AL-FARABI KAZAKH NATIONAL UNIVERSITY,
FACULTY OF BIOLOGY AND BIOTECHNOLOGY
DEPARTMENT OF BIOTECHNOLOGY

Lecture 1

INTRODUCTION TO SCIENTIFIC RESEARCH IN BIOLOGY AND BIOTECHNOLOGY

PhD, Acting Associate
Professor, Doktyrbay Gulina

Aim of the Lecture

To introduce students to the fundamental principles of scientific research in biology and biotechnology, explain the importance of scientific methods, and develop an understanding of the main stages of conducting scientific investigations.

SCIENTIFIC RESEARCH	RESEARCH METHODS	RESEARCH PROCESS
Principles and objectives	Observation and experiments	Planning and analysis

Expected Learning Outcomes

By the end of this lecture, students will be able to:

1. Define scientific research and explain its importance in biology and biotechnology.

2. Identify the main characteristics and principles of scientific research.

3. Describe the major stages of the scientific research process.

4. Differentiate between basic and applied research.

5. Explain the role of scientific methods in biological and biotechnological investigations.

What Is Scientific Research?

Scientific research is the systematic, planned investigation of natural phenomena. Observation and experiment generate evidence that can be tested, criticised and reproduced by other researchers. Its results are not final truths but provisional knowledge that stays open to revision when better evidence appears.



Scientific assumption

A statement accepted without testing. Example: "Cells always grow faster in a warmer incubator." It may sound plausible, but nothing has been measured.

Scientific evidence

Data from a controlled, repeatable measurement. Example: doubling time of E. coli in LB medium measured at 25, 37 and 42 °C, with replicates and stated uncertainty.

Why Research Matters in Biology

Living organisms	Cell theory (Schleiden, Schwann, Virchow; 1838–1858): all organisms consist of cells, and cells arise from pre-existing cells. Improved microscopy made this possible.
Cellular processes	Mitchell's chemiosmotic hypothesis (1961) explained how a proton gradient across a membrane drives ATP synthesis; it was later confirmed experimentally (Nobel Prize, 1978).
Genetics and heredity	Mendel's pea crosses (1866) showed discrete inheritance; Hershey and Chase (1952) indicated that DNA carries genetic information; Watson and Crick (1953) proposed the double helix.
Diseases	Koch's work on Bacillus anthracis (1876) linked a specific microbe to a specific disease and led to the criteria now called Koch's postulates.
Biological interactions	Studies of Vibrio fischeri bioluminescence revealed quorum sensing: bacteria coordinate gene expression according to population density using signalling molecules.

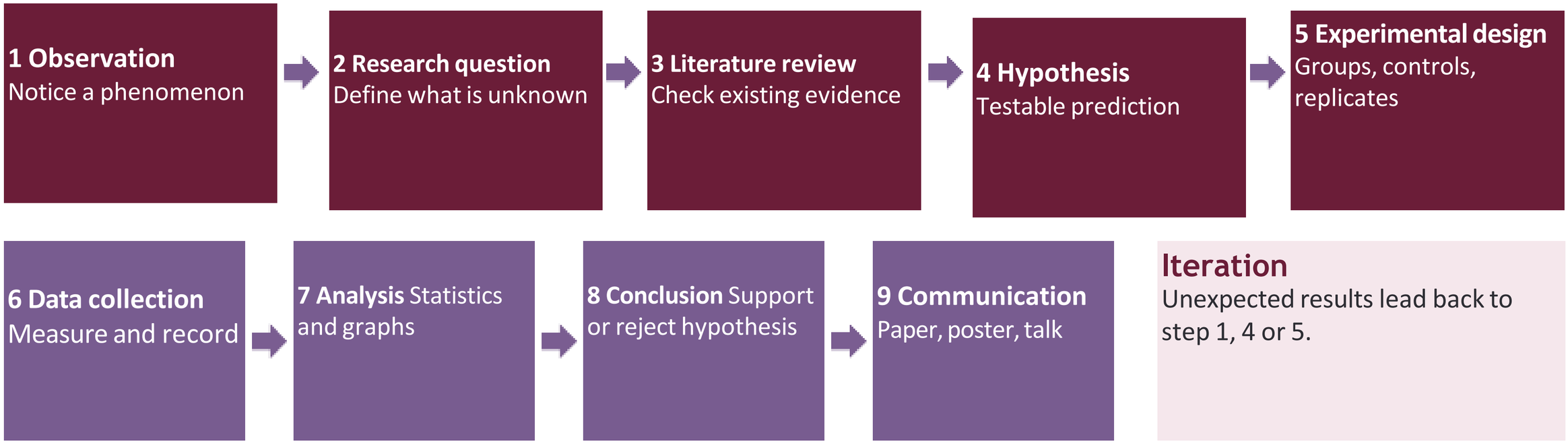
From Basic Discovery to Biotechnology

Restriction enzymes and DNA ligase (1970s)	Recombinant DNA cloning	**Recombinant human insulin; engineered microbes and crops**
Antigen recognition and immune memory (Jenner, Pasteur)	Attenuated, subunit and mRNA vaccine platforms	**Vaccines**
Thermostable DNA polymerase from Thermus aquaticus	PCR and real-time PCR	**Diagnostic tests, e.g. viral RNA detection**
Antibody specificity; hybridoma technique (1975)	Large-scale mammalian cell culture	**Monoclonal antibody biopharmaceuticals**

