties, making it living. Much of these concepts survived from the rise of the Sumerians around 4,000 BCE until the mid-19th century when Pasteur published his famous spontaneous generation experiment.[1] Louis Pasteur showed that life as a phenomenon is not a result of

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

CellPress
OPEN ACCESS

Cell Reports
Physical Science
Review

Stuart A. Harrison,[25] Seán F. Jordan,[26] Tony Z. Jia,[7,27] 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,[7,41] Harrison B. Smith,[7,27] Andrey Vieira,[42] and Joana C. Xavier[43]

[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
[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, Co. 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.2026.103212

ethereal interactions between inanimate matter—after brewing a nutrient-rich broth and bringing it to a boil in an "S" shaped flask (a process now called pasteurization), he showed that life can only start from other life.[1] The implication of his experiment is that the trait of being alive is hereditary and not self-generated. Pasteur's observations resulted in a new dilemma: if life is a matter of inheritance, then how can it have a beginning? These results demanded a different explanation for a problem that had been rooted in dogma. About the same time, in 1859, Charles Robert Darwin published the first edition of his book "On the Origin of Species," shedding some light on the consequences of inherited traits and setting the stage for evolutionary biology.[2] Darwin identified a mechanism for speciation and evolution through natural selection, but he avoided a serious explanation for life's origins.[3] In many ways, these two masterpieces of scientific inquiry paved the way to the scientific approach to life's origins used today.

Contemporary research into the origin of life (OoL) generally consists of various attempts to show how a process that was impossible in Pasteur's sterilized flask is possible on sterile planetary bodies. Individual researchers come to study OoL from a variety of different disciplinary backgrounds. Communication within the community is often hindered by the diversity of fields represented, each of which brings its own technical and methodological approaches.[4] A common challenge encountered by OoL researchers is the lack of knowledge about what techniques, methods, and data are available and/or typically used in other OoL-related disciplines. This is because OoL research has never constituted a discipline in its own right and borrows technical advances and insights from a variety of specialist fields. In a previous paper authored by some in our community, we articulated conceptual heterogeneity in OoL research.[5] Here, we articulate the methodological heterogeneity in the field—via experiments, models, and simulations–to help realize the goal of cross-pollinating knowledge among specialist fields within the community.

Cooperation and critical discourse between OoL researchers from different disciplines does not necessarily require that all researchers have an in-depth understanding of all related disciplines. However, each researcher should know at least the basics of different disciplines. Therefore, this article reviews the basic methodology used by different scientists working on the

CellPress
OPEN ACCESS

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

Experimental techniques are shown in light blue. Topics covered in part 2[7] are shown in gray circles. Gray lines connect related techniques within a given category, pink lines connect experimental and theoretical work, and green lines connect techniques to databases.

problem of life's origin. Topics are split into three broad categories: (1) experimental techniques, including analysis of small molecules, materials, and biopolymers; (2) database and data-driven computational resources; and (3) theoretical and modeling tools from quantum chemistry and thermodynamics to network methods and phylogenetics. Finally, we discuss emerging trends that integrate different components from the other sections. Each of these topics is incredibly diverse in its own right and might never be discussed together in another context. We embraced this heterogeneity to highlight the diversity of work required to understand the OoL and to illustrate how these can mutually inform each other. As portrayed in Figure 1, the goal of this review is to present the methodologies and techniques commonly used in OoL, rather than to be an in--depth review of any idea in particular or a synthesis of the current questions or research paradigms in the field, which at this stage can only be a fractured view of distant ideas.[6] In part 2,[7] we will focus on the theoretical approaches and modeling frameworks, covering molecular modeling and simulations, chemical thermodynamics, kinetics, network, protocells, information-theoretic approaches, and molecular phylogenetics. Herein, we focus on experimental techniques.

For each technique, we present basic introductory details and highlight a few examples relevant to OoL research. We anticipate many readers will find content in their area of expertise simplistic—this is the goal, to communicate the basics with interested scientists so they can expand their operational knowledge within the field and critically evaluate work within it. We aimed to write at a level accessible enough for scientists from different methodological backgrounds to understand and begin to formulate their own questions about the techniques covered. Sufficient citations are provided to guide the reader toward more depth wherever their interest may lie. We hope that this work serves as a catalyst for future post-disciplinary collaborations in OoL research.

EXPERIMENTAL TECHNIQUES FOR STUDYING THE ORIGIN(S) OF LIFE

Imagine, for an instant, that you're a new graduate student with ambitious goals of understanding the origin(s) of life (OoL). On your first day in a new lab, your advisor hands you a sealed test tube containing a mysterious substance. It appears to be a mixture of liquids and solids. You're told the container is air-tight and that the contents are not toxic or otherwise harmful. You're not told if the substance was produced in the lab, represents an ancient terrestrial sample, or if it's even from this planet. The principle investigator of your lab poses a series of questions: "What do you think this sample is?" "What is it made of?" "Where do you think it came from?" "Does it have any life-like properties?" "Was it ever—or could it ever be—alive?" You have access to a variety of instruments in the lab. What do you do next? How do you learn about this material?

As your investigation into the mysterious sample progresses, the diversity of questions and hypotheses that arise will require the use of specific analytical techniques tailored to your needs. For instance, if you're looking to understand the exact molecular structures of a subset of compounds within the sample, you might adopt spectroscopic and spectrometric methods, focusing on precise chemical and molecular targets. Should the sample suggest an environmental origin (as opposed to synthetic), techniques that characterize bulk elemental ratios or the diversity of compounds in complex media would be instrumental. If your analysis hints at biological signatures, perhaps in the form of peptides or RNA strands, technologies to sequence large molecules would become crucial. Each technique not only answers a specific set of questions derived from your initial hypotheses but also builds upon the findings from

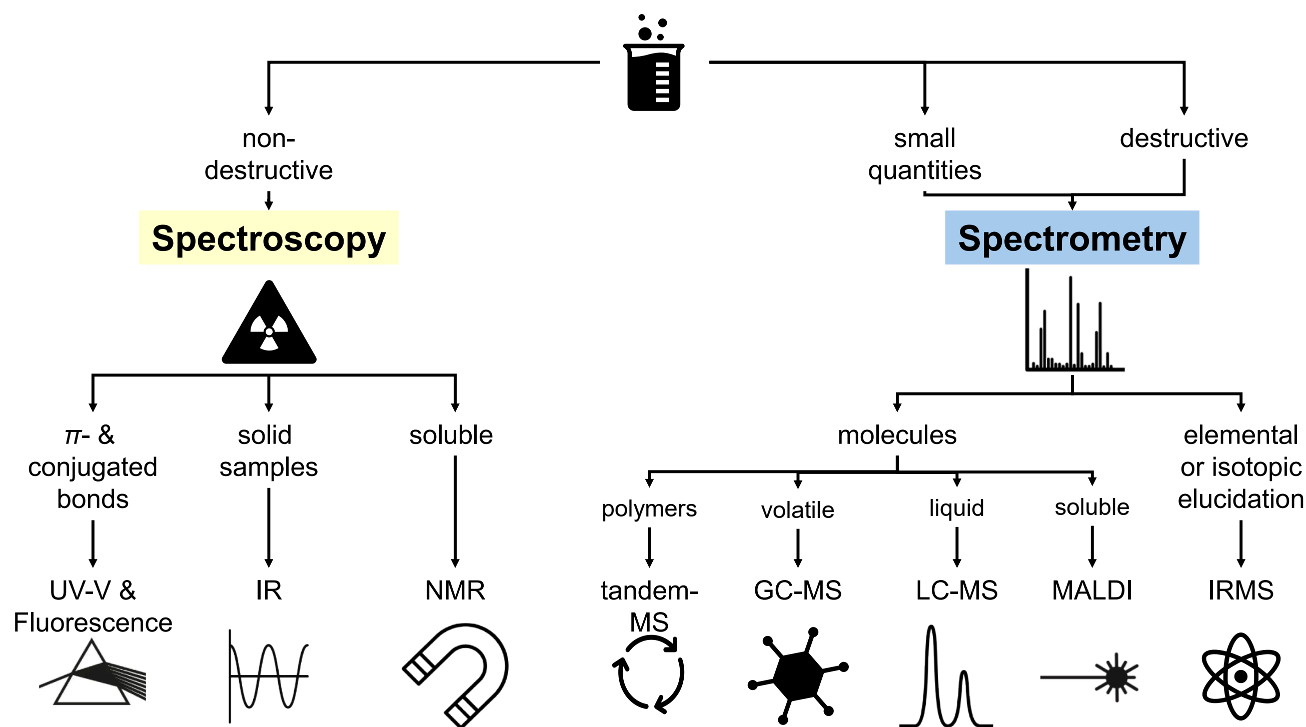

Figure 2. Simplified roadmap of the OoL analytical techniques targeting different compositions separated into two major categories

Spectroscopy and spectrometry described individually in "experimental techniques for studying the origin of life."

Generally, spectroscopy methods require higher amounts of sample but are non-destructive, while spectrometric methods can find analytes in very small concentrations, but the sample is often consumed during analysis.

other disciplines, enriching your understanding of the sample's potential origins and its implications in the study of life's beginnings. In this section, we review the basics of analytical techniques most commonly used in OoL research, including the physicochemical principles they employ and their strengths or limitations (Figure 2).

Spectroscopy

Spectroscopy deals with the interaction of matter with radiation as a means to characterize materials.[8] The type of radiation source, as well as the way materials transfer the energy (absorption, emission, scattering, photo- or chemi-luminescence, etc.), will determine the spectroscopic technique. A summary of different techniques is provided in Table 1. As a non-destructive technique,[8] spectroscopy is advantageous in the study of rare samples. In principle, a single sample could be analyzed by different spectroscopic techniques, permitting the identification and quantification of individual molecules in liquid, solid, or gaseous samples with little sample preparation required. High coverage of the electromagnetic field translates to precise identification of molecules. Because of its versatility, spectroscopy has been one of the most important tools for molecular identification in chemistry for decades.[8–10] Some techniques, such as ultraviolet-visible (UV-vis), fluorescence, and infrared spectroscopies and circular dichroism (CD), are inexpensive and taught in entry-level laboratory courses, which makes them some of the most accessible laboratory techniques for researchers just starting out.

While spectroscopy techniques are versatile, different techniques are specific to different features of compounds and materials. The resolution and sensitivity will depend on the sample preparation. Properties of the sample itself, such as the matrix, will influence the resolution and sensitivity of the method. In this context, matrix refers to the medium or environment in which a sample is embedded. Compared to other techniques, spectroscopic techniques have relatively low sensitivity, and hence detection and quantification can be a limitation. Accordingly, many methods require a strong signal/high concentration of elements/molecules. In many cases, we cannot resolve all spectra, and quantum chemical calculations (part 2 Molecular modeling and simulations: Quantum chemistry) are required to support empirical measurements and interpretation.[7,11]

In an OoL context, vis, UV, infrared (IR), Raman, and nuclear magnetic resonance (NMR) are commonly used spectroscopic tools. Spectroscopic analysis is often used for preliminary analysis of the bulk composition of a sample due to its non-destructive nature.

UV-vis and fluorescence spectroscopies

UV-vis light spectroscopy is a simple and inexpensive analytical procedure that is widely exploited in the fields of analytical chemistry and biotechnology. It is often one of the first techniques students learn when studying chemistry. The technique relies on the absorbance of light in the UV-vis light range (∼100–750 nm), where most molecules and ions absorb with different relative

CellPress
OPEN ACCESS

Table 1. Spectroscopic techniques used for OoL studies, with their advantages and disadvantages from a technical, sample, and physical-chemical interference points of view

Method	Pros	Cons	Common Targets
Ultraviolet-visible (UV-vis)	simple, inexpensive, fast acquisition	low sensitivity, matrix dependent	ions, organic, and inorganic molecules
Infrared (IR) by reflectance	simple, inexpensive, bulk analysis, fast acquisition	low sensitivity, matrix and mineral interferences	molecular bonds/dipole moment changes: primarily rotations, electrons, harmonic intramolecular vibrations, ionic bonds
Infrared by transmittance	mapping, bulk or μm-scale analysis, fast acquisition	matrix dependent, low resolution, low sensitivity at low wavelength	
Raman spectroscopy (RS)	mapping, coupled to microscopy (μm)	very low sensitivity for organic matter, fluorescence interferences	minerals, molecular polarization changes
Fluorescence/diode array	high sensitivity, high specificity, single or multi-wavelength analysis	standards required for product identification	organic and inorganic samples, pure and mixtures
Circular dichroism (CD)	high sensitivity	only works for polarized molecules, no residue fragment information	proteins, DNA
X-ray absorption	high resolution (nm-scale), element specific	inorganic only, no minerals, low sensitivity	atoms, ionic bonds, inorganic molecules, and molecular complexes
X-ray emission	high sensitivity/resolution, compound-specific, library to compare	no organic, low sensitivity, not quantitative	atoms, ionic bonds, inorganic molecules, and molecular complexes, minerals
Nuclear magnetic resonance (NMR)	high resolution, high versatility	high sample concentration required	organic and inorganic molecules, pure samples and mixtures, intra- and inter-molecular bonds, non-zero nuclear spin atoms/isotopes

intensities.[12] The absorbance spectrum of species can serve as a diagnostic test for their presence in solutions and can be used to relate the quantity of absorbance to the concentration when the absorbance coefficient is known.[13] Fluorescence spectroscopy is a related technique that monitors the emittance of light at a separate wavelength following absorbance. Fluorescence spectroscopy is more sensitive than UV-vis. Both techniques are typically employed in aqueous solutions, but gas-phase techniques also exist.[14] These techniques are often coupled to chromatographic techniques such as high-performance liquid chromatography (HPLC) (see Chromatography and hyphenated techniques: Liquid chromatography) to detect individual compounds in complex mixtures.[12]

Many analytical assays utilize the formation of a UV-absorbent or fluorescent molecule as a means of quantitative determination. When used for the identification of molecules, especially those with aromatic groups (e.g., nucleotides), it can be quite sensitive, enabling identification at micromolar concentrations. Thanks to the absorbance properties of common cofactors (e.g., NAD or ferredoxin), this technique can often be employed in enzymatic assays, and these concepts can also be transferred to enzyme-free reactions.[15] Due to its simplicity and straightforward sample preparation, UV-vis and fluorescence measurements are a good choice for "on-line" measurements in automated chemical systems (see part 2 Automation of laboratory experiments[7]).

UV-vis has been used in OoL to study the universality of ATP as energy currency,[16] to monitor the time-varying abundance of different peptide oligomers[17] and to detect chiral enantiomers.[18] Derivatization procedures are often necessary to couple strong UV absorbent or fluorescent groups to weakly absorbing molecules, which help to improve their detection (e.g., sugars[19] and amino acids[20]). UV-vis can be used to monitor the formation of inorganic iron sulfur clusters[21–23] and small molecule interactions in solution.[24] The analysis in Jordan et al.[21] is an example of how complementary analytical techniques enable more robust analysis, and critical review of experimental procedures, as an initial (and quick) analysis of using UV-vis revealed the existence of iron sulfur clusters, but follow-up analysis using NMR showed differences from this initial analysis, which implicated anaerobic degradation of the products.

