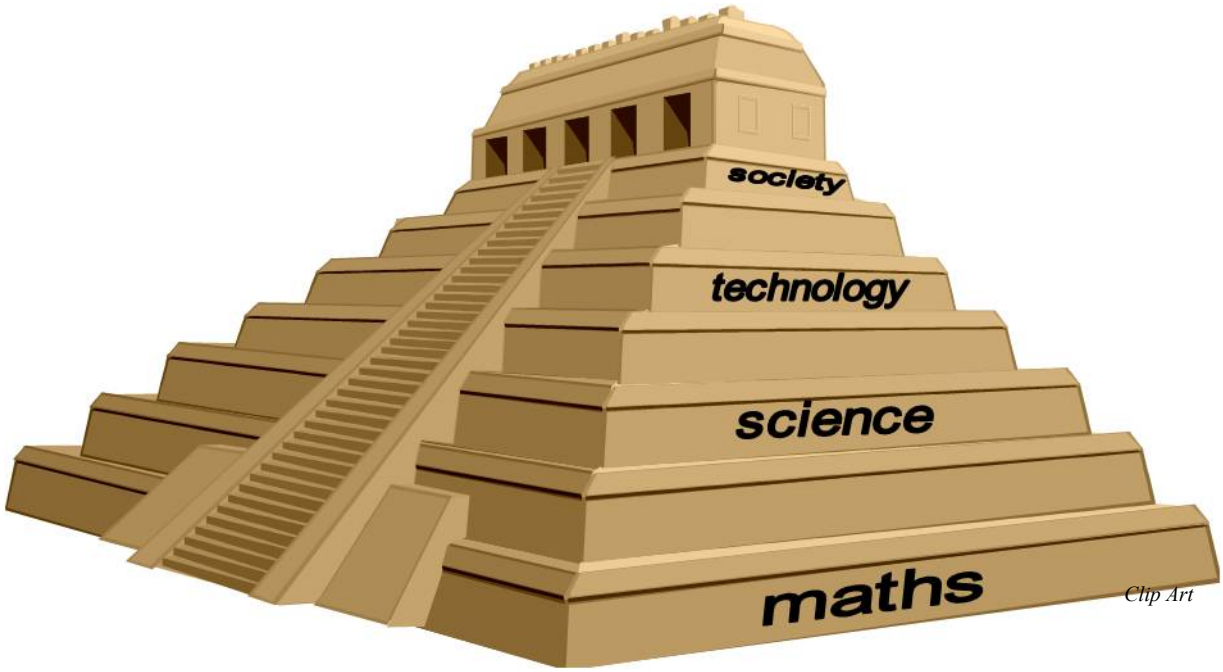
THE UNIVERSITY OF QUEENSLAND

Enabling SCIENCE

Introduction to Theory and Practice in Science



Clip Art

Quantitative skills underpin science and technology.
They are fundamental to theoretical understanding and practical applications.
Mathematics enables science!

Professor Peter O'Donoghue
Faculty of Science

Preface

The impetus for this booklet was provided by the introduction of a new first-year course at The University of Queensland in 2008 entitled "*Theory and Practice in Science*" (code SCIE1000) - a joint initiative of the Faculty of Science, the Faculty of Arts, and the Faculty of Social and Behavioural Sciences.

Few textbooks were available giving the right degree of breadth and depth considered appropriate for the topics covered in this course. Several generalist texts expounded upon philosophical and historical considerations or upon specific practical or technological applications. In comparison, there were many specialist textbooks available which provided detailed, even exhaustive, information about individual scientific disciplines or subdisciplines.

This booklet was therefore developed to give a multidisciplinary blend of concepts and knowledge at a level deemed appropriate for a generalist holistic course. It contains core conceptions in science, including many disciplines, but drawn heavily from chemistry, physics and biology (the "enabling" sciences). Contemporary anecdotes are used to illustrate concepts, and worked examples are given to demonstrate relevant mathematic skills.

Many students will have previously covered some of the material presented herein in their high school courses, but not in a uniform or integrated fashion. All students will have completed English and Mathematics (to be eligible for tertiary enrolment), but they will have disparate experience with Chemistry, Physics and Biology. Few students will have studied all these disciplines.

These notes are an attempt to provide an overview of the major issues in the natural sciences. They are not intended as replacements for other foundational courses, or as remediation for deficit areas of knowledge. They are provided as a one-stop shop for an integrated overview of information of which all scientists should be cognizant. It may be used as a course reader; to preview or to review material; to provide contextual relevance; to practice problem solving; and hopefully to engage students.

The material covered has been developed using many different sources; including textbooks, journal publications, fact sheets, and accredited web sites. Due acknowledgements and references are provided. The information is not presented as all-encompassing or encyclopaedic, but rather as introductory forays into relevant areas. The notes are fundamental and foundational to science – provided in an attempt to establish a stable platform upon which to build knowledge during your scientific career.

Prof Peter O'Donoghue

UQ POD Centre

TABLE OF CONTENTS

Chapter 1.


Clip Art

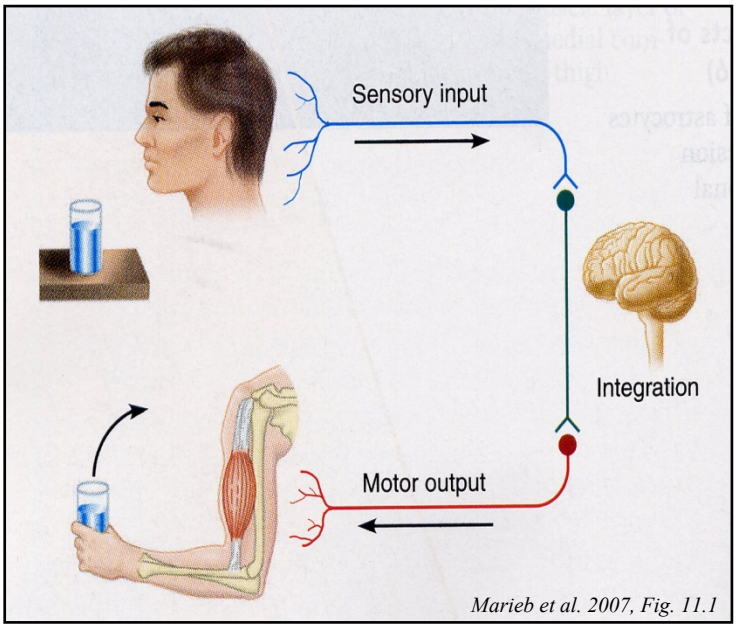
INTRODUCTION

Welcome to university and the course *"Theory and Practice in Science"*. Universities differ in many ways from secondary (high) schools because they place the onus for learning firmly on the student as a responsible independent adult learner. Even though there are many rules and regulations governing university courses and programs (and even professional codes of conduct), essentially the responsibility is on you, the student, to attend classes and examinations.

University staff schedule and conduct classes but usually they do not take roll-call (attendance) although your failure to attend small group classes (such as tutorials and laboratory classes) will be conspicuous to your tutor. Students must exercise some mental discipline as independent learners and resist the temptation to cut classes. If you do, you will be at a disadvantage to the rest of the class and will not perform as well. Many students have part-time employment to earn money for subsistence but they should not treat university as secondary to that employment. Full-time enrolment is exactly that. Although formal contact hours may only be 20-24 hours per week (5-6 hours per course for each of 4 courses), you are expected to undertake independent study (preview, review, extended readings, research, etc) on a matching basis (1 hour study for each hour of scheduled contact).

Your first onerous task will have been to organize your weekly schedule of classes. Ensure that you adhere to that schedule as it provides structure to the massive amount of content you will be exposed to. In most cases, it will have been arranged for tutorials and practicals to follow-on from lectures so that small group activities have direction and focus. During your degree, you will accumulate an extraordinary large amount of material: reference texts, recommended readings, lecture booklets, tutorial notes, practical guides, and a huge assortment of electronic files (documents, power-points, spreadsheets, databases, etc) accessed through intra-net and inter-net servers (such as Blackboard and Google). You must be organized and develop appropriate systems to sort, store and retrieve these materials. A cluttered desk does indicate a cluttered mind!

You MUST take notes during your classes. Simply listening and observing classes does not guarantee data retention or understanding. Turning all that sensory input into motor output by taking notes ensures that your brain has been engaged. The very act of writing involves many neural pathways and cognitive functions that serve to enhance comprehension and memory. Short-term and long-term memories involve different parts of the brain and sorting occurs during sleep. Periodic review of material ensures information persists in long-term memory, so taking notes during study (preview and review) facilitates better retention and understanding. Get into the habit of taking notes at every possible opportunity!



Marieb et al. 2007, Fig. 11.1

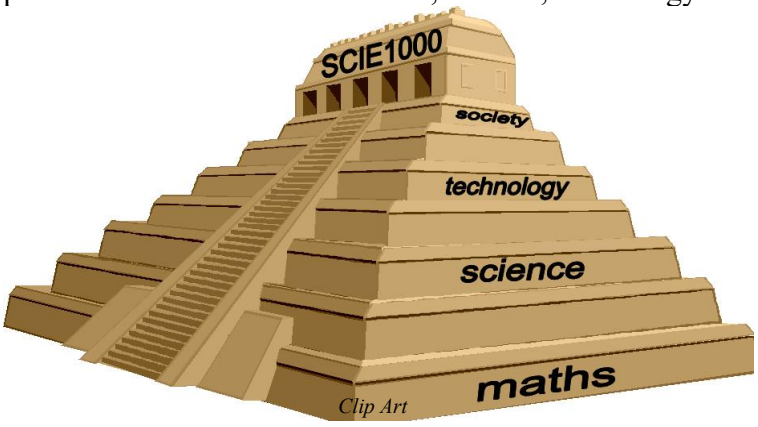
Why a Special Science Course?

Many governments and industries have commented on the recent global decline in the numbers of career scientists. Collectively, they are looking to higher educational authorities to redress this situation by providing appropriate training. In Australia, state and federal governments are exploring incentive schemes and have initiated policy changes. The Queensland Department of Education, Training and the Arts have developed a comprehensive paper entitled "Towards a 10-year plan for science, technology, engineering and mathematics (STEM) education and skills in Queensland".

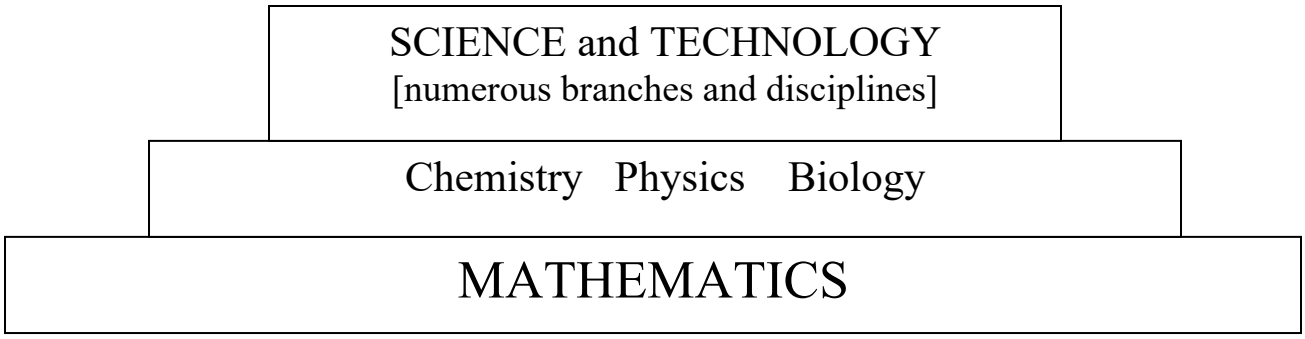
> *"The role of STEM cannot be underestimated in preparing Queenslanders for the challenges and opportunities of the future. The Queensland economy is booming. Strong demand for natural resources and the fastest-growing population in Australia are priming the rapid growth of Queensland's economy. However, our future prosperity cannot rely solely on the buoyancy of traditional industries and dynamic population growth. Global competition, market instabilities and changing trends in immigration are placing increasing pressures on the growth of the state's economy. To meet these challenges, Queensland needs to continue to encourage the emergence of new high-value, high-growth industries of the future and apply strategies to value-add to traditional industries. International experience demonstrates that high-growth economies are those that build upon strong foundations to move towards a knowledge-based economy. A workforce of scientifically and technologically literate people is key. With identified shortages across the engineering, science and medical professions, there is a growing need for students to specialize in STEM disciplines."* [www.education.qld.gov.au/projects/STEMplan]

A recent review (2007) of the Bachelor of Science (BSc) program at The University of Queensland concluded that not enough foundational courses and too many specialist courses were being offered. Science is based on observation, hypothetico-deductive logic, experimentation, critical interpretation and reproducibility. Scientists are trained to be innovative, honest, precise, rigorous and critical. However, the training and attributes we look for in scientists are not always evident in science education programs. Many programs tend to focus on content, especially theory, and graduating scientists may lack generic and specific scientific skills. The BSc program was revised to provide a stronger fresher (first-year) focus on the enabling sciences (mathematics, chemistry, physics and biology), including quantitative skill development in keeping with modern technologies. The course SCIE1000 was developed to focus on fundamental quantitative skills within the context of science. Mathematics is considered to be the primary enabling science as it is fundamental to all other branches.

We have adopted the ziggurat (Mesopotamian terraced pyramid) as our course icon to highlight the functional inter-dependencies between mathematics, science, technology and society.



Maths underpins science and technology; therefore students need a working knowledge of a range of topics, including algebra, calculus, discrete maths, probability and modeling. Maths is considered to be the primary enabling science as it is fundamental to all other branches. Other enabling sciences are considered to be chemistry, physics and biology; as they study matter, energy and life.



arithmetic [1+1=2] *algebra [a+b=c]* *functions [y=ax]* *calculus [dy/dx]*
trigonometry [sin/cos/tan] *equations [y=ax+b]* *differentiation [dy/dx]* *integration [$\Sigma f(x)dx$]*
periodic functions [y=sinx] *polynomials [ax^n]* *logarithms [lny]* *exponential functions [$y=e^x$]*
probability [P(E)] *statistics [$x \pm SE$]* *matrices [][]* *algorithms* *models*

The only way for students to obtain a working knowledge of mathematics is to combine theory with practice (and then practice, practice, practice). Conventional maths education seems to always generate the common lament from students: "When and where are we ever going to use this?" You will do so in this course, and thereby see the many and varied applications of maths in science.

Teaching programs need to focus on more than just concepts, by including applications, problem-solving, modeling, etc. All students should be familiar with year 12 Maths B curriculum (as it is a prerequisite for enrolment in science at this university) [Bolger K, Boggs R, Faragher R, Belward J 2002 Mathematics for Queensland. Year 12B. Oxford University Press.]. Maths B was not just content-based, but also focused on skill development, modeling, problem-solving and applications. Remember the KAPS and MAPS terminology:

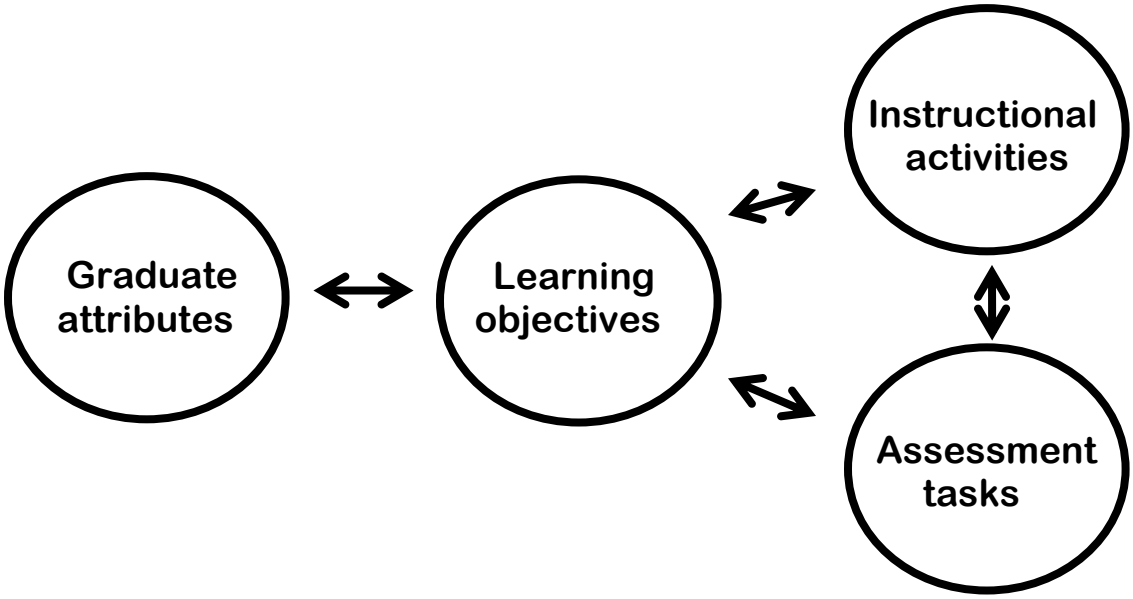
- KAPS = Knowledge and Procedures (definitions, rules, techniques)
- MAPS = Modeling and Problem Solving (representations, analyses, solutions)

Globally, improving maths education has been identified as a common goal within scientific communities. For example, the US National Research Council wants to "*Transform undergraduate education for future research biologists*" and has published a booklet entitled '*BIO2010*' [National Academies Press, 2003. ISBN 0-309-08535-7]. They consider the focus of educational programs in mathematics and computer science should be the acquisition and processing of data, quantitative analysis and display, modeling, prediction, program simulations, database access, search, retrieval, and *in silico* (computer) experiments.

The foregoing provides some of the background to the development, intent and extent of this course. Pertinent details of the course are available through the electronic course profile.

Course Profile [www.courses.uq.edu.au]

UQ has unilaterally adopted a higher education teaching and learning paradigm called the constructive alignment model, whereby desirable graduate attributes are articulated within specific course learning objectives, which are achieved through relevant instructional activities and assessment tasks. It is also known as the CIA model whereby Curriculum is linked to Instruction and Assessment.



This teaching and learning model ensures that all components are not isolated but inter-linked and integrated. The consequences of a fragmented course may not be immediately apparent, but imagine a lecturer not following the curriculum, someone setting the exams who had not attended the lectures, courses not adhering to market standards or employer expectations, etc. These situations can be avoided by making the process and content transparent to all stakeholders.

Graduate Attributes

Educational authorities recognize three domains of learning:

- psychomotor (about doing) involving skills (S)*;
- affective (about feeling) involving attitudes (A)*; and
- cognitive (about thinking) involving concepts (C)* and knowledge (K)*.

[*collectively known as the SACK model]

University courses must include more than content and technical procedures specific to any particular discipline. They must include generic skills applicable to vocation, employment, community and society. In the BSc program, UQ has articulated seventeen graduate attributes in five main categories, one category relevant to the field of study and four addressing general attributes. Staff cannot simply pay lip-service to these graduate attributes, but have to demonstrate where and how they are embedded in each course (matrices available through web-links in ECP).

BSc Graduate Attributes	
In-depth knowledge of field of study	knowledge, understanding, perspective
Effective communication	interaction, written & spoken communication, IT competency
Independence and creativity	work, learn, adapt, identify, create, innovate, solve
Critical judgement	define, analyze, critique, evaluate, reason, decide, reflect
Ethical and social understanding	responsibility, respect, appreciation, diversity

Learning Objectives

The broad aims of this course are to instil an appreciation of the quantitative skills and fundamental philosophies required for the practice of modern science, provide interdisciplinary contextual relevance, improve the mathematical and computational skills and communication skills of students and engage them in the "science community". Students will learn to:

- analyse the interdisciplinary nature of modern science;
- explain and demonstrate the importance of modelling in science;
- apply fundamental mathematical techniques to a range of scientific disciplines;
- design and write simple computer programs;
- interpret the philosophy of science and scientific thought;
- communicate science-based problems in a logical and appropriate style; and
- describe and discuss key issues in science, including social and ethical issues.

As you will notice, none of the learning objectives directly addresses specific content, neither scientific nor mathematical. For instance, they do not state that students will learn algebra, differential calculus, the laws of thermodynamics or the molecular structure of DNA. The learning objectives must be more than content-driven, they must include process. To this end, all the learning objectives begin with verbs (= doing words), such as analyse, explain, apply, design, etc.

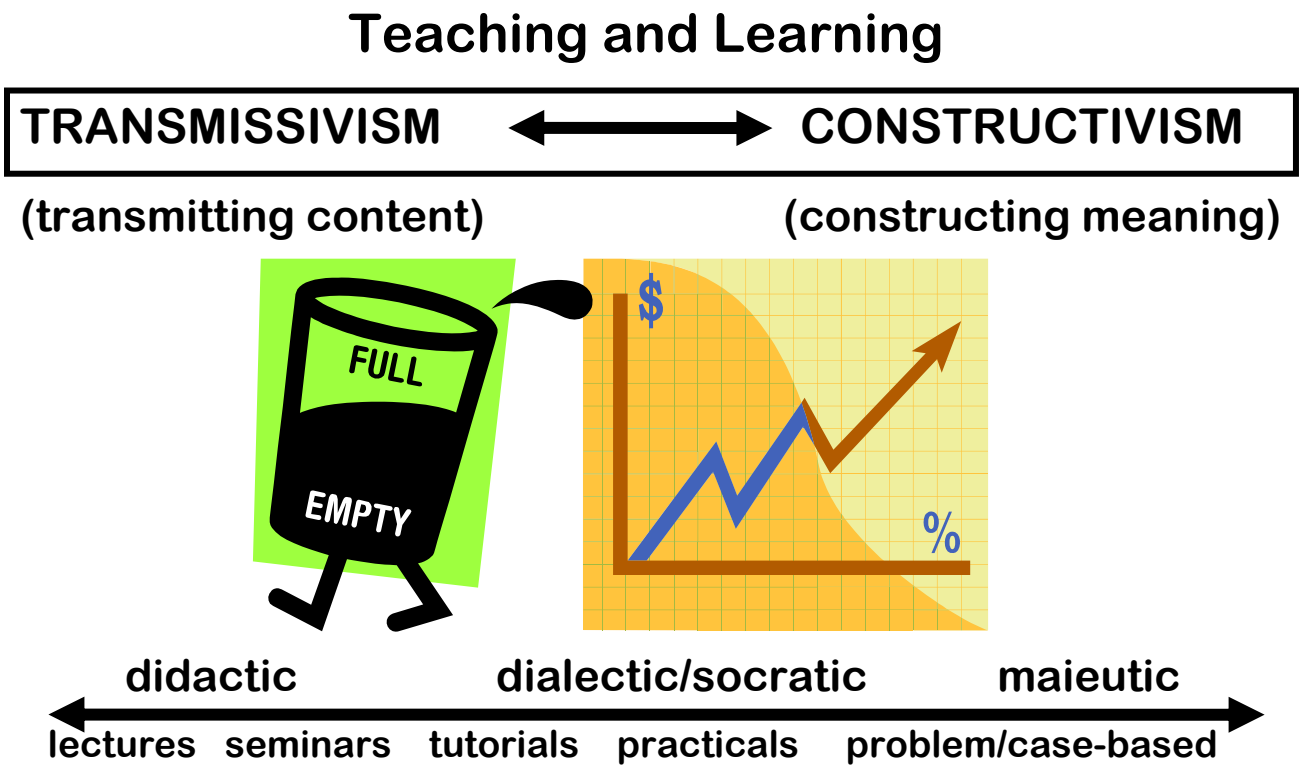
The scope, sequence and schedule of course work has been built around a logical progression of mathematical principles which have been contextualized with real relevant scientific topics from a diverse array of disciplines. Mathematical principles covered will include models, functions, equations, matrices, derivatives, optimization, numerical methods (for solving equations), exponentials, logarithms, integration, differential equations and discrete mathematics. Key scientific concepts will be drawn from chemical, physical, natural, earth and life science disciplines. Mathematically, the course can be likened to an *a la carte* (fixed menu) dinner, while scientifically, it is a smorgasbord. Hopefully, this framework will serve to reinforce our contention that maths underpins all the sciences.

Instructional Activities

There are many different modalities of education. All recognize the polarity of teaching (teacher-centred) and learning (student-centred) and attempt to reconcile these perspectives. Historically, teaching and learning occurred in small groups through question and discussion (the so-called Socratic method) and skilled trades were, and still are, taught through individual apprenticeships. With the Industrial Revolution came an educational revolution. Class sizes grew and methods of teaching large groups were introduced, notably in the form of lectures. While primary and

secondary schools have retained small class sizes to facilitate behaviour modification, tertiary institutions have embraced lecture formats as economical means for mass education. However, this does not mean lectures provide optimal learning opportunities for students.

Generally speaking, teaching and learning models form a continuum from what educational theorists call transmissivism (where knowledge is transmitted to students, such as in didactic lectures) to constructivism (where students construct meaning through dialogue). The former assumes the students' glass of knowledge can be filled by the teacher, while the latter recognizes that students already have some knowledge which must first be activated and validated before it can be built upon.



Students attend classes to learn, but what is learning? It is defined in many dictionaries simply as an increase in knowledge, but this covers many contexts, including acquisition, retention (memory), recognition (principles, ideas, concepts), cognition (making sense, understanding) and action (developing skills and competencies). Various strategies are used to avoid surface learning (atomistic in detail, isolated knowledge, limited understanding, quickly forgotten), foster deep learning (holistic in perspective, relational knowledge, good understanding, long-term retention) and develop and enhance student qualities (personal, social, philosophic, psychologic).