Types of Scientific Research

Type	Main aim	Typical question	Biotechnology example
Basic	Understand fundamental mechanisms	How does Cas9 recognise its target DNA?	Bacterial immunity research that led to CRISPR genome editing
Applied	Solve a defined practical problem	Which E. coli strain gives the highest enzyme yield?	Optimising fermentation conditions for protein production
Translational	Move findings from laboratory to clinic or industry	Is an edited-cell therapy safe and effective in patients?	CRISPR-edited cell therapy for sickle cell disease
Experimental	Manipulate a variable and measure its effect, with controls	Does 42 °C change growth rate compared with 37 °C?	Screening culture conditions in small bioreactors
Observational	Record variables without intervening	Which microbes differ between healthy and diseased gut?	Metagenomic surveys of hot springs that search for new enzymes

The Scientific Method as a Workflow



The method is iterative, not strictly linear. A rejected hypothesis is still a result: it narrows the possible explanations and usually produces a better question. Researchers also return to the literature as new findings appear.

Identifying a Research Problem

Knowledge gaps
Questions left open by the literature, such as the function of an uncharacterised gene.

Unexplained observations
Bacteria surviving phage infection led to the study of CRISPR systems.

Example: artemisinin supply
Artemisinin, a key antimalarial, was extracted from the plant Artemisia annua, and supply and price fluctuated.
Researchers engineered the yeast Saccharomyces cerevisiae to make artemisinic acid, a precursor, by adding pathway genes. The problem was a supply limit; the research question became metabolic: which pathway changes raise the yield?

Technological limits
Existing methods may be slow, costly or insensitive, e.g. culture-based detection that takes days.

Clinical and industrial
D^{n}r^{e}u^{e}g^{d}re^{s}sistance, low yields or unstable products create demand for targeted research.

Formulating a Research Question

System	Weak question	Strong question
Bacterial growth	Does temperature affect bacteria?	Does incubation at 25, 37 and 42 °C change the specific growth rate of E. coli K-12 in LB medium, measured as OD600?
Enzyme activity	Is amylase affected by conditions?	How does pH (4, 7, 9) affect the initial reaction rate of α-amylase at 37 °C?
Cell culture	Does the drug kill cells?	Does 0–50 µM compound X reduce the viability of HeLa cells after 48 h, measured by MTT assay?

Checklist for a strong question

Specific organism or system · measurable outcome with units · defined conditions and time · feasible with available methods · ethical · relevant to a known knowledge gap (FINER: feasible, interesting, novel, ethical, relevant).

Scientific Hypotheses

A hypothesis is a tentative, testable explanation that predicts a measurable outcome. It must be falsifiable: some possible result would contradict it. The researcher manipulates the independent variable and measures the dependent variable.

Null hypothesis (H_0)

The initial reaction rate of α-amylase is the same at pH 4, 7 and 9. Statistical tests ask whether the data are compatible with this statement.

Independent variable

Controlled by the researcher. Example: buffer pH (4, 7, 9).

Dependent variable

Measured response. Example: reducing sugar released per minute (A540, DNS assay).

Alternative hypothesis (H_1)

The rate differs between at least two pH values. Prediction: activity is highest near neutral pH, close to the enzyme's reported optimum.

From Hypothesis to Experiment

Research question
Does temperature influence the growth of E. coli?

Hypothesis
Growth rate is highest at 37 °C and lower at 25 °C and 42 °C.

Variables
Independent: temperature. Dependent: OD600, growth rate. Constant: medium, strain, shaking.

25 °C
Experimental group
3 independent cultures

37 °C
Control group: standard condition
3 independent cultures

42 °C
Experimental group
3 independent cultures

Expected measurements
OD600 every 2 h for 10 h (turbidity of the culture). Viable counts (CFU/mL) at selected times to confirm that turbidity reflects living cells. Derived values: growth rate μ and doubling time from the exponential phase. All flasks: same inoculum (OD600 = 0.05), 50 mL LB in 250 mL flasks, 200 rpm.

The Importance of Literature Review

Map current knowledge

Shows what is established, what is disputed and which models the field uses.

Identify research gaps

Contradictory results, untested conditions or missing organisms point to open questions.

Avoid unnecessary repetition

Reveals work already done; deliberate replication of key results is still valuable.

Develop hypotheses

Mechanisms found in related systems suggest testable predictions.

Select methods

Published protocols show which methods, controls and sample sizes are validated.

Search strategy: define keywords and synonyms, combine them with Boolean operators (AND, OR, NOT) and record the databases and dates searched.