IR spectroscopy

IR spectroscopy relies on energy absorbances that range from 900 nm to 1 mm.[25] Different wavelengths across this range stimulate different responses in molecules, and these different responses are suitable for different applications and searches. For instance, energy absorbances in the 30–1,000 μm (far-IR) range cause changes in the rotation of gas molecules. Meanwhile, absorbances in the 0.75–1.4 μm (near-IR) range can simultaneously excite electrons and induce harmonic vibrations in bonds. These interactions produce spectra with distinct properties according to the elements and molecular species analyzed.[25] For example, identification of organic compounds can often be accomplished by mid-IR (1.4–30 μm).[25] IR spectra can be simulated using quantum chemical calculations (see part 2 Molecular modeling and simulations: Quantum chemistry[7]), and this can often be useful when interpreting spectra or refining methods.[11] The sensitivity of IR measurements will depend significantly on the instrument details: less expensive instruments are available, but they are often unsuitable for dilute samples that are common in OoL research. Identifying individual

compounds will depend on the standards available, the complexity of the sample/mixture involved, and the amount of sample used in the analysis.[25]

IR observations can help characterize the geological materials and determine the presence of certain minerals (e.g., olivine)[26] or biogeological formations (e.g., within stromatolites).[27] It is possible to couple an IR source and detector with atomic force microscopy (AFM, see microscopy techniques), which enables the measurement of the IR spectrum of very small, specific areas in a rock sample. IR can characterize organic molecules in chemical standards, biological cultures, and environmental samples.[28] For example, it has been used on spaceflight analyses of primitive bodies (e.g., comets and asteroids) to identify small molecules (e.g., water and polyoxymethylene).[29] IR-only observations provide information on the functional groups rather than full molecular identification and are usually validated by other analytical techniques such as mass spectrometry. IR can be used to quantify kinetic rates or characterize the transformation mechanisms by the identification of products and intermediate compounds.[30] In OoL research, scientists have used IR coupled with X-ray surface analyses to study environmental samples (from 3.7 to 4.2 billion years ago) to identify possible molecules sheltered in those materials.[31] In a different study, IR analysis was used in combination with molecular modeling to analyze the activity of the amino acid alanine and demonstrate how the presence or absence of water molecules could change the ionic character of the amino acid, which is important for understanding its reactivity in abiotic conditions and its formation and preservation in space.[32]

RS

Raman spectroscopy (RS) analyses vibrational and rotational effects for materials in all phases (with molecular polarization changes in contrast to IR that detects dipole moment changes.[33] RS can be used to identify minerals in a matrix and some organic functional groups. It is more sensitive than IR or UV spectroscopy for elements that are fluorescent after a UV/IR/vis excitation, such as silicates and organics. RS can probe the chemical composition in different material surfaces, especially for inorganic matter, compared to IR spectroscopy. Hence, IR spectroscopy enables the characterization of organic compounds, while RS analyses primarily inorganic (minerals or salts) material. Both IR and RS can characterize organic-inorganic interactions either on the bulk sample or locally by mapping of the sample at the microscale (using microscopy combined with spectroscopy). One of the advantages of RS is that it can analyze a wide range of materials because it uses a broad wavelength range for excitation (from deep-UV, UV, and UV-vis to near-IR). Finally, for biological samples in OoL research, UV-vis and near-IR beamlines help take high-spatial-resolution pictures; however, the intense light source might destroy the biological material after analysis,[33] compared to RS.

In OoL, RS is used as a complementary technique to X-ray diffraction (XRD), Fourier transform infrared (FTIR), and surface microscopy. It has found applications in studying the role of amino acids and simulated prebiotic reaction mixtures in the formation of goethite,[34] the interaction of nucleobases with artificial seawater,[35] and the adsorption of amino acids on zeolites.[36] RS is usually not used exclusively but in conjunction with other techniques to establish the surface reaction of organics within a prebiotic context. RS is useful in the analysis of sediments; geobiological, organo-metallic, and organo-mineral complexes; as well as biofilms of extremophiles.

CD spectroscopy

CD is a type of spectroscopy that leverages the fact that materials can differentially absorb light of different polarizations.[37] CD can be used to determine molecular chirality, nucleic acid conformations, nucleic acid-ligand interactions, proteins' secondary and tertiary structure, as well as thermal stability, among others.[37,38] OoL studies have used this technique to differentiate chiral enantiomers,[39] to understand the unique self-assembly behavior of guanosine monophosphate nucleotides,[40] to evaluate the secondary structures of peptides and their significance in early Earth peptide chemistry,[41] and to investigate the role of transition metals in prebiotic oligomerization of depsipeptides[42] and the critical role of pH in non-thermal RNA strand separation and hybridization in the context of early Earth conditions.[43] More recently, gas phase CD has been applied to explain the plausible role of gaseous phase amino acids and their photo reactivity in the selection of L-amino acids in contemporary life forms.[44]

NMR spectroscopy

NMR spectroscopy detects changes in the local magnetic field of the atomic nuclei[45] when exposed to an external magnetic field. It relies on elements with an odd number of protons or neutrons, such as ^{1}H, ^{13}C, ^{15}N, or ^{31}P. The position and intensity of peaks in an NMR spectrum depend on the resonance frequency of the nucleus, its surrounding environment, and the strength of the applied magnetic field. NMR is an essential tool for structural characterization, providing direct information about atomic nuclei and correlations between them. It can be used both qualitatively and quantitatively.[46] However, compared to other techniques like mass spectrometry, NMR has relatively low sensitivity and typically requires millimolar concentrations ($\sim$1–10 mM), which is a major limitation. Additionally, achieving peak resolution can be challenging when analyzing complex reaction mixtures, often necessitating complementary techniques. NMR samples are prepared in dedicated tubes (often reusable), with the sample commonly dissolved in deuterated solvents. More specialized applications are gas-phase NMR[47] and solid-state NMR (ssNMR), where the latter is applied to solid samples, such as single crystals, amorphous solids, and solid-state biomolecules.[48]

NMR instruments are costly and typically require dedicated personnel for maintenance. Bench-top instruments that operate at lower frequencies are available but suffer from reduced sensitivity and spectral resolution, particularly for low-concentration samples. As a result, they are best suited for quick analysis of high-concentration samples. However, when peak assignment becomes challenging—such as with novel molecules, resolution issues, or large biomolecules—two-dimensional (2D) NMR techniques become indispensable.

The simplest and most commonly used technique is ^{1}H NMR spectroscopy. Recording a ^{1}H spectrum with standard 16 scans typically takes less than 2 min. The second most frequently used 1D NMR method is ^{13}C NMR, which generally requires 45 to

60 min per spectrum. Reaction kinetics and folding-unfolding studies of biomolecules can also be performed by acquiring a series of 1D NMR spectra at different time points, tracking nuclei such as ^{1}H, ^{13}C, or ^{31}P using a variable temperature unit. In addition to the standard 1D ^{13}C NMR, other auxiliary 1D NMR techniques, such as distortionless enhancement by polarization transfer (DEPT) and attached proton test (APT), provide additional structural insights.

2D NMR is an extension of 1D NMR, plotted on a 2D plane. Typically, both the horizontal and vertical axes display chemical shift values, showing correlations that are either homonuclear (such as ^{1}H–^{1}H) or heteronuclear (such as ^{1}H-^{13}C or ^{1}H-^{15}N). This allows the identification of correlations (interactions) between two nuclei, which may be attached covalently or interact through space via interactions known as the Overhauser effect. Several types of 2D NMR experiments exist, most of which have distinct acronyms such as correlation spectroscopy (COSY), exclusive correlation spectroscopy (ECOSY), total correlation spectroscopy (TOCSY), incredible natural-abundance double-quantum transfer experiment (INADEQUATE), heteronuclear single-quantum correlation spectroscopy (HSQC), heteronuclear multiple-bond correlation spectroscopy (HMBC), rotating-frame nuclear Overhauser effect spectroscopy (ROESY), nuclear Overhauser effect spectroscopy (NOESY), and several others. Additionally, each of these major 2D experiments has subtypes.

These techniques are invaluable for studying peptide folding and other biomolecular structures, often used alongside CD, XRD, and MD simulation models. Further elaboration on 2D NMR is beyond the scope of this manuscript, as more task-specific resources are available in the literature. For practical application, we provide a relevant case study (see A case study), demonstrating the combined use of 1D and 2D NMR with other analytical techniques to elucidate molecular structures.

Another type of 2D NMR experiment, distinct from those previously discussed, is diffusion-ordered spectroscopy (DOSY). This technique is used to determine the diffusion coefficients of molecules in a magnetic field. It is particularly useful for separating signals from different molecules or molecular aggregates within a mixture. Also known as pulsed-field gradient spectroscopy, DOSY can be performed at either low-resolution (suitable for simple mixtures with large differences in molecular sizes) or high-resolution (for species with very similar diffusion constants). For further reading on DOSY, we recommend consulting the relevant literature[49] and specific examples from OoL studies.[50,51] More complex techniques, such as three-dimensional (3D) and 4D NMR, exist and are typically used for protein structure determination and studying various folding phenomena in biological samples. These advanced experiments extend the capabilities of 2D NMR by allowing the detection of multiple nuclei correlations (e.g., ^{1}H, ^{13}C, and ^{15}N) within a sample.

In the study of OoL, NMR spectroscopy can analyze molecules containing the key biological elements CHNOPS (carbon, hydrogen, nitrogen, oxygen, phosphorus, and sulfur) using specific nuclei like ^{13}C, ^{1}H, ^{15}N, ^{17}O, ^{31}P, and ^{33}S. Among these, ^{13}C, ^{1}H, and ^{31}P are most frequently studied. Molecules associated with ^{29}Si are also examined within the OoL context, though less frequently. We hope to have conveyed the significant role that NMR plays not only in chemistry and physics but also in OoL research. We aim for this section to motivate new researchers to familiarize themselves with and utilize this powerful technique to address some of the most challenging questions in the field of OoL.

XRD

XRD is a highly specialized technique used to obtain atomic-level structural information from crystalline materials. It uses X-rays, which are short-wavelength electromagnetic waves that can interact with the interatomic spaces in crystals, as these spaces are comparable in size to the wavelength of X-rays. Detailed information about the structure of the material can be derived from directing X-rays onto a crystalline sample and measuring the angles at which these rays are diffracted.

The most common modes of XRD are single-crystal and powder XRD. A review of the incremental development of X-ray crystallography techniques and their general applications is outlined in Brooks-Bartlett and Garman.[52] XRD has applications in various fields such as materials chemistry for characterizing novel inorganic materials, in geology to identify mineralogy within rock matrices, in dairy science to study crystals in cheese, and in protein chemistry for structural determination, among others. The main drawback of single-crystal XRD is that it requires the sample to be both large and purely crystalline, rather than polycrystalline. Powder-XRD can address these issues, but it has its own limitations. It relies on known reference samples, which makes it less suitable for identifying novel or unknown materials. In contrast, single-crystal XRD is more effective for analyzing unknown samples.

OoL studies apply XRD to examine reaction mechanisms, molecular structures, and surface interactions of organics with geological substrates.[53,54] It has been extensively used in determining the biophysical properties of proteins and nucleic acids. One prominent example within OoL is its application in understanding the “last nucleotide problem” encountered in the non-enzymatic copying of a template, where the addition of the last nucleotide was inefficient, despite several attempts to resolve this.[55] XRD confirmed the mechanism of this templated non-enzymatic primer extension reaction using imidazole-activated nucleotides.[55] It was determined that an unstable structural complex between the template, primer, and the last incoming nucleotide prevented the stable binding of the nucleotide on the template, thus preventing the in-line chemical reaction. The authors used an analog (due to difficulties in crystallization of the original template-primer-nucleotide complex) of the imidazolium-bridged dinucleotide and demonstrated the conformational role of the template, primer, and the bridged-dinucleotide complex in aiding primer extension. This also highlighted differences in the slow kinetics of non-enzymatic ligation reactions compared to primer extension reactions, despite the same activation mechanism. Similarly, XRD has been used in structural evaluations associated with the substitution of non-canonical nucleobases (e.g., 2-thio-uridine instead of uridine) in short RNA strands and their base-pairing propensities in RNA duplexes,[56] and studies on RNA backbone linkage heterogeneity (i.e., 2′–5′ vs. 3′–5′ linkage) have shown that such variations might not destabilize the duplexes in a primitive RNA scenario.[57] XRD has also been used to study the interaction between organic and

inorganic materials. For example, XRD was used to characterize the formation of inorganic chemical gardens in the presence of salts and amino acids.[53] In another study, XRD patterns showed the competition of organics and dissolved salts for the occupation of interlayer vacancies in Na-montmorillonite clays.[54]

Chromatography and hyphenated techniques

Chromatography refers to a collection of techniques that separate compounds from a material or a mixture based on their properties prior to detection.[58] The fundamental principle behind chromatography is the differential interaction of components between two phases: a stationary phase and a mobile phase.[58] As the name suggests, the stationary phase is a solid material or liquid that does not move, and the mobile phase is a liquid or gas that moves through the stationary phase. After injection of volatiles/gas or solubles, the separation occurs as the sample is carried by the mobile phase through the stationary phase. Different components in the mixture interact differently with the stationary phase, causing them to move at different rates, for example, due to their polarity, molecular mass, steric hindrance, or other molecular properties.[58] This results in the separation of the mixture's components that leave the stationary phase at different times (retention times). Chromatography is a common technique to purify synthesized materials, but it can also be used for analysis. The result of a chromatographic analysis is a *chromatogram*, which shows the response of a detector (often a diode array detector [DAD]) as a Gaussian peak (corresponding to a molecule if the analytes were correctly separated) and identified by its retention time on the column. A well-separated sample contains non-overlapping peaks with distinct retention times. Optimizing a chromatographic method involves selecting appropriate stationary and mobile phases as well as the chromatographic program (i.e., eluting time, flow rate, temperature, and pressure) to improve the separation between the compounds. The most common analytical chromatography techniques are liquid chromatography (LC) and gas chromatography (GC). Chromatography is often used as an initial step in analysis, and combined in sequence or simultaneously with spectroscopic techniques (such as UV-vis or DAD) or mass spectrometry.[58] When two techniques are combined, it is common to hyphenate their abbreviations, such as GC-MS (for gas chromatography coupled to mass spectrometry). Often, hyphenated techniques provide one signal, with others providing separation power, as is the case for GC-MS. Combining two different techniques allows the limitations of individual techniques to be overcome with complementary approaches.

GC and LC instruments are very common instruments, and most chemistry labs have access to them. The instrument cost is relatively low (depending on which detector is used), and the instrument operations can be learned quickly. HPLC is a technique many researchers use with only a short training and without the requirement for additional support from a technician. While chromatography can be used for quantification, a major drawback of it is that it typically does not provide information about the identity of compounds, and validated standards are required for method development and analysis. However, the main advantage of chromatography is its flexibility and its ability to be coupled to other powerful techniques.

GC

In GC analysis, the mobile phase is usually an inert gas such as helium or hydrogen, and the solid phase is an inert solid.[58] A sample is introduced to the instrument, where it is ionized; the ions are then carried by the inert gas and interact with the stationary phase based on the properties of the solid phase and their boiling points. More volatile compounds will move faster and hit a detector sooner than less volatile compounds, creating a separation between the ions. The most commonly used detector for portable and laboratory GC is a mass spectrometer.