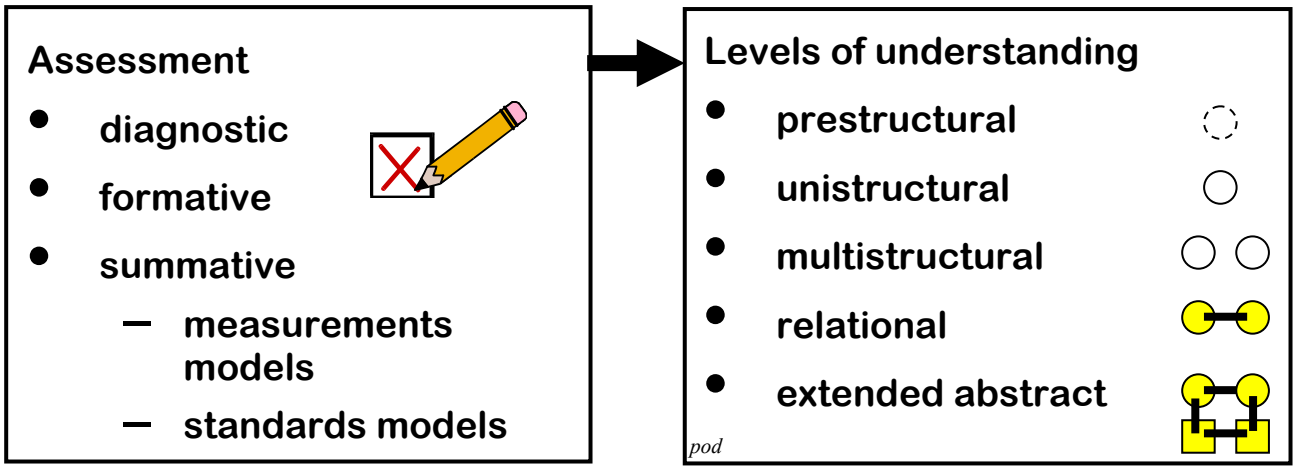
It is obvious that there is no single type of instructional activity that is universally suitable. Authorities recognize that a tailored teaching approach must be developed for each course, and even each cohort. Classes within individual courses are generally scheduled for five hours each week during semester, ranging from didactic lectures (traditional chalk-and-talk), dialectic lectorials (lectures with group discussions and activities), maieutic tutorials (small-group discussion groups) and interactive computer laboratories (individual and small-group exercises).

Timetables and schedules are available through SI Net [www.sinet.uq.edu.au] and class resources through BlackBoard [blackboard.elearning.uq.edu.au]. All classes will provide students with the opportunity to practice solving problems in the context of science, as befits contemporary workplace practices.

Assessment Tasks

While scientists concentrate on content and teachers on process, students focus on assessment. It has long been recognized that assessment drives learning. In the past, heavy emphasis has been placed on summative assessment tasks to measure learning rather that formative assessment to support learning. Assessment has traditionally been facilitated by 'measurement' models which rate individual performance against population normal distributions rather than by 'standards' models which rate performance against specific criteria. Courses should endeavour to 'assess for understanding'. This involves defining what we mean by 'understanding'.

Five hierarchical levels of understanding are recognized: pre-structural (don't get it); uni-structural (identify single elements); multi-structural (identify similar elements); relational (identify patterns); and extended abstract (generalize). Desirable learning outcomes should involve higher order understanding and assessment tools should evaluate cognitive, metacognitive and social competencies and affective dispositions. University courses should aim for at least multi-structural pattern recognition and relational thinking (typified by compare/contrast questions).



The emphasis of any course on quantitative skill development must be on the demonstration of those skills through problem-solving. Students should take every opportunity to work through examples, practice questions in their own time and practice questions under mock exam conditions - so that exam time will hold no surprises.

Like the Phoenician traders, examiners and students do not put all their goods on one ship; rather they have multiple modes of assessment spread throughout the semester. Small assignments are often due during semester and there is usually a centrally-controlled examination at the end of semester after swot vac (**s**tudy-**w**ith**o**ut-**t**eaching **vac**ation). The types of assignment and exam questions may include multiple choice questions (MCQs), short answer questions (SAQs) and extended writings (EW) in the form of mini-essays. Many questions will involve problem-solving and marks are allocated not only for final answers but also for process logic (working out pathways). Students should practice setting out their answers coherently. Survival depends on practice, practice, practice! This includes writing!

Scientific writing

Professional scientists are called upon to write 3 main types of documents: grant applications; scientific papers; and literature reviews. While the instructions given to authors by granting agencies, publishing houses and editors may differ, there are many common elements to these documents. All three are subject to peer review by independent referees to gauge integrity and quality, they generally adhere to a 'scientific' format, and they are mostly written in formal language (third person passive). Scientists conform to prescribed formats to get material published and they write for other scientists, rather than for the community at large. The growing demand to revise science communication has created jobs for science writers and knowledge brokers, third parties who are not science specialists but are trained communicators able to simplify and explain science to society.

Grant applications:

Governments make funds available each year for scientific research, often in applied priority areas. Various schemes have been developed for project grants, industry partnerships, exchange programs, cooperative research, institutional programs, etc. Securing funds is a fierce competitive process, with significant employment and career consequences. Individuals and organizations invest considerable resources in applying for grants but regrettably, the success rate is low (15-25%). It is, however, the only game in town, so every scientist plays. Grant applications require a catchy title (strong focus), a succinct summary (intelligible to laypersons), specific aims and objectives (logical and contextual), statements of expected outcomes (relevance and impact to society, industry, etc), an itemized budget (with justifications for expenditure) and finally, the actual project proposal (background, methodology, timetable). Applications are reviewed by independent referees, ranked according to specific criteria, and funded expediently (offers made until money runs out). Scientists are often judged on two criteria, grantsmanship (money-in) and authorship (papers-out).

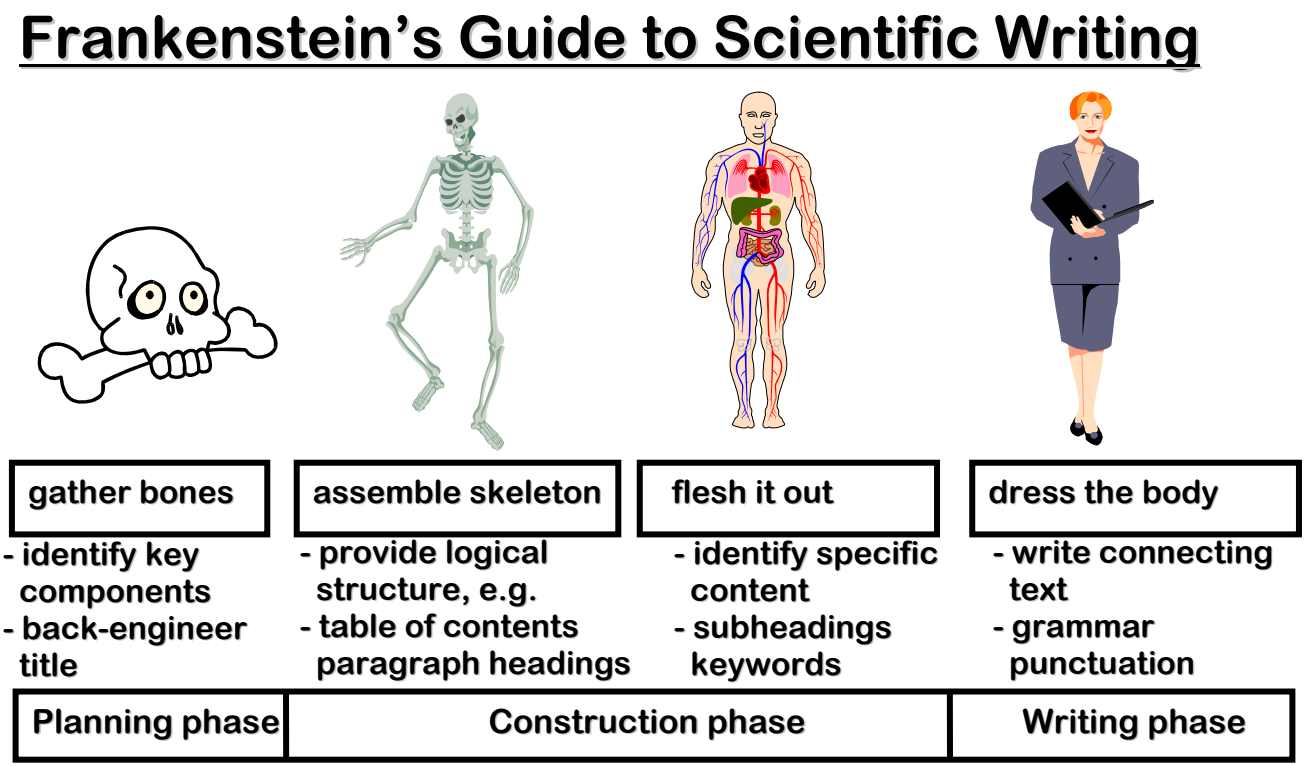
Scientific papers:

Libraries are filled with thousands of volumes of scientific journals which contain millions of scientific papers. Most journals are thematic and cater for specific disciplines and audiences. They vary widely in their distribution and availability, roughly half being parochial (regionally specific) and the remainder being international. Scientists submit manuscripts to the journal editors who then send them to independent referees for review. Articles may be rejected for many reasons (badly written, inexact science, falsehoods, not in context of journal, etc) or accepted for publication after major, minor or no revision. Few journals are free to the authors; most have page charges or reviewer fees. Even with the recent trend towards on-line editing and publishing, there are still fees payable to publishing houses. Scientists generally write two types of scientific papers: full research papers; or short contributions. Both follow an almost universal prescribed format: title (sufficiently descriptive in its own right), abstract (summary of whole paper), introduction (background and objectives), methodology (materials and methods used), results (observations and experimental findings), discussion (critical interpretation in context) and references (other works cited). Short communications are condensed versions of full papers. Scientists spend inordinate amounts of time crafting the abstract of their paper because they are reproduced in scientific databases and search engines, and are often the only thing other scientists read.

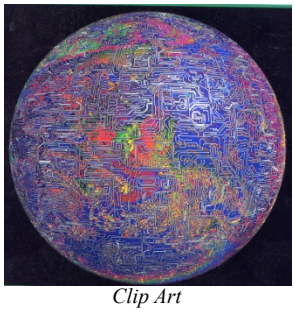
Literature reviews:

Due to the overwhelming number and diversity of publications available, many publishing houses now organize and publish literature reviews as journal articles, book chapters, and even whole books. Most reviews are commissioned articles whereby the authors have been invited to prepare and submit a treatise on a particular topic. The reviews can be critical analyses of contemporary issues, historical summaries of developing fields, personal opinion pieces, or collations of relevant resource materials (databases, catalogues, bibliographies, checklists, etc). Literature reviews provide an ideal starting point for novices to be introduced to a topic. Indeed, all higher degree students are required to write literature reviews to provide background and perspective to their research projects. Professional scientists use reviews to keep apace with recent developments and the authors achieve elevated status as recognized world authorities in the relevant field. Reviews follow various formats, but essentially they rationalize their existence (justify objectives), summarize previous studies (compare and contrast, pros and cons) and suggest future directions for research.

Scientific writing can be quite different from other forms of writing, but like all generic skills, it gets better with practice. Many scientists state that their major obstacles to writing are over-justification syndrome (trying to qualify everything), exception-syndrome (making generalizations but becoming bogged down with all the exceptions to the rule), forest-from-trees-syndrome (failing to properly identify scope and content of article), beginner's block (where to start) and common old writer's block (procrastinate - do something else instead). I liken the creative process of writing scientific papers to using Frankenstein's guide to building a body of knowledge: start with an assortment of bones (relevant items); construct a skeleton (logical framework); flesh it out (add substance); and dress it up (language).



Chapter 2.



Clip Art

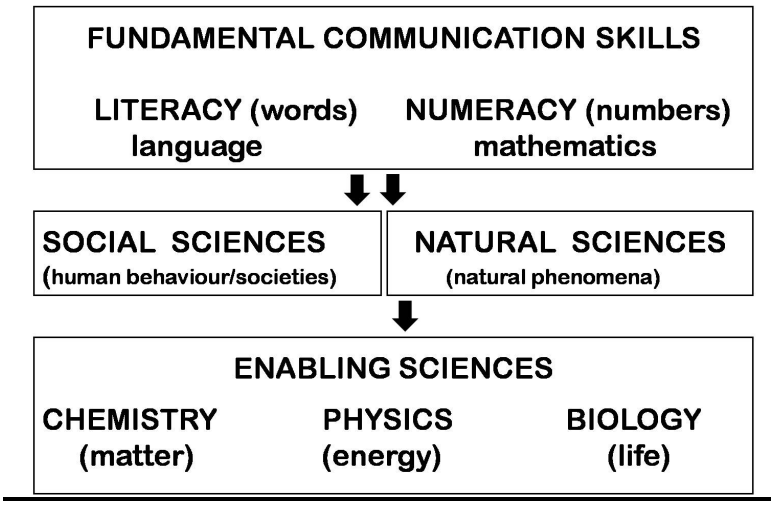
MAJOR CONCEPTS IN SCIENCE

Vocabulary list: science, chemistry, physics, biology, mathematics, energy, matter, life, numbers, formulae, units, models

What is Science?

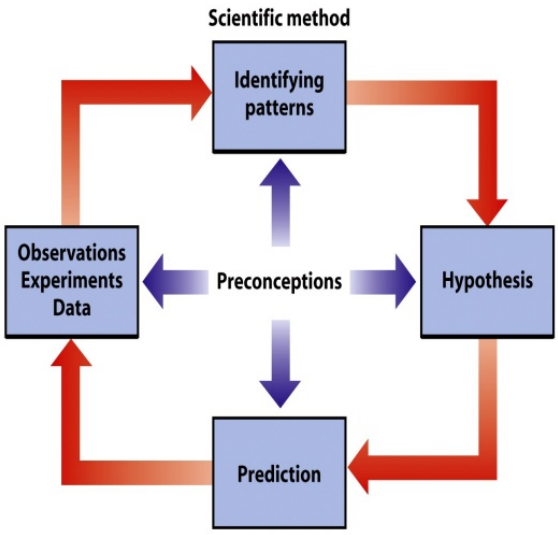
The word 'science' derives from the Latin '*scientia'*, meaning 'knowledge'. In an historical sense, it refers to any systematic knowledge or practice. The modern use of the term refers to a system of acquiring knowledge based on the scientific method, as well as to the organized body of knowledge gained through such research.

Fields of science are commonly classified into the natural sciences (which study natural phenomena, including biological life), and social sciences (which study human behaviour and societies). These groupings are empirical sciences, which means the knowledge must be based on observable phenomena and capable of being experimented for validity by other researchers working under the same conditions. Mathematics, which is sometimes classified within a third group of science called formal science, has both similarities and differences with the natural and social sciences. It is similar to empirical sciences in that it involves an objective, careful and systematic study of an area of knowledge; it is different because of its method of verifying its knowledge, using *a priori* rather than empirical methods. Major advances in formal science have often led to major advances in the physical and biological sciences. The formal sciences are essential in the formation of hypotheses, theories, and laws, both in discovering and describing how things work (natural sciences) and how people think and act (social sciences).



The history of science is marked by a chain of advances in technology and knowledge that have always complemented each other. Technological innovations bring about new discoveries and are bred by other discoveries which inspire new possibilities and approaches to longstanding science issues. Investing in science and technology is critical to ensuring prosperity and a high quality of life. Scientists are at the forefront of the development of scientific and technological innovations. Their primary objectives are to develop and conduct novel research that can be used to solve problems for private and public good. Experimental science is often differentiated from applied science, which is the application of scientific research to specific human needs, though the two are often interconnected.

Science must be considered to be content plus process. It is the collection of discovered knowledge as well as the process used to discover knowledge. It should therefore not be taught as a series of facts, but rather by explaining how material was discovered or developed over time, thereby providing an integrated and holistic appreciation as well as historical and contextual relevance. The process of observing the physical universe, framing experimental questions (hypotheses), analyzing and critically interpreting data, generating models, and making predictions is considered to constitute the scientific method, based on hypothetico-deductive logic. This method should not be regarded as a rigid template, but rather as a natural circular way of thinking, with no fixed starting point and no fixed end-point.



Trefil & Hazen, 2007; Fig. 1.2

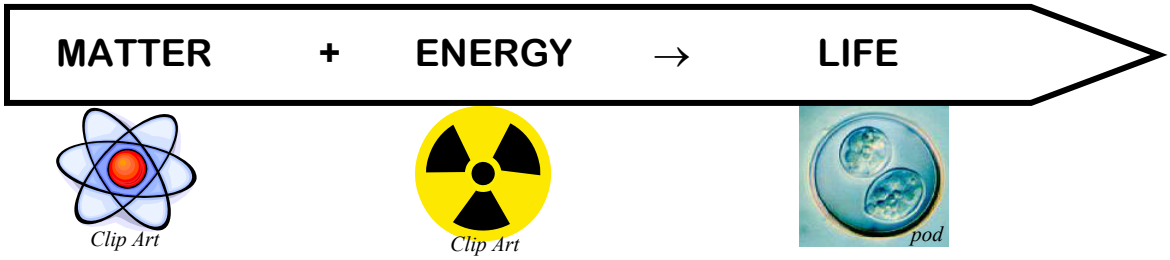
There are many principles and procedures associated with science but they do not constitute any universal doctrine. During the course component on the philosophy of science, you will reflect on many conceptions, misconceptions, theories, facts and myths which will challenge your critical reasoning skills. This is part of becoming as scientist. There are many websites espousing singular and general views on the epistemology, theory and history of science [e.g. http://amasci.com/miscon/myths10.html] [http://members.tripod.com/~touba/science.html]. While you should maintain an open mind and gather information, you should also display critical deductive, analytical and interpretative skills to avoid preconceptions based on misinformation becoming doctrine.

Of late, we have heard a lot about scientific ethics, particularly in the context of genetically-modified foods, and stem cell research. Many contentious issues continue to be debated - this is good because it indicates social awareness and responsibility. Ethics is an analytical and methodological inquiry into how moral judgements should be and are made. It is therefore prescriptive rather than descriptive: that is, it explores what we should do rather than what we can do. It seeks to identify moral concepts, rules, action-guides or principles of behaviour that provide a basis for a peaceful co-existence in society. Scientists have a personal and social responsibility to contribute and conform to legislation based on sound principles and high standards. Science has a lot to be proud of (such as massive improvements in health standards), but it also has a lot to answer for (such as industrial waste, pollution and greenhouse gas emissions)!

Can we trust science? It has been stated that “the truth of science lies in its reproducibility”, but what does that actually mean? Are scientific observations always accurate, do all experiments yield exactly the same results? They often contain inherent errors, mistakes, estimations, assumptions, uncontrolled variables, and many other factors which influence reproducibility. Scientists use a wide range of mathematical procedures to statistically analyze data, which effectively qualifies the term ‘reproducible’ to mean within 95% confidence limits, or 99% confidence limits, or better. During your second semester of study, you are required to complete the compulsory course STAT1201 “Analysis of Scientific Data” which is prerequisite to modern science. The current course SCIE1000 “Theory and Practice of Science” presages such analyses. It focuses on fundamental quantitative skills within the context of science.

Enabling Sciences

There are many different ways in which we could introduce fundamental and applied sciences. Most textbooks are dedicated to specific disciplines and therefore do not demonstrate how the sciences overlap and complement each other. They do not provide an interdisciplinary or multidisciplinary perspective or an integrated approach. The framework we have adopted covers the three fundamental enabling sciences of chemistry, physics and biology; the study of matter, energy and life.



Chemistry involves the study of matter - anything that takes up space and has mass. All matter consists of atoms and molecules which interact to produce many different compounds. Fundamental concepts address atomic and subatomic structure (periodic table), molecular structure (states, bonds, mixtures), reactions (types, energetics, equilibria, kinetics, dynamics), and analytical methods (mass spectrometry, magnetic resonance, diffraction, electron microscopy, separation).

atoms → molecules → reactions

subatomic *elements* *states* *solutions* *inorganic (metals, nonmetals)*
stoichiometry (equations) *redox* *organic (hydrocarbons)*
bonds *electrochemistry* *thermochemistry* *rates/equilibria*
nuclear chemistry *biochemistry*

Physics is the study of the physical universe, involving energy, matter, motion, time and space. It deals with celestial bodies, earthly objects, atomic and subatomic particles and various energy forms. Fundamental concepts address motion, dynamics and force laws (speed, velocity, acceleration, power, electromagnetism); conservation laws (energy, momentum, thermodynamics); wave laws (light, optics, X-rays, etc); and dynamical systems (fluids, networks).

energy → particles → planets → universe

optics *quantum mechanics* *classical mechanics* *astronomy*
electromagnestism *atomic physics* *force* *dynamics* *astrophysics*
waves *power* *motion*
relativity *thermodynamics* *momentum*

Biology is the study of life - the structure, function and inter-relationships of living organisms. Despite the extraordinary diversity of living organisms, they show remarkable unity at the molecular and cellular levels, reflecting their common ancestry. Fundamental concepts address molecular biology (biochemical building blocks), cellular biology (membranes, organelles), organismal biology (biodiversity, species richness) and environmental biology (communities, populations, ecosystems).

molecular biology → cellular biology → organismal biology → environmental biology

genetics/genomics *anatomy/histology* *zoology/botany*
biochemistry/proteomics *developmental biology/embryology*
microbiology *physiology/endocrinology* *ethology*
biotechnology *aquatic biology/limnology*
immunology *ecology* *environmental science*

These fields of study are not mutually exclusive, they exhibit manifold inter-relationships. Scientists require an integrated knowledge of matter, energy and life in order to understand and practice holistic science. For example, the study of living things requires knowledge of their chemical and physical composition and surroundings. A truly integrated approach to science (involving many different disciplines) is evident when considering some of the great fundamental and unifying themes in science:

- **Four fundamental forces hold everything together** - gravity, electromagnetism, the nuclear 'strong' force, and the decay 'weak' force (collectively called the unified field theory).
- **Energy is used to perform work** - it exists in many forms, it may be converted from one form to another, and it is conserved in closed systems
- **Chemicals are the building blocks of life** - hydro-carbons dominate on Earth (consistent with carbon-based life-forms on a water planet).
- **All life-forms have a unique genetic code** which undergoes replication (essential for inheritance) and transcription and translation (essential for protein synthesis and metabolism).
- **Cells are the basic units of life** – all living organisms have microscopic membrane-bound cells containing the genetic material and various organelles for energy transduction.
- **All living things evolve** – they change over time to better survive their environments through the processes of genetic recombination and natural selection.
- **Life-forms co-exist** - populations of organisms collect together in ecosystems where energy flows through while matter is recycled.

These scientific themes will be examined in detail in subsequent chapters, but for now we must establish that they all have common mathematic foundations.

NUMBERS

All knowledge is based on numeracy and literacy - numbers and words are fundamental to thinking, understanding, observing, recording, experimenting and communicating. Maths is all around you. You use it in everyday life, often without consciously thinking about it. How long did you sleep last night? What time did you arise? How much coffee did you drink for breakfast? etc. We use numbers and mathematical concepts all the time.

Look outside and consider the weather. If you were to describe it to someone-else, you would ultimately have to use mathematics. What meteorological parameters are you describing, what are their units, how are they measured, and what are the underlying scientific principles involved? Five examples may include:

- temperature measured in ºC using a thermometer (based on thermal expansion of fluid)
- air pressure measured in atmospheres using a barometer (based on gas pressure on diaphragm)
- rainfall measured in millilitres using a rain gauge (based on volume (height per surface area))
- humidity measured in % using a hygrometer (based on % water vapour saturation of air)
- wind measured in knots using a wind rose (based on velocity and direction of air)

Scientists quantify and enumerate data. We measure quantities and record numbers. We apply meaningful scales based on proven scientific principles. We assess maxima, minima, ranges, means, and variation. All measurements should also include a time dimension, because most things change over time. The weather changes not only diurnally (during daylight and night) but also daily, weekly, monthly, seasonally and yearly. Time itself is measured in minutes by a clock (based on the periodicity of the Earth's rotation).

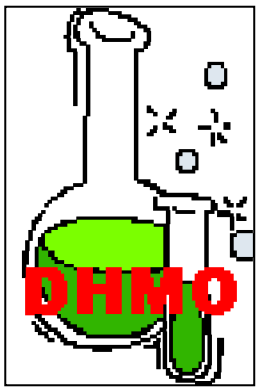
It is imperative that scientists establish a frame of reference to provide perspective to their observations and measurements. Consider the motion of a space shuttle. How many different movements can you identify?

There are four spatial movements: pitch, yaw, roll (rotation around x, y, z axes); and thrust (forward/backward) as well as temporal movement (over time). The extent to which these are perceived depends upon the perspective of the observer. Astronauts often comment that they only perceive motion when undergoing acceleration or deceleration, or when they can see an external object such as the Earth to give a frame of reference.

Perspective is so important for many reasons. The following two examples illustrate the dilemmas that can ensue when lacking perspective.

1. Google the chemical Dihydrogen Monoxide (DHMO)!

Websites describe it as "a colourless and odourless chemical compound, also referred to by some as Dihydrogen Oxide, Hydrogen Hydroxide, Hydronium Hydroxide, or simply Hydric acid. Its basis is the highly reactive hydroxyl radical, a species shown to mutate DNA, denature proteins, disrupt cell membranes, and chemically alter critical neurotransmitters. The atomic components of DHMO are found in a number of caustic, explosive and poisonous compounds such as Sulphuric Acid, Nitroglycerine and Ethyl Alcohol".