Finding Reliable Scientific Literature

Source	Coverage	Practical note
PubMed	Biomedical and life-science literature indexed by the U.S. National Library of Medicine	Free; MeSH terms improve precision
Google Scholar	Broad and multidisciplinary; includes theses and preprints	Free; coverage less transparent
Scopus	Abstract and citation database (Elsevier)	Subscription; citation analysis
Web of Science	Curated citation index (Clarivate)	Subscription; tracks citing papers

Peer-reviewed journals (e.g. Nature, Science, Cell) are the main route for primary research; start with a recent review, then read the primary papers it cites.

Primary article

Reports original data, methods and results. Evaluate it directly.

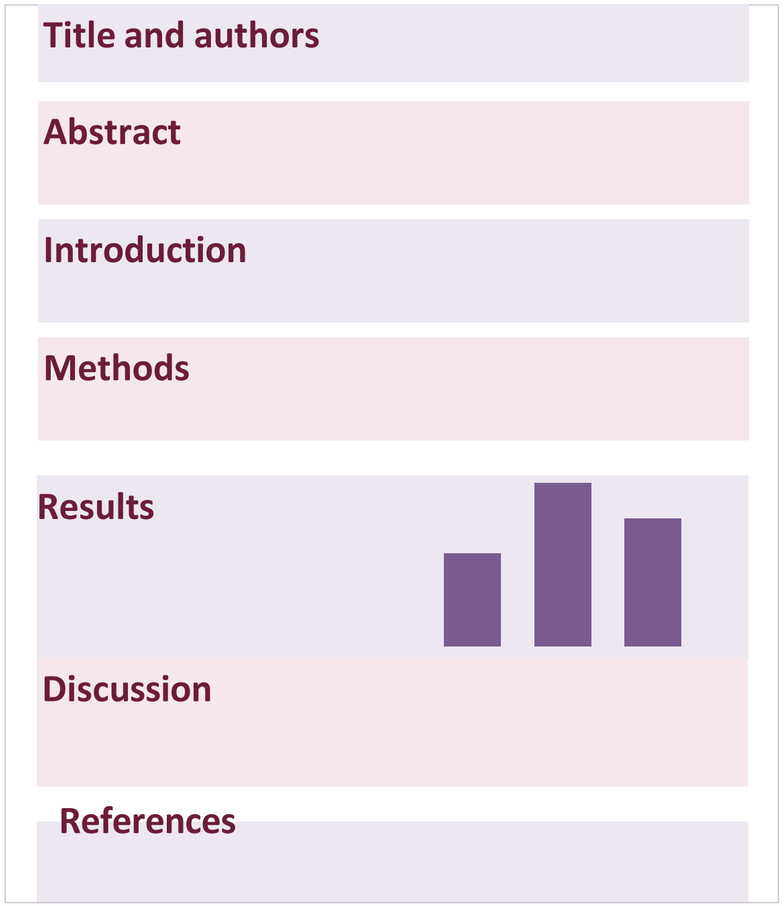
Review article

Synthesises many studies; systematic reviews use explicit search methods.

Peer review

Experts judge validity before publication; this reduces but does not remove errors.

Understanding a Scientific Research Paper



Abstract	Aim, methods, main results and conclusion
Introduction	Background, knowledge gap, hypothesis
Methods	Materials, protocols, statistics; must be repeatable
Results	Figures and tables, without interpretation
Discussion	Meaning, comparison with literature, limitations
Conclusion	Direct answer to the question; often the last paragraph
References	Sources cited, so claims can be checked
Supplementary	Extra datasets, protocols, uncropped images

Critical Evaluation of Scientific Literature

Criterion	Stronger evidence	Weaker evidence
Study design	Controlled, pre-specified hypothesis	Anecdote, no comparison group
Sample size	Justified, e.g. by power analysis	Very small n without justification
Controls	Positive and negative controls included	No controls
Statistics	Suitable test; effect size and CI reported	Only p-values; multiple testing ignored
Reproducibility	Methods and data shared; independently replicated	Incomplete methods; one experiment
Limitations	Discussed openly	Not acknowledged
Bias	Randomisation, blinding, declared conflicts	Selective reporting; undisclosed funding

Principles of Experimental Design

Controlled experiment
Change one factor and hold others constant, so that the effect can be attributed to it.

Experimental group
Receives the treatment; its response shows the effect under test.

Control group
Receives no treatment or the standard condition; provides the baseline.

Replication
Independent repeats estimate variability and show whether a result is stable.

Randomisation
Random allocation (e.g. flask positions in a shaker) prevents systematic bias.

Standardisation
Fixed protocols, calibrated instruments and the same reagent lots make groups comparable.

Variables in Biological Experiments

Independent variable
Temperature: 25, 37, 42 °C

Dependent variable
OD600 and growth rate μ (h^{-1})

Question
Does the manipulated factor change the measured response?

Controlled variables (kept constant)

Same E. coli strain and pre-culture age · same LB lot · starting OD600 = 0.05 · 50 mL in 250 mL flasks · 200 rpm shaking · same measurement times.

