

Origins of cells

International Baccalaureate · IB Diploma · Biology · HL

Scope and prerequisites

A membrane bubble is not automatically alive, and making an amino acid is not the same as making a cell. The origin-of-cells problem asks how chemistry could acquire compartments, catalysis and inherited information that could evolve together.

This is the HL-only A2.1 unit; there is no SL counterpart. It teaches early-Earth models, requirements for cellular life, prebiotic chemistry, vesicles, the RNA-world hypothesis, evidence for common ancestry, dating methods and hydrothermal-vent hypotheses. Prerequisites: water, membranes, nucleic acids and enzymes. All numerical datasets and timelines used for calculations below are original, invented teaching examples.

AHL: Early Earth and a testable origin problem

Early Earth's atmosphere contained very little free molecular oxygen compared with today's atmosphere. Consequently, there was little protective ozone, allowing more ultraviolet radiation to reach the surface. Models include higher carbon dioxide and methane concentrations than today's atmosphere, contributing to greenhouse warming. The precise composition varied across environments and time and is reconstructed from evidence; it should not be treated as one directly measured global gas recipe.

Energy from ultraviolet light, electrical discharges, heat or chemical gradients could drive reactions among simpler substances. Some resulting carbon compounds could supply ingredients for later biological chemistry. "Organic" describes a category of carbon chemistry; it does not mean that a compound must have been made by an organism. Conversely, showing that a reaction is chemically possible under one set of conditions does not show that those conditions existed everywhere or produced the first cells.

An origin hypothesis should generate predictions that evidence can test. For example, a proposed gas mixture and energy source might predict the production of particular compounds. Researchers can vary the mixture, include controls and analyse products. Reproducing every detail of the historical setting is impossible, and the earliest small compartments may have left no recognizable fossil. Those limitations make the problem difficult; they do not make every explanation equally supported.

AHL: What must a cellular system do?

A cell is the smallest structural unit capable of sustaining the integrated processes of life under suitable conditions. A cell exchanges matter and energy with its surroundings, maintains an internal organization, carries information and participates in

reproduction through its lineage. "Self-sustaining" does not mean isolated from an environment or able to make every required molecule from nothing.

A virus can carry genetic information and evolve, but it relies on a host cell's machinery and energy supply for reproduction and other functions. It has no autonomous cellular metabolism. These dependencies help explain why viruses are not classified as living cells in this course. A single feature such as movement is not a reliable living/non-living test: fire spreads and crystals grow without being cells.

Present-day cells arise by division of pre-existing cells. This observation describes current cell production; it does not by itself explain how the earliest cellular systems arose before a lineage of cells existed. **Abiogenesis** concerns that historical transition from non-living chemistry. It is different from claiming that modern, fully formed organisms routinely appear in sterile food or mud.

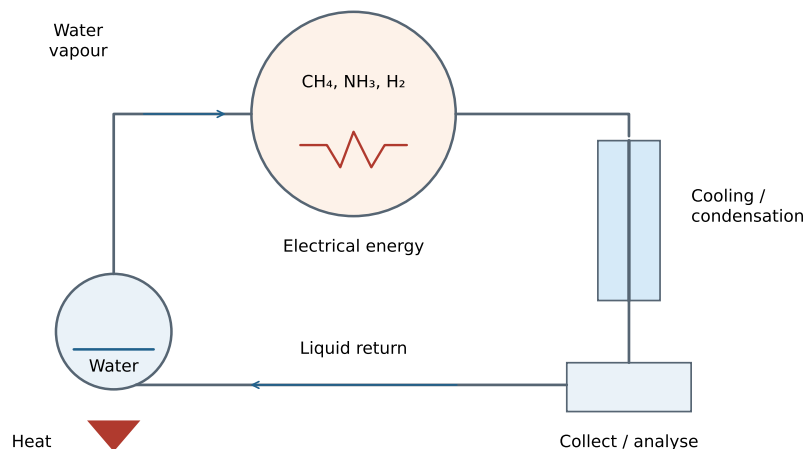
A plausible route needs interacting capacities:

- **Catalysis:** reactions must proceed sufficiently fast under the available conditions; a catalyst speeds a route without being used up overall.
- **Replication and heredity:** information-bearing molecules must be copied with enough fidelity to preserve useful features, while variation remains possible.
- **Self-assembly:** suitable components can organize into larger structures through their interactions, without a pre-existing cell individually placing every component.
- **Compartmentalization:** a boundary can keep interacting components together and make internal chemistry differ from the surroundings.