GC requires analytes to be volatile and may need a pre-treatment called derivatization, making the analyte more volatile by the addition of specific side chains. An example for derivatization in OoL research is the esterification of the carboxyl groups for the quantification of tricarboxylic acid cycle intermediates.[59] Depending on the detector and column used, GC analysis can have high sensitivity and resolution (down to the picogram per mL [ppt]). GC, and in particular GC-MS, has been used to demonstrate how key compounds in the citric acid cycle can be produced and broken down by iron ions in solution.[59] GC-MS enabled the identification and quantification of analytes after derivatization. It has also been used to analyze and quantify the products of Miller-Urey-type experiments, demonstrating how borosilicate glass (from a glass reactor) can play an important role in product distribution.[60]

LC

In LC experiments, materials in solution are injected into a liquid mobile phase, which then flows through a column with a solid stationary phase. The properties of the mobile phase and the column, combined with other parameters, such as the pressure and flow rate, will determine how compounds are separated. The difference between LC and HPLC has to do with the pressures, volumes, and flow rates used in the different instruments, which all affect the separation and runtime for experiments. Historically, HPLC referred to LC experiments with pressures exceeding $\sim$10 bars, and with smaller solid phase particles ($<\sim$10 μm), but the difference has been blurred, and most contemporary LC analysis is HPLC.[61]

There is a wide variety of columns for different separations. Commonly, columns are distinguished between normal phase, which means a polar stationary phase and a non-polar mobile phase or a reverse phase, where phases are reversed and the mobile phase is polar, moving over a non-polar stationary phase. Compared to GC, LC can handle a more diverse range of compounds by using different solvents, accommodating polar or nonpolar and acidic or alkaline species. LC typically offers shorter analysis time than GC. The importance of choosing the right chromatographic columns and preparation methods cannot be overstated, as these elements are critical for detecting compounds ranging from parts per billion to parts per trillion in both standard and mixture samples.

Often, HPLC is coupled with MS, as the combination of those techniques allows product identification and quantification in one injection. The UV-vis or DAD detector that can be commonly found on an HPLC, enables accurate quantification of a compound, while accurate analyte identification is possible by MS. The combination of HPLC-MS has been used for a wide range

Table 2. Different MS ionization techniques

Ionization method	Sample Media	Pros	Cons	Typical uses
ESI (electrospray ionization)	liquid	soft ionization, applicable for macromolecules	sensitive to contamination, multiple charges possible	very common ion source for LC, proteomics, lipidomics, metabolomics, complex mixtures
EI (electron ionization)	gas	simple instrumentation	limited to molecules below 600 Da, significant fragmentation of ions, sensitive to contamination	very common ion source for GC, organic mixtures, pure compounds, metabolomics
CI (chemical ionization)	gas	simple, soft ionization	not universal, neutral gas may introduce contamination	organic mixtures, pure compounds
MALDI (matrix-assisted laser desorption ionization)	solid (liquid)	soft ionisation, applicable for macromolecules	matrix selection, complex prep	proteomics solid samples
ICP (inductively coupled plasma ionization)	solid/liquid/gas	applicable for elemental analysis, fast analysis,	complexity of instruments and analysis, risk of contamination	inorganics isotope ratio/elemental analysis

Each technique provides pros and cons and are commonly used in different cases.

of sample analysis in OoL research, for example, tracking the abundance of depsipeptide oligomers through wet-drying cycles,[17] the analysis of formose reaction mixtures,[62,63] and the observation of abundance of replicators in a chemical Darwinian system.[64] These are only a few examples of the ubiquitous use of LC in prebiotic chemistry and other OoL research.

MS

MS analyses the mass-to-charge (*m/z*) ratio of ions to identify and quantify molecules in samples, revealing their composition and structure.[65] In MS, samples must be ionized, and the *m/z* ratios of the resulting ions are measured. These ratios help calculate the mass, enabling the identification of specific molecules in mixtures or the determination of molecular formulas. Additionally, the high resolution of modern MS technologies facilitates elemental and isotopic analysis. Different ionization techniques (Table 2) allow MS to analyze a wide range of materials, with its high sensitivity making it capable of analyzing molecules and mixtures in picomolar to millimolar quantities. Coupling MS with chromatography further enhances its sensitivity, enabling quantification of femtomoles of organic compounds that respond well to the chosen ionization source. Detection limits, reported as the limit of detection (LOD) and limit of quantification (LOQ), typically exceed 1 nanogram per mL (or ppb) and can reach as low as ppt, allowing for the detection of femtomoles in solutions.[66,67] However, the high sensitivity can lead to disadvantages such as susceptibility to sample contamination, high false-positive rates, and downstream challenges in data analysis.[68] Additionally, the ionization process, which transforms solid or liquid materials to the gas phase, can introduce non-linearities that can make absolute quantification difficult.[65] For this reason, MS is often coupled with other methods (such as UV-vis) to enable quantification and identification of molecules/elements in a single run.

MS instruments come in different varieties, but the main differences are the ionization and detection methods.[69] The ionization mode is an important factor to consider for the type of molecule being analyzed by MS. Most instruments can detect ions in either positive or negative mode; however, the detector cannot analyze both types of ions simultaneously. Molecules are more or less likely to gain or lose charge depending on their intrinsic properties, and therefore, a single sample can generate different signals depending on the ionization mode used. Electrospray ionization (ESI) applies a high voltage to the liquid sample as it flows out of a capillary, atomizing the sample into tiny, charged droplets. These will split into charged ions as the solvent evaporates, allowing the analytes to enter the gas phase. ESI is the most common ionization technique for samples in solution, particularly for metabolomics and proteomics analysis, but other ionization techniques, such as matrix-assisted laser desorption ionization (MALDI). A third common ionization mode is the electronic impact combined with GC (GC-MS), which can yield molecular fragmentation patterns that are used to identify molecules. MALDI is often coupled with time of flight (ToF-MS) to identify individual compounds.[69] With MALDI, the sample is required to be uniformly mixed in a *matrix* that absorbs the energy of the laser and converts it to heat energy so that the sample is not fragmented. The rapid heating allows for a small part of the matrix to be vaporized together with the sample, generating charged ions of various sizes. This technique can be used in OoL research to detect the polymerization of oligonucleotides from precursor monomers.[70] A limitation of MALDI is that each organic molecule and sample needs a specific matrix and a preparation method that may vary based on the study. This limitation also occurs in OoL research; for example, a study of polymerization reactions analyzed with MALDI is limited by mass resolution and discrimination.[71] Techniques such as MALDI have characterized highly conserved components of modern ribosomes, demonstrating their ability to catalyze peptide bond formation.[72]

The detection method (or mass analyzer and detector) describes how the ionized compounds are separated by their *m/z*

ratio and then detected by the instrument. The detection method will determine the mass range and resolution of measurements made using the technique. For example, in ToF-MS, the *m/z* ratio is determined by measuring the velocity of ions that are accelerated by an applied electric field. An example of using the ToF-MS is the detection of hopanoids as biomarker molecules in geological samples.[73]

A detailed understanding of an MS analysis requires an understanding of the ion source and the detection method. The price and performance characteristics of MS instruments can vary dramatically. Bench-top MS instruments exist, but often have restricted mass ranges, limited modes of operation, and lower resolution. High-resolution instruments with multiple analysis and detection methods are very expensive and require regular maintenance. Accordingly, many universities share MS instruments and employ technicians for instrument care and data analysis.

Quantification with MS is challenging, in part because the ionization process does not affect all molecules uniformly, leading to discrepancies between the actual number of molecules and the ions detected.[74] For quantification, introducing a carefully selected standard into the sample (often referred to as an internal standard) can help. This standard should closely mirror the behavior of the target molecule to ensure accurate comparisons in terms of ionization, molecular mass, separation, and detection performance.[75] While other techniques like isotope labeling exist for both absolute and relative quantification in MS, they are primarily developed for proteomics and may not apply broadly to other substances.[76] When coupled to a separation step (e.g., a chromatographic technique), MS can elucidate the identity of many compounds in complex mixtures.[62,77–79]

When comparing different MS instruments, it is important to consider key parameters like resolution and accuracy. Resolution refers to the instrument's ability to distinguish between closely spaced peaks in the spectrum, such as ions with very similar masses. Accuracy, on the other hand, refers to how closely the instrument's measurements of ion masses match the true or accepted values. Different instruments and even different settings on the same instrument can offer various benefits and trade-offs. High-resolution MS (HRMS) is a term usually used for mass spectrometers with ToF, orbitrap, or Fourier transform ion cyclotron resonance analyzers. Orbitrap MS instruments are known for their high resolving power across a broad mass range and maintain high mass accuracy. This capability allows accurate assignment of exact molecular formulas to detect ions and differentiate ions that have the same elements but different isotopes. In contrast, triple-quadrupole MS instruments, which typically have lower resolution, are more suited for robust relative quantification of known ions.[45,65] HRMS techniques require the help of a technician and several months of method development and post-processing analysis, as the analysis leads to the detection of hundreds of compounds. In the following subsections, we discuss the application of MS to identify specific molecules as well as elemental and isotopic compositions.

Molecular analysis

MS is commonly used to confirm the identity of individual molecules in a sample. This process often requires coupling MS with another technique and relies on additional data about the molecules, such as their mass and their propensity to fragment upon ionization. Identification typically utilizes reference spectra from extensive databases, although in some cases, simulated spectra can also be used.[80] One of the most common MS techniques in this regard is MS coupled with chromatography (GC-MS or HPLC-MS) to identify and quantify organic compounds (specifically in untargeted metabolomics' studies).[81] For example, GC-MS has been used to analyze complex mixtures like those from Miller-Urey experiments,[60] and conducted time series analysis of the formose reaction.[60,63] ESI-MS has been used to study the selective formation of peptide bonds on mineral surfaces,[82] and for the analysis of formose reaction products, a known combinatorial explosion that entails hundreds of compounds in solution.[62]

A powerful way to identify small molecules with MS is through tandem mass spectrometry (MS/MS), in which the analytes are fragmented more than once.[65] In the first mass analysis step (which is identical to conventional MS) "parent ions" (MS^1) are identified. These parents are then fragmented and reanalyzed by a second mass analysis step (MS^2). With some instruments, this process can be repeated several times (MS^n), but since the molecule is fragmented in each step, leading to more fragments and fewer ions overall, more than two fragmentation steps are less common. MS/MS provides information about the ionized analyte via its fragmentation or degradation pattern. The fragmentation pattern can be used to assign the molecular composition and often the structure of the molecule. Typically, this technique depends on databases of reference spectra, which exist for conventional biopolymers such as proteins.[83] However, it can also be used to sequence non-conventional polymers, such as depsipeptides, which have been used in OoL research and cannot be sequenced using traditional sequencing techniques or traditional proteomics methods.[84,85]

Elemental and isotopic analyses

MS is commonly used for the identification of molecules in sample mixtures, but it can also be used for isotopic and elemental elucidation. Isotope-ratio MS (IR-MS) enables the identification of isotopes in bulk materials or individual molecules. IR-MS can discriminate between nuclei such as ^{2}H and ^{1}H, ^{13}C and ^{12}C, ^{15}N and ^{14}N, or ^{18}O and ^{16}O.[86] The determination of the relative abundance of a non-radiogenic element provides information about isotopic fractionation in a sample due to the origin of the material. This can be useful for distinguishing biological processes from abiotic backgrounds, because biological processes tend to be enriched in lighter isotopes of carbon and nitrogen.[86] Similarly, the relative abundance of radiogenic nuclei and isotopic proxies can be used to date material in the different strata of a sediment (e.g., $^{14}C/^{12}C$, $^{18}O/^{16}O$, $^{40}Ar/^{39}Ar$, $^{87}Sr/^{86}Sr$, and $^{206}Pb/^{204}Pb$).[87] IR-MS has been used to analyze extraterrestrial objects (comets, asteroids, planets, natural satellites, etc.), which can yield insights into the composition of the early solar system and therefore of Earth, providing information about Earth's early history, before, and during OoL.[88]

Inductively coupled plasma MS (ICP-MS) ionizes and analyzes a sample to produce atomic (or small molecular) ions.[89] Unlike the other ionization methods described here, this method is

high temperature and atomizes most small organic molecules, meaning it is unsuitable for the analysis of most biomolecules; however, it is well suited for the analysis of trace elements and isotopes of those elements. The coupling of ICP with MS facilitates efficient ionization of the analyte, leading to exceptionally high sensitivity, and enables the analysis of substances at very low concentrations (down to ppt concentrations with high sensitivity and relatively high resolution according to the MS detector), including the detection of trace elements in geological samples.[90] This technique has been used to trace the origin of the elements in meteorites.[91] Compared with alternative methods for measuring trace metal abundances, the advantage of ICP-MS is that it is more sensitive, can measure multiple elements simultaneously, and has relatively simple sample preparation requirements. The major drawback is that ICP-MS instruments are highly specialized and, therefore, expensive to purchase and operate.

MS techniques are used across many scientific domains and in the OoL community because they offer a wide range of benefits and flexibility. They are particularly powerful when combined with complementary techniques and large reference datasets. For this reason, MS has continued to play a central role in chemistry, proteomics, and metabolomics. We anticipate it will continue to play a central role in OoL research.

Microscopy techniques

Microscopy techniques are tools to investigate what the sample "looks like" in the most direct way. They involve the examination of objects and their details using microscopes, enabling the visualization of structures and features at scales beyond the naked eye's capabilities. Some microscopy techniques can be combined with spectroscopy, enriching the physical insights with chemical information. Raman and IR spectroscopy are most commonly used for this purpose.

Microscopy techniques can be used in OoL research[92] to visualize supramolecular structures assembled from primitive chemicals or chemical ensembles, such as protocells and other primitive compartments,[92,93] nanofibers and molecular self-assemblies,[94,95] and mineral structures (both terrestrial[96] and extra-terrestrial[97]). Here, rather than providing a detailed review on the entire breadth of microscopy techniques available to researchers, we briefly highlight a selection of microscopy techniques commonly used in OoL research.

Light and fluorescence microscopy

Light microscopy (also known as brightfield microscopy) illuminates a sample (typically prepared on a transparent surface). As the light traverses the sample, it can be absorbed, scattered, or deflected, which alters the light's path before it is magnified by an objective lens and captured by a detector or camera. The resulting pattern in the detector forms the image we see.[98] A series of images acquired over time can be strung together to create a "movie" of a sample. However, depending on the physical characteristics of the sample, it might not always be feasible to distinguish between features of the material and noise or other artifacts. In these situations, alternative methods exploit light's inherent properties (such as phase shifts or changes in the light's polarity after it interacts with the sample). These methods include phase contrast microscopy,[99] differential interference contrast (DIC) microscopy,[100] and polarization microscopy,[101] each of which has been applied to OoL studies for both static image and movie acquisition (e.g., Kindt et al.; Jia and Fraccia[102,103]).