Confounding variables (uncontrolled)

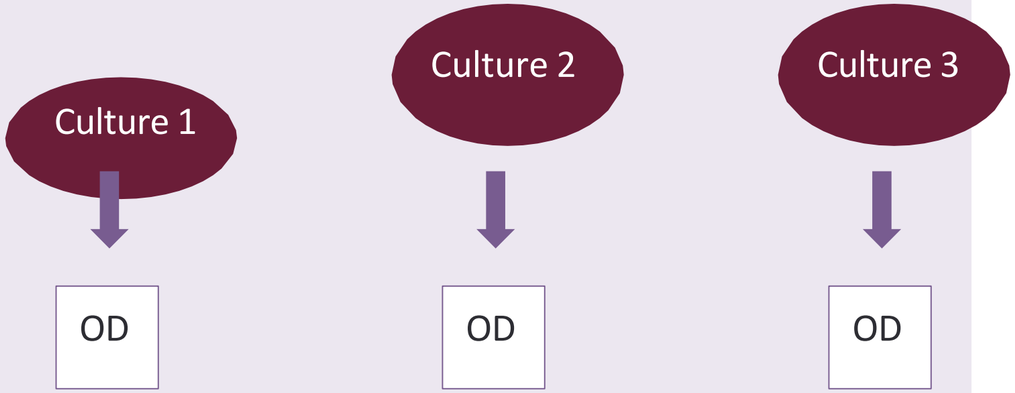
Hidden factors that differ between groups: 25 °C flasks always on the top shelf with different aeration, an incubator with a different calibration, an older pre-culture. Remedy: randomise positions, share one pre-culture, verify temperatures.

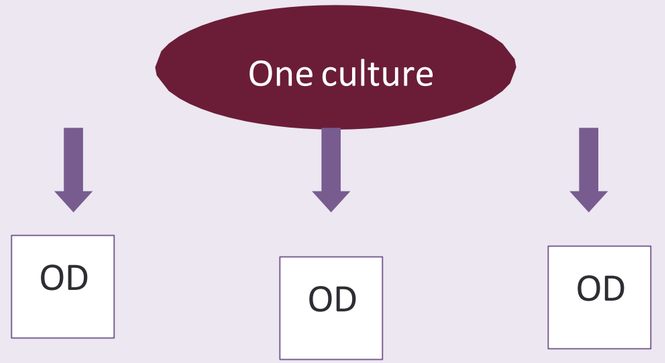
Controls in Biological Research

Control	Purpose	PCR	Enzyme assay	Diagnostic test
Positive	Shows the method can work	Known template with working primers	Enzyme at known activity	Known positive sample
Negative	Detects contamination or background	No-template control (water)	Reaction without enzyme or substrate	Known negative sample
Experimental (vehicle)	Isolates the treatment effect	Reaction with buffer only	Solvent without inhibitor	Sample matrix without analyte
Biological (internal)	Checks sample quality and normalises	Reference gene, e.g. GAPDH	Total protein per sample	Human RNase P gene in SARS-CoV-2 RT-PCR

A result is interpretable only if the controls behave as expected. A positive signal in the no-template control, for example, invalidates the whole PCR run.

Biological and Technical Replicates





Independent samples

Separate cultures, animals or patients. They capture true biological variability and define n.

Repeated measurements

Same sample measured several times. They estimate pipetting and instrument error, are averaged and do not raise n.

Common Experimental Techniques in Biology

Light microscopy
Cell shape, size and motility; staining adds contrast. Resolution limit about 200 nm.

Cell culture
Growth of cells in vitro in defined medium at 37 °C; used for viability and drug-response assays.

PCR
Exponential amplification of a chosen DNA sequence; detects a gene or pathogen.

Gel electrophoresis
Separates DNA or protein by size in an electric field; sizes are read against a ladder.

Spectrophotometry
Absorbance follows the Beer–Lambert law: OD600 for turbidity, A260 for nucleic acids.

Centrifugation
Separates by size and density; report relative centrifugal force (× g), not only rpm.

Molecular Biology Research Techniques

DNA ANALYSIS

DNA extraction Lyse cells; remove protein, RNA

→

Quality check
A260/A280 ≈ 1.8

→

PCR
Amplify target with primers

→

Gel analysis
Check product size

→

Sequencing
Sanger or NGS

GENE EXPRESSION ANALYSIS

RNA extraction A260/A280 ≈ 2.0; check integrity

→

Reverse transcription
RNA to cDNA

→

qPCR
Fluorescence per cycle gives Ct

→

Relative expression Normalise to reference gene (2^−ΔΔCt)

Sanger vs. NGS

Sanger sequencing reads one fragment at a time; next-generation sequencing reads millions in parallel.

Why qPCR needs care

Results depend on primer efficiency, stable reference genes and no-template controls.

Cell Culture and Microbiological Research

System	Medium	Conditions	Readout
E. coli K-12	LB broth	37 °C, 200 rpm, aerobic	OD600; CFU/mL
S. cerevisiae	YPD	30 °C, shaking	OD600; cell counts
HeLa or HEK293 cells	DMEM + 10% fetal bovine serum	37 °C, 5% CO_2, humidified	Trypan blue count; MTT assay

Sterile technique

Laminar-flow cabinet, 70% ethanol, aseptic handling; media sterilised by autoclave (121 °C) or 0.22 µm filter.