A compartment that grows but cannot transmit an inherited functional difference is not yet a complete model of Darwinian cellular evolution. Conversely, a replicating molecule in solution does not automatically explain how a stable, regulated cell forms. Investigations can solve parts of the problem without having solved the whole transition.

AHL: Evaluate the Miller–Urey experiment

In the classic spark-discharge investigation, heated water supplied vapour to a circulating system containing methane, ammonia and hydrogen. Electrical sparks supplied energy. Cooling condensed vapour, and collected liquid could be analysed. The apparatus modelled selected possible early conditions rather than reproducing the entire early planet.



The diagram shows a laboratory model and its material flow, not a photograph of an early-Earth environment.

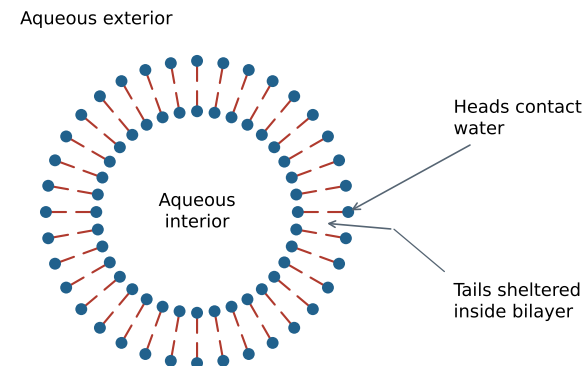
Amino acids and other carbon compounds were detected among the products. This supports the possibility of abiotic synthesis of some building blocks under the tested conditions. It did not produce living cells, prove a complete origin pathway, or show that finished proteins and genomes arose spontaneously in that apparatus.

Evaluate the method through a claim–evidence–limit chain. Supplying a defined gas mixture and an energy source makes the conditions testable. Collection and chemical analysis identify products. Controls without the energy input, checks on starting materials and procedures to exclude biological contamination help distinguish synthesis from contamination. Repeatability strengthens a chemical claim.

The historical gas mixture is a limitation when applying the result to all early environments: the true composition and suitable energy inputs remain questions. A successful model reaction demonstrates feasibility within its conditions. A failed trial with a different mixture would limit that trial's pathway; it would not alone disprove every abiogenesis hypothesis. Do not confuse scientific evaluation with saying an experiment was either “perfect proof” or “worthless”.

AHL: Boundaries can form by self-assembly

Suitable amphipathic molecules have a water-interacting region and a water-avoiding region. In appropriate aqueous conditions, fatty acids can aggregate into bilayers and closed vesicles. Water-avoiding tails are sheltered within the membrane while water-interacting heads contact the aqueous regions on either side. The resulting compartment can enclose a water-based interior.



A boundary can self-assemble.
Heredity and metabolism need further explanation.

Self-assembly can create a boundary; it does not by itself create heredity or a working metabolism.

A boundary can keep catalysts and reactants close, alter concentrations and permit a different internal chemical environment. It must also allow adequate exchange of matter and energy. A perfectly impermeable container would prevent nutrient entry and waste removal; a boundary that leaks everything would lose useful components.

Assembly depends on molecular structure and conditions such as pH, concentration and dissolved ions. An observed vesicle is therefore evidence that a certain kind of compartment can form under particular conditions. It is not proof that every early membrane had the composition of a modern phospholipid membrane, or that the vesicle is already alive. Growth, division and the inheritance of contents must be connected to the chemistry inside.

AHL: RNA as an early information-and-catalysis system

Modern cells divide tasks among DNA, RNA and proteins. That division creates a historical question: how could hereditary information be used before the complete modern translation system existed? The **RNA-world hypothesis** proposes an earlier stage in which RNA had central roles in both information storage and catalysis.