In certain situations, direct light cannot be used to visualize samples, often due to a lack of contrast with their background. In these cases, a technique known as fluorescence microscopy can be used to detect fluorescence signals in samples, allowing for their visualization.[104] In fluorescence microscopy, light of a specific *excitation* wavelength is directed onto a sample, causing fluorescent molecules within the sample to emit light at a different, specific wavelength. This emitted light pattern is then collected by an objective lens and captured by a detector, creating an image of the sample. Some chemicals and materials are naturally fluorescent, including specific protocells,[105] lipids and proteins,[106] RNA and DNA,[107] as well as other organic compounds useful for biogeochemistry studies[108] and fluorescent mineral studies.[109] For OoL studies, samples from environmental sources or synthesized in the lab that contain these naturally fluorescent materials can often be imaged directly, or with minimal processing and sample preparation, using a fluorescence microscope. However, samples without such natural fluorescence often need a "fluorescent tag" to be added into the system to be visualized by a fluorescence microscope. These "tags" can be small molecules or proteins that can be non-covalently incorporated into the sample (e.g., fluorescent thioflavin T labeling of peptide and DNA[110] or fluorescent intercalation into nucleic acid duplexes to identify species or quantify potential prebiotic oligoribonucleotides[111,112]). Another option to fluorescently label a sample is to covalently link a fluorescent molecule to a component of the sample so that when the sample is produced, the fluorescence is already intrinsically apparent within the sample: examples of this include green fluorescent protein (GFP) fusions[113,114] or fluorescein labeling of nucleic acids.[115] Such fluorescence microscopy techniques can also take advantage of fluorescence transfer of pairs of fluorescent molecules (i.e., the fluorescence emission wavelength of one molecule equating to the excitation wavelength of another) to visualize molecular interactions between fluorescently tagged components in a sample through fluorescence resonance energy transfer microscopy.[116] Single-molecule particle tracking methods using fluorescence microscopy have also been used to analyze the diffusion[117] and coalescence/wetting[118] properties of membraneless protocells.

Confocal microscopy and optical coherence tomography

Using a high intensity of direct light for fluorescence excitation, known as epifluorescence, works well for general sample imaging under a microscope. However, this method may not always provide sufficient detail about the sample's spatial layout, especially in the vertical (*z* direction) dimension in thicker samples. To overcome this limitation and achieve precise vertical resolution, confocal microscopy is utilized.[119,120] This advanced technique has been particularly valuable in OoL research, such as exploring the hypothesis of giant vesicle colonies[121] or examining interactions between organic compounds and minerals.[122] Similar to epifluorescence microscopy, confocal microscopy also uses

incident light to stimulate fluorescence in a sample. However, a confocal microscope uses a laser to target and excite specific regions within the sample's horizontal dimension (width and length). This targeted approach reduces the risk of damaging the sample. Photobleaching, the fading of fluorescence due to prolonged light exposure, can lead to the deterioration of the sample or its fluorescent signal, but it can be minimized through the controlled use of laser light in confocal microscopy.[123,124] It is possible to enhance image clarity and depth resolution in confocal microscopy through a component known as a pinhole. This pinhole ensures that only light from the precisely focused plane of the sample contributes to the image, effectively filtering out unfocused light. By adjusting the pinhole to a small size, the microscope captures high-resolution images of extremely thin cross-sections of the sample along its vertical axis (*z* direction). Sequentially imaging slices at various depths allows for the assembly of a detailed 3D reconstruction of the sample, achieving exceptional spatial resolution.

Confocal microscopy not only enhances imaging precision but also supports dynamic studies of complex systems. A notable example is fluorescence recovery after photo bleaching (FRAP).[125] In this technique, the microscope's excitation laser is intensified to deliberately bleach a targeted area within a fluorescent sample. Subsequently, the recovery of fluorescence within this area is observed over time. The speed at which fluorescence returns can reveal key physical characteristics of the system, such as molecule exchange rates between compartments,[126] the diffusion rates of fluorescent molecules,[127] or the phase (solid or liquid) of the sample.[128] This method provides valuable insights into the molecular dynamics and structural properties of the system under study.

EM

Electron microscopy (EM) uses a beam of accelerated electrons to generate images. The resolution attained is significantly higher than that of light microscopy due to the electrons' short wavelength and, depending on the electron microscope type and operating mode, can extend into the picometer (pm) range. EM has a broad range of applications. There are primarily two types of EM–scanning (SEM) and transmission (TEM). SEM generates images by sweeping a focused electron beam across the specimen and detecting the energy variation resulting from electron-specimen interactions.[129] On the other hand, TEM involves transmitting an electron beam through a specimen to produce a high-resolution image.[130] It can be used to examine the specimen's nano to atomic-scale internal structure, including the arrangement of organic molecules, the nano-structure of inorganic minerals, and elemental distributions for investigating organic origins or dating rocks and fossils (although interpretation of these data is a source of debate).[131–133] For TEM analysis, the specimen must be prepared as an ultrathin foil (<100 nm thick) to enable electron transmission and image projection. This method can also accommodate nanoparticulate matter placed on small grid plates for electron beam exposure. TEM typically offers higher resolution than SEM but requires more extensive sample preparation due to its focus on interior details; whereas SEM, which images the specimen's surface, allows for the examination of more intact samples with less preparation.

Recent research has utilized SEM or TEM in conjunction with spectroscopy to explore the oceanic lithosphere.[134,135] These studies focus on the relationship between catalytically active minerals and nearby abiotic carbonaceous materials, which are thought to offer free energy for chemical synthesis.[119] Through high-resolution imaging, SEM and TEM reveal the detailed structure of rocks, emphasizing the role of nanocrystalline minerals in catalysis and how the porous nature of these minerals may enhance the synthesis of abiotic organic compounds by confining reactions at the nanoscale.

In the context of studying the OoL through geological perspectives, it becomes essential to examine the surface chemistry of minerals and rock formations that possess catalytic properties. The advancement and accessibility of *in situ* liquid and/or gas-cell TEM, along with atomic-resolution aberration-corrected TEM, enable the direct observation of catalytic sites on mineral surfaces during experiments that mimic prebiotic hydrothermal conditions. These cutting-edge EM techniques open the door to gaining novel insights into the physical and chemical processes involved in the synthesis of prebiotic organic compounds.

AFM

AFM generates images of surfaces by moving a small, sharp probe over the material. The probe has a cantilever attached to it and is deflected by the structure of the surface. The deflection can be measured, usually with a laser, and this gives insights into the properties of the surface, including its shape. The probe is often less than a few nanometers in width at the tip, and this can create high-resolution images of the sample. AFM can generate images at a higher resolution than optical microscopy, approaching the detail of EM. It has some advantages over electron microscopy, because its sample preparation is less complicated than what is required for EM methods. The properties of the probe can be chosen to measure more than just the topography of the material, such as electrical or magnetic properties of the surface. AFM has been used to characterize the hypothesized ancient microfossils,[120,121] and such analysis can often be complemented with RS to characterize the shape and composition of materials. AFM can also be coupled directly to IR spectroscopy.[120] AFM has also been used to analyze aggregates formed from the composition of non-canonical nucleobases.[122] AFM is still widely used in many labs, particularly in materials science; however, optical and EM are much more common.

Genomic sequencing

Nucleic acids play a fundamental role in modern biology, and it has been postulated that primitive systems on early Earth also incorporated nucleic acids[136] or nucleic acid-like molecules.[137] While previously mentioned techniques can provide extensive insight into nucleic acids and amino acids, identifying the exact order in which these encode information, a process known as sequencing, is crucial for deepening our understanding of both the primitive and contemporary aspects of functional polymers.[125]

Sanger sequencing

Sanger sequencing, a foundational method developed in the 1970s, revolutionized genetic analysis by enabling the determination of the precise order of nucleotides in a DNA

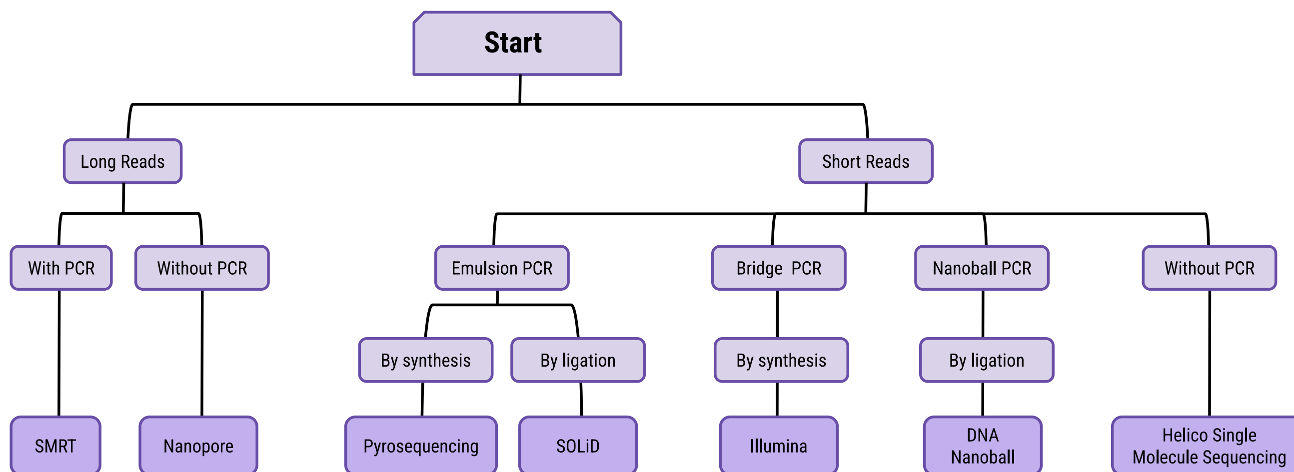

Figure 3. An overview of various next-generation sequencing technologies, showcasing their underlying principles
Figure adapted from Satam et al.[128]

molecule.[138,139] This technique relies on DNA synthesis, where a DNA strand is copied in the presence of normal nucleotides (i.e., deoxynucleotides) and special versions of nucleotides called dideoxynucleotides. In the copy of DNA, while the DNA strand is elongated using deoxynucleotides, the incorporation of a dideoxynucleotide terminates the DNA chain. These dideoxynucleotides, specific to one of the four DNA bases (A, T, C, or G), are each tagged with a distinct fluorescent marker. By copying the DNA in a mixture containing these terminating nucleotides, a series of DNA fragments of varying lengths is produced, each ending at the position of its respective base. These fragments are then separated by length using gel electrophoresis, and their sequence is read by detecting the fluorescent tags. The result is a readable, linear sequence of bases, revealing the genetic information contained within the DNA sample. Sanger sequencing has been instrumental in countless genetic research projects, including the Human Genome Project, offering insights into the fundamental blueprints of life.

The main advantages of Sanger sequencing are its high accuracy, cost-effectiveness, and the simplicity of its well-established protocol. In OoL research, Sanger sequencing is commonly used to identify enriched variants at the end of the *in vitro* selection processes[126,127] (see part 2 Synthetic biology: Protocells[7]). However, Sanger sequencing also comes with notable limitations. Its lower throughput, in comparison to high-throughput sequencing (HTS) technologies (see Genomic sequencing: High-throughput sequencing), limits its applicability for extensive genomic projects. Additionally, it is best suited for analyzing sequences up to about 900 bases, which constrains its utility for longer DNA fragments.

HTS

The advent of HTS techniques has revolutionized genomic research by enabling the analysis of millions or even billions of bases swiftly and cost-effectively. Unlike Sanger sequencing, which focuses on single nucleic acid sequences, HTS can simultaneously process a vast array of nucleic acids from mixed samples. Over the past two decades, several HTS methods have emerged, including long-read techniques like single-molecule real-time (SMRT) and Nanopore sequencing, and short-read methods such as pyrosequencing, SOLiD, Illumina, DNA Nanoball, and Helico single-molecule sequencing, among others. Figure 3 summarizes these technologies, categorizing them by read length, PCR usage, and sequencing method (synthesis or ligation).[128] Among these, sequencing by synthesis and nanopore sequencing have become particularly prominent in OoL studies, chosen for their optimal balance of cost, read length capabilities, and throughput. Importantly, both of these platforms are equipped to analyze both DNA and RNA, directly or indirectly, addressing a critical need in some OoL studies.

Sequencing by synthesis, most commonly used on Illumina sequencing platforms, requires a DNA library consisting of many copies of each sequence from the original sample, typically generated by polymerase chain reaction (PCR). While DNA can be amplified directly by PCR, RNA must first undergo reverse transcription to synthesize complementary DNA (cDNA) before amplification, a process known as reverse transcription PCR (RT-PCR).[140] Once these amplified libraries are prepared, they are subjected to a flow cell, where sequences attach to its surface. A flow cell can accommodate up to billions of clusters, enabling billions of sequences to be analyzed in parallel, which highlights the high throughput of this technique.

However, the vast scale of sequencing data generated in a single run requires significant processing to produce accurate and usable information. One potential issue during library preparation is the introduction of adapter bias; in a typical library preparation, defined short sequences called adapters are ligated to RNA or cDNA of interest for its amplification or sequencing. However, adapter ligation exhibits differential efficiency depending on target RNA or cDNA sequences, leading to unequal representation of sequences in the final library. The ligation step in RNA sequencing library preparation is a recognized source of bias, driving improvements in enzyme technologies and library construction methods.[141] This bias can

be further amplified during subsequent PCR amplification, ultimately affecting the accuracy and reliability of sequencing data. Nevertheless, certain techniques during the library preparation phase, such as sample barcoding,[142] can help accurately quantify, categorize, separate, or increase the throughput of sequences within a library. Barcoding also helps to minimize sequencing noise and bias.

Sequencing by synthesis techniques have been heavily used in the OoL field to study the landscape profiles of evolving primitive nucleic acids that have a specific function (such as aptamers, RNA strands that bind to small molecules[143]) that can catalyze reactions (such as in ribozymes[144]) and that can polymerize (such as in replicating systems[145]). Sequencing studies on systems combining genotype with phenotype information by linking peptides with nucleic acids through display methods have also afforded more detail into evolving functional peptide landscapes[146] or in the synthesis and evolution of ribosomes[147].

We note that while sequencing by synthesis is typically associated with Illumina sequencing platforms, a similar principle is also utilized in MGI sequencing platforms, which employ sequencing by combinatorial probe-anchor synthesis (cPAS) combined with DNA nanoball technology.[148] Recent studies have shown that the sequencing quality, accuracy, and sensitivity of both platforms are comparable, while MGI platforms can operate at lower costs per data.[149] In OoL research, sequencing by cPAS has been used to study non-enzymatic ligation and recombination dynamics in short random RNA mixtures.[150]

While the sequencing techniques mentioned above have provided significant advances in understanding nucleic acid evolution and function both within and outside OoL studies, these techniques do not allow the sequencing of modified or non-canonical nucleotides, which may have been relevant on the early Earth.[151] Although mechanisms to sequence non-canonical nucleotides, such as by mass ladder analysis[152] or "indirect" sequencing techniques requiring extra library preparation steps,[153] have been used, they still require further optimization for widespread and general use and require further development to efficiently and effectively detect modified nucleotide bases.[154]

One such technique that has recently been used for such purposes is nanopore sequencing,[155–157] which directly sequences analyte nucleic acids (as opposed to sequencing by synthesis, which indirectly sequences nucleic acids by producing a complementary strand to the analyte).[128,158] First, the nucleic acid to be analyzed passes through a nanopore, which is composed of a protein in a synthetic polymer membrane. While passing through the pore, one by one, each of the bases elicits an ionic current. This current is unique for each base, due to their differences in electronic structure, and the sequence of the nucleic acid is inferred based on the string of different ionic currents that appear when the nucleic acid transits through the pore. Some OoL research has used nanopore sequencing to study the oligomerization of nucleic acids under simulated early Earth geological conditions, such as in hydrothermal fields.[106,159,160] It should be noted that nanopore sequencing exhibits relatively high error rates,[161] limiting its application, especially in sequencing highly random DNA/RNA mixtures such as evolving populations as mentioned above. Improving error-correction methods in the future could help overcome this challenge, though ongoing improvements in algorithms and chemistry are also working to mitigate these issues.