Measuring growth and viability

OD600 measures turbidity, which includes dead cells; CFU counts viable, culturable bacteria. Trypan blue enters cells with damaged membranes; MTT reflects metabolic activity.

Growth phases

Lag, exponential, stationary and death.

Growth rate comes from the exponential phase.

Collecting Scientific Data

Qualitative data

Descriptions without numbers: colony morphology, presence of a gel band, cell shape. Record with images and consistent terms.

Quantitative data

Numerical measurements: OD600, CFU/mL, band intensity, enzyme rate. Always record units, instrument and replicate number.

Laboratory notebook and data integrity

Record date, objective, protocol, deviations, reagent lot numbers and raw data at the time of work. Data integrity is summarised as ALCOA: records are Attributable, Legible, Contemporaneous, Original and Accurate. Never overwrite entries; correct them with a dated, signed note. Keep raw files (images, instrument output) separate from processed data.

Understanding Biological Data

Culture	OD600 at 6 h (37 °C)
Culture 1	0.82
Culture 2	0.78
Culture 3	0.90
Culture 4	0.85
Culture 5	1.45 (possible outlier)
Mean ± SD (all, n = 5)	0.96 ± 0.28
Mean ± SD (without #5, n = 4)	0.84 ± 0.05

Mean and SD

The mean describes the centre; SD describes spread of individual values. SE = SD/√n describes precision of the mean.

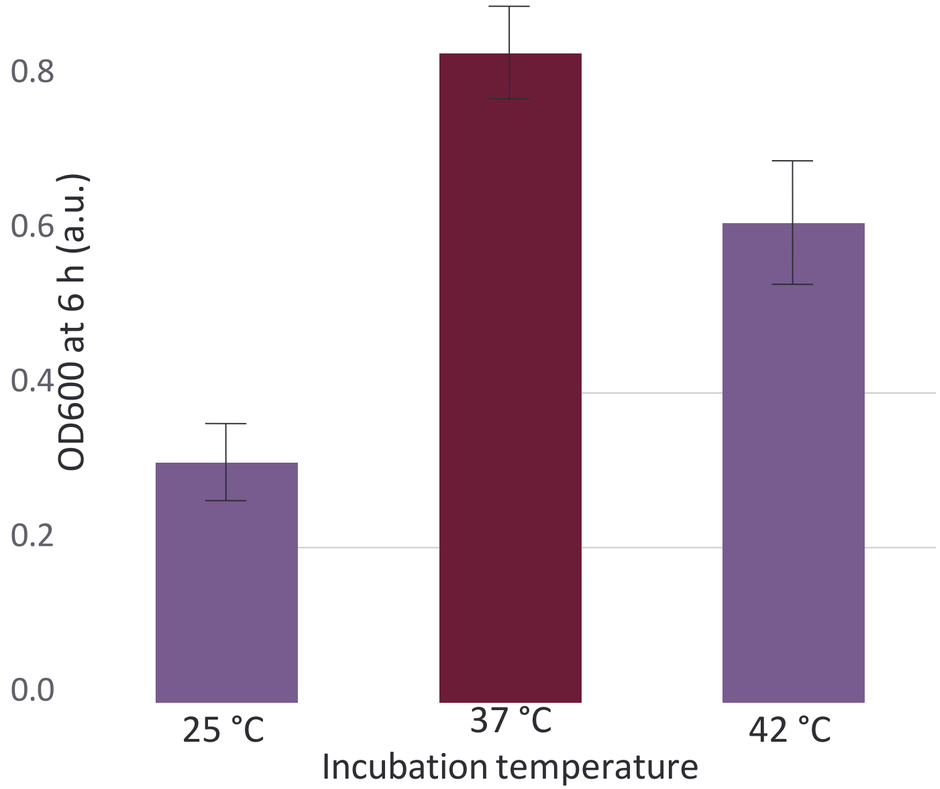
Variability and uncertainty

Biological variability is real difference between cultures; measurement uncertainty comes from the instrument and pipetting.

Outliers

One value inflates the SD fivefold. Investigate the cause; remove only with a pre-specified, documented criterion.

Data Visualization in Scientific Research



Graph	Use when
Scatter plot	Relationship between two continuous variables
Line graph	Change over time, e.g. growth curves
Bar chart	Few groups; show error bars and n
Box plot	Distribution: median, quartiles, outliers
Microscopy image	Morphology; include scale bar and staining

Label axes with units, state what error bars represent, and show individual points when n is small.