RNA can carry information in its sequence and serve as a template for complementary copying. Some RNA molecules, called **ribozymes**, can catalyse reactions. The ribosome's peptide-bond-forming centre is based on ribosomal RNA, showing that catalytic RNA is relevant to modern biology as well as a laboratory possibility.

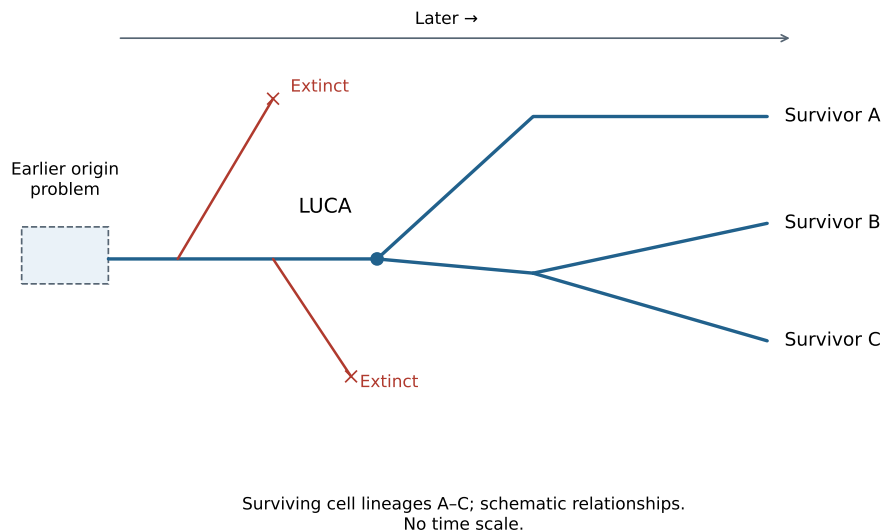
These properties make a dual-role system plausible. They do not establish that an isolated RNA molecule can already carry out every step of self-replication, metabolism and membrane maintenance without assistance. Researchers still investigate nu-

cleotide supply, polymer formation, copying accuracy, stability and the connection to compartments. “RNA can act as a template” is a more precise claim than “every RNA molecule copies itself”.

A useful model links copying with variation and selection. If different sequences produce different catalytic effects, and compartments retain those differences through reproduction, some lineages could leave more descendants under particular conditions. This describes how selection could improve a heritable system; it does not mean the molecules deliberately seek progress or that evolution has a predetermined goal.

AHL: LUCA is not the first cell

The last universal common ancestor, or **LUCA**, is the shared ancestral lineage from which all present-day cellular life descends. It need not have been the first living system, the only organism alive at its time, or one uniquely identifiable fossil cell. Earlier lineages could have existed and left no surviving descendants. Other forms could also have been outcompeted or lost; common ancestry of survivors does not prove that only one origin event was chemically possible.



The schematic shows ancestry, not a dated family tree; LUCA is the common ancestor of surviving lineages.

Evidence includes the broadly conserved genetic code and shared genes involved in fundamental information-processing functions. Compare sequences from many organisms and several independent genes, with methods that account for change. Similarities that form a consistent, branching pattern support common ancestry more strongly than one superficial resemblance.

An **inference** is an explanation drawn from observations and reasoning. A

modern sequence does not directly display the whole history of its lineage. Gene transfer between lineages, gene loss, different evolutionary rates and uncertainty in ancient reconstructions complicate the inference. The evidence remains informative when these limits are recognized and alternative explanations are tested.

■ Worked example: a similarity is not a date

An invented alignment has 18 matching sites out of 24 comparable sites:

$$\text{identity} = \frac{\text{matching sites}}{\text{compared sites}} \times 100\% = \frac{18}{24} \times 100\% = 75\%.$$

This describes similarity in that alignment. It does not specify the rate of change, which differences are repeated changes at one site, or the relevant calibration. It therefore cannot alone give a date for LUCA or prove that one sampled organism was another’s direct ancestor.

AHL: How do scientists estimate ancient dates?

Different evidence answers different dating questions. Fossils and chemical traces can show that life existed **by** a certain time, provided their biological origin and geological context are established. They rarely identify the instant when life first appeared. Older evidence may be absent because rocks were destroyed, altered or never preserved the relevant features.