Some municipalities in America began enquiries into how to ban this dangerous substance in their locality, until someone pointed out that it was a web-prank and that DHMO was actually water!

2. Consider the monetary conundrum (propagated in school yards)!

Three businessmen go out for lunch. After the meal, the waiter brings the bill for $30. They each pay $10. However, the cashier notes a mistake as the bill should have only totalled $25. The cashier gives the waiter $5 in change to return to the businessmen. The waiter is not good at maths so he gives each of the businessmen $1 in change and he pockets the remaining $2. This means each businessman paid $9 (originally $10 but received $1 change). Now $9 times 3 equals $27. Add to this the $2 the waiter kept gives $29. What happened to the other $1?

If you follow the money, you can see that nothing actually goes missing ($25 in cash register, $3 with customers, $2 with waiter). It all depends on your perspective!

3. Magic numbers!

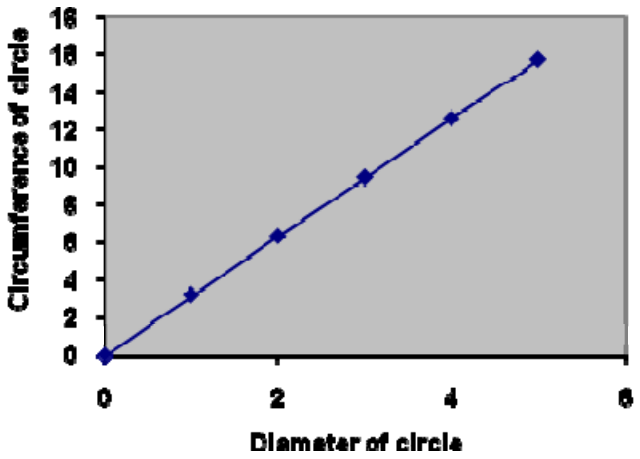
This simple mathematical trick can be used to discover any number a person thinks of, and find out their age as well. Ask a person to secretly pick a number of any size, then double it, then add five, then multiply the result by 50, then add 1760 (this key number changes yearly, in 2011 it will be 1761), then subtract the year in which they were born, and then tell you the final answer. The last two numerals in the answer tell you the person's age this year, and the remaining numerals tell you the number they first thought of. Try it with several people and with several numbers in different orders of magnitude. It works!

The 'magic' in the calculations can be unmasked by using some simple algebra. Let the original number be x. The calculation then becomes [50(2x+5) + 1760 – year of birth] which simplifies to [(100x) + (2010 – year of birth)]. All you are really doing is asking the person to multiply their number by 100 (so it will appear unchanged in the final answer), and subtract their year of birth from the current year (thereby giving their current age). No magic!

Formulae

Scientists may describe relationships between entities in several ways: using words, equations, symbols or graphical representations. They often use formulae (mathematical equations with numbers and symbols) to illustrate and define relationships between various components (variables, parameters, factors) of a system. For example, the relationship between the circumference and diameter of a circle can be expressed in the following ways:

Words: The circumference of a circle is equal to its diameter multiplied by the constant pi.
Equation: circumference = diameter x constant (pi)
Symbols: $C = d \times \pi$
Graphics:



However, no units are given for any of the variables. The formula should include units: e.g.

C (in metres) = d (in metres) x π (no units for this constant)

All equations must be balanced; that is, the units given either side of the equation must be equal. Making sure they do balance provides a checking mechanism for your calculations, especially for complex equations with many units involved. You can also derive the units for a variable by knowing those of its component parameters; that is, you can work out the units on one side of an equation by knowing that they must equal those of the other side of the equation. Consider the units for density. The density of an object is equal to its mass divided by its volume:

density = mass / volume $[\rho = m / v]$

where mass is given in grams, and volume is given in cm^3 (or ml for liquids)

The units will be g/cm^3 $(= g.\ cm^{-3})$ or g/ml $(= g.\ ml^{-1})$

There are various systems used for giving units to parameters. Scientists have adopted seven Standard International (SI) units as the basis for all measurements:

- length in metres (m)
- mass in kilograms (kg)
- amount in moles (mol)
- time in seconds (s)
- temperature in Kelvin (K)
- electric current in amperes (A, amp)
- luminous intensity in candela (cd)

While most of us may be familiar with many of these SI base units, few may know the precise definition for each unit:

BASE SI UNITS				
Physical quantity		Units		
name	symbol	name	symbol	definition
Length	l	metre	m	1 m = distance travelled by light in free space in 1 / 299,792,458 of 1 s (previously 1 / 10,000,000 of distance from equator to north pole)
Mass	m	kilogram	kg	1 kg = mass of solid cylinder of platinum-iridium alloy known as International Prototype Kilogram (~ mass of 1 L of water)
Amount of substance	n	mole	mol	1 mol = amount of substance containing as many elemental entities as there are atoms in 12 g of carbon-12
Time	t	second	s	1 s duration of 9,192,631,770 periods of radiation corresponding to transition between two hyperfine levels of the ground state of caesium-133 atom (originally defined in terms of period of revolution of Earth around sun)
Temperature	T	kelvin	K	based on thermodynamic temperature scale where absolute zero (the theoretical absence of all thermal energy) is zero K (1 K ≡ 1°C ≡ 1.8°F)
Electric current	I	ampere	A (amp)	1 A = amount of electric charge passing a point per unit time (6.242×10^8 electrons per second)
Luminous intensity	lv	candela	cd	1 cd = monochromatic radiation of frequency 540×10^{12} Hz with radiant intensity of 1/683 watt per steradian (~ light emitted by single candle)

Many units are given with prefixes which act as decimal multipliers. For example, units of length often include prefixes indicating multipliers to the power of 10; as:

kilometre	1 km = 10^3 m
metre	1 m
centimetre	1 cm = 10^{-2} m
millimetre	1 mm = 10^{-3} m
micrometer	1 μm = 10^{-6} m
nanometre	1 nm = 10^{-9} m

Remember: $10^0 = 1$
$10^1 = 10$
$10^2 = 100$
$10^{-1} = 1/10 = 0.1$
$10^{-2} = 1/10^2 = 1/100 = 0.01$
$10^{\frac{1}{2}} = 10^{0.5} = \sqrt{10}$

Number	Prefix	Symbol
10^{24}	yotta	Y
10^{21}	zetta	Z
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10^{1}	deca	da
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a
10^{-21}	zepto	z
10^{-24}	yocto	y

1 billion = 1 milliard = 1,000 million = 10^9 (giga-)

1 trillion = 1 million million = 10^{12} (tera-)

By convention, the Greek (and sometimes Latin) alphabets are used for scientific notation:

Greek alphabet name	Upper case symbol	Lower case symbol
Alpha	A	α
Beta	B	β
Gamma	Γ	γ
Delta	Δ	δ
Epsilon	E	ε
Zeta	Z	ζ
Eta	H	η
Theta	Θ	θ
Iota	I	ι
Kappa	K	κ
Lambda	Λ	λ
Mu	M	μ
Nu	N	ν
Xi	Ξ	ξ
Omicron	O	ο
Pi	Π	π
Rho	P	ρ
Sigma	Σ	σ
Tau	T	τ
Upsilon	Y	υ
Phi	Φ	φ
Chi	X	χ
Psi	Ψ	ψ
Omega	Ω	ω

Number	Greek prefix	Latin prefix
0.5	hemi	semi
1	mono	uni
1.5	-	sesqui
2	di	bi
3	tri	ter
4	tetra	quadri
5	penta	quinque
6	hexa	sexi
7	hepta	septi
8	octa	octo
9	ennea	nona
10	deca	deci
many	poly	multi

Scientific formulae used throughout this document include the following:

Density = mass (kg) / volume (m^3)	[$\rho = m / v$]
Molar mass = mass (g) / amount (mol)	[$M = m / n$]
Molarity = amount of solute (mol) / volume of solution (L)	[$c = n / v$]
Molality = amount of solute (mol) / mass of solvent (kg)	[$molal = n / m$]
Pressure x volume = amount x temperature x gas constant R	[$P\ V = n\ R\ T$]
Displacement = change in position (m)	[$\Delta S = S_{final} - S_{initial}$]
Velocity = change in displacement (m) / change in time (s)	[$v_{av} = \Delta S / \Delta t = (S_f - S_i) / (t_f - t_i)$]
Acceleration = change in velocity (m s^{-1}) / change in time (s)	[$a_{av} = \Delta v / \Delta t = (v_f - v_i) / (t_f - t_i)$]
Momentum = mass (kg) x velocity (m/s)	[$p = m\ v$]
Force (newtons, N) = mass (kg) x acceleration (m/s^2)	[$F = m\ a$]
Force of gravity = ($mass_1$ x $mass_2$ x constant G) / (distance)^2	[$F = (m_1 \text{ x } m_2 \text{ x } G) / d^2$]
Work (joules, J) = force (N) x distance (m)	[$W = F\ d$]
Power (watts, W) = work, energy (J) / time (s)	[$P = W / t$]
Energy (J) = power (W) x time (s)	[$E = P\ t$]
Potential energy (J) = mass (kg) x g (m s^{-2}) x height (m)	[$E_P = m\ g\ h$]
Kinetic energy (J) = ½ x mass (kg) x [velocity (m s^{-1})]^2	[$E_K = ½\ m\ v^2$]
Efficiency (%) = [($temp_{hot}$ - $temp_{cold}$) / $temp_{hot}$] x 100	[$\%E = \{(T_h - T_c) / T_h\}$ x100]
Electrostatic force (N) = ($charge_1$ x $charge_2$ x k) / (distance)^2	[$F = (c_1 \text{ x } c_2 \text{ x } k) / d^2$]
Voltage (volts) = current (amps) x resistance (ohms)	[$V = I\ R$]
Electric power (watts) = current (amps) x voltage (volts)	[$P = I\ V$]
Velocity of wave (m/s) = wavelength (m) x frequency (Hz)	[$v = \lambda f$]
Energy = mass x (speed of light)^2	[$E = mc^2$]
Degrees Fahrenheit = (1.8 x degrees Celsius) + 32	[°F = (1.8 x °C) +32]
Temperature (kelvin) = degrees Celsius + 273	[K = °C + 273]

The units for more complicated quantities in science and nature can always be derived in terms of the seven base SI units. For example, the units for velocity are m/s (= m.s^{-1}) which are derived from the SI base units for the integral components given in the definition of velocity as distance (m) over time (s). Likewise, the units for all quantities can be derived from the SI base units (examples shown in following table).

[By convention, units named after people (the unit for power is named after James Watt) are given as uppercase abbreviations (W) but are spelled out in lowercase (watt), with the exception of Celsius.]

DERIVED UNITS				
Physical quantity		Units		
name	symbol	name	symbol	definition
Area	A	hectare	ha	= 10^4 m^2
Volume	V	litre	L	= 10^3 mL = 10^3 cm^3 = 10^{-3} m^3
Gallon (Imperial, US)		gallon	gal	1 US gal = 0.83 imperial gal = 3.785 L
Density	ρ			kg m^{-3} (g cm^{-3})(g mL^{-1})
Ton			ton	1 short ton = 2,000 lb = 907 kg 1 long ton = 2,240 lb = 1,016 kg
Tonne			tonne	1 metric ton = 1 tonne = 1,000 kg
Molar mass	*M*			g mol^{-1}
Molarity	*c* (M)			mol L^{-1}
Molality				mol kg^{-1}
Displacement	S			m
Velocity	v			m s^{-1} 1 knot = 1 nautical mile/hr = 0.514 m s^{-1}
Acceleration	a			m s^{-2}
Frequency	*f*	hertz	Hz	s^{-1} (oscillations per second)
Force	F	newton	N	kg m s^{-2} (= J m^{-1})
Energy (Work)	*E*	joule	J	kg m^2 s^{-2} (= N m) (= W s) ***1 kWh = 3,600,000 J (3.6 MJ) 1 kWh = 86 kcal = 86 x 10^{-6} toe (ton of oil equivalent)
Power	P	watt	W	kg m^2 s^{-3} (= J s^{-1}) ***1 kW = 24 kWh/d 1 kWh/d ~ 40 W
Horsepower		horsepower	hp	1 hp = 0.75 kW (= 18 kWh/d)
calorie		calorie	cal	(2500 kcal = 10,000 kJ = 3 kWh) NOTE: 1 diet Calorie = 1,000 cal
Barrel (oil)		barrel of oil		1 barrel = 6.1 GJ = 1,700 kWh
Pressure	P	pascal	Pa	kg m^{-1} s^{-2} (= N m^{-2})
Pressure	P	bar	bar	= 10^5 Pa
Pressure	P	atmosphere	atm	= 101,325 Pa
Pressure	P	torr	torr	= 133.32 Pa (= 1 mm Hg)
Pressure	P	millimetres mercury	mm Hg	133.32 Pa (= 1 torr)
Electric charge	Q	coulomb	C	= charge on 6.3 x 10^{18} electrons (A s)
Electric potential difference	U	volt	V	kg m^2 s^{-3} A^{-1} (= J A^{-1} s^{-1})
Electric resistance	R	ohm	Ω	kg m^2 s^{-3} A^{-2} (= V A^{-1})
Electric conductance	G	siemen	S	kg^{-1} m^{-2} s^3 A^2 (= Ω^{-1})
Electric capacitance	C	farad	F	A^2 s^4 kg^{-1} m^{-2} (= A s V^{-1})
Magnetic flux	φ	weber	Wb	kg m^2 s^{-2} A^{-1} (= V s)
Inductance	L	henry	H	kg m^2 s^{-2} A^{-2} (= V A^{-1} s)
Magnetic flux density	B	tesla	T	kg s^{-2} A^{-1} (= V s m^{-2})

Despite the simplicity of the decimal system, other systems of units are still used throughout the world. Many countries still use British Standard units, such as miles, yards, feet, inches, pounds, ounces, degrees Fahrenheit, etc. Scientists are often called upon to convert units from one system to another. We use formulae to convert units, as illustrated by the following method, called the factor-label method (sometimes called dimensional analysis). Using this technique, two basic mathematic rules apply: first, any number divided by itself equals one, and second, any number multiplied by one remains unchanged.

To convert 6 metres into feet, we need to look up the appropriate conversion ratio for the units, in this case, 1 ft ~ 0.3 m (make sure the units are the correct ones).

Using the first mathematical rule, this conversion ratio can be written as either of two fractions:
1 (ft) / 0.3 (m) or 0.3 (m) / 1 (ft) (both fractions = 1 as any number divided by itself = 1)

Using the second rule, our measurement can be multiplied by one of these fractions to achieve the appropriate conversion (from m to ft):

6 (m) x 1 (ft)/0.3 (m) = 6 (ft) / 0.3 = 20 (ft)

[NOTE that the m units cancel out leaving ft. If the wrong fraction is used, the units do not cancel out but become hopelessly confused (thus acting as a check mechanism):

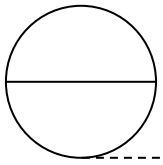
e.g. 6 (m) x 0.3 (m)/1 (ft) = 1.8 (m^2 / ft) (which is clearly wrong!!!)]

Scientists should always include units in their calculations and use them to not only check their working-out and but also to check their answers for sense, proportion and logic.

Special numbers

There are several special values that occur in mathematics which are used frequently in scientific calculations. Two of the most frequently used are $\boldsymbol{\pi}$ (*pi*) and ***e*** (Euler's number).

Pi (π) is incorporated into many scientific formulae, particularly where calculations involve geometric and/or trigonometric expressions. The Greek letter π (spelt *pi* in text) was adopted for the number from the Greek word for perimeter. *Pi* is a mathematic constant defined as the ratio of any circle's circumference to its diameter; that is,

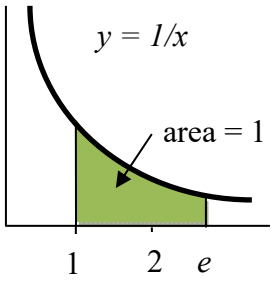


$$\boldsymbol{\pi = C/d}$$

This ratio is constant, regardless of the circle's size. If a circle has twice the diameter of another circle, it will also have twice the circumference. Alternatively, π can also be defined as the ratio of a circle's area (A) to the square of its radius (r); that is, $\boldsymbol{\pi = A/r^2}$. These definitions are based on Euclidean plane geometry, which can be problematic when *pi* occurs in areas of mathematics that do not involve geometry. Mathematicians therefore often prefer to define *pi* without reference to geometry, instead selecting one of its analytical properties as a definition. A common choice is to define π as twice the smallest positive x for which $cos(x) = 0$.

Pi is an irrational number, meaning that its value cannot be expressed exactly as a fraction involving integers. Consequently, its decimal representation never ends or repeats. It is also a transcendental number, meaning that no finite sequence of algebraic operations on integers (powers, roots, etc) can be equal to its value. The numerical value of π truncated to 5 decimal places = 3.14159

Euler's number (*e*) is a mathematical constant derived from work on logarithmic functions and differential calculus. It is the unique real number such that the area under the curve $y = 1/x$ over the interval $1 \leq x \leq e$ is exactly 1. Consequently, the area under the curve for the interval $1 \leq x \leq e^t$ is t. Also, the function e^x has the same value as the slope of the tangent line, for all values of x.



More generally, the only functions equal to their own derivatives are of the form Ce^x, where C is a constant. The function e^x so defined is called the exponential function, and its inverse is the natural logarithm, or logarithm to the base e. The number e is commonly defined as the base of the natural logarithm. It was called Euler's number and denoted by the letter e, possibly because it is the first letter of the word 'exponential'.

The number e is irrational; it is not a ratio of integers (root of a linear polynomial). It is also transcendental; it is not a root of any polynomial with integer coefficients. The numerical value of e truncated to 5 decimal places = 2.71828

Constants

Many mathematical constants have been found in science, and will be covered in relevant sections.

Fundamental Constants	Symbol	Value
Avogadro's number	N_A	6.022×10^{23} mol^{-1}
Faraday's constant	F	96,486 C mol^{-1}
Universal gravitational constant	G	6.67×10^{-11} m^3 s^{-2} kg^{-1} (N m^2 kg^{-2})
Universal electrostatic constant	K	9.0×10^9 N m^2 C^{-2}
Permittivity of vacuum (= $1/(4\pi K)$)	ε_0	8.85×10^{-12} C^2 N^{-1} m^{-2} (kg^{-1} m^{-3} s^4 A^2)
Planck's constant	h	6.626×10^{-34} J s
Speed of light (in vacuum)	c_0	3.00×10^8 m s^{-1}
Permeability of vacuum	μ_0	$4\pi \times 10^{-7}$ m s^{-2} A^{-2}
Rydberg constant	R_H	1.10×10^7 m^{-1}
Gas constant	R	1.99 cal K^{-1} mol^{-1} 82.1 cm^3 atm K^{-1} mol^{-1} 0.0821 L atm K^{-1} mol^{-1}
Boltzmann's constant	k	1.38×10^{-23} J K^{-1}
Mass of proton	m_p	1.673×10^{-27} kg (= 1836.1 m_e)
Mass of neutron	m_n	1.675×10^{-27} kg
Mass of electron	m_e	9.11×10^{-31} kg
Charge of electron	e	1.602×10^{-19} C (= charge of proton)
Bohr radius (of hydrogen atom)	a_0	5.29×10^{-11} m

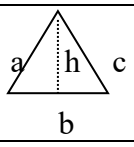
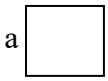
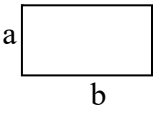
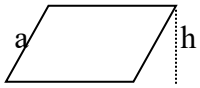
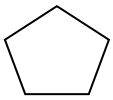
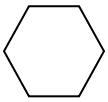
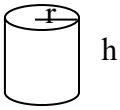
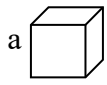
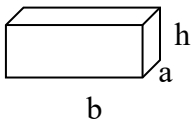
ASTRONOMICAL CONSTANTS	Symbol	Value
Astronomical unit	AU	1.496×10^{11} m
Hubble's constant	H	~ 20 km / s / Mly
Light year	ly	9.46×10^{15} m (= 6.324×10^{4} AU)
Mass of Sun	M_{sun}	1.99×10^{30} kg
Mass of Mercury	$M_{mercury}$	3.30×10^{23} kg
Mass of Venus	M_{venus}	4.87×10^{24} kg
Mass of Earth	M_{earth}	5.97×10^{24} kg
Mass of Moon	M_{moon}	7.35×10^{22} kg
Mass of Mars	M_{mars}	6.42×10^{23} kg
Mass of Jupiter	$M_{jupiter}$	1.90×10^{27} kg
Mass of Saturn	M_{saturn}	5.68×10^{26} kg
Mass of Uranus	M_{uranus}	8.66×10^{26} kg
Mass of Neptune	$M_{neptune}$	1.03×10^{26} kg
Mass of Pluto	M_{pluto}	1.10×10^{22} kg
Radius of Sun	R_{sun}	6.96×10^{5} km
Radius of Mercury	$R_{mercury}$	2.44×10^{3} km
Radius of Venus	R_{venus}	6.05×10^{3} km
Radius of Earth	R_{earth}	6.38×10^{3} km
Radius of Moon	R_{moon}	1.74×10^{3} km
Radius of Mars	R_{mars}	3.40×10^{3} km
Radius of Jupiter	$R_{jupiter}$	7.15×10^{4} km
Radius of Saturn	R_{saturn}	6.03×10^{4} km
Radius of Uranus	R_{uranus}	2.55×10^{4} km
Radius of Neptune	$R_{neptune}$	2.53×10^{4} km
Radius of Pluto	R_{pluto}	1.14×10^{3} km
Orbital radius of Earth (around Sun)		1.5×10^{8} km
Orbital radius of Moon (around Earth)		3.8×10^{5} km
Acceleration of gravity on Earth	g	9.81 m s^{-2}
Acceleration of gravity on Moon		1.62 m s^{-2}
Area of Earth's surface		5×10^{8} km^2

MATHEMATICS

The following information is provided as a short-hand refresher course for fundamental mathematical operations which are presumed to be familiar to the reader.