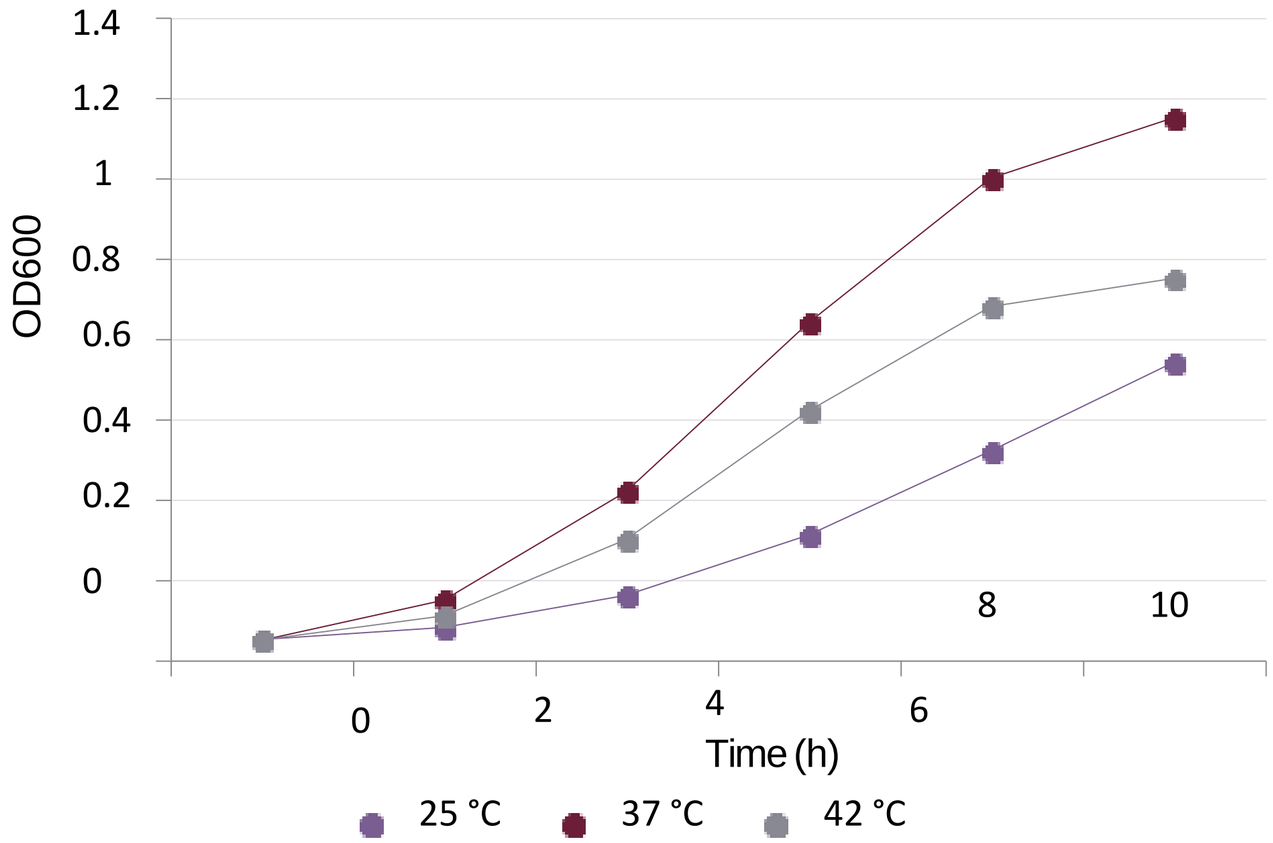
Statistical Analysis in Biology

Method	Use	Example
Hypothesis test	Compare data with a null hypothesis	H_0: mean growth rate equal at 37 and 42 °C
P-value	Probability of data at least this extreme if H_0 is true	p = 0.03
Confidence interval	Range of plausible values for an estimate	95% CI for a mean difference
t-test	Compare two group means	37 °C vs. 42 °C
ANOVA	Compare three or more means	25, 37, 42 °C together
Correlation	Strength of association	OD600 vs. CFU/mL

Significance is not

'Ampᵖ-voalʳuᵗeᵃⁿisᶜneᵉot the probability that H_0 is true. With large samples, a 2% difference can give $p < 0.001$ and still be biologically trivial. Report effect size and confidence interval, and ask whether the change matters for the organism or process.

Interpreting Experimental Results



Pattern and comparison

37 °C reaches the highest OD600; 25 °C grows slowest. This supports the hypothesis.

Unexpected result

42 °C still grows. Heat stress limits, not stops, growth; check strain and thermometer before explaining.

Causation

Temperature was manipulated, so a causal claim is justified; correlations in observational data are not.

Reproducibility and Reliability

Reproducibility
Independent researchers obtain consistent results using the same methods on new samples.

Repeatability
The same operator and equipment give consistent results in repeated measurements.

Error	Example
Systematic	Uncalibrated pipette; pH meter drift
Random	Pipetting variation; biological variation
Gross	Sample mix-up; contamination

Standard operating
pWrroittceendpurorteoscols reduce operator-dependent variation.

Documentation
Complete methods, raw data and reagent details allow others to repeat the work.

Why it matters in biotechnology
In a survey of 1,576 researchers, over 70% had failed to reproduce another scientist's experiment. A process that cannot be reproduced cannot be validated for manufacturing.