Because each base elicits a unique ionic current in nanopore sequencing, this means that not only can canonical bases be detected, but non-canonical bases and base modifications can also be identified based on each of their respective unique ionic currents. The nanopore system can theoretically be optimized to detect nearly any non-canonical base or base modification,[162] as well as even amino acids in a peptide,[138] and variations of nanopore sequencers containing inorganic nanopores or electrodes have also been developed.[139] However, given that there are hundreds of known base modifications of RNA alone,[152] not to mention a very large number of non-canonical bases,[163,164] the resolution required to distinguish ionic currents of all of these bases within the same nucleic acid polymer may be practically challenging to achieve. Such thorough analyses may only be possible through further technological development.

Another potential advantage of nanopore sequencing is its ability to read long sequences (e.g., over 10,000 bases) without fragmenting them to around 500 bases or less, which is necessary when using sequencing by synthesis techniques. However, in some OoL research, another long-read sequencing technology, SMRT, has been chosen to analyze evolving populations of genomic RNA (approximately 2,000 bases) due to its accuracy after error correction.[145,165] Sequencing by synthesis techniques have not been applied to these studies because the fragmentation process disrupts the combination of mutations within the same genomes.

In summary, while Sanger sequencing is best suited for projects requiring high accuracy in the sequencing of short DNA fragments (such as precise mutation detection, validating sequences obtained by other methods, or sequencing small genomes), HTS technologies are ideal for large-scale genomic studies (such as whole-genome sequencing, transcriptome analysis, and metagenomics), where the goal is to process a vast amount of genetic material quickly and cost-effectively. In addition to selecting the appropriate sequencing technology for specific research goals, it is crucial to address challenges related to transparency and reproducibility in sequencing workflows. These challenges are particularly evident when custom scripts are used, especially those tied to patents or proprietary methods. To mitigate these issues, one potential solution involves using repositories with restricted access under academic licenses, sharing simplified versions of code, and providing detailed workflow descriptions. These measures can balance reproducibility with intellectual property concerns while fostering scientific openness.

A case study

OoL researchers have a wide variety of analytical techniques at their disposal. The choice of the technique depends on several factors such as availability, ease of access, costs, and technical knowledge. We have introduced the basics of analytics throughout this section, and now shall consider examples to demonstrate how a combination of the aforementioned techniques can be employed to address research questions and ambiguities that arise from specific techniques. We will discuss a key article as an example, which is available as open access.[122]

The paper considers the possibility of non-biological nucleotides. Structurally, a nucleotide monomer can be simplified into three components—a sugar, a nucleobase, and a phosphate. The sugar is connected to the nucleobase by a glycosidic bond, and the phosphate is connected to the sugar by an ester bond. Many studies in the OoL consider how biological nucleotides came to be, and if alternative chemical systems are possible, or if modern nucleic acids were preceded by simpler compounds. Cafferty et al.[122] show that under aqueous reaction conditions, melamine and barbituric acid (BA), two nucleobase-like molecules, form glycosidic linkages with ribose and ribose-5-phosphate (R5P), resulting in non-canonical nucleosides and nucleotides, respectively. The nucleotides of melamine (MMP) and barbituric acid (BMP) were also found to form supramolecular assemblies within the crude reaction mixtures.

The study by Cafferty et al.[122] used a combination of techniques to investigate if BA and melamine could form nucleotides with R5P. The study utilizes two specific types of chromatography for purification of nucleotides from the reaction mixture, ion-exchange chromatography (BMP with anion exchange and MMP with cation exchange), and HPLC. UV-vis spectroscopy was used for the detection and characterization of MMP and BMP nucleotides at 260 nm in a sodium dihydrogen phosphate buffer (pH 7). HRMS was used for the confirmation of BMP and MMP masses. NMR spectroscopy was used for structural and stereochemical characterization of the nucleotides. CD spectroscopy was used to analyze the supramolecular assemblies. AFM was used for imaging of the supramolecular topographies.

HPLC coupled to an MS is a good technique to determine the product mass (m/z), especially because it allows detection at micromolar concentrations. The authors in this case used HPLC-MS to confirm the product masses and reported that the product separation on the HPLC was detected using UV spectroscopy at 260 nm. The authors confirmed the m/z of the products BMP and MMP nucleotides in the negative ion mode (m/z 339.0243 and 337.0659, respectively). While HPLC-MS provides a mass, it does not give any information on the stereochemistry, conformation, configuration, or structure. If a product has two isomers of the same m/z, they would not be differentiated by HPLC-MS unless a column and method were designed to separate them. Therefore, in this study, NMR was used to determine the structure. Optimized reactions generated products in concentrations that were undetectable by NMR, so a bulk scale column chromatography was performed to isolate the products and pool them together to obtain sufficient concentrations for NMR analysis.

Once sufficient quantities of BMP/MMP were obtained, the authors used ^{1}H NMR to confirm the products. The authors used 2D NMR (see "NMR spectroscopy") techniques like HSQC, HMBC, COSY, and TOCSY in conjunction to confirm the proton signal assignments and glycosidic bond formation between C1′ of the sugar and C5 of the non-canonical nucleobases. 1D rotating frame Overhauser effect (ROE) was used to confirm the anomeric ratios (α vs. β anomers). Note that canonical nucleobases have a C1′–N glycosidic bond, which is less stable than a C1′–C glycosidic bond observed with BMP in this specific study.

The authors used CD (see "CD spectroscopy") to observe any possible supramolecular assemblies in the mixtures of BMP and MMP. In the case of amino acids and peptides, CD is the go-to technique as the peptide secondary structures have unique signature-curves.[166] To further study supramolecular assembly, the authors used a type of AFM (see "AFM") and, in conjunction with the CD observations, concluded that BMP and MMP nucleotides assemble as hexad stacks through non-covalent interactions and give rise to linear strands as observed under AFM.

As one can see how a combination of techniques was required to establish a simple non-canonical nucleotide synthesis and how one can glean structural information to study a molecule in such precise detail. From a chromatographic point of view, the authors have used anion exchange chromatography for BMP and cation exchange chromatography for MMP. The reason for not using the same type of column for both types of nucleotide is the ability of the molecules to possess net surface charges as a function of pH. BA would more easily lose a proton at the pKa of the amines to have a net negative charge and bind to a cationic resin surface, whereas melamine would be more easily protonated on its amines and bind to an anionic surface. The authors have also used HPLC to obtain very high-purity compounds. Sometimes, flash chromatography is used for the initial purification of the crude reaction mixture, followed by the use of HPLC for the separation of any closely eluting molecules. The authors report HRMS in the BMP/MMP case; however, it might be beneficial to provide detailed information about the fragmentation conditions used for the analysis, so that it aids folks who might want to replicate or adapt the protocol.

The authors used CD to study the base-pairing phenomenon. While XRD or NMR give more detailed structural information, CD is a relatively quick technique providing more flexibility with solvents, temperatures, pH, and ionic strength of the samples. CD can be used to obtain signature curves of nucleic acids arising from secondary structures.[167] Molecular dynamic simulation could also be used to some degree in predicting secondary structure phenomena of nucleic acids (see part 2, "molecular modelling and simulations").[7,168]

As one can see, analysis is a bottleneck in experimental OoL, and thus we hope that this review helps any new-comer to comprehend the challenges and develop approaches that can improve our understanding of complex prebiotic chemical networks, often termed as dirty reactions. This section (and the review), in general, does not intend to show a fixed method to solve analytical barriers, but its intent is to demonstrate how these analytical tools are used in OoL to understand and study reactions.

The experimental techniques outlined here have been used extensively in OoL research in recent decades. Having an operational understanding of them can help OoL scientists understand and contextualize experimental results, possible sources of error, and alternative interpretations of data. As with all science, there may be new experimental techniques that will dominate the field in the coming years, but these are the primary sources of empirical data to date.

DATABASES IN OoL

Characterizing the mysterious sample with which your advisor provided you will be no easy task. As you may have noticed in "experimental techniques for studying the origin of life," many

experimental approaches require reference spectra, standards, or other information to be useful. Much of this information is housed in online databases. In some cases, these databases are new or recently developed, such as the Open Reaction Database (ORD). Many tasks in OoL research are feasible or easier only if some relevant databases are readily available. For example, proteomics and metabolomics would not be possible without databases like MassBank and U.S. National Institute of Standards and Technology (NIST).[169,170] Similarly, inferring possible sequences of ancient or even prebiotic proteins requires databases of protein sequences in various microbes. Understanding the data that are already available to you will help you understand your sample better and help you craft analyses with the appropriate controls. Thanks to the development of many techniques described in "experimental techniques for studying the origin of life" and other efforts of previous researchers, there are already multiple online databases that may help you achieve your specific OoL research goals.

Note that most of the databases were not developed to solve problems in OoL; for example, chemists developed some reaction databases fitting the needs of chemical engineering instead of prebiotic chemistry research. Not until recently did researchers start to develop databases dedicated to specific OoL problems (e.g., the Astrobiology Habitable Environments Database [AHED][171]). In addition, some OoL research goals may require data that are still beyond what modern techniques can achieve (e.g., reaction rate constants of thousands of reactions under prebiotically plausible environments). Therefore, many of the most popular databases lack the metadata and context required to unambiguously characterize key processes involved in the OoL. Large chemical databases can give the misleading impression that they contain information about the most relevant compounds; however, the space of chemical compounds is vast, and the space of chemical reactions is far larger, even without considering different experimental or environmental parameters.

Despite limitations, various databases have proven useful in OoL research in recent years, such as multiple works that will be mentioned in part 2, and emerging trends.[7] Here, we review these databases and highlight how they have been used in the context of OoL. We have broken these into two broad categories: (1) physical and chemical databases that contain data that may or may not be relevant to Earth's biochemistry, and (2) biochemical and biological databases that contain information relevant to Earth's biochemistry, including biochemical reactions, gene sequences, and protein structures.

Physical and chemical databases

Access to large-scale, standardized physical and chemical data can help build theoretical models of abiogenesis, identify candidate experimental systems for synthesizing biomolecules and life-like molecular systems, locate missing links in research paradigms, and predict environmental conditions that may better facilitate OoL and more. Databases of chemical species and reactions can be roughly categorized into two types: databases of experimentally confirmed reactions and databases of rule-based (or algorithm-generated) reactions (see part 2, "chemical thermodynamics, kinetics, and networks: chemical reaction networks" for more information about reaction network generation[7]). In this section, we focus on general physical-chemical data not linked to specific living organisms. For chemical data linked to living organisms, see "biochemical and biological databases."

Some physical and chemical databases used in OoL research are summarized in Table 3. These databases were used in tasks, such as comparison between computationally generated libraries of molecules and databases of empirically confirmed molecules and detecting autocatalytic cycles among abiotic reactions.

To algorithmically generate molecular structures, there are multiple software tools available, including MolGEN (closed source),[173] OMG (open source),[174] and Surge (open source).[175] Rule-based/algorithmically generated reaction databases are currently rare; however, one example is the AllChemy database.[176] The AllChemy database is a partially open-access database generated by machine-learning algorithms (requires registration; some of the features mentioned in the publication are not available through the online portal). Another example is MØD, which provides a software package for graph-based cheminformatics that can be used to generate a rule-based reaction network by specifying a set of reaction rules.[177] To learn more about these methods, see part 2 "Chemical thermodynamics, kinetics, and networks: Chemical reaction networks."[7]

Biochemical and biological databases

Biochemical and biological data are primarily composed of genomic sequences, the enzymes encoded in those sequences, and the molecules and reactions implicated by those enzymes. Sequence databases can contain experimentally confirmed or computationally predicted metadata. For example, in UniProt, a star indicates if the descriptions of a protein have been manually curated or just inferred computationally; the latter occurs when a protein sequence matches another in the database that has been experimentally characterized.[178] Other types of data include experimental standard measurements for identifying known metabolites or enzyme structure databases to map protein sequences to plausible structures and functions.[179–181]

Following extensive work on the functional annotation of genomes and linking enzymology and genomics, large-scale biochemical databases that include both biological and biochemical data have arisen; for example, the Kyoto Encyclopedia of Genes and Genomes (KEGG)[182] has been used extensively to investigate the origins of early metabolic networks[183,184] (for more detail on these, see part 2 Chemical thermodynamics, kinetics, and networks and molecular phylogenetics[7]). However, the annotations of compounds and reactions are often not detailed enough, and therefore, current biochemical databases are often insufficient for detailed explorations of the biochemical space. For example, it is often unclear if a reaction is one-step or multi-step and, in some cases, a multi-step reaction and its corresponding single-step reactions all have records in the database, which creates duplicates. In other cases, the distinction between reactants, reagents, and catalysts is unclear, reactions are not mass-balanced and/or lack stoichiometric information, and different chemical species share the same name (e.g., starch and glycogen). Kinetic data are almost always absent, reaction conditions (pH, temperature, salts, and buffers) are mostly

Table 3. Examples of physical and chemical databases commonly used in OoL of life research

Database name	Website	Contents	Notes	OoL use cases
Reaxys	reaxys.com	chemical compounds, materials, reactions, patents, and bibliographic information	commercial license sometimes "reactants" and "reagents" are not rigorously distinguished, and some reaction equations are not balanced	Cleaves et al.[164]
U.S. National Institute of Standards and Technology (NIST) Standard Reference Data (SRD)	nist.gov/srd	collection of databases; spectral, thermochemical, physical, and structural properties for chemical compounds; and kinetic and thermodynamic data for chemical reactions	49 free SRD databases and 41 fee-based SRD databases	Schrimpe-Rutledge et al.[83]; Tang et al.[172]
CHEMnetBASE	chemnetbase.com	collection of interactive databases and dictionaries for chemical species, e.g., CRC Handbook of Chemistry and Physics	commercial license interactive features (e.g., visualization tools, customizable workspace)	–
CAS database	cas.org/cas-data/cas-reactions	single- and multi-step reactions yield data, detailed reaction conditions, defined substance roles, and experimental procedures.	commercial license	–
Open Reaction Database (ORD)	open-reaction-database.org	chemical reactions, reactants, products, catalysts, conditions, and yields, along with experimental procedures and observations	open access	–

missing or expressed in non-compatible formats (poor metadata), and often the phase of a chemical species involved in the reaction is not specified. In the case of rule-based/algorithm-generated datasets, the reliability of predictions about what reactions are possible may not be high, and measures of uncertainty are lacking. All of these problems are worse when we consider chemical reactions not included in modern biochemistry, many of which may be relevant for understanding prebiotic chemical processes or understanding the structure of biochemistry by comparing it to alternative possibilities. Nevertheless, one should keep in mind that the lack of data could be largely because techniques of our time are still not advanced enough to obtain that data and not because developers were unwilling to include that data.