Order of operations = BODMAS
- Brackets
- Operations (exponents, roots, logarithms, trigonometric functions)
- Division
- Multiplication
- Addition
- Subtraction

Geometry (perimeter, area, volume)

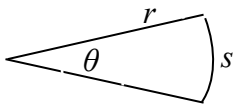
Type	Shape	1-dimension perimeter (P) circumference (C)	2-dimensions area (A)	3-dimensions volume (V)
triangle		P = a+b+c	A = ½ bh	-
square		P = 4a	$A = a^2$	-
Rectangle		P = 2(a+b)	A = ab	-
Parallelogram		P = 2(a+b)	A = bh	-
Pentagon		P = 5a	$A = 1.72a^2$	-
Hexagon		P = 6a	$A = 2.6a^2$	-
Circle		C = 2πr	$A = \pi r^2$	-
Sphere		-	$A = 4\pi r^2$	$V = {}^4/_3\ \pi r^3$
Cylinder		-	$A = 2\pi rh$	$V = \pi r^2 h$
Cube		-	$A = 6a^2$	$V = a^3$
Rectangular block		-	A = 2(ab+bh+ah)	V = abh

Trigonometry

arc length and angle

angle θ in radians = s / r

2π radians = 360° (1 radian ~ 57.3°)



right triangle

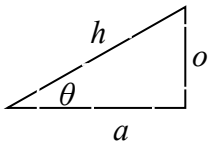
Pythagorean theorem $a^2 + o^2 = h^2$

$\sin \theta = o/h$ (opposite / hypotenuse) $\Rightarrow \theta = \sin^{-1} (o/h)$

$\cos \theta = a/h$ (adjacent / hypotenuse) $\Rightarrow \theta = \cos^{-1} (a/h)$

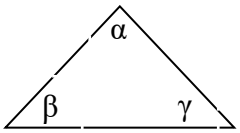
$\tan \theta = o/a$ (opposite / adjacent) $\Rightarrow \theta = \tan^{-1} (o/a)$

[SOH CAH TOA = Some Old Hippies Can't Always Hide Their Old Age]



general triangle

$\alpha + \beta + \gamma = 180° = \pi$ radians



identities

$\tan \alpha = \sin \alpha / \cos \alpha$

$\sin (-\alpha) = -\sin \alpha$

$\cos (-\alpha) = \cos \alpha$

$\sin (\alpha + \beta) = \sin \alpha \cos \beta + \sin \beta \cos \alpha$

$\cos (\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$

$\sin (2\alpha) = 2 \sin \alpha \cos \alpha$

$\cos (2\alpha) = \cos^2 \alpha - \sin^2 \alpha$

Algebra

fractions

$(a/b)\,(c/d) = ac / bd$

$(a/b) / (c/d) = ad / bc$

$1 / (1 / a) = a$

exponents

$a^0 = 1$

$a^1 = a$

$a^m\, a^n = a^{m+n}$

$a^m \div a^n = a^{m-n}$

$(a^m)^n = a^{mn}$

$a^m = 1/a^{-m}$

$a^{-m} = 1/a^m$

logarithms

if $y = 10^x$, then $x = \log_{10} y$

if $y = e^x$, then $x = \ln y$

base of natural logarithm $\ln x = 2.3026 \log x$

$\log (ab) = \log (a) + \log (b)$

$\log (a/b) = \log (a) - \log (b)$

$\log (a^n) = n \log (a)$

$\ln 1 = 0$

$e^0 = 1$

$\ln e^x = x$

$e^a\, e^b = e^{a+b}$

Functions

linear	$y = mx + c$	gradient = m = rise / run = $\Delta y / \Delta x$
quadratic	$y = ax^2 + bx + c$ there are 2 solutions to $y = 0$;	$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
periodic	$y = a + b.\sin(c.x)$	a = vertical shift b = amplitude c = period
power	$y = ax^p$	note that logarithmic transformation yields linear relationship $\log y = \log a + p \log x$
exponential	$y = ce^{kx}$	exponential growth when k is +ve exponential decay when k is -ve
surge	$y = ax^p e^{-kx}$	combination of power and exponential fns

discrete change (step-wise)

- unconstrained (unlimited)
  - geometric — $G_{i+1} = (1+r)\, G_i$
  - stage-classified [matrices] — $P_{i+1} = T\, x\, P_i$
- constrained (carrying capacity)
  - logistic map — $L_{i+1} = L_i + rL_i\, [1 - (L_i/K)]$

Calculus

Derivatives

if $y = x^n$, then $y' = nx^{n-1}$
if $y = \sin kx$, then $y' = k \cos kx$
if $y = \cos kx$, then $y' = -k \sin kx$
if $y = e^{kx}$, then $y' = ke^{kx}$

Product rule: $(uv)' = u'.v + u.v'$

Quotient rule: $\left(\frac{u}{v}\right)' = \frac{u'.v - u.v'}{v^2}$

Chain rule: $\frac{dy}{dx} = \frac{dy}{du} \bullet \frac{du}{dx}$

When estimating a solution to $f(t) = 0$,

a single step of Newton's method involves the calculation:

$$t_{i+1} = t_i - \frac{f(t_i)}{f'(t_i)}$$

When estimating a solution to a differential equation $y'(t) = \ldots$,

a single step of Euler's method involves the calculation:

$$y_{new} = y_{old} + (y' \bullet h) \text{ where } h = \text{step size}$$

Integrals

$$\int x(dx) = \frac{1}{2}x^2 + C \qquad \text{c = constant}$$

$$\int \frac{1}{x}(dx) = \ln x + C$$

$$\int x^n(dx) = \frac{x^{n+1}}{n+1} + C \qquad n \neq -1$$

$$\int e^x(dx) = e^x + C$$

$$\int \sin x(dx) = -\cos x + C$$

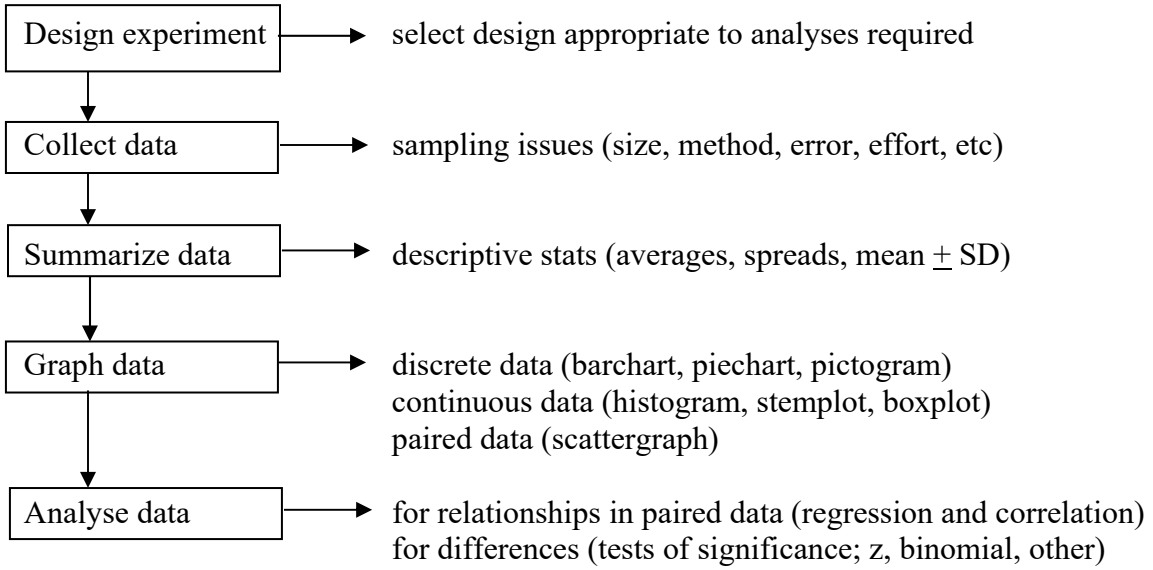
$$\int \cos x(dx) = \sin x + C$$

Common functions used in calculus

Differentiation		Integration	
Function/relationship	symbols	Function/relationship	symbol
displacement, $s(t)$ → velocity, $v(t)$	$v(t) = \frac{ds}{dt}$	velocity, $v(t)$ → displacement, $s(t)$	$s = \int v(t)dt$
velocity, $v(t)$ → acceleration, $a(t)$	$a(t) = \frac{dv}{dt}$	acceleration, $a(t)$ → velocity, $v(t)$	$v = \int a(t)dt$
mass, $m(x)$ → density, $\rho(x)$	$\rho(x) = \frac{dm}{dx}$	density, $\rho(x)$ → mass, $m(x)$	$m = \int \rho(x)dx$
population, $P(t)$ → growth, $r(t)$	$r(t) = \frac{dP}{dt}$	growth, $r(t)$ → population, $P(t)$	$P = \int r(t)dt$

Statistics

Terminology	Definition
Hypothesis testing	inferential procedure that uses data to evaluate the credibility of an hypothesis
Null hypothesis (H_0)	hypothesis stating that an independent variable (x) has no effect on the dependent variable (y)
Alternative hypothesis (H_1)	hypothesis stating that an independent variable (x) has an effect on the dependent variable (y)
Independent variable (x)	variable that you control, or choose (plotted on x-axis of graphs)
Dependent variable (y)	variable that you are measuring (plotted on y-axis of graphs)
Type I error	rejection of null hypothesis when it is true
Type II error	failure to reject null hypothesis when it is false
Variance	mean of sum of squared amounts by which values deviate from the mean value
Normal distribution	bell-shaped curve that represents distribution of values of variable about the mean
One sample t-test	procedure tests whether mean of a single variable differs from a specified constant
Independent samples t-test	procedure that compares means for a single variable between two independent sample groups
Paired samples t-test	procedure that compares the means of two variables for a single group
Chi-squared test (χ^2)	procedure to test a hypothesis about how well a sample population fits the population proportions specified by the null hypothesis
Regression	procedure for finding the best line of fit for the relationship between two variables such that $y = ax + b$
Analysis of variance (ANOVA)	procedure that produces a one-way analysis for a quantitative dependent variable by a single independent variable



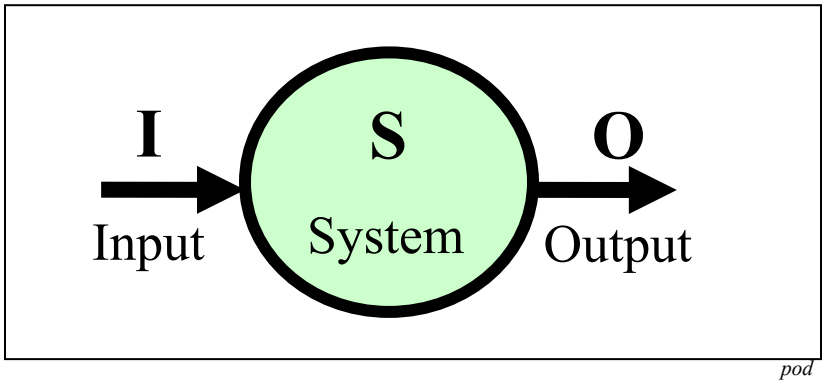
Analyses to fit Y by X		X	
		Continuous data (numeric)	Nominal (discrete) Ordinal (categorical)
Y	Continuous (numeric)	Scatterplot Regression lines	One way ANOVA, Means
	Nominal (discrete) Ordinal (categorical	Logistic regression	Contingency table

Models

We frequently use models in science to represent complex systems. They are used to mimic natural systems and construct virtual (theoretical) systems. They facilitate the compartmentalization of information and parameters into discrete entities and allow scientists to contemplate transition and equilibrium states. There are three primary technical uses of models in science:

- to control (to constrain or manipulate a system to produce a desirable condition)
- to understand (to examine and explain a system or a theory)
- to predict (to forecast the future or some state that is currently unknown)

Consider the simple ISO model:



pod

Type of problem	Given	To find	Uses of model
Instrumentation	S and O	I	CONTROL
Synthesis	I and O	S	UNDERSTAND
Analysis	I and S	O	PREDICT

Haefner, 2005

With the advent of computers, models have become increasingly sophisticated and they can incorporate hundreds of variables and perform billions of computations. Despite their growing complexity, models are still purpose-built - they have been designed to answer specific questions about I, S or O. Throughout this course, we will consider various models applicable to matter, energy and life; models developed for global warming, carbon trading, carbon dating, exponential population growth, logistic growth, enzyme activity, drug clearance, and disease prediction.

Chapter 3.

MATTER AND MOLECULES


Clip Art

Vocabulary list: chemistry, solids, liquids, gases, density, mass, volume, atoms, molecules, atomic number, atomic mass, Avogadro's number, mole, molar mass, concentration, molarity, molality, Ideal Gas Equation, Combined Gas Laws

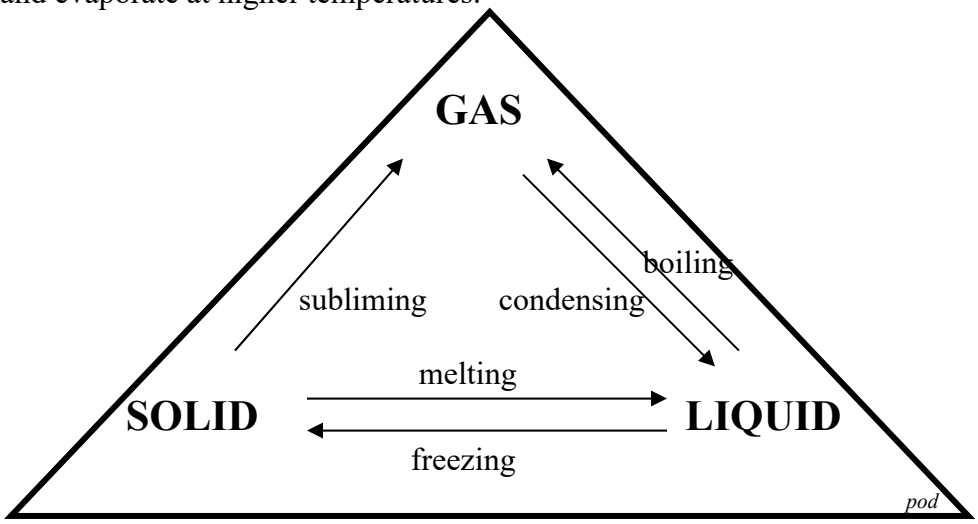
Chemistry

Chemistry is the science concerned with the study of matter - anything that takes up space and has mass. All matter consists of atoms and molecules which interact in many different ways producing many different compounds. Several types of chemistry are recognized (organic, physical, analytical, biological) but they all involve the same fundamental concepts:

- atomic and subatomic structure (periodic table)
- molecular structure (states, bonds, mixtures)
- reactions (types, energetics, equilibria, kinetics, dynamics)
- analytical methods (mass spec, NMR, diffraction, EM, separation)

States of Matter

Matter can exist in three different states: solids, liquids and gases. These states have markedly different characteristics: solids are generally denser than liquids which in turn are denser than gases; solids have a fixed shape while liquids and gases do not; and solids and liquids have a fixed volume while gases do not. Changes in state are temperature dependent; substances solidify at lower temperatures and evaporate at higher temperatures.



The density (ρ) of an object is proportional to its mass (m) and inversely proportional to its volume (v); according to the formula:

density (g/mL) = mass (g) / volume (mL) [$\boldsymbol{\rho = m / v}$]

Solids are denser than liquids and gases because their mass is compressed into a smaller volume; that is, their atoms and molecules are packed closer together.

Moles

All matter is composed of atoms which have a central nucleus (containing protons and neutrons) surrounded by electrons. The atoms of every element have a characteristic number of protons, given as the atomic number (superscript in periodic table). Every atom also has mass, measured in atomic mass units (amu) (subscript in periodic table).

	IA (1)	IIA (2)	IIIA (3)	IVA (4)	VA (5)	VIA (6)	VIIA (7)	VIIIA (8)
1	1 **H** 1.008							2 **He** 4.003
2	3 **Li** 6.941	4 **Be** 9.012	5 **B** 10.81	6 **C** 12.01	7 **N** 14.01	8 **O** 16.00	9 **F** 19.00	10 **Ne** 20.18
3	11 **Na** 22.99	12 **Mg** 24.31	13 **Al** 26.98	14 **Si** 28.09	15 **P** 30.97	16 **S** 32.07	17 **Cl** 35.45	18 **Ar** 39.95
4	19 **K** 39.10	20 **Ca** 40.08	31 **Ga** 69.72	32 **Ge** 72.64	33 **As** 74.92	34 **Se** 78.96	35 **Br** 79.90	36 **Kr** 83.80

atomic number

chemical symbol

atomic mass

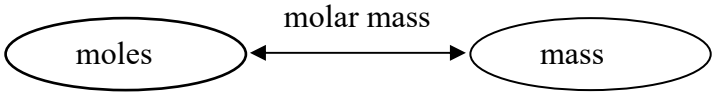
pod

Atomic mass units have been scaled against the carbon-12 (^{12}C) atom ('arbitrarily' selected as a common solid stable element on Earth) whereby 1 amu equals $^1/_{12}$ of the mass of a carbon-12 atom, that is, 1 amu = 1.66×10^{-24} g. When this definition is reversed, we obtain 1 g = 6.022×10^{23} amu.
As 1 amu = $^1/_{12}$ mass of a carbon-12 atom, the number of atoms in 1 g of carbon-12 = $^1/_{12} \times (6.022 \times 10^{23})$ atoms. This means 12 g of carbon-12 contains 6.022×10^{23} atoms.

The SI unit for the amount of a substance is the mole (mol) which is defined as the amount of substance that contains the same number of specified entities as there are atoms in 12g of carbon-12. This is called Avogadro's number [$N_A = 6.022 \times 10^{23}$ mol^{-1}].
It denotes that 1 mole of any substance contains 6.022×10^{23} entities.
e.g. one mole of lead (Pb) contains 6.022×10^{23} atoms of Pb
one mole of carbon dioxide (CO_2) contains 6.022×10^{23} molecules of CO_2
one mole of sucrose ($C_{12}H_{22}O_{11}$) contains 6.022×10^{23} molecules of $C_{12}H_{22}O_{11}$

Obviously, the atomic masses of all these substances differ, but 1 mole of each contains the same number of entities (atoms or molecules). While this situation appears cumbersome and complicated, it actually allows very simple calculations to be performed without having to revert to complex base units.

Moles can be converted to mass, and vice versa, using their relationship with molar mass (also called molecular mass, formula mass, and sometimes molecular weight):

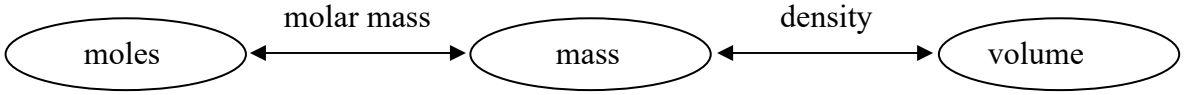


The molar mass of a substance is derived from its chemical formula and it equals the combined mass of all the constituent atoms. For example, the molecular formula for ethylene is C_2H_4 , comprising 2 carbon and 4 hydrogen atoms. According to the periodic table, carbon atoms have an atomic mass of 12.01 g/mol and hydrogen atoms have an atomic mass of 1.008 g/mol, so the molar mass of ethylene = (2 x 12.01) + (4 x 1.008) = 28.05 g/mol. This means 1 mole of ethylene = 28.05 g.

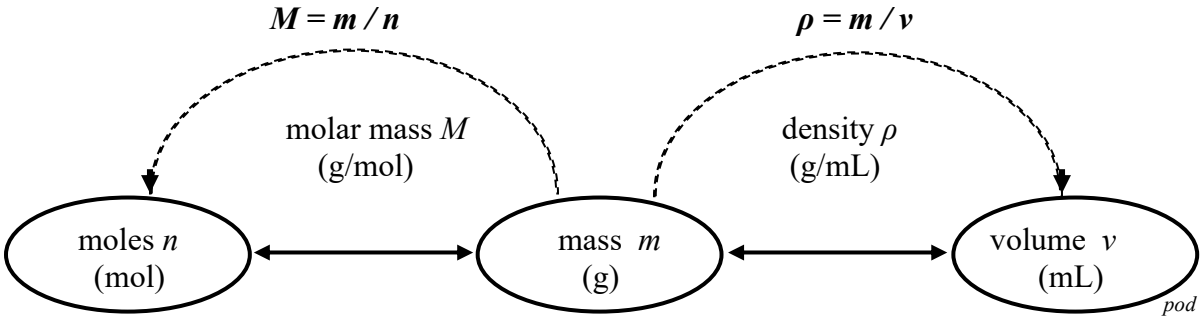
The relationship between moles (n), mass (m) and molar mass (M) is given by the formula:

molar mass (g/mol) = mass (g) / amount (mol) [$\boldsymbol{M = m / n}$]

We can add another dimension to this chain by incorporating density (= mass / volume)



We can summarize the inter-connected chain of relationships as follows:



This chain contains 5 variables which are inter-dependent. By knowing 2 or 3 variables, the remainder can be calculated using algebraic variations of the two formulae. Students require a working knowledge of the mathematical principles involved in solving problems using these equations. Most mathematical questions involve problem-solving skills and are not simply a series of arithmetic calculations. Students should begin by identifying the variables that are given, as well as the unknown variable that is required. They should then decide on the formula(e) linking the variables and show their inter-relationships. The formula(e) can then be reorganized to yield an equation for the unknown variable. Students should conduct dimensional analyses by including units for all variables (converting them to standard units as necessary). Answers should be checked for sense (do the units resolve properly?) and proportion (is the answer too large or too small?).

Question: At 25ºC, the density of water is 1.00 g/mL. At this temperature, what volume will be occupied by 3.89 mol of water? (Note that molar mass of H_2O = 18.02 g/mol)
Answer: We are given density = 1 g/mL amount = 3.89 mol molar mass = 18.02 g/mol temperature = 25ºC (but remains unchanged and not used in calculations) First, use molar mass formula to calculate mass $m = M\,n$ = 18.02 g/mol x 3.89 mol = 70.02 g Next, use density formula to calculate volume $v = m / \rho$ = 70.02 g / 1 g/mL = 70.02 mL

Concentrations

Molarity (or molar concentration) is the unit of concentration used by scientists for aqueous solutions. It denotes the amount of substance in a particular volume of solution, expressed as the number of moles of solute per litre of solution (mol/L).

$$\text{molarity (mol/L)} = \frac{\text{amount of solute (mol)}}{\text{volume of solution (L)}} \qquad [\ \boldsymbol{c = n / v}\ ]$$

For example, a one-molar (1M) solution of sucrose ($C_{12}H_{22}O_{11}$ molar mass = 342.3 g/mol) consists of 342.3 grams of sucrose dissolved in enough water to bring the final total volume to 1 litre.
[Remember we can convert moles to mass by knowing the molar mass of a compound, that is, we can substitute $n = m / M$ in the molarity formula to give $c = m / (M\ v)$]. It is often convenient to think of molarity in terms of grams per litre (g/L) where the molar mass of a chemical (g/mol) gives the number of grams that must be present in 1 litre of solution to give a 1M solution.

Question: How many grams of sucrose are required to make 10 litres of a 0.25M solution? (the molar mass of sucrose = 342.3 g/mol)
Answer: We are given the molar mass (M) of sucrose and told the molarity required (c) and final volume (v). We need to calculate the mass (m) of sucrose required. First use molarity formula to calculate amount $n = c\ v$ = 0.25 mol/L x 10 L = 2.5 mol Next, use molar mass formula to calculate mass $m = M\ n$ = 342.3 g/mol x 2.5 mol = 855.75 g [Alternatively, we can consider that a 1M solution has 342.3 g in 1 L (equivalent to 3,423 g in 10 L) and we require a 0.25M solution which is 3,423 x 0.25 = 855.75 g]

There are problems dealing with the molarity of many solutions because their volume changes with temperature. As solutions heat up, their volume will increase so their molarity will decrease. To overcome this thermal variation, another method has been developed for expressing solution composition. Molality (or molal concentration) has been defined as the amount of solute (mol) per kilogram (kg) of solvent. It is temperature independent as mass does not depend on temperature.

$$\text{molality (mol/kg)} = \frac{\text{amount of solute (mol)}}{\text{mass of solvent (kg)}}$$

Question: What mass of sodium chloride would have to be dissolved in one kilogram of water to prepare a solution with a molality of 0.15 mol/kg? (molar mass of salt = 58.44 g/mol)
Answer: We are given the molar mass (M) of salt and told the molality required and final mass of water (m). We need to calculate the mass (m) of salt required. First use molality formula to calculate amount $n = molality$ x $m_{(solvent)}$ = 0.15 mol/kg x 1 kg = 0.15 mol Next, use molar mass formula to calculate mass $m_{(solute)} = M\ n$ = 58.44 g/mol x 0.15 mol = 8.76 g [Alternatively, we can consider that we require 1 kg of solvent to contain 0.15 mol of salt, so we use the molar mass formula to convert 0.15 mol to mass $m = nM$ = 0.150 mol x 58.44 g/mol = 8.76 g]

Volume, Temperature and Pressure

The behaviour of gases is even more complicated because gases expand to occupy all the space of their containers. This implies that all gas atoms and molecules move freely around within their containers but they exert a pressure when they collide with the walls of their container. The pressure (*P*) exerted by a gas is dependent on the amount of gas present (*n*), the volume in which it is contained (*V*) and its temperature (*T*). The volume of a gas is inversely proportional to its pressure (Boyle's Law) but directly proportional to its temperature (Charles' Law) and amount present (Avogadro's Law). These features can be combined in one equation, the Ideal Gas Equation (with stipulated units):

pressure (atm) x volume (L) = amount (mol) x temperature (K) x constant R (L atm / mol K)

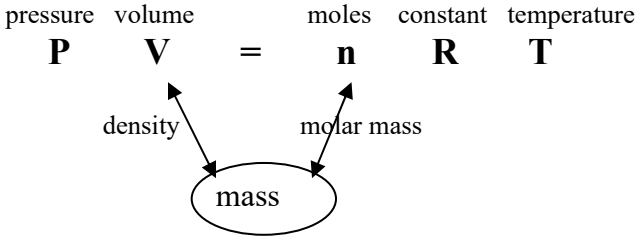
$$P\,V = n\,R\,T$$

where *P* = pressure (in atmospheres, atm),
V = volume (in litres, L),
n = amount (in moles, mol),
T = temperature (in Kelvin, K), and
R = a constant BUT with specific units (R = 0.0821 L atm / mol K).

This situation is often confused when other units are used to quantify parameters. They can either be converted back to the units given above, or the value of *R* (and its units) can be converted accordingly. Other common measures include:

pressure: 1 atm = 760 mm Hg = 760 torr = 101,325 Pa = 101.325 kPa = 1.01325 bar
volume: 1 L = 1,000 mL = 1,000 cm^3 (or can be calculated as mass over density)
amount: *n* = mol (or can be calculated as mass over molar mass)
temperature: K = ºC + 273.15

The Ideal Gas Equation has 4 variables and one constant. When given 3 variables, you can calculate the missing fourth. [Remember to use the correct units for each variable and remember that you can convert mass into volume using density, and mass into moles using molar mass].



Question: A quantity of 0.25 mol of a gas occupies a volume of 545 cm^3 at a temperature of 25ºC. Calculate the pressure of the gas? [The constant *R* = 0.0821 L atm / mol K.]
Answer: We are given amount *n* = 0.25 mol volume *V* = 545 cm^3 [convert to = 0.545 L] temperature *T* = 25ºC [convert to = (25 + 273.15) K = 298.15 K] constant *R* = 0.0821 L atm / mol K Use the Ideal Gas Equation to calculate *P = nRT / V* = (0.25 mol)(0.0821 L atm / mol K)(298.15 K) divided by (0.545 L) = 11 atm

Question: A canister containing 1.0 mole of gas is subjected to a pressure of 740 torr. The molar mass of the gas is 28.02 g/mol. At what temperature will that gas have a density of 1.25 g/L?

Answer: We are given
amount n = 1.00 mol
pressure P = 740 torr [which we convert to = 740/760 atm = 0.9737 atm]
constant R = 0.0821 L atm / mol K

First, we need to calculate volume by using
molar mass formula to calculate mass $m = M\,n$ = 28.02 g/mol x 1.0 mol = 28.02 g
and then using density formula to calculate volume V = m/ ρ = 28.02 g / 1.25 g/L = 22.4 L

Now use the Ideal Gas equation to calculate
$T = PV / nR$ = (0.9737 atm)(22.4 L) divided by (1.00 mol)(0.0821 L atm / mol K) = 266 K

When dealing with changing conditions, comparing equations can often reveal specific relationships between parameters. The relationship between state 1 and state 2 can be expressed as:

$$\frac{P_1V_1}{P_2V_2} = \frac{n_1RT_1}{n_2RT_2}$$

The constant R cancels out, and the amount of gas in most systems does not vary, so $n_1 = n_2$; yielding:

$$\frac{P_1V_1}{P_2V_2} = \frac{T_1}{T_2}$$

This equation shows the relationships between parameters in what is called the Combined Gas Laws.