Research Ethics in Biology and Biotechnology

Scientific integrity

Fabrication is inventing data. Falsification is manipulating materials, equipment, processes or results. Plagiarism is presenting others' work as one's own. Honest reporting also means disclosing conflicts of interest and negative results.

Protecting people and animals

Human research follows the Nuremberg Code (1947), the Declaration of Helsinki and the Belmont Report: respect for persons, beneficence, justice. Animal research applies the 3Rs (Replace, Reduce, Refine) and needs ethics-committee approval.

Informed consent

Participants understand risks and benefits and agree voluntarily.

3Rs

Replace, Reduce, Refine the use of animals.

Biosafety

Risk assessment for pathogens and GMOs.

Biosafety and Laboratory Good Practices

Level	Typical agents	Key controls
BSL-1	Non-pathogenic E. coli K-12	Standard microbiological practice
BSL-2	S. aureus; human cell lines	Biosafety cabinet for aerosols; restricted access
BSL-3	Mycobacterium tuberculosis	Controlled airflow; respiratory protection
BSL-4	Ebola virus	Maximum containment suit or Class III cabinet

Good Laboratory Practice (GLP)

Quality system for planning, performing, recording and archiving non-clinical safety studies so results are traceable.

1 Risk assessment: agent, procedure, exposure route

2 Containment: biosafety cabinet, ventilation

3 PPE: lab coat, gloves, eye protection

4 Decontaminate: 70% ethanol, autoclave at 121 °C

5 Waste: separate sharps, biohazard, chemical

Case Study: Recombinant Human Insulin

1 Medical problem
Animal insulin: limited supply, immune reactions

2 Gene design
Sequence known (Sanger); A and B chain genes made

3 Cloning
Genes inserted into plasmids with restriction enzymes, ligase

4 Expression
E. coli produced each chain in culture

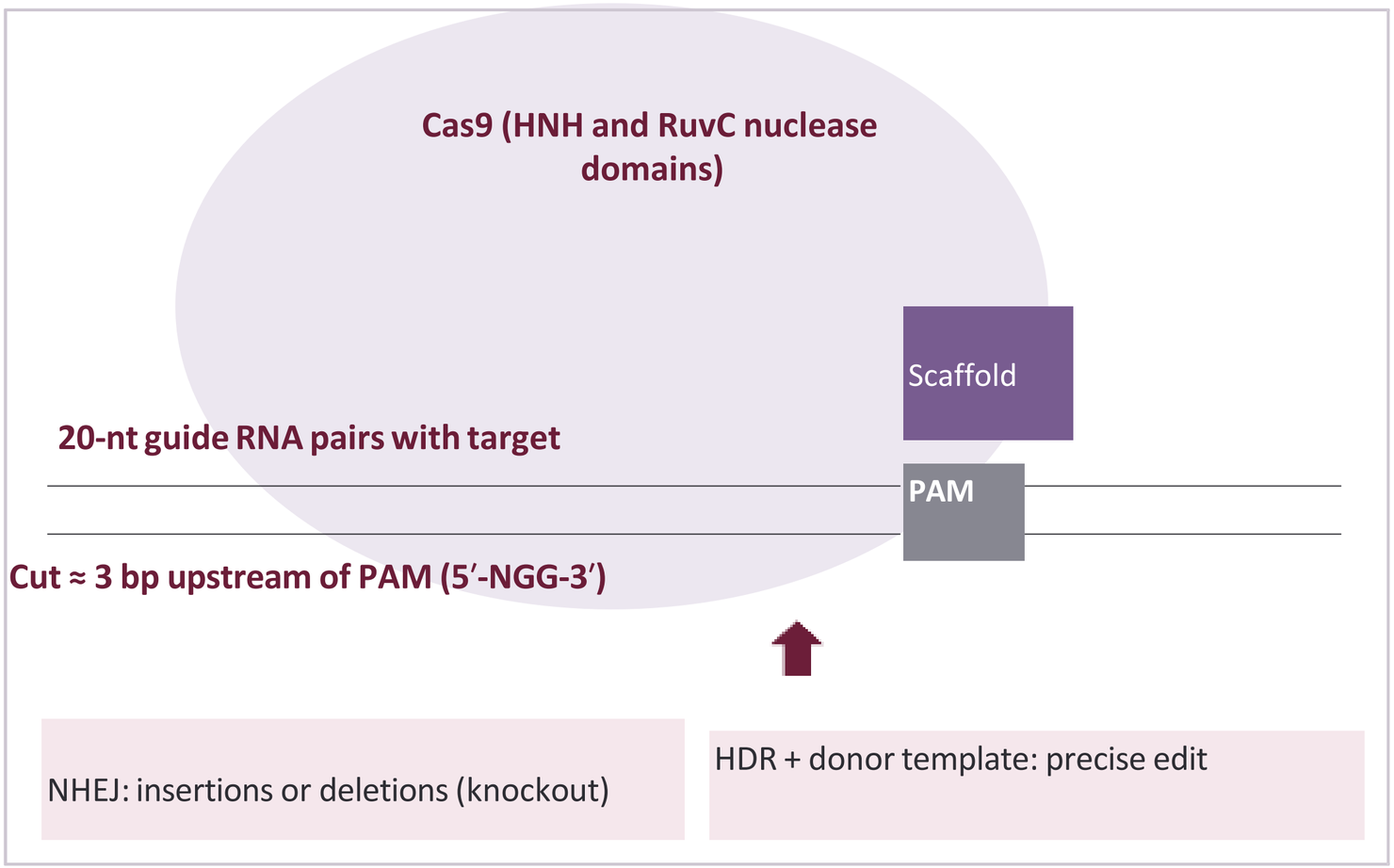
5 Purification
Cells lysed; chains purified, joined by disulfide bonds