However, advances have been made with what is available today. The KEGG database has been filtered for reactions from anaerobic prokaryotes, potentially closer to primordial metabolism.[184,185] In a similar manner, the MetaCyc database has been manually filtered for core metabolic reactions to reconstruct a prebiotically plausible autotrophic metabolism.[186]

The location, references, and other useful information of multiple databases of biological and biochemical data commonly used in OoL research are summarized in Table 4. These databases were used in multiple tasks, such as reconstructing ancient metabolic networks, inferring the structure of ancient proteins, and exploring the origin of translation machinery.[183,184]

CONCLUSIONS

The analytical foundations of OoL research draw on mature techniques from biology, geochemistry, and analytical chemistry, many of which are now broadly accessible to experimentalists entering the field. Adapting these methods to abiotic, prebiotic, or synthetic biological systems remains challenging and often requires time-intensive protocol development. However, existing literature and databases provide valuable templates for accelerating progress. Importantly, no single technique can capture the full complexity of any given material or process—there is no one-size-fits-all tool for molecular characterization. For this reason, complementary approaches and multimodal analyses are essential for reliable detection and interpretation. Yet, the diversity of data types and analytical workflows continue to challenge cross-disciplinary integration. Establishing shared standards for data quality, accessibility, and reproducibility will be essential for

Table 4. Examples of biochemical and biological databases commonly used in origin of life research

Database name	Website	Content/notes	OoL use cases
KEGG (Kyoto Encyclopedia of Genes and Genomes)	kegg.jp	genomes, genes, pathways, enzymes, biochemical reactions, compounds, and more.	Goldford et al.[183]; Xavier et al.[184]
BRENDA	brenda-enzymes.org	enzymes, reactions, compounds, metadata.	Mulkidjanian et al.[187]; Rivas et al.[188]
NCBI (National Center for Biotechnology Information) Genome	ncbi.nlm.nih.gov/genome	genomes can be filtered for reference high-quality genomes (RefSeq); includes resources, metadata, tools (e.g., BLAST), and is integrated with external databases	Weiss et al.[189]; Xavier et al.[190]
JGI	genome.jgi.doe.gov	genomes	Gagler et al.[191]
RCSB/PDB	rcsb.org	protein structures	Edwards et al.[192]; Longo et al.[193]
AlphaFold	alphafold.ebi.ac.uk	protein structures predicted by artificial intelligence; includes an algorithm for predicting new structures	Alvarez-Carreño et al.[194]
UniProt	uniprot.org	protein sequences, associated functional information, and metadata.	Lake et al.[195]
MetaCyc	metacyc.org	metabolic pathways verified experimentally with metadata; some functions and modules can be accessed free of charge, while others require a paid subscription	Harrison et al.[186]; Sousa et al.[196]
LUCApedia	eebgroups.princeton.edu/lucapedia	organizes information from different databases and publications on possible reactions, cofactors, and proteins present in the Last Universal Common Ancestor (LUCA)	Goldford et al.[183]; Blanco et al.[197]
metaXCMS	xcmsonline.scripps.edu	any HRMS data; used in metabolomics to reconstruct primitive metabolomic pathways and biomolecules	Tautenhahn et al.[198]; Trapp et al.[199]; Sidebottom et al.[200]; Misra et al.[201]

connecting empirical results to the theoretical frameworks that seek to explain them. The methodologies summarized here provide the empirical foundation upon which predictive and computational models—discussed in part 2[7]—can be built to bridge the observation with theory in the quest to understand how life emerges from nonlife.

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

CellPress
OPEN ACCESS

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. Pasteur, L. (1861). Mémoire sur les corpuscules organisés qui existent en suspension dans l'atmosphère. In Examen de la doctrine des générations spontanées (France: Académie des sciences).
2. Darwin, C. (2003). The Origin of Species, 150th Anniversary Edition (Penguin).
3. Peretó, J., Bada, J.L., and Lazcano, A. (2009). Charles Darwin and the origin of life. Orig. Life Evol. Biosph. *39*, 395–406.
4. Aydinoglu, A.U., and Taşkın, Z. (2018). Origins of Life Research: a Bibliometric Approach. Orig. Life Evol. Biosph. *48*, 55–71.
5. 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.
6. Lane, N., and Xavier, J.C. (2024). To unravel the origin of life, treat findings as pieces of a bigger puzzle. Nature *626*, 948–951.
7. Asche, S., Bautista, C., Boulesteix, D., Champagne-Ruel, A., Mathis, C., Markovitch, O., Peng, Z., Adams, A., Dass, A.V., et al. (2026). OoLEN. What it takes to solve the Origin(s) of Life: An integrated review. Part 2–Theoretical methods and emerging trends. Cell Rep. Phys. Sci. *7*, 103211. https://doi.org/10.1016/j.xcrp.2026.103211.
8. Bienz, S., Bigler, L., Fox, T., and Meier, H. (2021). Spectroscopic Methods in Organic Chemistry (Georg Thieme Verlag KG (Stuttgart)). https://doi.org/10.1055/b000000049.
9. Perkins, W.D. (2020). Sample Handling in Infrared Spectroscopy—An Overview. Practical Sampling Techniques for INFRARED ANALYSIS (CRC Press (UK)), pp. 11–53. https://doi.org/10.1201/9781003068044-2.
10. Connor, M.C., and Shultz, G.V. (2021). Problem Solving Using NMR and IR Spectroscopy for Structural Characterization in Organic Chemistry. In Problems and Problem Solving in Chemistry Education (The Royal Society of Chemistry), pp. 166–198. https://doi.org/10.1039/9781839163586-00166.
11. Morzan, U.N., Alonso de Armiño, D.J., Foglia, N.O., Ramírez, F., González Lebrero, M.C., Scherlis, D.A., and Estrin, D.A. (2018). Spectroscopy in Complex Environments from QM-MM Simulations. Chem. Rev. *118*, 4071–4113.

12. (2007). Ultraviolet–visible (UV-Vis) spectroscopy. In Tutorial Chemistry Texts, R.J. Anderson, D.J. Bendell, and P.W. Groundwater, eds. (Royal Society of Chemistry), pp. 7–23.
13. Mayerhöfer, T.G., Pahlow, S., and Popp, J. (2020). The Bouguer-Beer-Lambert Law: Shining Light on the Obscure. ChemPhysChem *21*, 2029–2046.
14. Samir, B., Kalalian, C., Roth, E., Salghi, R., and Chakir, A. (2019). Gas-phase UV absorption spectra and OH-oxidation kinetics of 1H-1,2,3-triazole and pyrazole. RSC Adv. *9*, 27361–27368.
15. Brabender, M., Henriques Pereira, D.P., Mrnjavac, N., Schlikker, M.L., Kimura, Z.-I., Sucharitakul, J., Kleinermanns, K., Tüysüz, H., Buckel, W., Preiner, M., and Martin, W.F. (2024). Ferredoxin reduction by hydrogen with iron functions as an evolutionary precursor of flavin-based electron bifurcation. Proc. Natl. Acad. Sci. USA *121*, e2318969121.
16. Pinna, S., Kunz, C., Halpern, A., Harrison, S.A., Jordan, S.F., Ward, J., Werner, F., and Lane, N. (2022). A prebiotic basis for ATP as the universal energy currency. PLoS Biol. *20*, e3001437.
17. C, M., Frenkel-Pinter, M., Smith, K.H., Rivera-Santana, V.F., Sargon, A.B., Jacobson, K.C., Guzman-Martinez, A., Williams, L.D., Leman, L.J., Liotta, C.L., et al. (2022). Water-Based Dynamic Depsipeptide Chemistry: Building Block Recycling and Oligomer Distribution Control Using Hydration-Dehydration Cycles. JACS Au *2*, 1395–1404.
18. Yu, J., Darú, A., Deng, M., and Blackmond, D.G. (2024). Prebiotic access to enantioenriched amino acids via peptide-mediated transamination reactions. Proc. Natl. Acad. Sci. USA *121*, e2315447121.
19. Haas, M., Lamour, S., and Trapp, O. (2018). Development of an advanced derivatization protocol for the unambiguous identification of monosaccharides in complex mixtures by gas and liquid chromatography. J. Chromatogr. A *1568*, 160–167.
20. Fürst, P., Pollack, L., Graser, T.A., Godel, H., and Stehle, P. (1990). Appraisal of four pre-column derivatization methods for the high-performance liquid chromatographic determination of free amino acids in biological materials. J. Chromatogr. *499*, 557–569.
21. Jordan, S.F., Ioannou, I., Rammu, H., Halpern, A., Bogart, L.K., Ahn, M., Vasiliadou, R., Christodoulou, J., Maréchal, A., and Lane, N. (2021). Spontaneous assembly of redox-active iron-sulfur clusters at low concentrations of cysteine. Nat. Commun. *12*, 5925.
22. Bonfio, C., Valer, L., Scintilla, S., Shah, S., Evans, D.J., Jin, L., Szostak, J.W., Sasselov, D.D., Sutherland, J.D., and Mansy, S.S. (2017). UV-light-driven prebiotic synthesis of iron-sulfur clusters. Nat. Chem. *9*, 1229–1234.
23. Boyd, J.M., Drevland, R.M., Downs, D.M., and Graham, D.E. (2009). Archaeal ApbC/Nbp35 homologs function as iron-sulfur cluster carrier proteins. J. Bacteriol. *191*, 1490–1497.
24. Khaled, M.A., Mullins, D.W., Jr., Swindle, M., and Lacey, J.C., Jr. (1983). Complexes of polyadenylic acid and the methyl esters of amino acids. Orig. Life *13*, 87–96.
25. Thompson, J.M. (2018). Infrared Spectroscopy (Jenny Stanford Publishing).
26. Chukanov, N.V., and Chervonnyi, A.D. (2016). Infrared Spectroscopy of Minerals and Related Compounds (Springer).
27. Perri, E., Tucker, M.E., and Spadafora, A. (2012). Carbonate organo-mineral micro- and ultrastructures in sub-fossil stromatolites: Marion lake, South Australia. Geobiology *10*, 105–117.
28. Odularu, A.T. (2020). Worthwhile Relevance of Infrared Spectroscopy in Characterization of Samples and Concept of Infrared Spectroscopy-Based Synchrotron Radiation. Journal of Spectroscopy *2020*, 1–11. https://doi.org/10.1155/2020/8869713.
29. Fornasier, S., Mottola, S., Keller, H.U., Barucci, M.A., Davidsson, B., Feller, C., Deshapriya, J.D.P., Sierks, H., Barbieri, C., Lamy, P.L., et al. (2016). Rosetta's comet 67P/Churyumov-Gerasimenko sheds its dusty mantle to reveal its icy nature. Science *354*, 1566–1570.
30. Altwegg, K., Balsiger, H., and Fuselier, S.A. (2019). Cometary chemistry and the origin of icy solar system bodies: The view after*Rosetta*. Annu. Rev. Astron. Astrophys. *57*, 113–155.
31. Brown, A.J., Walter, M.R., and Cudahy, T.J. (2005). Hyperspectral imaging spectroscopy of a Mars analogue environment at the North Pole Dome, Pilbara Craton, Western Australia. Aust. J. Earth Sci. *52*, 353–364. https://doi.org/10.1080/08120090500134530.
32. Ghassemizadeh, R., Moore, B., Momose, T., and Walter, M. (2019). Stability and IR Spectroscopy of Zwitterionic Form of β-Alanine in Water Clusters. J. Phys. Chem. B *123*, 4392–4399.
33. Larkin, P. (2017). Infrared and Raman Spectroscopy: Principles and Spectral Interpretation (Elsevier).
34. Carneiro, C.E.A., Ivashita, F.F., de Souza, I.G., de Souza, C.M.D., Paesano, A., da Costa, A.C.S., di Mauro, E., de Santana, H., Zaia, C.T.B.V., and Zaia, D.A.M. (2013). Synthesis of goethite in solutions of artificial seawater and amino acids: a prebiotic chemistry study. Int. J. Astrobiol. *12*, 149–160. https://doi.org/10.1017/s1473550413000013.
35. Baú, J.P.T., Anizelli, P.R., de Santana, H., da Costa, M.F., and Zaia, D.A.M. (2019). An experimental and theoretical vibrational study of the interaction of cytosine and uracil with artificial seawaters: A prebiotic chemistry experiment. Vib. Spectrosc. *101*, 92–99.
36. Carneiro, C.E.A., de Santana, H., Casado, C., Coronas, J., and Zaia, D.A.M. (2011). Adsorption of amino acids (ALA, CYS, HIS, MET) on zeolites: fourier transform infrared and Raman spectroscopy investigations. Astrobiology *11*, 409–418.
37. Miles, A.J., Janes, R.W., and Wallace, B.A. (2021). Tools and methods for circular dichroism spectroscopy of proteins: a tutorial review. Chem. Soc. Rev. *50*, 8400–8413.
38. Ranjbar, B., and Gill, P. (2009). Circular dichroism techniques: biomolecular and nanostructural analyses- a review. Chem. Biol. Drug Des. *74*, 101–120.
39. Meierhenrich, U. (2008). Stereochemistry for the study of the origin of life. Amino Acids and the Asymmetry of Life (Springer Berlin Heidelberg), pp. 17–46.
40. Cassidy, L.M., Burcar, B.T., Stevens, W., Moriarty, E.M., and McGown, L.B. (2014). Guanine-centric self-assembly of nucleotides in water: an important consideration in prebiotic chemistry. Astrobiology *14*, 876–886.
41. Brack, A., and Orgel, L.E. (1975). Beta structures of alternating polypeptides and their possible prebiotic significance. Nature *256*, 383–387.
42. Frenkel-Pinter, M., Sargon, A.B., Glass, J.B., Hud, N.V., and Williams, L.D. (2021). Transition metals enhance prebiotic depsipeptide oligomerization reactions involving histidine. RSC Adv. *11*, 3534–3538.
43. Mariani, A., Bonfio, C., Johnson, C.M., and Sutherland, J.D. (2018). pH-Driven RNA Strand Separation under Prebiotically Plausible Conditions. Biochemistry *57*, 6382–6386.
44. Meinert, C., Garcia, A.D., Topin, J., Jones, N.C., Diekmann, M., Berger, R., Nahon, L., Hoffmann, S.V., and Meierhenrich, U.J. (2022). Amino acid gas phase circular dichroism and implications for the origin of biomolecular asymmetry. Nat. Commun. *13*, 502.
45. Silverstein, R.M., Webster, F.X., Kiemle, D.J., and Bryce, D.L. (2014). Spectrometric Identification of Organic Compounds (John Wiley & Sons).
46. Simmler, C., Napolitano, J.G., McAlpine, J.B., Chen, S.-N., and Pauli, G.F. (2014). Universal quantitative NMR analysis of complex natural samples. Curr. Opin. Biotechnol. *25*, 51–59.
47. Marchione, A.A., and Conklin, B. (2016). Gas phase NMR for the study of chemical reactions: Kinetics and product identification. Gas Phase NMR (The Royal Society of Chemistry), pp. 126–151.
48. Baldus, M. (2022). Biological solid-state NMR: Integrative across different scientific disciplines. J. Struct. Biol. X *6*, 100075.
49. Jeannerat, D., and Furrer, J. (2012). NMR experiments for the analysis of mixtures: beyond 1D 1H spectra. Comb. Chem. High Throughput Screen. *15*, 15–35.

CellPress
OPEN ACCESS

50. Dass, A.V., Wunnava, S., Langlais, J., von der Esch, B., Krusche, M., Ufer, L., Chrisam, N., Dubini, R.C.A., Gartner, F., Angerpointner, S., et al. (2023). RNA oligomerisation without added catalyst from 2′,3′-cyclic nucleotides by drying at air-water interfaces. ChemSystemsChem *5*, e202200026. https://doi.org/10.1002/syst.202200026.

51. Bapat, N.V., Paithankar, H., Chugh, J., and Rajamani, S. (2020). NMR analysis of nucleotide π-stacking in prebiotically relevant crowded environment. Commun. Chem. *3*, 51.