When the temperature is constant, $\boldsymbol{P_1V_1 = P_2V_2}$ [which is Boyle's Law]

When the pressure is constant, $\boldsymbol{V_1 / V_2 = T_1 / T_2}$ [which is Charles' Law]

When the volume is constant, $\boldsymbol{P_1 / P_2 = T_1 / T_2}$ [which is Gay-Lussac's Law]

Question: A gas has a volume of 2.5 L at a temperature of 40°C. What will be the volume of the gas if its temperature is raised to 75°C while its pressure is kept constant?

Answer: We are given
volume in state 1 = V_1 = 2.5 L
temperature in state 1 = T_1 = 40°C [convert to = (40 + 273.15) K = 313.15 K]
temperature in state 2 = T_2 = 75°C [convert to = (75 + 273.15) K = 348.15 K]

As pressure and amount (moles) are kept constant, the Combined Gas Law can be simplified to Charles' Law $V_1 / V_2 = T_1 / T_2$

We can solve for $V_2 = (V_1T_2) / T_1$ = (2.50 L)(348.15 K) divided by (313.15 K) = 2.78 L

Chapter 4

Anon.

MOTION AND ENERGY

Vocabulary list: physics, motion, displacement, velocity, acceleration, gravity, energy, work, power, potential energy, kinetic energy, thermal energy, heat, electricity, magnetism, current, voltage, resistance, wavelength, frequency, amplitude

Physics

Physics is the science concerned with the study of the physical universe, involving energy, matter, forces, motion, heat, light, time, space and other phenomena. It is extremely broad in scope dealing with celestial bodies, earthly objects, atomic and subatomic particles and various energy forms. Core fields include astrophysics, classical mechanics, quantum mechanics, thermodynamics, electromagnetism and relativity. Theories hold true within fields but sometimes not between fields. For example, classical mechanics correctly describes the motion of objects in everyday experience, but it breaks down at the atomic scale, where it is superseded by quantum mechanics, and at speeds approaching the speed of light, where relativistic effects become important.

Motion kinematics

Many objects move. They exhibit changes in position over time. They include inanimate objects (ranging in size from barely-visible specks of dust to enormous arrays like star constellations) and animated life-forms (growing plants and motile animals). Their movement may be barely perceptible (growing grass), apparently rapid (dragon-fly wing-beats), invisibly fast (fired bullet) or incomprehensibly astronomical (speed of light). Mankind has observed motion throughout history and certain relationships became apparent. The Italian scientist Galileo Galilei (1564-1642) studied moving objects and conducted a series of experiments that helped formalize our knowledge of motion into 3 concepts: displacement, velocity and acceleration.

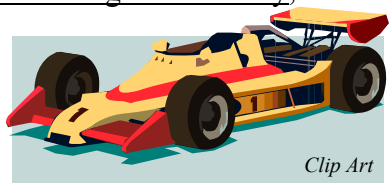
Displacement: The relative positions of two objects are commonly measured using known reference points to create a scale. We commonly refer to the interval between objects as 'distance' and measure it in metres. A single object can also change its position or location, and we often refer to this as the distance traveled. However, it is better to refer to it as 'displacement' (S), that is, the change in the position of an object from an initial state i to a final state f, mathematically represented as:

$$\textit{delta } S \quad = \quad \Delta S \quad = \quad S_f - S_i$$

Velocity: When time is taken into consideration, an object can be perceived to have traveled a specific distance in a particular time interval. We often refer to this as 'speed' and measure it as distance traveled divided by the time it took to travel (in metres per second, kilometers per hour, miles per hour, etc). However, scientists recognized that the direction of travel was also important, so they incorporated it into the definition of 'velocity' which is the displacement of an object in a particular direction divided by the time taken (velocity is said to be a 'vector' because it involves a directional component). The calculation of average velocity considers changes in both displacement and time for an object to move from an initial state i to a final state f:

$$v_{av} \quad = \quad \Delta S / \Delta t \quad = \quad (S_f - S_i) / (t_f - t_i)$$

Acceleration: When considering objects in motion, their speed or velocity was also seen to change over time. Objects picking up speed (increasing their velocity) are undergoing 'acceleration', while those slowing down as decelerating. We have all experienced these changes in velocity when traveling in cars, trains and planes – the surge of acceleration pushing us back into our seats and the deceleration of braking pitching us forward. Acceleration is a measure of the rate of change in velocity, and is simply defined as the change in velocity over time taken. Because we are considering the rate of change of a rate, we are considering time twice. This gives the units of acceleration as metres per second per second, or metres per second squared ($m/s^2 = m.s^{-2}$). The calculation of average acceleration involves determining changes in velocity and time from an initial state *i* to final state *f*:



$$a_{av} = \Delta v / \Delta t = (v_f - v_i) / (t_f - t_i)$$

Question: An athlete runs due east and covers 50 metres in 8 seconds. He stops, turns around and walks back to the start in 40 seconds. What is his average velocity while running, his average velocity while walking, and the average velocity for the round trip?
Answer: Average velocity while running = (50 - 0) / (8 - 0) = 50/8 = 6.25 m/s Average velocity while walking = (0 - 50) / (48 - 8) = -50/40 = -1.25 m/s** Average velocity for round trip = 0 / 48 = 0** **Velocity has a direction component which may be perceived along a linear axis (180^0) as positive (forward) or negative (backwards).

Question: An airplane has an average acceleration of 6 m/s^2 during take-off. How long does it take for the plane to reach a velocity of 216 km/h?
Answer: Rearrange $a_{av} = \Delta v / \Delta t$ to give $\Delta t = \Delta v / a_{av}$ = (216 - 0 km/h) / (6 m/s^2) = (60 m/s) / (6 m/s^2) = 10 s

Newton's Laws

The English scientist Isaac Newton (1642-1727) synthesized the work of Galileo and others into statements of the basic principles that govern the motion of everything in the universe. He developed three fundamental Laws of Motion and one Law of Universal Gravitation.

First Law of Motion: "*An object at rest will remain at rest, and an object in motion will remain in motion, unless acted upon by an external force*". This law recognizes that things stay the same unless something disrupts that stasis. The tendency to stay unchanged is called inertia. You have to apply force to get a stationary object to move, or to change its motion. This law recognizes two types of motion, uniform motion (= velocity) and changing motion (= acceleration). The force required to produce a change in motion depends on the size of the object (large objects are difficult to stop) as well as its velocity (you would not try to stop a speeding bullet). Objects are said to possess "momentum", which is defined as the product of an object's mass times its velocity;

Momentum (kg-m/s) = mass (kg) x velocity (m/s) [$\boldsymbol{p = m\ v}$]

Second Law of Motion: "*The acceleration of an object is directly proportional to the force applied to it, and inversely proportional to its mass*." This law extends the concept of force being required to change motion. Applying force to an object causes acceleration. The greater the force, the greater the acceleration. However, greater force is required to accelerate larger objects because of their greater mass. These relationships are given in the definition of force as follows [note that the units are kilogram-metres per second squared, where one kg-m / s^2 is called the newton (N)]:

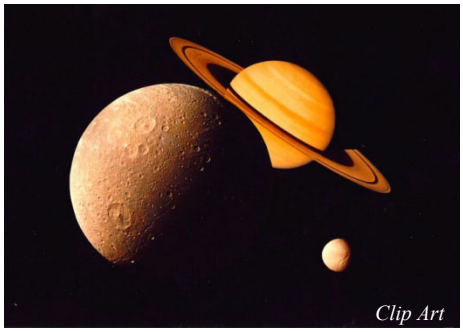
Force (newtons, N) = mass (kg) x acceleration (m/s^2) [$\boldsymbol{F = m\ a}$]

Third Law of Motion: "*For every action, there is an equal and opposite reaction*." This law is less intuitive that the others. We tend to think of our world in terms of causes and effects rather than opposing reactions. We think of the forceful damage done to a car when it hits a tree, rather than the tree providing an opposing force to stop the car. Forces always act simultaneously in pairs. Your weight is exerting a force on your chair, while your chair is exerting an equal and opposite force to support you. Indeed, your weight is a measure of the force required to counter-balance the gravitational pull of the Earth on your body. At the Earth's surface, if an object is dropped and allowed to fall freely, it will accelerate at a rate known as 'gravity', $g = 9.8\ m/s^2$ (or 32 feet/ s^2).

Newton's Law of Universal Gravitation: "*The attractive force (gravity) between any two objects is proportional to their masses and inversely proportional to the square of the distance between them*." Gravitational forces are not restricted to the Earth, but are found throughout the universe between any two objects. The greater the objects, the greater the force. The closer the objects, the greater the force.

Force of gravity = ($mass_1$ x $mass_2$ x constant G) / $(distance)^2$ [$\boldsymbol{F = (m_1 \times m_2 \times G) / d^2}$]

Let us consider the constant G and its units. The units either side of every equation must balance. On the left-hand side, gravity is given as a force which is measured in newtons (= kg-m / s^2). On the right-hand side, mass is measured in kg and distance in m, so we obtain [(kg^2 / m^2) x 'G']. The units for G resolve to the complicated mix of m^3/sec^2-kg (or $N\text{-}m^2/kg^2$). The constant G is known as the universal gravitational constant, it applies to any



two masses anywhere in the universe. In 1798, the English scientist Henry Cavendish (1731-1810) first measured G in an experiment where he suspended small lead balls near large fixed lead spheres and measured the twisting force (torque) on the suspension wire. By substituting the known values into Newton's Law of Universal Gravitation, he obtained a value for $G = 6.67 \times 10^{-11}$ m^3/sec^2-kg (or 6.67×10^{-11} N-m^2/kg^2). This constant holds true universally.

Energy

Energy is defined as the ability to do work, and it exists in several forms: as;

- potential energy (stored energy);
- kinetic energy (associated with movement);
- radiant energy (associated with light);
- thermal energy or heat (= kinetic energy of atoms and molecules);
- chemical energy (stored in the bonds between atoms); and
- nuclear energy (bound within the nucleus of an atom).

Energy may be converted from one form to another, such as radiant energy from the sun being converted to heat, and the potential energy in a battery being converted to light in a torch. One of the most important laws in science is the law of conservation of energy (also called the First Law of Thermodynamics) which states that, "*even though energy can be converted from one form to another, the total amount of energy in a closed system remains constant*".

We all know that energy can be used to perform work. Our muscles use chemical energy to stand and walk, our appliances use electrical energy to heat water and cook food. Work is defined as the application of energy over distance, according to the formula:

Work (joules) = force (newtons) x distance (metres) [$\boldsymbol{W = F\,d}$]

In the metric system where force is measured in newtons (N), work is measured in newton metres (N m). This unit was named 'joule' after the English scientist and one joule was defined as the amount of work done when a force of one newton is exerted through a distance of one metre. The measurement of force conforms to Newton's second law of motion which states that force is proportional to mass times acceleration ($\boldsymbol{F = m\,a}$).

Question:
How much work do you do when you carry your 5 kg computer up the lecture room stairs (5 metres) in Brisbane (where gravity = 9.8 m/s^2)?

Answer: We are given
distance d = 5 m,
mass m = 5 kg, and
acceleration a = gravity g = 9.8 m/s^2

First we can calculate force $F = m\,a = m\,g$ = 5 kg x 9.8 m/s^2 = 49 kg m/s^2 = 49 newtons

Then we can calculate work $W = F\,d$ = 49 N x 5 m = 245 N m = 245 joules

Another dimension is added to the concept of doing work (or utilizing energy) when we introduce a temporal element, that is, the time taken to do work. This is known as power, the rate at which energy is used, and it is defined as the amount of work done over time; shown by the formula:

Power (watts) = work (joules) / time (seconds) [$\boldsymbol{P = W / t}$]

In the metric system, the units of measurement for power are watts, named after the inventor of the steam engine. One watt is defined as the expenditure of one joule of energy in one second. The formula for power can be restated as:

work, or energy (joules) = power (watts) x time (seconds)

This equation is used by power companies to calculate your energy consumption and charge you for usage. However, they transform the units of power from watts to kilowatts and the units of time from seconds to hours, thus deriving units of kilowatt-hours (kWh) which appear on your electricity bills.

Question: Consider the cost of lighting your house. You have twenty 60 watt light globes which you switch on from 6 pm to midnight. Your electricity provider charges you 14 cents per kilowatt hour, so how much does it cost to light your house for one week?
Answer: Note that you need to calculate work in kilowatt hours (instead of joules) You are given power P = 20 globes at 60 watts each = 1200 watts = 1.2 kW and time t = 6 pm to midnight every night for one week = 6 x 7 = 42 hours. First we can calculate work $W = P\,t$ = 1.2 kW x 42 h = 50.4 kWh Then we can calculate the cost = 50.4 kWh x 14 cents per kWh = $7.06

Having examined energy in terms of work and power, let us consider several different types of energy and their properties.

Potential energy

Many objects possess considerable reserves of potential energy, energy which is stored but is capable of doing work when released. It may occur as:

- gravitational potential energy (e.g. that stored in a pile-driver and released when it falls);
- chemical potential energy (e.g. that stored in petrol and released when combusted);
- electrical potential energy (e.g. that stored in a battery and released when switched on);
- elastic potential energy (e.g. that stored in a clock spring and released when uncoiled); and
- magnetic potential energy (e.g. stored in a compass magnet and released when suspended).

Objects have gravitational potential energy because the force of gravity gives them the capability of exerting their own force. The amount of gravitational potential energy possessed by an object lifted above the Earth's surface is equal to the amount of work required to lift it to its position. This is equal to its weight times its height (remembering from Newton's second law of motion that force = mass x acceleration, in this context, weight = mass x gravity); thus giving the equation:

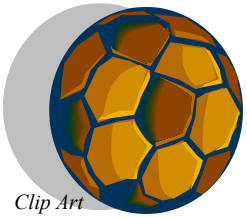
Wikipedia, 2008, 'Devil's Marbles'

gravitational potential energy (joules) = mass (kg) x gravity (m/s^2) x height (m) [$E_P = m\ g\ h$]

Question: Remember carrying your computer up the stairs. We calculated the work required to carry a 5 kg computer up the stairs (5 metres) of the lecture room in Brisbane (where gravity = 9.8 m/s^2). What is the gravitational potential energy contained in that computer over that height?
Answer: We are given mass m = 5 kg gravity g = 9.8 m/s^2 height h = 5 m We can calculate $E_P = m\ g\ h$ = 5 kg x 9.8 m/s^2 x 5 m = 245 kg m^2/s^2 = 245 joules. [This is exactly the same as the work required to lift it to that position] Wow, don't you love physics!

Kinetic energy

This is the energy associated with motion - a soccer ball, a car, a rifle bullet, a rocket - all have kinetic energy when in motion. Two things should be apparent just thinking about these objects: first, the larger they are, the more energy they have; and second, the faster they travel, the more energy they have. These relationships are combined into the equation for kinetic energy, which equals the mass of the object times the square of its speed, times the constant ½:



Clip Art

kinetic energy (joules) = ½ x mass (kg) x [speed (m/s)]^2 [$E_K = ½\ m\ v^2$]

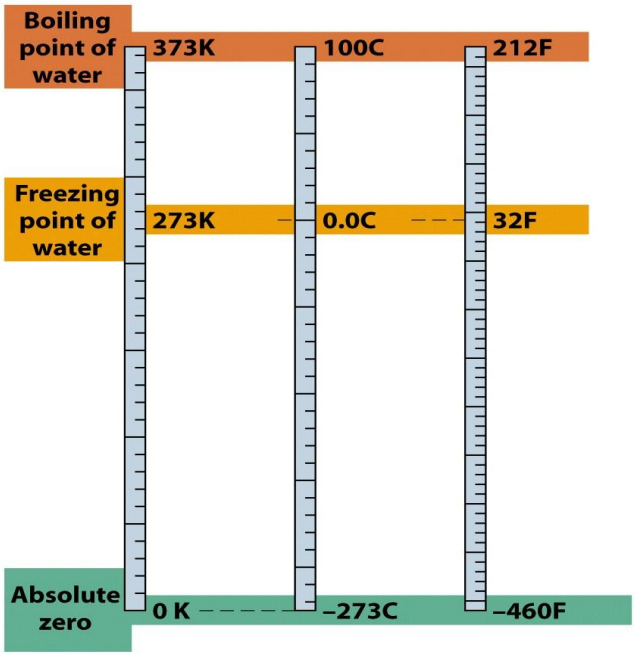
Question: What is the kinetic energy of a 200 gram cricket ball bowled at your middle-stump at 80 metres per second? Compare this to a 50 gram tennis ball served at your head at 120 metres per second.
Answer: We are given the cricket ball's mass m = 200 g = 0.2 kg, and its velocity v = 80 m/s, so we can calculate the kinetic energy of the cricket ball $E_K = ½\ m\ v^2$ = ½ x 0.2 x 80^2 = 640 joules Similarly, we can calculate the kinetic energy of the tennis ball, E_K = ½ x 0.05 x 120^2 = 360 joules

Thermal energy

Energy is interconvertible; it can change from one form to another. However, this conversion is never totally efficient, some is lost as heat. Heat (or thermal energy) is the kinetic energy associated with the motion of atoms and molecules. This energy of motion can be transferred from warmer to cooler objects, thus facilitating a temperature change.

We have seen that temperature can be measured in degrees Celsius, degrees Fahrenheit or Kelvin. Heat is also often measured in calories, defined as the amount of heat required to raise the temperature of one gram of water one degree Celsius.

[NOTE: do not confuse this calorie (cal) with the calorie unit (Cal) associated with diet and nutrition, as one Cal = 1,000 cal].



Trefil & Hazen, 2007; Fig. 4.2

While energy is inter-convertible, there are restrictions on how it shifts from one form to another. The Second Law of Thermodynamics accounts for 3 properties of energy flow:

- directionality "*Heat will not flow spontaneously from a cold to a hot object*"
- efficiency "*The conversion of energy is never totally efficient, some is lost as heat*"
- stability "*The state of disorder (entropy) of a system increases with time*".

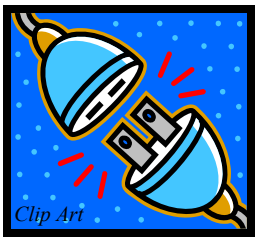
This law can be used to calculate the maximum possible efficiency of any engine by comparing the temperature differences (in Kelvin) between its hot and cold temperature-reservoir components:

$$\text{efficiency (\%)} = [(\text{temp}_{hot} - \text{temp}_{cold}) / \text{temp}_{hot}] \times 100$$

Coal-fired power plants generate high energy steam of ~500 K, while the ambient atmospheric temperature into which waste heat is dumped is ~300 K. The maximum possible efficiency of the plant is therefore = [(500-300) / 500] x 100 = 40%. This means that more than half the energy produced is dumped as waste heat due to the inefficiency of the system (despite the best efforts of our engineers).

Electromagnetism

Electricity and magnetism are different aspects of a major phenomenon present in the universe, the electromagnetic force. Collectively, four major forces are recognized within the unified field theory (colloquially known as the Theory of Everything): namely, gravity, electromagnetism, the 'strong' force (holding atomic nuclei together), and the 'weak' force (responsible for processes such as beta decay that tears elementary particles apart). Electromagnetism describes the interaction of charged particles with electric and magnetic fields. The forces involved are quite powerful - consider that a comb charged with static electricity easily lifts a piece of paper against the gravitational force of the entire planet.



Electricity is the movement of electrons from one object to another. The unit of electric charge, the coulomb (C), was defined by scientists as equal to the charge on 6.3 x 10^{18} electrons. Electricity may take the form of static electricity or an electrical current. Static electricity is stationary, remaining with an object. Objects that accumulate excess electrons become negatively charged, while those losing electrons become positively charged. Like charges repel, while unlike charges attract. Coulomb's law quantifies the electrostatic force between two objects as proportional to the product of their charges and inversely proportional to the square of the distance between them:

electrostatic force (newtons) = k x [1st charge x 2nd charge (in coulombs)] / distance^2 (metres)

where k = Universal Electrostatic (Coulomb) Constant (= 9 x 10^9 newton metre^2/coulomb^2)

Electrical current is associated with the movement of electrons, direct current (DC) involving their one-way movement, and alternating current (AC) involving electron flow that reverses periodically (two way pulse consistent with a sine function). Characteristics associated with electrical circuits are current, voltage and resistance.

- Current (in amperes) is the amount of electric charge that passes through the circuit. One amp equals the flow of one coulomb per second past a given point in the circuit.
- Voltage (in volts) is a measure of the pressure produced by the energy source to push the current through the circuit; the greater the voltage, the faster the current flows.
- Resistance (in ohms) measures the difficulty the current experiences in passing through the circuit. The higher the resistance, the more electrical energy is converted to heat, which reduces the efficiency of the system.

Relationships between resistance, current and voltage are given in Ohm's law:

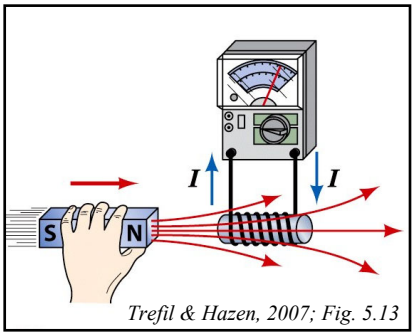
voltage (volts) = current (amps) x resistance (ohms) [$\boldsymbol{V = I\,R}$]

The load of an electric circuit is the place where useful work gets done, such as the filament of a light globe, the heating element in your oven, or the electric motor of your vacuum cleaner. The power used by the load depends both on the current flowing and the size of the voltage:

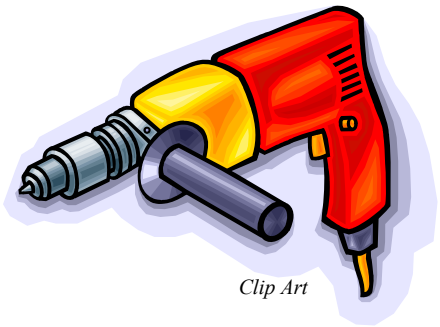
electric power (watts) = current (amps) x voltage (volts) [$\boldsymbol{P = I\,V}$]

Question: Your CD player has a resistance of 60 ohms. When plugged into a 240 volt outlet, how much current will flow, and what is the power consumption?
Answer: We are given voltage V (= 240 volts) and resistance R (= 60 ohms), so we can calculate current $I = V / R = 240 / 60 = 4$ amps. We now have I (= 4 amps) and V (= 240 volts), so we can calculate power consumption $P = I V = 4 \times 240 = 960$ watts

Magnets are metallic objects that attract other metallic objects, such as iron. They exhibit force fields similar to electric fields, characterized by opposite poles (termed north and south). Magnetic lines of force extend between the poles, creating a magnetic (dipole) field. Magnetic and electrical forces share many features. In 1831, Joseph Henry in America and Michael Faraday in England discovered how to produce a current from magnetism, simply by bringing a magnet near a coil of wire. Furthermore, it was found that the current alternated whenever the magnet was rotated or whenever the coil of wire was rotated. All dynamos and electric motors operate according to these simple principles. Dynamos use kinetic energy to rotate coils of wire in magnetic fields to produce electricity, while electric motors do the reverse, that is, they pass electricity through coils of wire in magnetic fields to produce kinetic energy of rotation. Magnetic fields can also be created by the motion of electrical charges, e.g. electromagnets use coils of wire to produce magnetic fields whenever electricity runs through the wire.



Trefil & Hazen, 2007; Fig. 5.13


Clip Art

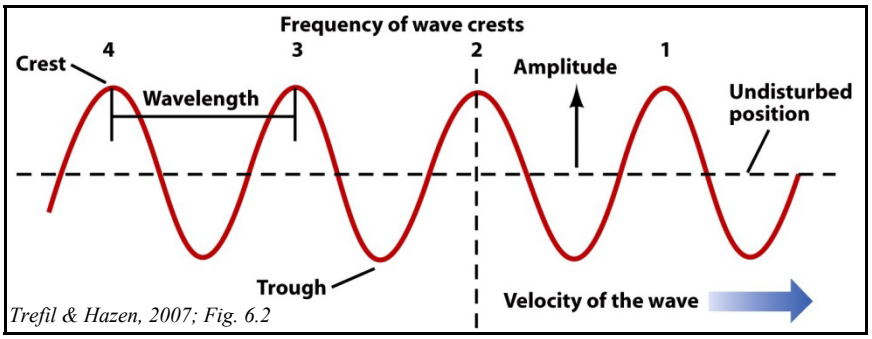
Electromagnetic radiation

Electromagnetism is fundamental to many real-world phenomena. Aside from gravity, most of the forces in everyday experience are the result of electromagnetism. Gamma rays, X-rays, light, microwaves and radio waves are all examples of electromagnetic radiation. This energy is transmitted in the form of a wave, an oscillating disturbance that carries energy from place to place without requiring matter to travel across the intervening distance.

Consider an ocean wave, where individual molecules of water oscillate up and down, but do not travel back or forth along the wave. The energy of the wave is transferred through the oscillation, and not through matter traveling along the wave. While many waves we are familiar with can involve matter (such as earthquakes through solids, ocean waves through liquids, and sound waves through air), electromagnetic waves are not dependent on matter and can travel through an absolute vacuum at the speed of light.

Most waves have a periodic cycle, which show highs (crests) and lows (troughs) over time. The shape of many waves conforms to the trigonometric function $y = \sin x$, and its sinusoidal (wave-like) form is shown below. The axis about which the sine wave oscillates may exhibit vertical and/or horizontal shift, which results in the inclusion of adjustments and constants into the base equation. For

example, the equation for the height of tides is given by the equation $y = A \sin [B (x - C)] + D$, which includes four parameters adjusting for amplitude, period, horizontal (phase) shift and vertical shift.