6 Quality control
Purity, identity and potency tested

7 Medical use
FDA approval of human insulin (Humulin), 1982

Modern production oken expresses proinsulin in E. coli or yeast and converts it to insulin enzymatically.

Case Study: CRISPR-Cas9 Gene Editing



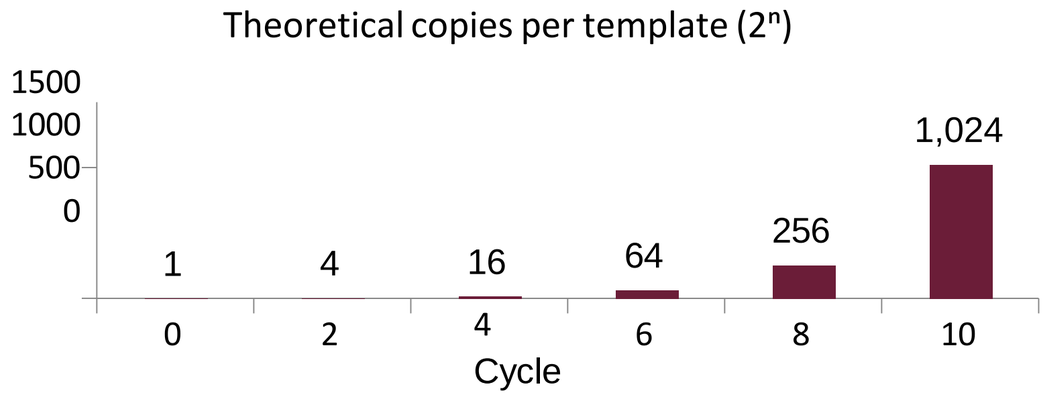
Discovery · CRISPR repeats act as adaptive immunity against phages (Barrangou, 2007).
Programming · A single guide RNA directs Cas9 to a chosen sequence (Jinek, 2012).
Recognition · Cas9 binds a PAM, then the guide pairs with adjacent DNA.
Cleavage · HNH and RuvC domains cut each strand.
Repair · NHEJ or HDR; used in human cells since 2013 (Cong).

Case Study: PCR in Molecular Diagnostics

1 Denaturation
94–98 °C: strands separate

2 Annealing
50–65 °C: primers bind

3 Extension
72 °C: polymerase copies



Components and detection
Template DNA, two primers flanking the target, thermostable Taq polymerase, dNTPs, Mg^{2+} buffer; 25–35 cycles. Products are seen on a gel, or by fluorescence in real-time PCR. RT-qPCR detects viral RNA via a cycle threshold (Ct).

Communicating Scientific Research

Experiments
Data and figures

Manuscript
IMRaD structure

Submission
Journal choice

Screening
Editor checks scope

Peer review
Independent expert assessment

Revision
Reply to reviewers

Publication
Citable record with DOI

Posters and talks
Early results are discussed at conferences, giving feedback before submission.

Collaboration
Joint projects combine methods; authorship records contributions.

Open science
Preprints, open access and FAIR data in repositories such as GenBank.

Conclusion: From Scientific Questions to Biological Innovation

Question
Research begins with a meaningful, testable question grounded in the literature.

Design
Controls, replicates and randomisation make experiments reliable.

Analysis
Critical analysis weighs variability, statistics and limits of interpretation.

Ethics
Honesty, informed consent, animal welfare and biosafety are essential.

Innovation
Discoveries become technologies when evidence is robust.

Insulin
Recombinant human insulin, 1982

PCR diagnostics
Detection of pathogens

CRISPR editing
Research and therapy

References

Alberts, B., Heald, R., Johnson, A., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2022). *Molecular biology of the cell* (7th ed.). W. W. Norton.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*(7604), 452–454. https://doi.org/10.1038/533452a Blainey, P., Krzywinski, M., & Altman, N. (2014). Replication. *Nature Methods, 11*(9), 879–880. https://doi.org/10.1038/nmeth.3091 Bustin, S. A., et al. (2009). The MIQE guidelines. *Clinical Chemistry, 55*(4), 611–622. https://doi.org/10.1373/clinchem.2008.112797

Cong, L., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science, 339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Goeddel, D. V., et al. (1979). Expression in Escherichia coli of chemically synthesized genes for human insulin. *PNAS, 76*(1), 106–110. https://doi.org/10.1073/pnas.76.1.106

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Jinek, M., et al. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816–821. https://doi.org/10.1126/science.1225829