52. Brooks-Bartlett, J.C., and Garman, E.F. (2015). The Nobel science: One hundred years of crystallography. Interdiscip. Sci. Rev. *40*, 244–264.

53. Borrego-Sánchez, A., Gutiérrez-Ariza, C., Sainz-Díaz, C.I., and Cartwright, J.H.E. (2022). The Effect of the Presence of Amino Acids on the Precipitation of Inorganic Chemical-Garden Membranes: Biomineralization at the Origin of Life. Langmuir *38*, 10538–10547.

54. Villafañe-Barajas, S.A., Baú, J.P.T., Colín-García, M., Negrón-Mendoza, A., Heredia-Barbero, A., Pi-Puig, T., and Zaia, D.A.M. (2018). Salinity Effects on the Adsorption of Nucleic Acid Compounds on Na-Montmorillonite: a Prebiotic Chemistry Experiment. Orig. Life Evol. Biosph. *48*, 181–200.

55. Zhang, W., Tam, C.P., Walton, T., Fahrenbach, A.C., Birrane, G., and Szostak, J.W. (2017). Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex. Proc. Natl. Acad. Sci. USA *114*, 7659–7664.

56. Sheng, J., Larsen, A., Heuberger, B.D., Blain, J.C., and Szostak, J.W. (2014). Crystal structure studies of RNA duplexes containing s(2)U:A and s(2)U:U base pairs. J. Am. Chem. Soc. *136*, 13916–13924.

57. Shen, F., Luo, Z., Liu, H., Wang, R., Zhang, S., Gan, J., and Sheng, J. (2017). Structural insights into RNA duplexes with multiple 2´-5´-linkages. Nucleic Acids Res. *45*, 3537–3546.

58. Robards, K., and Ryan, D. (2021). Principles and Practice of Modern Chromatographic Methods (Academic Press).

59. Muchowska, K.B., Varma, S.J., and Moran, J. (2019). Synthesis and breakdown of universal metabolic precursors promoted by iron. Nature *569*, 104–107.

60. 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.

61. Zeger, V.R., Thapa, B., Shamsaei, D., DeLair, J.F., Taylor, T.L., and Anderson, J.L. (2025). Ionic Liquids in Analytical Chemistry: Fundamentals, Technological Advances, and Future Outlook. Anal. Chem. *97*, 4793–4818. https://doi.org/10.1021/acs.analchem.5c00264.

62. Colón-Santos, S., Cooper, G.J.T., and Cronin, L. (2019). Taming the Combinatorial Explosion of the Formose Reaction via Recursion within Mineral Environments. ChemSystemsChem *1*, e1900014. https://doi.org/10.1002/syst.201900014.

63. 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.

64. Liu, K., Blokhuis, A., van Ewijk, C., Kiani, A., Wu, J., Roos, W.H., and Otto, S. (2024). Light-driven eco-evolutionary dynamics in a synthetic replicator system. Nat. Chem. *16*, 79–88.

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

66. Kondrat, R.W., McClusky, G.A., and Cooks, R.G. (1978). Multiple reaction monitoring in mass spectrometry/mass spectrometry for direct analysis of complex mixtures. Anal. Chem. *50*, 2017–2021.

67. Amer, B., Deshpande, R.R., and Bird, S.S. (2023). Simultaneous Quantitation and Discovery (SQUAD) Analysis: Combining the Best of Targeted and Untargeted Mass Spectrometry-Based Metabolomics. Metabolites *13*, 648. https://doi.org/10.3390/metabo13050648.

68. 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.

69. de Hoffmann, E., and Stroobant, V. (2013). Mass Spectrometry: Principles and Applications (John Wiley & Sons).

70. Burcar, B.T., Barge, L.M., Trail, D., Watson, E.B., Russell, M.J., and McGown, L.B. (2015). RNA Oligomerization in Laboratory Analogues of Alkaline Hydrothermal Vent Systems. Astrobiology *15*, 509–522.

71. Šponer, J.E., Šponer, J., Kovařík, A., Šedo, O., Zdráhal, Z., Costanzo, G., and Di Mauro, E. (2021). Questions and Answers Related to the Prebiotic Production of Oligonucleotide Sequences from 3',5' Cyclic Nucleotide Precursors. Life *11*. https://doi.org/10.3390/life11080800.

72. Bose, T., Fridkin, G., Davidovich, C., Krupkin, M., Dinger, N., Falkovich, A.H., Peleg, Y., Agmon, I., Bashan, A., and Yonath, A. (2022). Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds. Nucleic Acids Res. *50*, 1815–1828.

73. Steele, A., Toporski, J.K.W., Avci, R., Guidry, S., and McKay, D.S. (2001). Time of flight secondary ion mass spectrometry (ToFSIMS) of a number of hopanoids. Org. Geochem. *32*, 905–911.

74. Wilm, M. (2011). Principles of electrospray ionization. Mol. Cell. Proteomics *10*, M111.009407.

75. Yang, S., Mu, L., Feng, R., and Kong, X. (2019). Selection of Internal Standards for Quantitative Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis Based on Correlation Coefficients. ACS Omega *4*, 8249–8254.

76. Calderón-Celis, F., Encinar, J.R., and Sanz-Medel, A. (2018). Standardization approaches in absolute quantitative proteomics with mass spectrometry. Mass Spectrom. Rev. *37*, 715–737.

77. Cohen, Z.R., Kessenich, B.L., Hazra, A., Nguyen, J., Johnson, R.S., MacCoss, M.J., Lalic, G., Black, R.A., and Keller, S.L. (2022). Prebiotic Membranes and Micelles Do Not Inhibit Peptide Formation During Dehydration. Chembiochem *23*, e202100614.

78. Buch, A., Glavin, D.P., Sternberg, R., Szopa, C., Rodier, C., Navarro-González, R., Raulin, F., Cabane, M., and Mahaffy, P. (2006). A new extraction technique for in situ analyses of amino and carboxylic acids on Mars by gas chromatography mass spectrometry. Planet. Space Sci. *54*, 1592–1599.

79. 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.

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

81. Premstaller, A., and Oefner, P.J. (2003). Denaturing high-performance liquid chromatography. Methods Mol. Biol. *212*, 15–35.

82. Bedoin, L., Alves, S., and Lambert, J.-F. (2020). Origins of life and molecular information: Selectivity in mineral surface-induced prebiotic amino acid polymerization. ACS Earth Space Chem. *4*, 1802–1812.

83. 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.

84. 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.

85. Forsythe, J.G., Yu, S.-S., Mamajanov, I., Grover, M.A., Krishnamurthy, R., Fernández, F.M., and Hud, N.V. (2015). Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. Angew. Chem. Int. Ed. Engl. *54*, 9871–9875.

86. Muccio, Z., and Jackson, G.P. (2009). Isotope Ratio Mass Spectrometry. Analyst *134*, 213–222.

87. Westall, F., Brack, A., Fairén, A.G., and Schulte, M.D. (2023). Setting the geological scene for the origin of life and continuing open questions about its emergence. Front. Astron. Space Sci. *9*, 1095701.

88. Schmitt-Kopplin, P., Gabelica, Z., Gougeon, R.D., Fekete, A., Kanawati, B., Harir, M., Gebefuegi, I., Eckel, G., and Hertkorn, N. (2010). High molecular diversity of extraterrestrial organic matter in Murchison meteorite

revealed 40 years after its fall. Proc. Natl. Acad. Sci. USA *107*, 2763–2768.

89. Meermann, B., and Nischwitz, V. (2018). ICP-MS for the analysis at the nanoscale – a tutorial review. J. Anal. At. Spectrom. *33*, 1432–1468. https://doi.org/10.1039/c8ja00037a.

90. Jenner, G.A., Longerich, H.P., Jackson, S.E., and Fryer, B.J. (1990). ICP-MS — A powerful tool for high-precision trace-element analysis in Earth sciences: Evidence from analysis of selected U.S.G.S. reference samples. Chem. Geol. *83*, 133–148.

91. Shinotsuka, K., Hidaka, H., and Ebihara, M. (1995). Detailed abundances of rare earth elements, thorium and uranium in chondritic meteorites: An ICP-MS study. Meteoritics *30*, 694–699.

92. Jia, T.Z., Chandru, K., Hongo, Y., Afrin, R., Usui, T., Myojo, K., and Cleaves, H.J., 2nd. (2019). Membraneless polyester microdroplets as primordial compartments at the origins of life. Proc. Natl. Acad. Sci. USA *116*, 15830–15835.

93. Fares, H.M., Marras, A.E., Ting, J.M., Tirrell, M.V., and Keating, C.D. (2020). Impact of wet-dry cycling on the phase behavior and compartmentalization properties of complex coacervates. Nat. Commun. *11*, 5423.

94. Pappas, C.G., Shafi, R., Sasselli, I.R., Siccardi, H., Wang, T., Narang, V., Abzalimov, R., Wijerathne, N., and Ulijn, R.V. (2016). Dynamic peptide libraries for the discovery of supramolecular nanomaterials. Nat. Nanotechnol. *11*, 960–967.

95. Yu, K., Hidaka, K., Sugiyama, H., Endo, M., Matsumura, S., and Ikawa, Y. (2022). A Hexameric Ribozyme Nanostructure Formed by Double-Decker Assembly of a Pair of Triangular Ribozyme Trimers. Chembiochem *23*, e202100573.

96. Russell, M.J., and Ponce, A. (2020). Six "Must-Have" Minerals for Life's Emergence: Olivine, Pyrrhotite, Bridgmanite, Serpentine, Fougerite and Mackinawite. Life *10*, 291. https://doi.org/10.3390/life10110291.

97. Pilorget, C., Okada, T., Hamm, V., Brunetto, R., Yada, T., Loizeau, D., Riu, L., Usui, T., Moussi-Soffys, A., Hatakeda, K., et al. (2021). First compositional analysis of Ryugu samples by the MicrOmega hyperspectral microscope. Nat. Astron. *6*, 221–225.

98. Thorn, K. (2016). A quick guide to light microscopy in cell biology. Mol. Biol. Cell *27*, 219–222.

99. Barer, R. (1947). Some Applications of Phase-Contrast Microscopy. J. Cell Sci. *88*, 491–500.

100. Heydarian, H., Yazdanfar, P., Zarif, A., and Rashidian, B. (2020). Near Field Differential Interference Contrast Microscopy. Sci. Rep. *10*, 9644.

101. Oldenbourg, R. (1996). A new view on polarization microscopy. Nature *381*, 811–812.

102. Kindt, J.T., Szostak, J.W., and Wang, A. (2020). Bulk Self-Assembly of Giant, Unilamellar Vesicles. ACS Nano *14*, 14627–14634.

103. Jia, T.Z., and Fraccia, T.P. (2020). Liquid Crystal Peptide/DNA Coacervates in the Context of Prebiotic Molecular Evolution. Crystals *10*, 964.

104. Combs, C.A. (2010). Fluorescence microscopy: a concise guide to current imaging methods. Curr Protoc Neurosci *Chapter 2*, Unit2.1.

105. (2021). Microscopy techniques for protocell characterization. Polym. Test. *93*, 106935.

106. 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.

107. Trinks, H., Schröder, W., and Biebricher, C.K. (2005). Ice and the origin of life. Orig. Life Evol. Biosph. *35*, 429–445.

108. Wünsch, U.J., Acar, E., Koch, B.P., Murphy, K.R., Schmitt-Kopplin, P., and Stedmon, C.A. (2018). The Molecular Fingerprint of Fluorescent Natural Organic Matter Offers Insight into Biogeochemical Sources and Diagenetic State. Anal. Chem. *90*, 14188–14197.

109. Robbins, M.A. (2013). The Collector's Book of Fluorescent Minerals (Springer Science & Business Media).

110. Jing, S., Liu, Q., Jin, Y., and Li, B. (2021). Dimeric G-Quadruplex: An Effective Nucleic Acid Scaffold for Lighting Up Thioflavin T. Anal. Chem. *93*, 1333–1341.

111. Kolbeck, P.J., Vanderlinden, W., Gemmecker, G., Gebhardt, C., Lehmann, M., Lak, A., Nicolaus, T., Cordes, T., and Lipfert, J. (2021). Molecular structure, DNA binding mode, photophysical properties and recommendations for use of SYBR Gold. Nucleic Acids Res. *49*, 5143–5158.

112. Bowler, F.R., Chan, C.K.W., Duffy, C.D., Gerland, B., Islam, S., Powner, M.W., Sutherland, J.D., and Xu, J. (2013). Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. Nat. Chem. *5*, 383–389.

113. Snapp, E. (2005). Design and use of fluorescent fusion proteins in cell biology. Curr. Protoc. Cell Biol. *Chapter 21*, 21.4.1–21.4.13.

114. Lopez, A., and Fiore, M. (2019). Investigating Prebiotic Protocells for A Comprehensive Understanding of the Origins of Life: A Prebiotic Systems Chemistry Perspective. Life *9*, 49. https://doi.org/10.3390/life9020049.

115. Yan, F., Fan, K., Bai, Z., Zhang, R., Zu, F., Xu, J., and Li, X. (2017). Fluorescein applications as fluorescent probes for the detection of analytes. Trends Analyt Chem *97*, 15–35.

116. Nahalka, J., and Hrabarova, E. (2021). Prebiotic Peptides Based on the Glycocodon Theory Analyzed with FRET. Life *11*, 380. https://doi.org/10.3390/life11050380.

117. Chen, C., Li, P., Luo, W., Nakamura, Y., Dimo, V.S., Kanekura, K., and Hayamizu, Y. (2021). Diffusion of LLPS Droplets Consisting of Poly(PR) Dipeptide Repeats and RNA on Chemically Modified Glass Surface. Langmuir *37*, 5635–5641.

118. Chen, C., Jia, H., Nakamura, Y., Kanekura, K., and Hayamizu, Y. (2022). Effect of Multivalency on Phase-Separated Droplets Consisting of Poly(PR) Dipeptide Repeats and RNA at the Solid/Liquid Interface. ACS Omega *7*, 19280–19287.

119. Russell, M.J., Hall, A.J., and Martin, W. (2010). Serpentinization as a source of energy at the origin of life. Geobiology *8*, 355–371.

120. Hassenkam, T., Andersson, M.P., Dalby, K.N., Mackenzie, D.M.A., and Rosing, M.T. (2017). Elements of Eoarchean life trapped in mineral inclusions. Nature *548*, 78–81.

121. Kempe, A., Schopf, J.W., Altermann, W., Kudryavtsev, A.B., and Heckl, W.M. (2002). Atomic force microscopy of Precambrian microscopic fossils. Proc. Natl. Acad. Sci. USA *99*, 9117–9120.

122. Cafferty, B.J., Fialho, D.M., Khanam, J., Krishnamurthy, R., and Hud, N.V. (2016). Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. Nat. Commun. *7*, 11328.

123. Chakrova, N., Canton, A.S., Danelon, C., Stallinga, S., and Rieger, B. (2016). Adaptive illumination reduces photobleaching in structured illumination microscopy. Biomed. Opt. Express *7*, 4263–4274.

124. Boudreau, C., Wee, T.-L.E., Duh, Y.-R.S., Couto, M.P., Ardakani, K.H., and Brown, C.M. (2016). Excitation Light Dose Engineering to Reduce Photo-bleaching and Photo-toxicity. Sci. Rep. *6*, 30892.