Rock ages can be constrained using radiometric methods and the relationships between rock layers. A date for the surrounding geological material is not automatically a direct date for the formation of every object inside it. Investigators must rule out younger contamination, later fluid movement and deformation that could change the interpretation.

A chemical signature, such as a carbon-isotope pattern compatible with biological processing, can be suggestive but needs testing against non-biological explanations. Shapes resembling cells or filaments also require contextual and chemical evidence. A convincing interpretation combines lines of evidence rather than treating one familiar-looking shape as decisive.

Molecular clocks use accumulated sequence differences, calibrated with other information, to estimate divergence times. They require assumptions about evolutionary rates, relationships and calibrations. Ancient reconstructions are uncertain and can change when the model or evidence changes. The estimated age of a common ancestor is not simply the age of the oldest fossil found so far.

To appreciate scale, an invented interval from 3.8 to 3.2 billion years ago lasts 0.6 billion years, or 600 million years. A diagram that compresses that interval into a centimetre can hide an enormous number of generations. This numerical example illustrates elapsed time; it is not a proposed date for life’s origin or LUCA.

AHL: Why investigate hydrothermal settings?

Hydrothermal systems circulate water through rock and create environments with mineral surfaces and chemical gradients. Some vent settings offer differences in redox conditions or proton concentrations that could provide energy for reactions. Mineral pores can offer small spaces that concentrate substances. A useful hypothesis must connect the proposed conditions to particular chemistry rather than saying only that vents are warm.

Ancient seafloor hydrothermal precipitates contain structures and chemical associations that researchers have interpreted as evidence of early microorganisms. Biological and non-biological processes can sometimes make similar structures, and the rocks' ages and alteration histories need careful evaluation. Such evidence can support life in an ancient hydrothermal environment without proving the exact birthplace of the first living system.

Genomic reconstructions also investigate whether conserved sequences and inferred metabolisms are consistent with an early ancestor using chemical energy in a vent-related setting. These are model-based inferences from surviving organisms, not a direct reading of LUCA's entire environment. Fossil context and genomic evidence are different lines of reasoning; agreement can strengthen a hypothesis, but shared assumptions must not be counted as independent confirmation.

AHL: Distinguish later endosymbiosis

Endosymbiosis is an association in which one organism lives within another. Evidence that mitochondria and chloroplasts descend from bacterial lineages includes their own DNA, bacterial-type ribosomes and division within cells. Several consistent comparisons are more persuasive than one feature alone.

This model addresses the later history of eukaryotic cell organization. It is not a complete explanation for the origin of the first cells: bacteria and a host lineage already existed in the proposed association. It also does not mean every organelle descended from an engulfed bacterium. Keep the question being answered distinct from the wider origin-of-life problem.

Practice

All calculations use invented teaching data. Answer with an observation, inference and limitation where relevant.

1. Link little free oxygen to ozone and ultraviolet exposure on early Earth. Why should one laboratory gas recipe not be treated as a directly measured universal atmosphere?
2. Explain why a virus is not classified as a living cell here. Why are crystal growth and fire spreading insufficient evidence of cellular life?

3. Explain how current cells arising from pre-existing cells differs from the historical origin problem. Name four capacities a plausible early cellular system must bring together.
4. Describe the roles of water heating, the gas mixture, electrical energy, cooling and collection in the classic spark-discharge model. State what the product evidence supported and what it did not show.
5. Suggest one control and one contamination check for a model chemical-synthesis investigation. Explain why failing to produce cells is not itself a failure to test building-block synthesis.
6. Explain how fatty acids can form a vesicle in suitable conditions. Why does its boundary need both retention and exchange, and why is the vesicle not automatically alive?
7. Explain how RNA can be relevant to both heredity and catalysis. Give one modern example and two unresolved requirements of an early RNA-based cellular model.
8. Distinguish LUCA from the first living system and from the only organism alive at its time. Give two kinds of evidence for common ancestry.
9. An alignment has 21 matches among 28 compared sites. Calculate identity. Repeat for 16 out of 20, and explain why neither percentage alone dates a common ancestor.
10. Explain how fossils, rock dating and molecular clocks answer related but different questions. Why does the oldest known evidence usually give a minimum age for the existence of life rather than an exact origin date?
11. An invented geological interval runs from 3.8 to 3.2 billion years ago. Find its duration in millions of years. Explain one way a small timeline can mislead a reader about evolutionary timescales.
12. Explain two potentially useful features of a hydrothermal environment. Distinguish evidence for ancient life at a vent from proof that all life originated there.
13. Explain how evidence for mitochondrial endosymbiosis differs from evidence for abiotic synthesis. Why does a single matching gene not prove the complete origin of life?