Trefil & Hazen, 2007; Fig. 6.2

All waves have distinct characteristics, including:

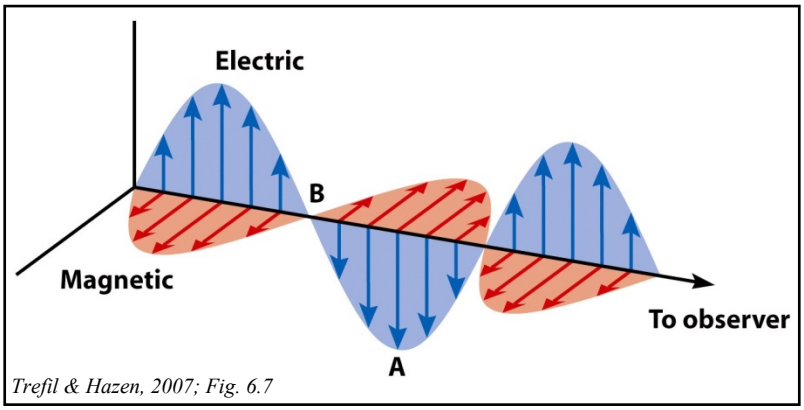
- wavelength (distance between crests of successive waves)
- frequency (number of waves past a given point every second, one per second = 1 Hertz, Hz)
- velocity (speed and direction of wave)
- amplitude (height of wave crest above undisturbed position)

Obviously, the speed of a wave is dependent on its wavelength and frequency, such that:

velocity (m/s) = wavelength (m) x frequency (Hz) $[\ v = \lambda f\ ]$

Question: The surf report predicts a 2 metre swell with waves travelling on-shore at 15 metres per second hitting the coast every 10 seconds. What is the wavelength of the waves?
Answer: We are given three parameters: amplitude (2 metres), velocity (v = 15 m/s), and frequency (f = 1 wave per 10 seconds = 1/10 Hz = 0.1 Hz). We only require the latter two parameters for our calculation using the formula $v = \lambda f$, which rearranges to solve for wavelength $\lambda = v / f = 15 / 0.1 = 150$ m

There are many examples we could explore about wave forms, how seismic waves or earthquakes are measured on the Richter scale, how musical notes have specific wavelengths, how bird song fluctuates, how aircraft create sonic booms, etc. However, most examples involve waves that move through solid, liquid or gaseous media, and we can see or visualize the entities involved. It is much more difficult to visualize electromagnetic waves which do not involve any medium whatsoever. Electromagnetic waves are made up of electrical and magnetic fields that fluctuate together; once started, they are self-perpetuating, even in a vacuum. Consider electrical and magnetic fields as force fields, without substance but with specific properties. In electromagnetic waves, these force fields are arranged perpendicular to each other, and perpendicular to the direction of the wave. It is the force fields, not particulate matter, that wax and wane with regular periodicity. The waves may indeed begin when charged particles or electrical charges accelerate, but once begun, they no longer depend on the source that emitted them.

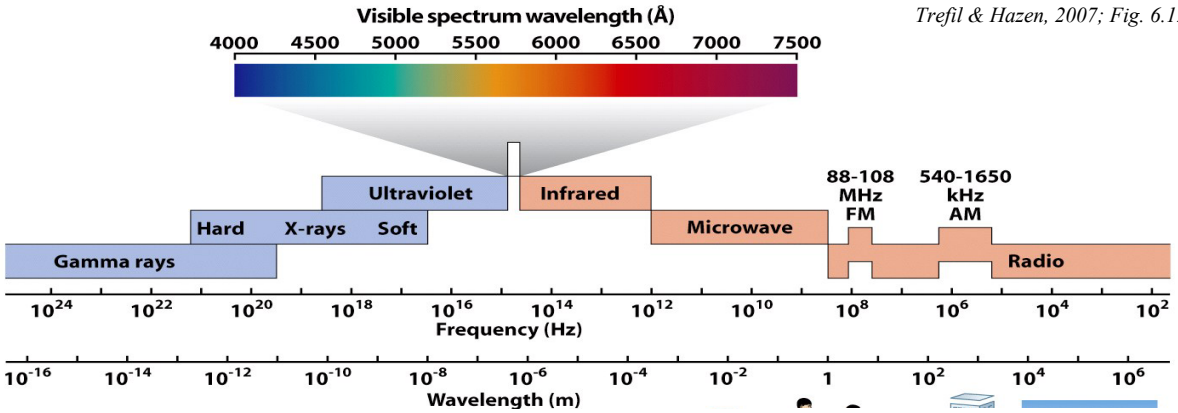


Trefil & Hazen, 2007; Fig. 6.7

There are two important concepts to consider when examining electromagnetic waves: namely, speed and energy. Scientists used equations developed for electricity and magnetism (collectively known as Maxwell's equations, and involving known universal constants) to predict the speed of electromagnetic waves. The answer was always the same; it was the speed of light, namely:

c = 300,000 km/s = 3 x 10^8 m/s = 186,000 miles/s

The nature of light had long puzzled scientists, but it was found to be a form of electromagnetism. The speed of electromagnetic waves depends entirely on the nature of the interactions between electrical and magnetic forces, so all electromagnetic waves, regardless of their wavelength or frequency, travel at the same speed, the speed of light. Electromagnetic waves transfer energy in the form of radiation. The more energy put into accelerating charged particles or electrical charges, the more energy contained in the electromagnetic wave. High-frequency short-wavelength waves transfer more energy that low-frequency long-wavelength waves (e.g. sunlight has more energy than radio waves). The electromagnetic spectrum is very diverse, with wavelengths ranging from millions of metres to smaller than subatomic particles.



Trefil & Hazen, 2007; Fig. 6.12

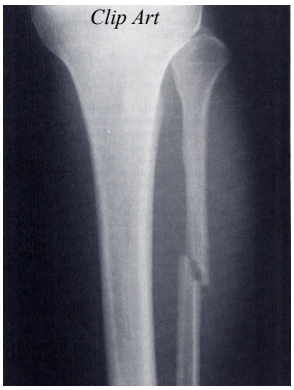
Question: In the middle of winter, you are sitting in front of your fire with a nice glass of claret in hand. You notice the coals and embers in the fire are glowing red. Deep in the recesses of your memory, you recall that the average wavelength of red light is 680 nanometres. What is the frequency?
Answer: You are given one variable, wavelength (λ = 680 nm = 6.8 x 10^{-7} m). You are dealing with light, which has a constant velocity (speed) of c = 3 x 10^8 m/s Using the formula $v = \lambda f$, we can rearrange to solve for frequency $f = v / \lambda$ = 3 x 10^8 / 6.8 x 10^{-7} = 4.4 x 10^{14} Hz WOW factor: Remember that 1 Hz = one cycle per second. For the fire to glow red, the source (charged particles or electric charge) has to oscillate more than 400 trillion times per second (that is, 400,000,000,000,000 times). All in your fireplace!

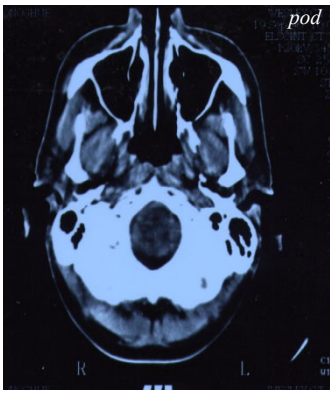
Question: You are getting cold so you stoke up your fire by raking the coals. Why does it become hotter and change colour?
Answer: Raking the coals exposes fresh fuel which burns faster generating more heat. This increases the frequency of oscillation of the electromagnetic source resulting in the production of light with a lower wavelength (~570 nm for yellow light, ~440 nm for blue light).

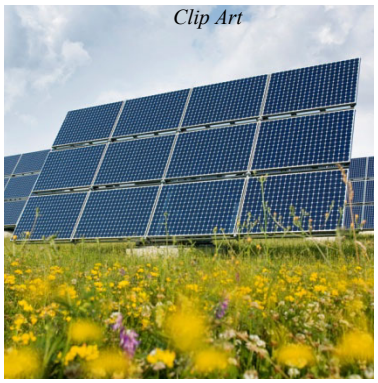
When electromagnetic radiation encounters a physical object, the wave may be transmitted through the object, it may be reflected by the object, it may be absorbed (along with its energy), or it may be absorbed and then re-emitted.

We are constantly bombarded by electromagnetic radiation in the form of gamma rays, X-rays, ultraviolet radiation, visible light, infrared radiation, microwaves, and radio waves. Some pass straight through us, some are reflected, some are absorbed and could be re-emitted (not necessarily in the same form).

Various forms of radiation are used for medical imaging, to see inside our bodies in the event of trauma (e.g. X-ray of broken bone) or possible disease (computerized axial tomography (CAT) scan of head).

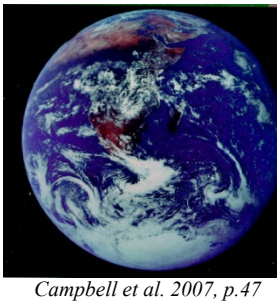






Solar energy is vital to life on Earth - we are dependent on electromagnetic radiation from our Sun. We have also come to depend on electromagnetic radiation in countless ways every day, in radio and television, in computers and telephones, for heating and lighting, in microwave and convection ovens, for medical imaging, diagnostics and cancer treatment. The future of the planet, and humanity, depends on developing new and better ways to produce and utilize electromagnetic radiation.

Chapter 5.



Campbell et al. 2007, p.47

CLIMATE CHANGE

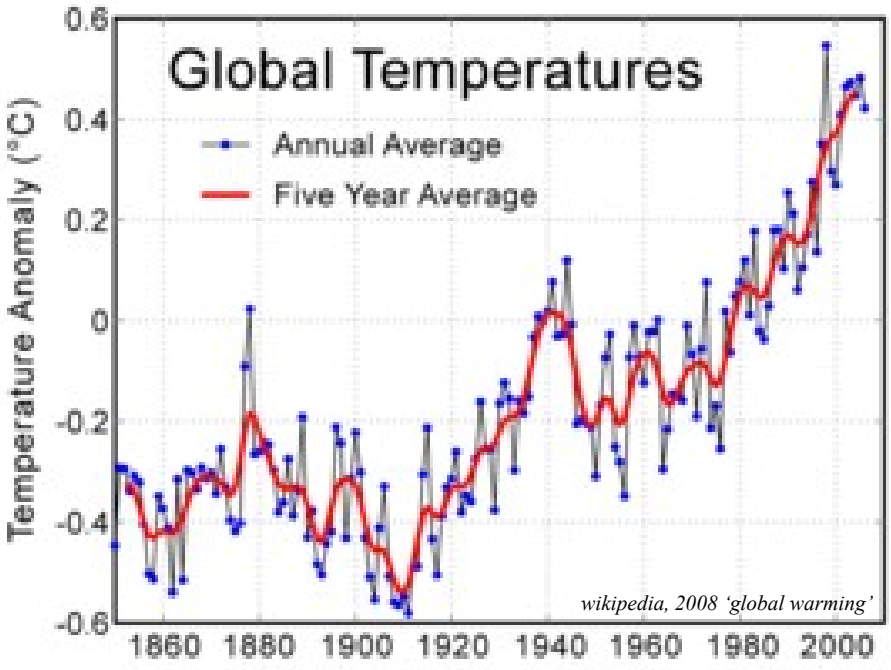
Vocabulary list: climate change, global warming, weather, temperature, Keeling curve, solar energy, greenhouse effect, greenhouse gases, El Nino, Southern Oscillation, Kyoto Protocol

Energy Flow

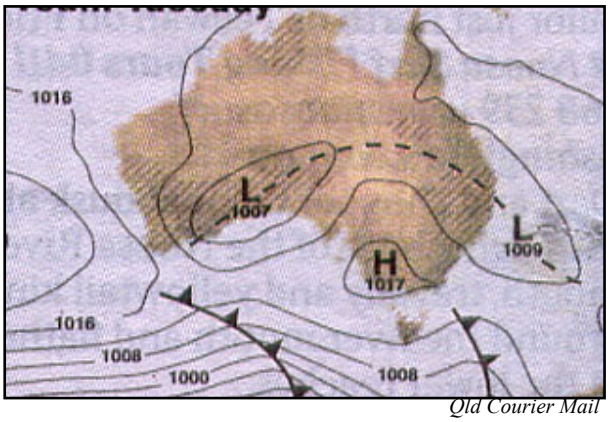
Previous chapters have examined the various states of matter and the different types of energy. The properties and relationships between mass and energy have been studied for many years and several laws have been developed. Perhaps the best known example is Albert Einstein's famous equation $E = mc^2$ which states that "*every object at rest contains potential energy equivalent to the product of its mass multiplied by a constant (the speed of light squared)*". The First Law of Thermodynamics concerns the conservation of energy and states that, even though energy can be converted from one form to another, "*The total amount of energy in a closed system remains constant*". However, the Earth is not a closed system. It constantly receives new energy from the Sun and loses energy to outer space. The resultant energy balance significantly affects our climate.

Climate

The climate of the Earth has continually changed throughout history, the best known examples being the Ice Ages which had profound effects on the Earth and its inhabitants. The Earth is currently experiencing a period of global warming as evidenced by increasing temperatures. While it has been argued that this may be part of a natural cycle of warming and cooling, there is considerable evidence to show that it has been exacerbated (in onset and severity) by human activity ever since the Industrial Revolution (especially by greenhouse gas emissions).



The climate of a region is the average weather experienced over many years, while weather is the atmospheric and oceanic conditions experienced daily/weekly. We recognize many components of weather, including temperature, humidity, air pressure, rainfall and wind conditions. Watch the next weather forecast on television and they will show maps of isotherms (temperature gradients), isobars (pressure gradients with high and low pressure systems), wind indicators (direction and velocity), isohyets (rainfall patterns) as well as satellite images of cloud formations.



Qld Courier Mail

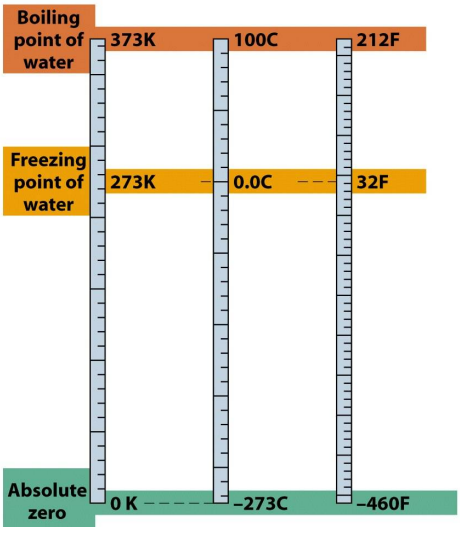
Temperature

The most significant weather parameter would have to be temperature because it influences all the other parameters. Air pressure varies due to differences in air temperature, wind is air flowing from high to low pressure systems, water vapour condenses at low temperatures forming clouds and ultimately precipitation. Temperature is the parameter most often cited as evidence for global warming but the consequences of temperature increases will include more catastrophic weather patterns. But what do we actually mean by temperature?

All objects in the universe are at a temperature above absolute zero. They possess some internal energy - the kinetic energy of atoms which are in constant random motion as they move around and vibrate. Atoms move more rapidly at higher temperatures, hence they have more kinetic energy. Heat (thermal energy) is that energy that moves spontaneously from a warmer to a cooler object, either by conduction (energy transfer through collisions of atoms), convection (transport of warmer atoms) or radiation (travel of infrared energy and other forms of light). The two laws of thermodynamics stipulate the conservation of energy and the directionality of energy flow.

Various devices have been developed to measure temperature, including the classical mercury thermometer and modern digital sensors. They incorporate materials whose properties change with temperature (especially through expansion with heat). The volume of mercury in a thermometer, for example, is directly proportional to temperature (over a specified pressure range). [The relationships between gas volume, temperature and pressure are given in Charles' and Boyle's Laws].

Thermometers use these relationships to display the temperature on a graduated numeric scale. Three different scales are commonly used (degrees Celsius, degrees Fahrenheit, Kelvin) but they have all been calibrated against two standard reference points, the boiling and freezing points of water.



Trefil & Hazen, 2007; Fig. 4.2

Temperature data can be converted from one system to another using the following formulae:

$$^{o}C = 5/9\ (^{o}F - 32)$$

$$K = {}^{o}C + 273$$

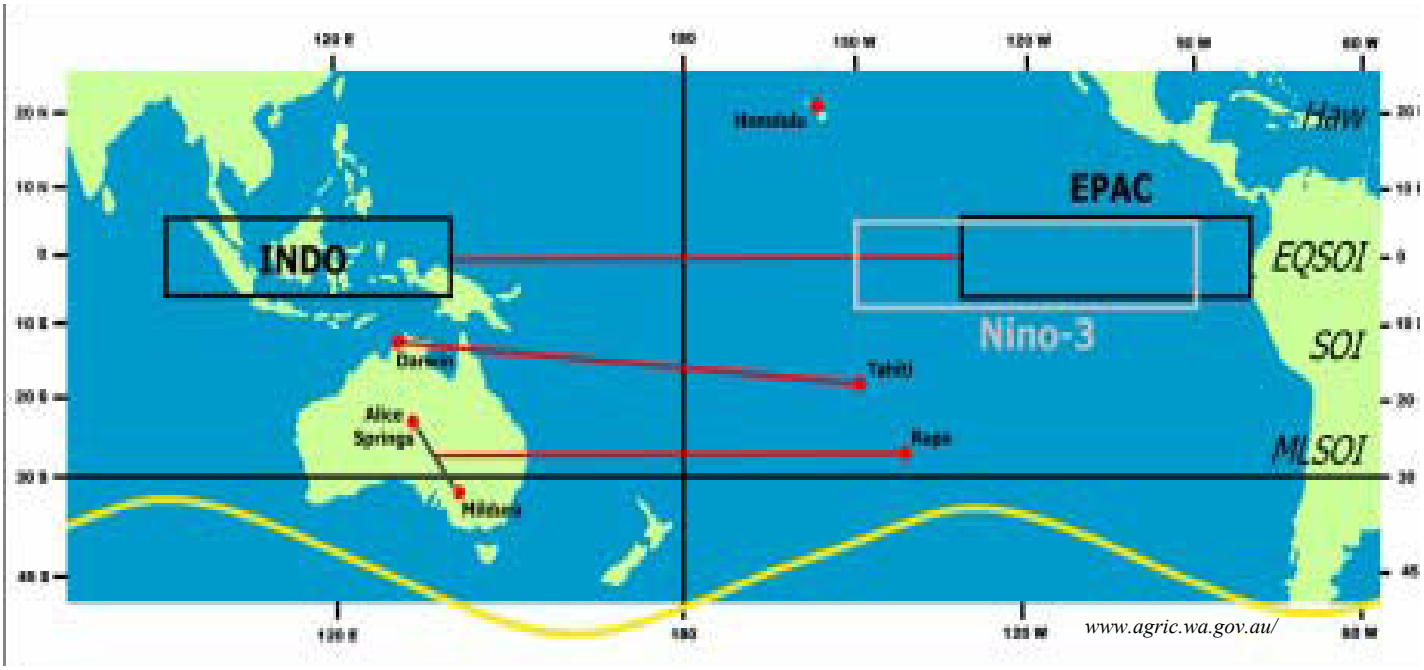
Ambient (air) temperature records have been available for many decades and there are now numerous devices located throughout the world at meteorological stations to record air and water temperature. From these records, meteorologists forecast the weather for the next few days. Longer term analyses of data have also revealed various climatic patterns.

Perhaps the best known example which is currently used for climate prediction across the Pacific and Australasia is the ENSO model.

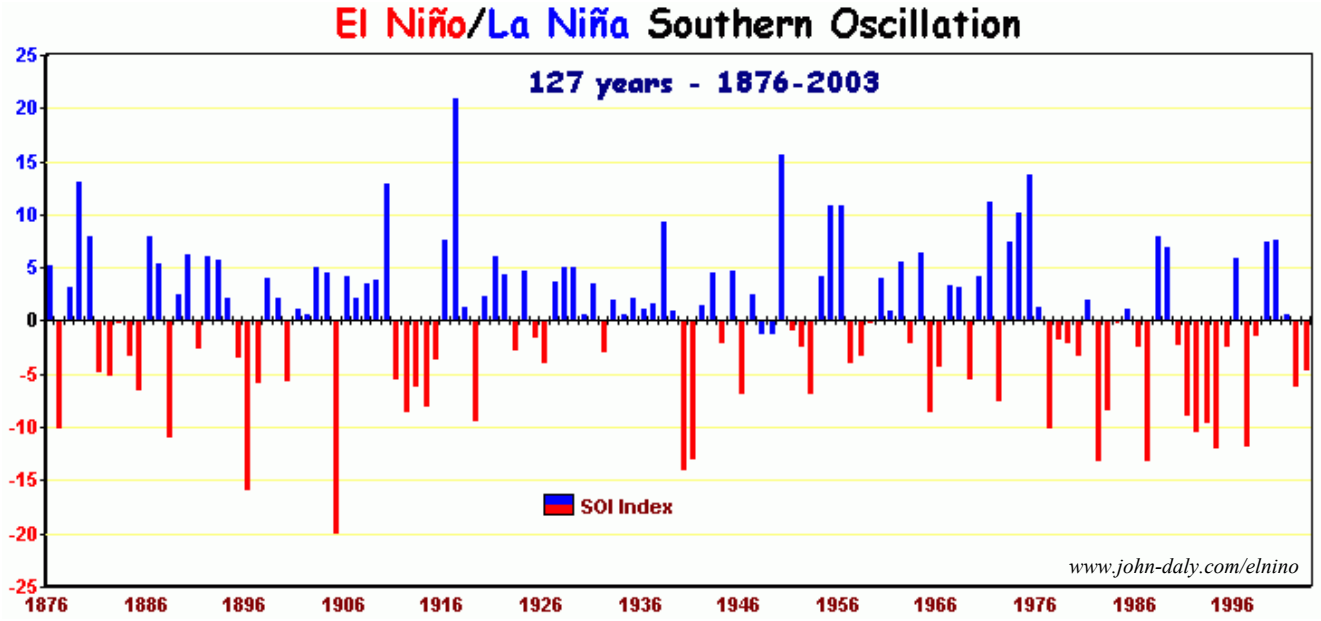
El Nino - Southern Oscillation (ENSO)

Periodic fluctuations in sea surface temperature were noticed as early as 1726 in the eastern Pacific Ocean (along the west coast of South America). Every 2-7 years (average of 4), warm surface waters caused extensive fish mortalities which in turn reduced local seabird populations. This crippled the local economy which relied on fishing as well as harvesting bird guano for export as superphosphate fertilizer. This oceanic flux became known as El Nino (EN) and it has been associated with cyclic slackening of westward trade winds in the Pacific, reduced monsoons, droughts in Australia and cyclones in Polynesia.

Similarly, periodic fluctuations in atmospheric pressure (at sea-level) were noticed since 1904 between South America and Australasia. These changes were associated with contrary weather events across the Pacific and the atmospheric flux became known as the Southern Oscillation (SO) index.



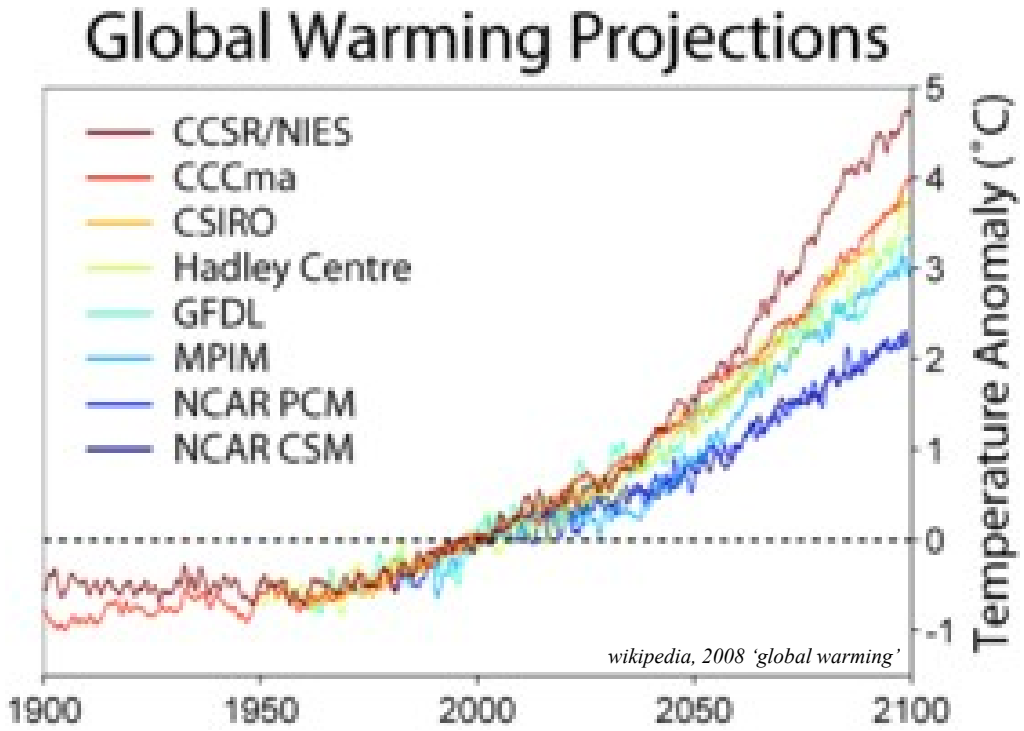
In the 1960's, both events were found to be coincidental and they were combined into the ENSO model which is applicable to periodic warming of the Pacific Ocean (1/4 the circumference of the Earth). It recognizes the oceanic-atmospheric correlation between the sea surface temperature in equatorial Pacific (between latitudes 5ºS-5ºN and longitudes 90ºW-150ºW) and the negative of the sea level pressure difference between Tahiti and Darwin. The ENSO model has been used to predict agricultural productivity, forecast extreme weather events, allocate resources for disaster planning, etc.