125. Jia, T.Z., Nishikawa, S., and Fujishima, K. (2022). Sequencing the origins of life. BBA Adv. *2*, 100049.

126. Pressman, A., Moretti, J.E., Campbell, G.W., Müller, U.F., and Chen, I.A. (2017). Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. Nucleic Acids Res. *45*, 10922.

127. Seelig, B., and Szostak, J.W. (2007). Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature *448*, 828–831.

128. Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R.P., Banday, S., Mishra, A.K., Das, G., and Malonia, S.K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. Biology *12*, 997. https://doi.org/10.3390/biology12070997.

129. Ul-Hamid, A. (2018). A Beginners' Guide to Scanning Electron Microscopy (Springer).

130. Franken, L.E., Grünewald, K., Boekema, E.J., and Stuart, M.C.A. (2020). A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. Small *16*, 1906198.

131. De Gregorio, B.T., Stroud, R.M., Nittler, L.R., Alexander, C.M., Bassim, N.D., Cody, G.D., Kilcoyne, A.D., Sandford, S.A., Milam, S.N., and Nuevo, M. (2009). Isotopic and chemical variation of organic nanoglobules in primitive meteorites. Meteorit. Planet. Sci. *48*, 904–928.

132. Phoenix, V.R., Konhauser, K.O., School of Earth Sciences; Adams, D.G., and Bottrell, S.H. (2001). Role of Biomineralization as an Ultraviolet Shield: Implications for Archean Life, *29* (Geological Society of America), pp. 823–826.

133. Rasmussen, B., Muhling, J.R., Suvorova, A., and Fischer, W.W. (2021). Apatite nanoparticles in 3.46–2.46 Ga iron formations: Evidence for phosphorus-rich hydrothermal plumes on early Earth. Geology *49*, 647–651.

134. Sforna, M.C., Brunelli, D., Pisapia, C., Pasini, V., Malferrari, D., and Ménez, B. (2018). Abiotic formation of condensed carbonaceous matter in the hydrating oceanic crust. Nat. Commun. *9*, 5049.

135. Ménez, B., Pisapia, C., Andreani, M., Jamme, F., Vanbellingen, Q.P., Brunelle, A., Richard, L., Dumas, P., and Réfrégiers, M. (2018). Abiotic synthesis of amino acids in the recesses of the oceanic lithosphere. Nature *564*, 59–63.

136. Higgs, P.G., and Lehman, N. (2015). The RNA World: molecular cooperation at the origins of life. Nat. Rev. Genet. *16*, 7–17.

137. Hud, N.V. (2018). Searching for lost nucleotides of the pre-RNA World with a self-refining model of early Earth. Nat. Commun. *9*, 5171.

138. Brinkerhoff, H., Kang, A.S.W., Liu, J., Aksimentiev, A., and Dekker, C. (2021). Multiple rereads of single proteins at single-amino acid resolution using nanopores. Science *374*, 1509–1513.

139. Ohshiro, T., Matsubara, K., Tsutsui, M., Furuhashi, M., Taniguchi, M., and Kawai, T. (2012). Single-molecule electrical random resequencing of DNA and RNA. Sci. Rep. *2*, 501.

140. Ohan, N.W., and Heikkila, J.J. (1993). Reverse transcription-polymerase chain reaction: An overview of the technique and its applications. Biotechnol. Adv. *11*, 13–29.

141. Jackson, T.J., Spriggs, R.V., Burgoyne, N.J., Jones, C., and Willis, A.E. (2014). Evaluating bias-reducing protocols for RNA sequencing library preparation. BMC Genom. *15*, 569.

142. Shiroguchi, K., Jia, T.Z., Sims, P.A., and Xie, X.S. (2012). Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. Proc. Natl. Acad. Sci. USA *109*, 1347–1352.

143. Gold, L., Janjic, N., Jarvis, T., Schneider, D., Walker, J.J., Wilcox, S.K., and Zichi, D. (2012). Aptamers and the RNA world, past and present. Cold Spring Harb. Perspect. Biol. *4*, a003582. https://doi.org/10.1101/cshperspect.a003582.

144. 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.

145. Furubayashi, T., Ueda, K., Bansho, Y., Motooka, D., Nakamura, S., Mizuuchi, R., and Ichihashi, N. (2020). Emergence and diversification of a host-parasite RNA ecosystem through Darwinian evolution. eLife *9*, e56038. https://doi.org/10.7554/eLife.56038.

146. Blanco, C., Verbanic, S., Seelig, B., and Chen, I.A. (2020). EasyDIVER: A Pipeline for Assembling and Counting High-Throughput Sequencing Data from In Vitro Evolution of Nucleic Acids or Peptides. J. Mol. Evol. *88*, 477–481.

147. Hammerling, M.J., Fritz, B.R., Yoesep, D.J., Kim, D.S., Carlson, E.D., and Jewett, M.C. (2020). In vitro ribosome synthesis and evolution through ribosome display. Nat. Commun. *11*, 1108.

148. Fehlmann, T., Reinheimer, S., Geng, C., Su, X., Drmanac, S., Alexeev, A., Zhang, C., Backes, C., Ludwig, N., Hart, M., et al. (2016). cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs. Clin. Epigenetics *8*, 123.

149. Jeon, S.A., Park, J.L., Kim, J.-H., Kim, J.H., Kim, Y.S., Kim, J.C., and Kim, S.Y. (2019). Comparison of the MGISEQ-2000 and Illumina HiSeq 4000 sequencing platforms for RNA sequencing. Genomics Inform. *17*, e32.

150. Mizuuchi, R., and Ichihashi, N. (2023). Minimal RNA self-reproduction discovered from a random pool of oligomers. Chem. Sci. *14*, 7656–7664.

151. Wolk, S.K., Mayfield, W.S., Gelinas, A.D., Astling, D., Guillot, J., Brody, E.N., Janjic, N., and Gold, L. (2020). Modified nucleotides may have enhanced early RNA catalysis. Proc. Natl. Acad. Sci. USA *117*, 8236–8242.

152. Zhang, N., Shi, S., Jia, T.Z., Ziegler, A., Yoo, B., Yuan, X., Li, W., and Zhang, S. (2019). A general LC-MS-based RNA sequencing method for direct analysis of multiple-base modifications in RNA mixtures. Nucleic Acids Res. *47*, e125.

153. Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell *149*, 1635–1646.

154. Alfonzo, J.D., Brown, J.A., Byers, P.H., Cheung, V.G., Maraia, R.J., and Ross, R.L. (2021). A call for direct sequencing of full-length RNAs to identify all modifications. Nat. Genet. *53*, 1113–1116.

155. Xu, L., and Seki, M. (2020). Recent advances in the detection of base modifications using the Nanopore sequencer. J. Hum. Genet. *65*, 25–33.

156. Stephenson, W., Razaghi, R., Busan, S., Weeks, K.M., Timp, W., and Smibert, P. (2022). Direct detection of RNA modifications and structure using single-molecule nanopore sequencing. Cell Genom. *2*, 100097.

157. Furlan, M., Delgado-Tejedor, A., Mulroney, L., Pelizzola, M., Novoa, E.M., and Leonardi, T. (2021). Computational methods for RNA modification detection from nanopore direct RNA sequencing data. RNA Biol. *18*, 31–40.

158. Deamer, D., Akeson, M., and Branton, D. (2016). Three decades of nanopore sequencing. Nat. Biotechnol. *34*, 518–524.

159. Da Silva, L., Maurel, M.-C., and Deamer, D. (2015). Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. J. Mol. Evol. *80*, 86–97.

160. DeGuzman, V., Vercoutere, W., Shenasa, H., and Deamer, D. (2014). Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization. J. Mol. Evol. *78*, 251–262.

161. Delahaye, C., and Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. PLoS One *16*, e0257521.

162. Ledbetter, M.P., Craig, J.M., Karadeema, R.J., Noakes, M.T., Kim, H.C., Abell, S.J., Huang, J.R., Anderson, B.A., Krishnamurthy, R., Gundlach, J.H., and Romesberg, F.E. (2020). Nanopore Sequencing of an Expanded Genetic Alphabet Reveals High-Fidelity Replication of a Predominantly Hydrophobic Unnatural Base Pair. J. Am. Chem. Soc. *142*, 2110–2114.

163. Hoshika, S., Leal, N.A., Kim, M.-J., Kim, M.-S., Karalkar, N.B., Kim, H.-J., Bates, A.M., Watkins, N.E., Jr., SantaLucia, H.A., Meyer, A.J., et al. (2019). Hachimoji DNA and RNA: A genetic system with eight building blocks. Science *363*, 884–887.

164. Cleaves, H.J., 2nd, Butch, C., Burger, P.B., Goodwin, J., and Meringer, M. (2019). One Among Millions: The Chemical Space of Nucleic Acid-Like Molecules. J. Chem. Inf. Model. *59*, 4266–4277.

165. Mizuuchi, R., Furubayashi, T., and Ichihashi, N. (2022). Evolutionary transition from a single RNA replicator to a multiple replicator network. Nat. Commun. *13*, 1460.

166. Nanda, J., Rubinov, B., Ivnitski, D., Mukherjee, R., Shtelman, E., Motro, Y., Miller, Y., Wagner, N., Cohen-Luria, R., and Ashkenasy, G. (2017). Emergence of native peptide sequences in prebiotic replication networks. Nat. Commun. *8*, 434.

167. Bishop, G.R., and Chaires, J.B. (2003). Characterization of DNA structures by circular dichroism. Curr. Protoc. Nucleic Acid Chem. *Chapter 7*, 7.11.1–7.11.8.

168. Šponer, J., Banáš, P., Jurečka, P., Zgarbová, M., Kührová, P., Havrila, M., Krepl, M., Stadlbauer, P., and Otyepka, M. (2014). Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. J. Phys. Chem. Lett. *5*, 1771–1782.

169. Semerjian, H.G., and Burgess, D.R., Jr. (2022). Data programs at NBS/NIST: 1901–2021. J. Phys. Chem. Ref. Data *51*, 011501.

170. Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. J. Mass Spectrom. *45*, 703–714.

171. AHED Home. [cited 21 Jan 2025]. Available: https://ahed.nasa.gov/.

172. 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.

173. Meringer, M., and Schymanski, E.L. (2013). Small Molecule Identification with MOLGEN and Mass Spectrometry. Metabolites *3*, 440–462.

174. Peironcely, J.E., Rojas-Chertó, M., Fichera, D., Reijmers, T., Coulier, L., Faulon, J.-L., and Hankemeier, T. (2012). OMG: Open Molecule Generator. J. Cheminform. *4*, 21.

175. McKay, B.D., Yirik, M.A., and Steinbeck, C. (2022). Surge: a fast open-source chemical graph generator. J. Cheminform. *14*, 24.

176. 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*. https://doi.org/10.1126/science.aaw1955.

177. Andersen, J.L., Flamm, C., Merkle, D., and Stadler, P.F. (2016). A software package for chemically inspired graph transformation. Graph Transformation (Springer International Publishing), pp. 73–88.

178. UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. *49*, D480–D489.

179. Wensien, M., von Pappenheim, F.R., Funk, L.-M., Kloskowski, P., Curth, U., Diederichsen, U., Uranga, J., Ye, J., Fang, P., Pan, K.T., et al. (2021). A lysine-cysteine redox switch with an NOS bridge regulates enzyme function. Nature *593*, 460–464.

180. Tu, C., Li, J., Young, R., Page, B.J., Engler, F., Halfon, M.S., Canty, J.M., Jr., and Qu, J. (2011). Combinatorial peptide ligand library treatment followed by a dual-enzyme, dual-activation approach on a nanoflow liquid chromatography/orbitrap/electron transfer dissociation system for comprehensive analysis of swine plasma proteome. Anal. Chem. *83*, 4802–4813.

181. Butler, C.A., Veith, P.D., Nieto, M.F., Dashper, S.G., and Reynolds, E.C. (2015). Lysine acetylation is a common post-translational modification of key metabolic pathway enzymes of the anaerobe Porphyromonas gingivalis. J. Proteomics *128*, 352–364.

182. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

183. Goldford, J.E., Hartman, H., Smith, T.F., and Segrè, D. (2017). Remnants of an Ancient Metabolism without Phosphate. Cell *168*, 1126–1134.e9.

184. 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.

185. 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.

186. Harrison, S.A., Palmeira, R.N., Halpern, A., and Lane, N. (2022). A biophysical basis for the emergence of the genetic code in protocells. Biochim. Biophys. Acta. Bioenerg. *1863*, 148597.

187. Mulkidjanian, A.Y., and Galperin, M.Y. (2009). On the origin of life in the zinc world. 2. Validation of the hypothesis on the photosynthesizing zinc sulfide edifices as cradles of life on Earth. Biol. Direct *4*, 27.

188. Rivas, M., Becerra, A., Peretó, J., Bada, J.L., and Lazcano, A. (2011). Metalloproteins and the pyrite-based origin of life: a critical assessment. Orig. Life Evol. Biosph. *41*, 347–356.

189. Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., and Martin, W.F. (2016). The physiology and habitat of the last universal common ancestor. Nat. Microbiol. *1*, 16116.

190. 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.

191. Gagler, D.C., Karas, B., Kempes, C.P., Malloy, J., Mierzejewski, V., Goldman, A.D., Kim, H., and Walker, S.I. (2022). Scaling laws in enzyme function reveal a new kind of biochemical universality. Proc. Natl. Acad. Sci. USA *119*, e2106655119. https://doi.org/10.1073/pnas.2106655119.

192. Edwards, H., Abeln, S., and Deane, C.M. (2013). Exploring fold space preferences of new-born and ancient protein superfamilies. PLoS Comput. Biol. *9*, e1003325.

193. Longo, L.M., Jabłońska, J., Vyas, P., Kanade, M., Kolodny, R., Ben-Tal, N., and Tawfik, D.S. (2020). On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment. eLife *9*. https://doi.org/10.7554/eLife.64415.

194. 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.

195. Lake, J.A., Larsen, J., Tran, D.T., and Sinsheimer, J.S. (2018). Uncovering the Genomic Origins of Life. Genome Biol. Evol. *10*, 1705–1714.

196. Sousa, F.L., Nelson-Sathi, S., and Martin, W.F. (2016). One step beyond a ribosome: The ancient anaerobic core. Biochim. Biophys. Acta *1857*, 1027–1038.

197. Blanco, C., Bayas, M., Yan, F., and Chen, I.A. (2018). Analysis of evolutionarily independent protein-RNA complexes yields a criterion to evaluate the relevance of prebiotic scenarios. Curr. Biol. *28*, 526–537.e5.

198. Tautenhahn, R., Patti, G.J., Kalisiak, E., Miyamoto, T., Schmidt, M., Lo, F.Y., McBee, J., Baliga, N.S., and Siuzdak, G. (2011). metaXCMS: second-order analysis of untargeted metabolomics data. Anal. Chem. *83*, 696–700.

199. Teichert, J.S., Kruse, F.M., and Trapp, O. (2019). Direct prebiotic pathway to DNA nucleosides. Angew. Chem. Int. Ed. Engl. *131*, 10049–10052. https://doi.org/10.1002/ange.201903400.

200. Sidebottom, A.M., and Carlson, E.E. (2015). A reinvigorated era of bacterial secondary metabolite discovery. Curr. Opin. Chem. Biol. *24*, 104–111.

201. Misra, B.B., Fahrmann, J.F., and Grapov, D. (2017). Review of emerging metabolomic tools and resources: 2015-2016. Electrophoresis *38*, 2257–2274.