■ Worked answers

1. Little molecular oxygen means little ozone could form, so more ultraviolet radiation could reach the surface. Early atmospheric models are reconstructed and vary across settings and time; a defined gas recipe tests a possible chemical setting rather than directly measuring the whole ancient atmosphere.
2. A virus relies on host cellular machinery and energy for reproduction and lacks autonomous cellular metabolism. Crystals and fire can grow or spread without the integrated hereditary, compartmentalized metabolic organization of a cell.

3. Cell division describes how cells arise today, not how the first cellular lineage originated. An origin model must connect catalysis, replication/heredity, self-assembly and compartmentalization with exchange of matter and energy.
4. Heating supplies vapour; the methane/ammonia/hydrogen mixture supplies reactants; sparks provide energy; cooling condenses material; collection allows analysis. Amino-acid production supports abiotic building-block synthesis under those conditions, not production of living cells or a complete historical pathway.
5. A matched no-energy-input control tests whether the energy source affects products. Analyse clean starting materials and blanks, and use procedures that exclude biological contamination. A building-block experiment can answer its chemical question without generating a complete cell.
6. Amphipathic molecules can arrange with heads contacting water and tails sheltered in a bilayer that closes around an aqueous interior. Retention keeps useful components together; exchange supplies reactants and removes products. A boundary alone supplies neither inherited functional information nor an integrated metabolism.
7. RNA sequences can store information and act as copying templates; some folded RNAs catalyse reactions. Ribosomal RNA contributes to peptide-bond formation. Early models still need workable nucleotide supply and accurate copying, for example, as well as suitable stability and connection to compartments.
8. LUCA is ancestral to surviving cellular lineages, not necessarily the first cell or the only contemporary organism. The conserved genetic code and shared fundamental genes support common ancestry; earlier or competing lineages may have become extinct.
9. $I = 21/28 \times 100 = 75\%$; $I = 16/20 \times 100 = 80\%$. Identity alone supplies no calibrated change rate and does not account for multiple changes, gene transfer or the full relationship among samples.
10. Fossils/chemical traces can establish life's presence in context; radiometric and stratigraphic methods constrain geological ages; calibrated molecular clocks estimate divergence times from sequences. An older origin could predate the oldest preserved, discovered evidence, so that evidence usually establishes life existed by a time.
11. $\Delta t = (3.8 - 3.2) \times 1000 = 600$ million years. A short drawn interval can conceal many generations; apparent spacing in an unscaled diagram must not be read as a measured duration.
12. Chemical/proton gradients could supply energy, and pores/mineral surfaces could concentrate reactants or support reactions. Biological-looking structures and suitable chemistry in an ancient deposit can support life there, but neither identifies the unique origin site nor rules out alternative environments or abiotic structures without further tests.
13. Endosymbiosis concerns existing cell lineages combining in later eukaryotic evolution; abiotic synthesis concerns chemical building blocks forming without organ-

isms. One gene comparison has a limited history and possible alternative explanations. A complete origin model needs several processes and convergent lines of evidence, not one similarity.

Exam tips

- State the question a model actually tests. Building blocks, compartments, common ancestry and eukaryotic organelles are distinct explanatory targets.
- Separate observations from inferences; add a limitation without discarding valid evidence. Feasibility under defined conditions is not proof of a unique historical pathway.
- LUCA is the ancestor of surviving lineages, not automatically the first cell. Similarity percentages are not dates without a model and calibration.
- For early-life claims, connect chemical, geological and sequence evidence while considering contamination, alteration and non-biological alternatives.

References: IB A2.1; Miller, original research; Dodd and colleagues, hydrothermal evidence.