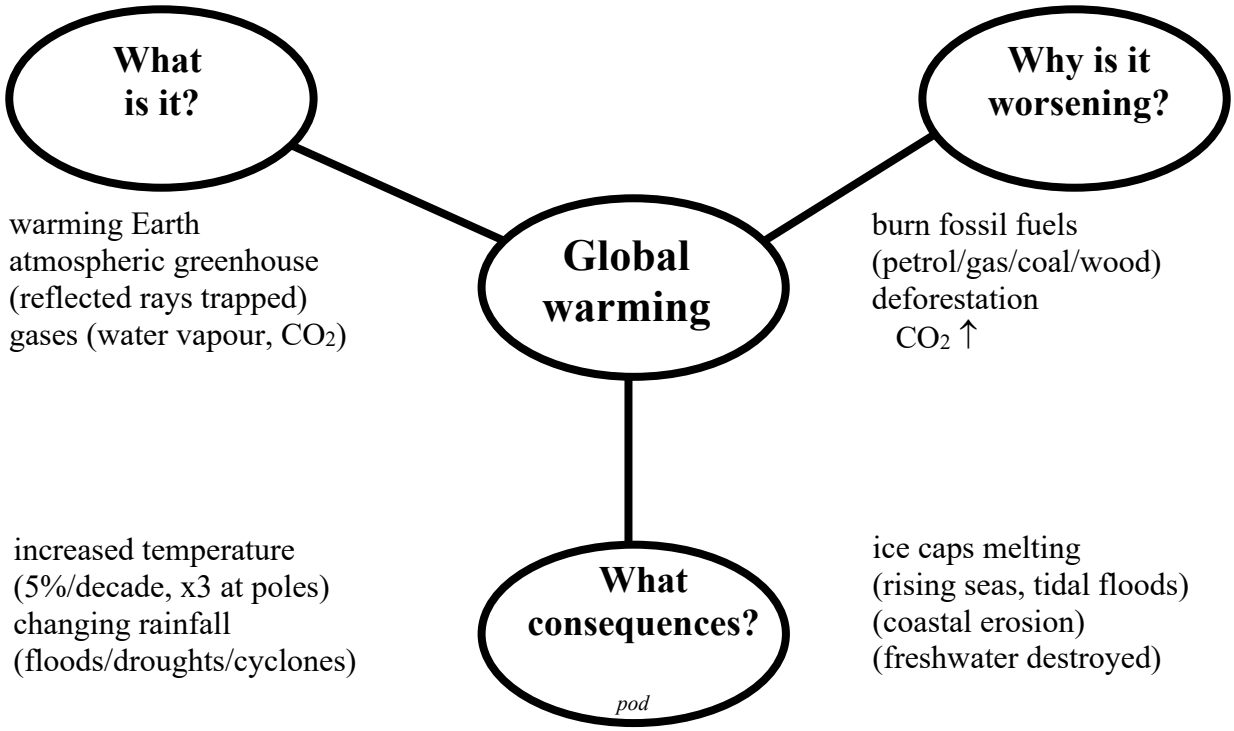
In Australia, a negative SOI means there is an El-Niño under way, and therefore drought conditions can be expected in eastern Australia. The more negative the number, the further south the drought will extend (www.john-daly.com/elnino).

Global Warming

Despite advances in meteorological research and mathematical modelling, few can accurately predict the impact of global warming. Scientists are still debating the extent of temperature changes based on a variety of climatic and environmental models. It is an extremely complex system and many computer models have been developed using a huge number of parameters. Eight projections are shown in the following graph:



Several years ago, there was considerable debate about whether global warming was in fact occurring, but today there is general acceptance by scientists, governments and the general populace that it is a reality. However, the debate still rages as to how much warming has occurred, how much more will occur, why it occurred, what will be the consequences and what can be done about it? The following concept map identifies major concerns:

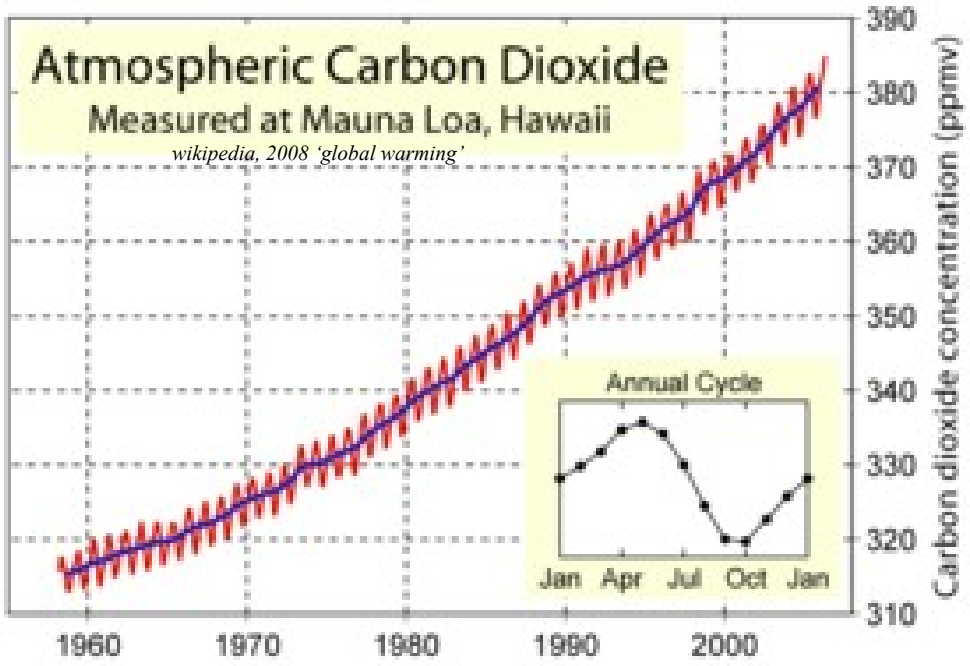


The Earth receives ~ 1.8 x 10^{17} joules of energy from the Sun every second, but ~ ¼ is reflected back into space and another ¼ is absorbed by gases in the atmosphere. The remainder reaches the Earth's surface where it is reflected or absorbed. Only a small portion of solar energy is taken up by living organisms, the majority is absorbed by soil, water and air and re-emitted (mostly as heat). The temperature of the Earth's surface is determined largely by the greenhouse effect, that is, the process by which atmospheric gases differentially absorb and emit radiation. The atmosphere is largely transparent to the Sun's incoming visible light and ultraviolet radiation (small wavelengths from 10^{-6}-10^{-10} m) which warms the surface, but it is somewhat opaque to the resultant outgoing infrared (heat) energy (larger wavelengths from 10^{-4}-10^{-6} m). This energy is trapped within the atmosphere thus warming the planet. Naturally occurring greenhouse gases have a mean warming effect of about 33 °C (59 °F), without which the Earth would be an uninhabitable ice planet.

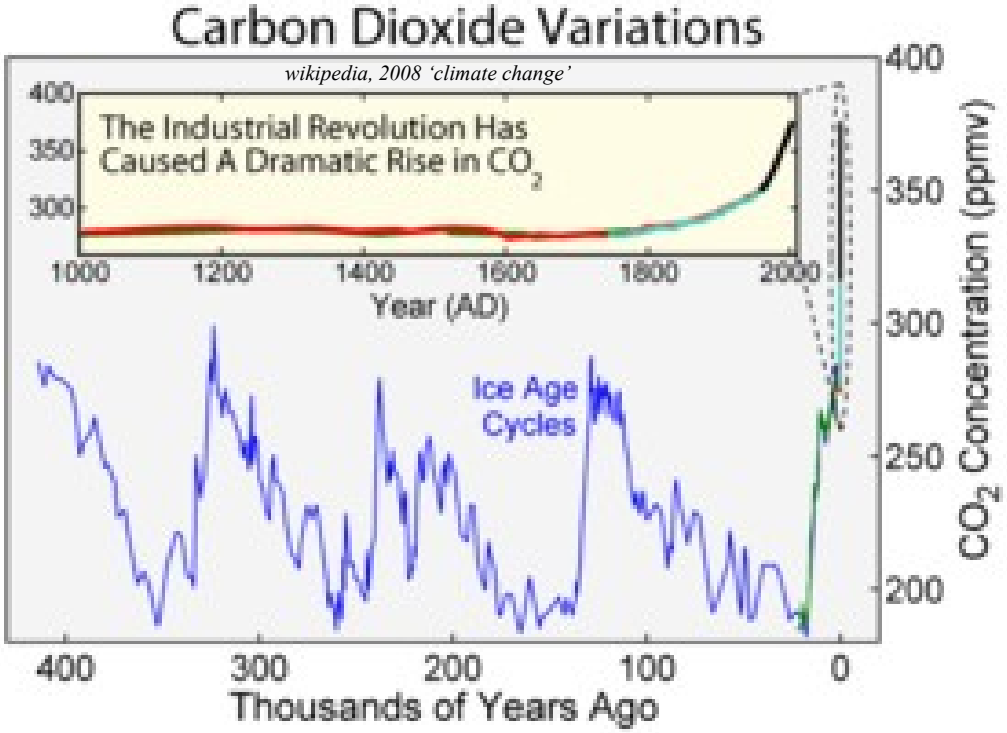
Greenhouse gases

There are over 30 greenhouse gases, the major ones being: water vapour, which causes about 36–70% of the greenhouse effect (not including clouds); carbon dioxide (CO_2), which causes 9–26%; methane (CH_4), which causes 4–9%; and ozone (O_3), which causes 3–7%. The atmospheric concentrations of carbon dioxide and methane have increased by 31% and 149% respectively above pre-industrial levels since 1750. The present atmospheric concentration of carbon dioxide is about 383 parts per million (ppm), and future levels are expected to rise to 541 to 970 ppm by the year 2100.

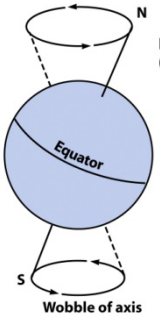
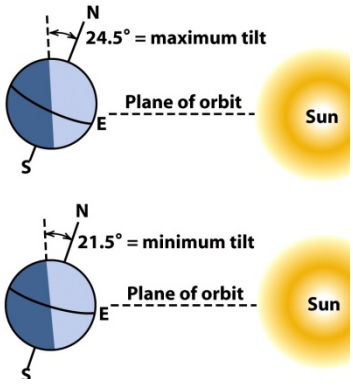
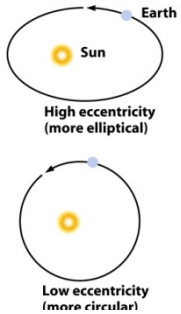
Monthly carbon dioxide measurements show an overall upward trend, with annual variation related to the seasonal uptake of carbon dioxide by the deciduous perennial forests of the Northern Hemisphere (see the following 'famous' Keeling curve).



When viewed from a historical perspective, recent changes in carbon dioxide concentrations appear well outside the range of 'normal' variation over the last 400,000 years [historical estimates obtained through various combinations of carbon dating fossils, ice cores, ocean sediments, coral growth].



Cyclic changes in climate have also been associated with variations in solar radiation (decadal changes in sun-spots and solar-flares) and periodic changes in the Earth's orbit (Milankovitch cycles involving axial wobble, axial tilt and orbital eccentricity).

Cycle	Cause	Effect	Amplitude	Periodicity
Precession	axial wobble 	shifting seasons	360°	19,000 - 24,000 years (23,000)
Obliquity	axial tilt 	intensity of seasons (esp. latitudinal)	22.1-24.5°	41,000 years
Eccentricity	orbital shape (planetary influences) 	intensity of seasons	0.005-0.058 eccentricity	95,000 yr 136,000 yr 413,000 yr (combined 100,000)

Since the Industrial Revolution, however, the extent and rate of change in carbon dioxide concentration has significantly accelerated. The burning of fossil fuels (coal, oil, petroleum, gas) accounts for ~ ¾ of the increase in carbon dioxide over the past 20 years. Most of the rest is due to changes in land-use; in particular, deforestation and biomass burning. In addition, atmospheric methane has almost doubled over this period due to the intensive animal husbandry of herbivores with fermentative digestive processes (esp. cattle and sheep). Ozone levels have also been significantly affected by the industrial production of halogenated compounds as well as volatile organic compounds (butane, propane).

Consequences

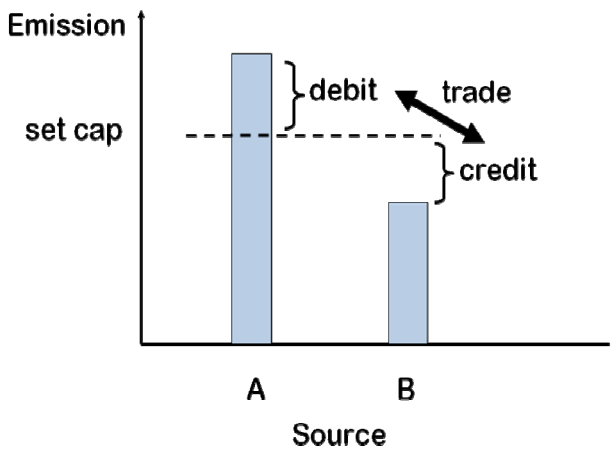
The impact of global warming is far-ranging and will have major consequences; including glacial retreat, Arctic shrinkage, and worldwide sea level rise. The hydrological cycle will become more vigorous and erratic with changes in precipitation patterns resulting in flooding and drought. The generation of more pressure cells will lead to more extreme weather events, such as hurricanes, cyclones, typhoons, etc. Other predictions with global consequences include slowing of the Gulf Stream, the demise of Amazon rainforests, and the release of gas hydrates from the sea-floor.

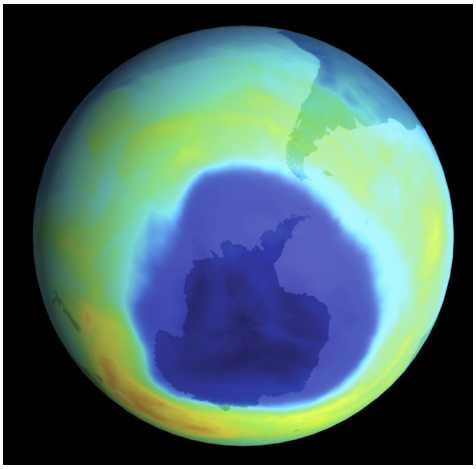
Collectively, these changes will have profound effects on the Earth's biota, resulting in both species extinctions and expansions. Conventional agriculture yields will decline due to reduced soil moisture, greater evaporation, poor thermotolerance, increased stress and susceptibility to pests, predators and pathogens. Tropical diseases will spread as host populations become displaced and vector species become more widespread and abundant. As living conditions change and resources become strained, there will be significant economic and social upheavals throughout the world.

There is ongoing political and public debate regarding what actions should be taken by nations, states, corporations and individuals to curtail future warming or adapt to its expected consequences. Environmental groups encourage action by consumers, communities and organizations. Businesses are exploring increasing energy efficiency and moving towards alternative fuels. A quota on worldwide fossil fuel production has been suggested, citing a direct link between fossil fuel use and carbon dioxide emissions.

An important innovation has been the development of greenhouse gas emissions trading through which companies, in conjunction with government, agree to cap their emissions or to purchase credits from those below their allowances. Free trade schemes provide powerful economic motivating forces for change by industries, governments, communities and individuals. Emissions trading schemes include:

- pricing models (fix price, allow emissions to vary) (e.g. taxation model of UK)
- quantity models (fix emissions, allow price to vary) (e.g. top-down "cap & trade" model of European Union; bottom-up "baseline & credit" model of US), and
- hybrid models (fix both price & emissions) (e.g. "safety-valve" model of China).





International goodwill and collaboration played a significant role in the phasing-out of chlorofluorocarbon (CFC) refrigerants and propellants by industry during the last two decades. CFC's were found to destroy atmospheric ozone (one Cl atom can destroy 10^5 O_3) leading to huge 'holes' in the ozone layer over polar regions in the 1970's. Ozone is vital to life on our planet as it helps block ultraviolet solar radiation (damaging to DNA resulting in mutation and cancer). Twenty concerned countries met at the Vienna Convention in 1985 and by 1987 had signed an international pledge, known as the Montreal Protocol, to phase-out industrial CFC use. In this instance, global action was highly successful and ozone levels are being restored.

The world's primary international agreement aimed at reducing greenhouse gas emissions is the Kyoto Protocol. Scientists around the world expressed strong concern over greenhouse gas levels at the Villach Conference in 1985, they predicted global warming at the Toronto Conference in 1988, and demanded a call-to-arms at the Rio Summit in 1992. The Kyoto Protocol proposed global trading schemes and target emissions in 1997 but the draft agreements were not ratified until 2005 when 55 countries signed. Australia signed in 2007 after a change in federal government. Regrettably, Australia is guilty of the highest emission rate per capita, due to our high dependence on coal for energy and on fossil fuels for transportation around our large continent.

As of 2009, over 160 countries have signed the Kyoto Protocol, but these countries still only account for ~55% of global greenhouse gas emissions. The 6 greenhouse gases of immediate concern are carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), sulphur hexafluoride (SF_6), hydrofluorocarbons (HFC) and perfluorocarbons (PFC) [paradoxically, usage of the latter two increased as CFCs were phased-out]. Several countries continue to procrastinate, citing extensive economic hardship in trying to comply with targets. They need to be reminded that while most resources dwindle around them, "political will is a renewable resource!"

Even if immediate action was taken to stabilize greenhouse gas levels, global warming and sea level rises are expected to continue for more than a millennium - the delay in reaching equilibrium due to the large heat capacity of the oceans. Despite the immensity of the task, humanity has a strong moral, ethical, social and environmental responsibility to leave a lasting heritage to future generations.

Question: Domestic livestock contribute significantly to greenhouse gas emissions, especially ruminants which rely on fermentative digestive processes to derive nutrition from ingested plant materials. Show balanced chemical equations for three types of anaerobic glycolysis (catabolic breakdown of glucose $C_6H_{12}O_6$)!

Answer:
Three types of anaerobic glycolysis

- alcohol fermentation

$C_6H_{12}O_6 \rightarrow 2\ C_2H_5OH + 2\ CO_2$
(glucose) (ethanol) (carbon dioxide)

- lactic acid fermentation

$C_6H_{12}O_6 \rightarrow 2\ C_3H_5O_3 + H_2$
(glucose) (lactate) (hydrogen)

- methanogenesis

$C_6H_{12}O_6 \rightarrow 3\ CH_4 + 3\ CO_2$
(glucose) (methane) (carbon dioxide)

Question: Develop a simple function to calculate the annual methane production by all the sheep and cattle in Australia (include units!).

Model answer:

Annual methane production = (methane production per cow per day x number cows x 365 days)
+ (methane production per sheep per day x number sheep x 365 days)

methane production can be measured in Litres or grams (don't care which but must be consistent)
[methane density = 0.717 Kg/m^3 , that is 1 Kg ~ 1400 L or 1 g ~ 1.4 L]

approximations* daily methane production per cow ~ 180 g/day (~ 250 L/day)
number cows ~ 33 million (ranges from 20-45 million)
daily methane production per sheep ~ 30 g/day (~ 42 L/day)
number sheep ~ 116 million (accept anywhere from 100-175 million)

*huge ranges are given for these values (plus or minus 400%) - such is science!

Chapter 6.

DATING ROCKS AND RELICS

Vocabulary list: dendrochronology, palaeomagnetism, radioactive isotopes, exponential decay, nuclear chemistry, half-life, carbon dating.

Chronology

Time is an important dimension within our universe. Things change over time and we have developed many sophisticated mechanisms for chronometry (time-keeping). We measure the time of day in specific increments (seconds, minutes, hours), we live by the diurnal cycle (day, night), we calendar the passing of seasons (weeks, months) and we take stock of our finite life-spans (years, decades). Our modern technologies operate at hyper-speed (in milli-seconds and micro-seconds) and we contemplate events and histories in astronomical terms (over millennia and millions of years). Various scientific techniques are used to calculate the ages of particular physical objects, whether they be animal (cadavers, skeletons), vegetable (trees, timber) or mineral (rocks, including fossils). No single technique is universally applicable to dating objects over all time periods.

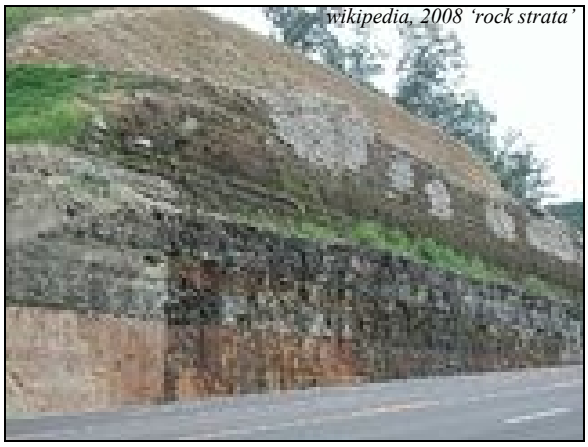
Tree rings

Dendrochronology is the science by which archaeologists can date wooden artifacts, but only over the last few centuries. The oldest known trees are the massive redwood *Sequoia* spp. which may be up to 3,000 years old. As trees grow, they lay down annual growth rings which can be observed in cross-sections through the trunks. The age of a tree can therefore be determined by counting the number of annual rings, provided of course that the date of felling is known.

When the date of felling is not known, the age of the tree (or timber) can often be determined by observing variations in the size of the annual rings. Trees grow more in favourable seasons and lay down more wood in these good years compared to poor growth years, thus generating wide and narrow rings respectively. The pattern of rings thereby provides a 'fingerprint' for that time period, so by comparing the sequence obtained for a particular piece of timber with a reference catalogue of sequences of known age, the timber can often be precisely dated. Regrettably, there are many gaps in the reference catalogue and it only extends backwards in time for the last several thousands of years.

Rock needles

Palaeomagnetic dating is used to date rocks which formed during periods in Earth's history when the Earth's magnetic field was reversed. Such reversals have occurred some 280 times during the last 10 million years. Rocks formed from molten volcanic magma contain minerals which act like miniature compass needles thereby providing a record of the prevailing magnetic field. Rocks formed before and after will exhibit different patterns of mineral alignment so fingerprint sequences can be determined for rock strata according to their palaeomagnetic stripes.

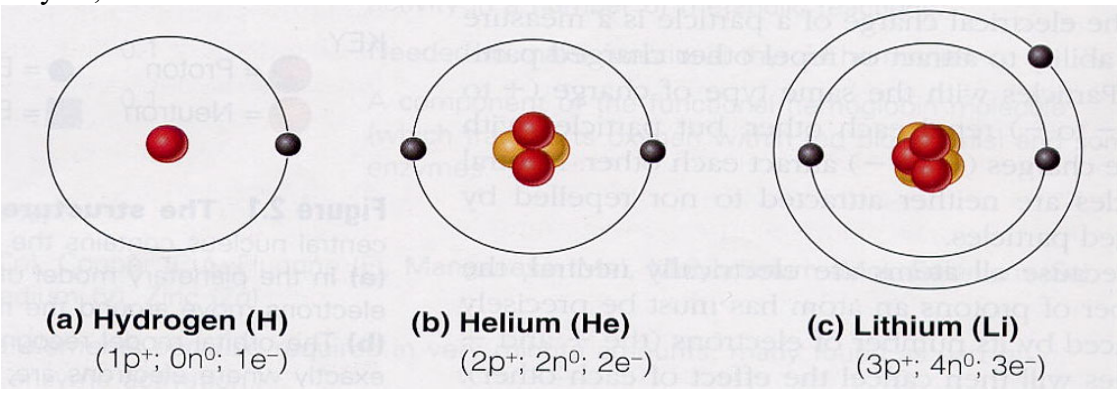

wikipedia, 2008 'rock strata'


wikipedia, 2008 'rock strata'

These sequences are compared to reference catalogues thereby allowing the ages of rocks to be estimated. The technique lacks precision because the magnetic stripes represent periods of unequal duration. Magnetic field reversals take several thousands of years to occur and they last for several thousands of years.

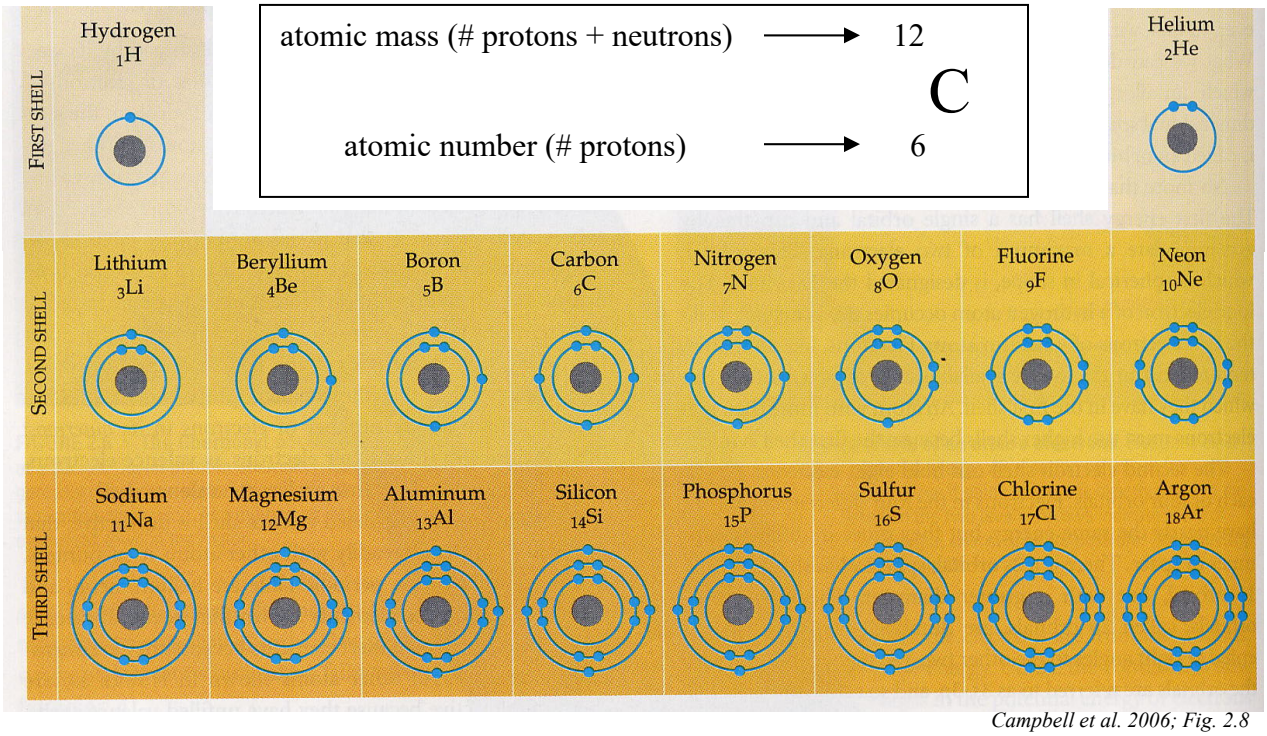
Radioactivity

Radiological or radiometric dating is a more precise technique which can be applied to determine the age of a geological deposit or an archaeological find. It is based on the rate of decay of radioactive isotopes contained within samples, that rate being constant for specific isotopes. Before exploring the mathematical principles involved, we need some understanding of the components and scientific principles involved. All matter is made of atoms which possess a central nucleus (containing protons, each with a single positive charge, and neutrons, which are uncharged or neutral) surrounded by an 'orbiting' cloud of extremely small electrons (each with a single negative charge) arranged into different layers, or shells.



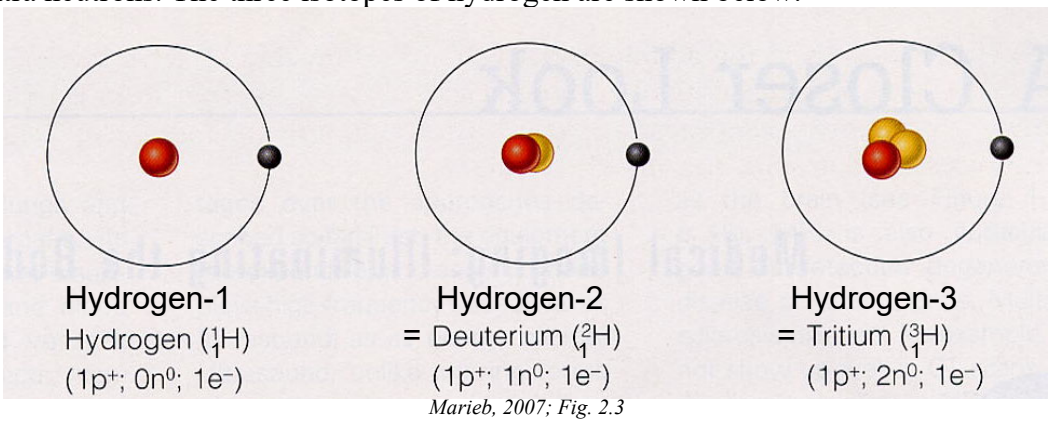
Marieb, 2007; Fig 2.2

Over 100 elements are recognized on the basis of their atomic structure. The periodic table lists them according to the number of protons in their nuclei; carbon has 6 protons, oxygen has 8, etc.



Campbell et al. 2006; Fig. 2.8

Different isotopes of each element can be produced depending on the number of neutrons present in the nucleus; e.g. normal carbon has 6 protons and 6 neutrons (giving a total 'mass number' of 12 and hence the name carbon-12) (alternate notation ${}^{12}_{6}\mathrm{C}$) compared to carbon-14 (${}^{14}_{6}\mathrm{C}$) which has two extra neutrons. The three isotopes of hydrogen are shown below.



Marieb, 2007; Fig. 2.3

Not all isotopes are stable, many are unstable (radioactive) and their structure will change by various means (including neutron-proton replacement, electron capture, alpha decay, and beta decay). While it is impossible to predict exactly when the nucleus of an unstable isotope will change, the statistical likelihood can be calculated and expressed as an exponential decay rate (half-life). The half-life of strontium-90 (${}^{90}_{38}\mathrm{Sr}$) is 28 years, that of carbon-15 (${}^{15}_{6}\mathrm{C}$) is 2.4 seconds, and that of uranium-238 (${}^{238}_{92}\mathrm{U}$) is 4.5 billion years. Various objects can therefore be dated by the measuring and comparing the amounts of radioactive isotopes present.

Question: Name the six different types of radioactive decay shown by isotopes.
Describe the decay particles, and give an example for each in the form of a chemical equation.

[use the convention of showing isotopes in the form ${}^{238}_{92}U$
where superscript = number of protons + neutrons, and subscript = number of protons]

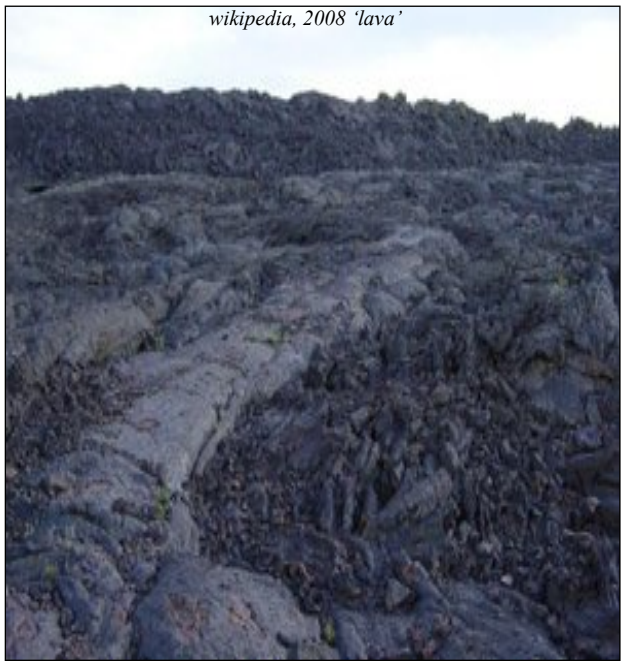
[electrons = ${}^{0}_{-1}e$, neutrons = ${}^{1}_{0}n$, gamma emission = ${}^{0}_{0}\gamma$]

Answer:

Type	Decay particles	Example (others given in Blackman *et al.* 2008. Chemistry. Wiley. pp1048-1059
alpha decay	2 protons plus 2 neutrons (${}^{4}_{2}He$)	${}^{238}_{92}U \rightarrow {}^{234}_{90}Th + {}^{4}_{2}He$
beta decay	electron (${}^{0}_{-1}e$) plus anti-neutrino ($\overline{\nu_e}$)	${}^{3}_{1}H \rightarrow {}^{3}_{2}H + {}^{0}_{-1}e + \overline{\nu_e}$
gamma decay	high energy photons (${}^{0}_{0}\gamma$) accompanies alpha decay or beta decay	${}^{234}_{90}Th \rightarrow {}^{234}_{91}Pa + {}^{0}_{-1}e + \overline{\nu_e} + {}^{0}_{0}\gamma$ ${}^{230}_{90}Th \rightarrow {}^{226}_{88}Ra + {}^{4}_{2}He + {}^{0}_{0}\gamma$
positron emission	positron = positively-charged electron (${}^{0}_{1}e$) plus neutrino (ν_e)	${}^{54}_{27}Co \rightarrow {}^{54}_{26}Fe + {}^{0}_{1}e + \nu_e$
neutron emission	neutron (${}^{1}_{0}n$)	${}^{7}_{2}He \rightarrow {}^{6}_{2}He + {}^{1}_{0}n$
electron capture	electron (${}^{0}_{-1}e$) plus neutrino (ν_e) plus X-rays	${}^{50}_{23}V + {}^{0}_{-1}e \rightarrow {}^{50}_{22}Ti + \nu_e$ + X-ray

Isotopes with long half-lives are used to date rocks and fossils of great antiquity while those with shorter half-lives are used to date younger materials.

Volcanic rocks often contain potassium-40 ($^{40}_{19}\mathrm{K}$) which decays to argon-40 ($^{40}_{18}\mathrm{Ar}$) with a half-life of 1.25 billion years. From the moment of formation (crystallization/solidification of molten lava), the parent isotope decays at a constant rate while the daughter isotope becomes trapped and accumulates in the crystal (it is freed only when the rock sample is melted). By determining the ratio of the two isotopes, the age of the rock can be calculated. If there are equal amounts of potassium-40 and argon-40, half the potassium-40 must have decayed so the age of the rock equals the half-life of the isotope (i.e. 1.25 billion years). While dateable crystals are usually found in volcanic rock, fossils are usually found in sedimentary rocks. Fossils are therefore often dated indirectly by dating the volcanic rocks that sandwich their strata. Other isotopes used to date rocks include uranium-238 (half-life of 4.5 billion years) and rubidium-87 ($^{87}_{37}\mathrm{Rb}$) (half-life of 49 billion years).



wikipedia, 2008 'lava'

Living organisms constantly take up carbon from their environments and use it as chemical building blocks. Plants take up carbon from the atmosphere for photosynthesis and animals ingest it as part of their food web. Most of the carbon consists of the stable isotope carbon-12 but a small amount consists of the unstable isotope carbon-14 which decays with a half-life of 5,730 years. When plants and animals die, they no longer take up fresh supplies of carbon. The amount of carbon-12 in the dead tissues will remain constant while the amount of carbon-14 will decline.

wikipedia, 2008 'mummy'

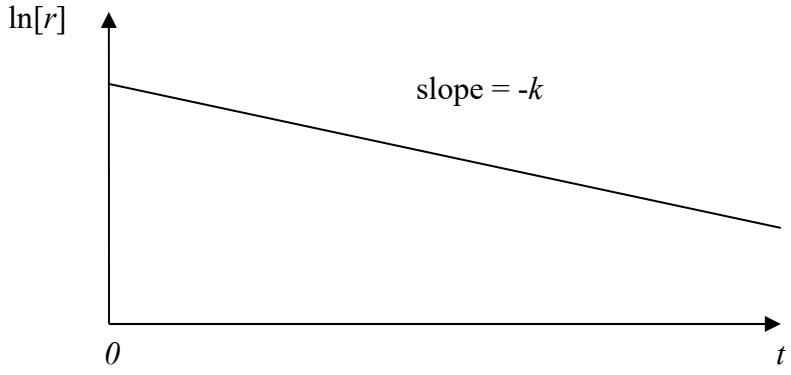
wikipedia, 2008 'mammoth'

The actual age of organic remains (such as wooden or bone relics from tombs) is calculated by comparing the ratio of carbon-14 to carbon-12 in the remains with the ratio in contemporary samples. Carbon dating has frequently been used to estimate the ages of many relics of human civilizations (such as wooden items, clothing, tools) as well as fragments of biological specimens that are not fossilized (bones, hair, teeth, tissues, etc). The technique has gained certain notoriety in television dramas on archaeology, forensics, and myth-busters. Objects more than 50,000 years old, however, have too little carbon-14 left to measure accurately so this dating scheme cannot be used to date rocks and fossils. The ratio of carbon-14 to carbon-12 in contemporary biological samples is 1.33×10^{-12}. The ratio decreases by a factor of 2 every 5,730 years, according to the exponential function:

$$[r_t] = [r_0]\ e^{-k\,t}$$

where r_0 = the carbon-14 to carbon-12 ratio of the historical sample,
r_t = the carbon-14 to carbon-12 ratio of a contemporary sample,
k = slope (= 1.21×10^{-4} yr^{-1} for carbon-14 half-life of 5,730 years), and
t = age of the sample in years.

First order kinetic treatment of the relationship between ratio and time transforms the exponential curve into a straight line, as evidenced by plotting the natural logarithm of ratio versus time.



The equation for the line is: $\ln [r_t] = -k\,t + \ln [r_0]$

which rearranges to $k\,t = \ln [r_0] - \ln [r_t]$

$= \ln [r_0 / r_t]$ [remember the second log law, - = /]

giving $t = \ln [r_0 / r_t] / k$

$= \ln [r_0 / r_t] / (1.21 \times 10^{-4})$ yr

This equation is used to carbon date organic remains. For example, charcoal fragments from an Aboriginal rock shelter in the Northern Territory were found to have a carbon-14 to carbon-12 ratio of 1.08×10^{-12}. The same ratio in living objects is 1.33×10^{-12}. How old were the charcoal fragments?

We are given $r_0 = 1.33 \times 10^{-12}$ and $r_t = 1.08 \times 10^{-12}$
We know the slope $k = 1.21 \times 10^{-4}$ yr^{-1} for carbon-14 half-life of 5,730 years.
We can therefore calculate the age of the sample t $= \ln [r_0 / r_t] / k$

$= \ln [1.33 \times 10^{-12} / 1.08 \times 10^{-12}] / (1.21 \times 10^{-4})$

$= (2.08 \times 10^{-1}) / (1.21 \times 10^{-4})$

$= 1.72 \times 10^{3}$ yr

$= 1{,}720$ yr

Question: Russian palaeontologists found a preserved mammoth frozen in Siberian ice and estimated that only one-quarter of the original carbon-14 remained in tissue samples. How old was the mammoth?

Answer:
The half-life of carbon-14 is 5,730 years, so after 5,730 years, only half the original carbon-14 would remain, and after another 5,730 years, only half of that would remain (= one quarter of original). Thus the mammoth was 2 x 5,730 = 11,460 years old.

Question: The Shroud of Turin, claimed by some to have been used to bury Christ, was found to have a carbon-14 to carbon-12 ratio of 1.22×10^{-12}. How old is the Shroud?

Answer:
We are given
$r_0 = 1.33 \times 10^{-12}$
$r_t = 1.22 \times 10^{-12}$
and we know the slope $k = 1.21 \times 10^{-4}$ yr^{-1} for carbon-14 half-life of 5,730 years.

We can therefore solve for
age of sample t $= \ln [r_0 / r_t] / k$
$= \ln [1.33 \times 10^{-12} / 1.22 \times 10^{-12}] / (1.21 \times 10^{-4})$
$= (0.0863) / (1.21 \times 10^{-4})$
$= 713.45$ yrs

Chapter 7.


pod

LIVING ORGANISMS

Vocabulary list: biology, life, cell, microscopic, metabolism, reproduction, growth, nutrients, solvent, pH, lipid, protein, carbohydrate, nucleotide, DNA

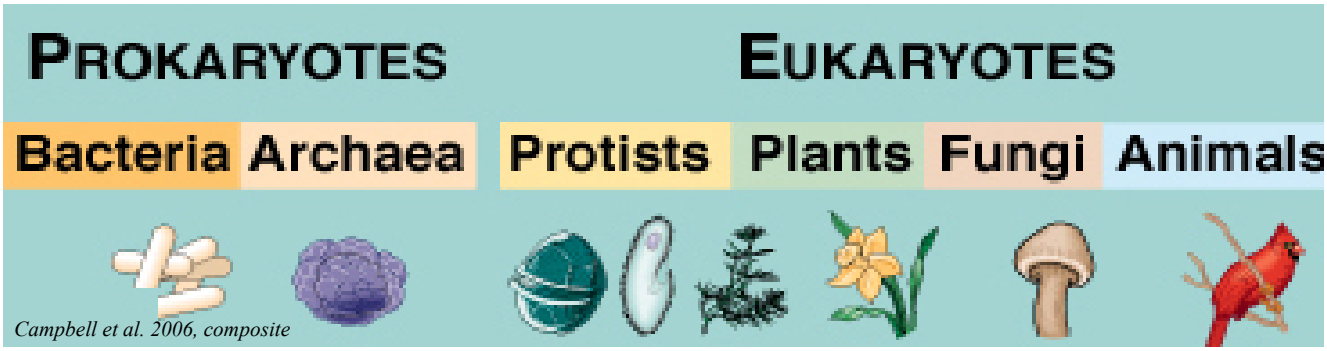
Biology

Biology is the science concerned with the study of life - the structure, function and co-existence of living organisms. Despite the extraordinary biodiversity of living things, they show remarkable unity at the molecular and cellular levels, reflecting their common ancestry, including:

- common building blocks (molecules = sugar, fat, protein, nucleotides);
- collective partnerships (cellular endosymbioses, membranes, organelles);
- diversification (organismal biodiversity, species richness, relative abundance); and
- inter-relationships (ecology, communities, populations, systems).

What is life?

We have previously examined various properties of matter and energy. Both are vitally important to living organisms, providing substance and sustenance. A diverse array of biota occurs on Earth. Two empires of organisms are recognized, the Prokaryotes (without discrete nuclei) and the Eukaryotes (with distinct membrane-bound nuclei), which are further subdivided into 6 Kingdoms, comprising the Bacteria, Archaea, Protists, Plants, Fungi and Animals.



Properties common to these organisms are a carbon- and water-based cellular form with complex organization and heritable genetic information. They undergo metabolism, possess a capacity to grow, respond to stimuli, reproduce and, through natural selection, adapt to their environment in successive generations. The various functions considered necessary for life include:

- ingestion (take up food materials);
- digestion (breakdown of ingested foodstuffs);
- metabolism (chemical reactions for energy production and biosynthesis);
- excretion (removal of waste products);
- movement (locomotion, contractility, propulsion, peristalsis);
- responsiveness (sensory input, motor output);
- reproduction (asexual/sexual); and
- growth (change in developmental form and/or size).

In order to achieve these functions, organisms must live within relatively narrow environmental ranges delimited by upper and lower tolerable limits of various physico-chemical parameters; including:

- water (universal solvent necessary for chemical reactions);
- nutrients (chemical substances used for energy and cell building);
- oxygen (presence/absence to support aerobic/anaerobic metabolic reactions);
- temperature (thermotolerance and homeostasis); and
- atmospheric pressure (barotolerance and gas exchange).

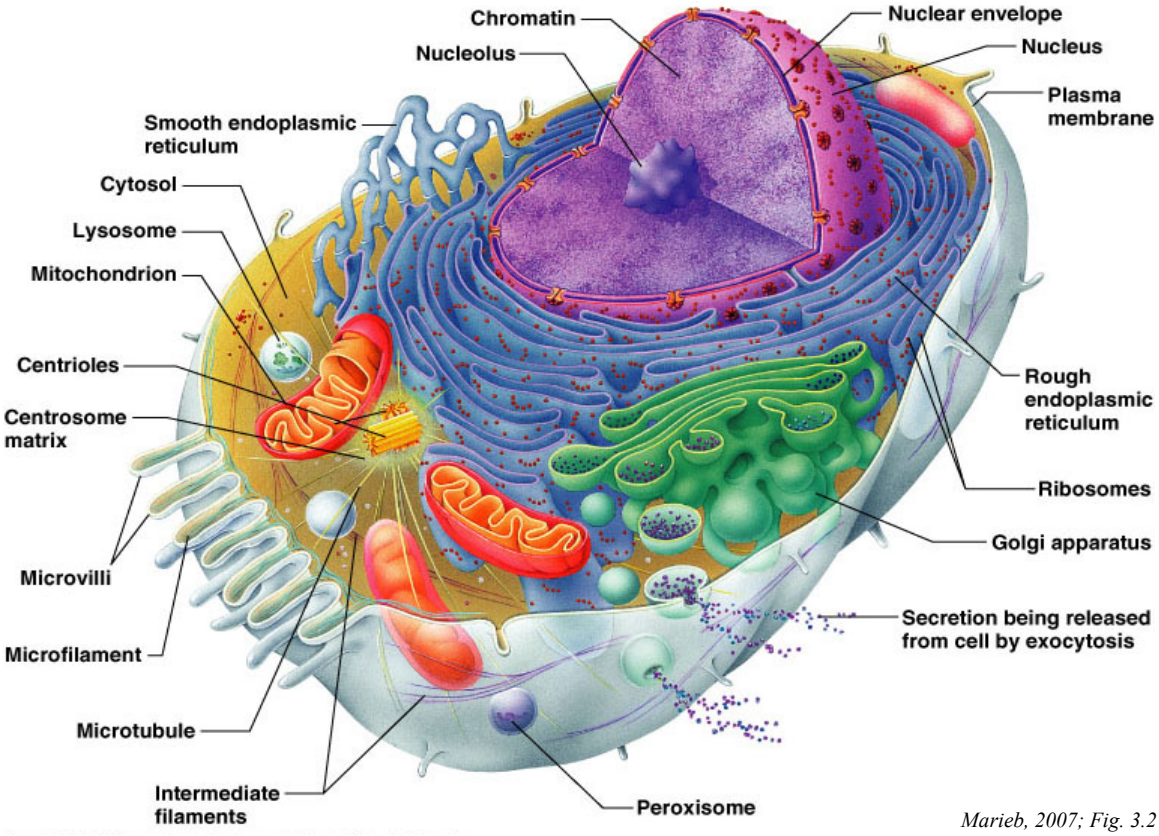
Common features

Let us (flippantly, but conveniently) contend that living organisms are 'self-replicating, membrane-bound, microscopic bags of sugary, proteinaceous water'. As we dissect this statement, we will uncover several basic features of living organisms and see that structure and function are interdependent.

1. Why bags? Cells are the basic units of life. These bags preserve the structural integrity of the organism and maintain the boundary between the external and internal environment. Many life forms persist as unicellular organisms, while others exist as complex multicellular organisms with aggregates of cells forming specialized tissues and organs.

All cells exhibit three basic features:

- they are bound by cytoskeletal elements (to provide form, and sometimes motility);
- they have internal organelle systems (to meet metabolic and developmental requirements);
- they have centralized genetic material (to process information).



Marieb, 2007; Fig. 3.2

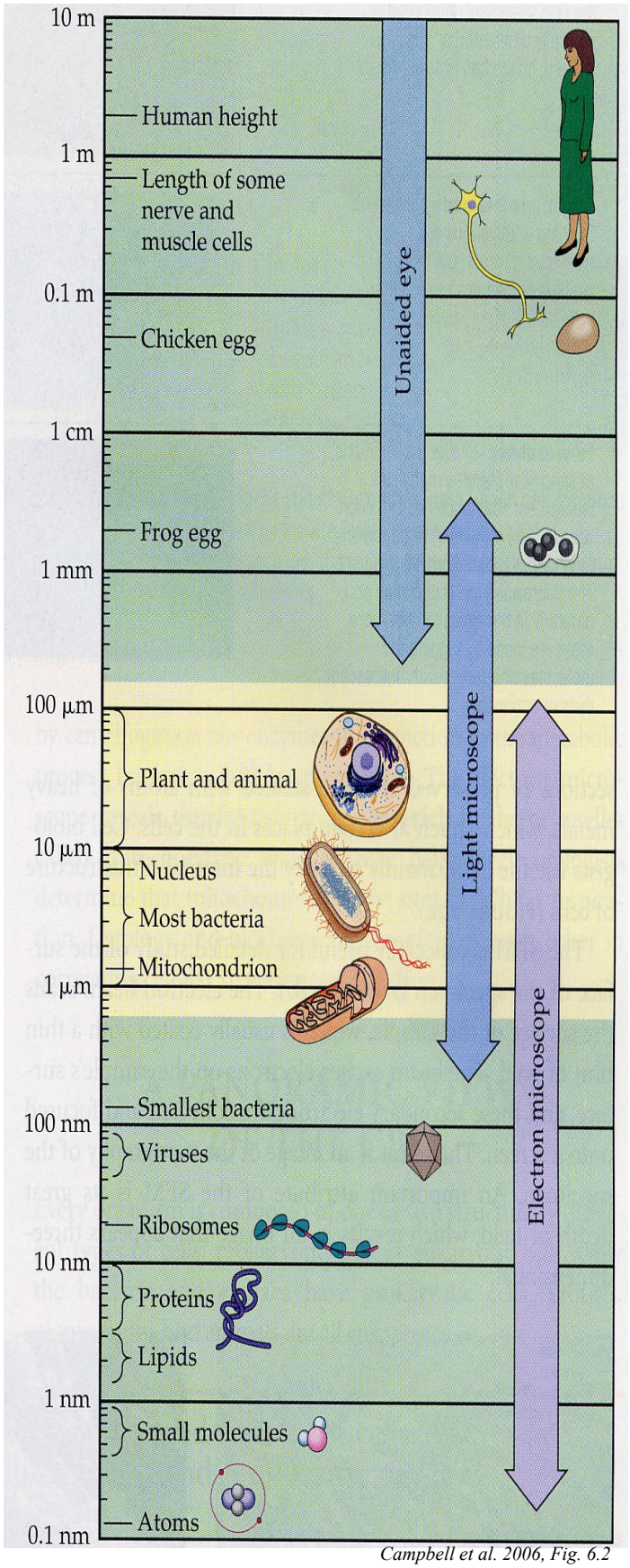
2. <u>Why microscopic</u>? Living organisms occur in a wide rage of sizes. Compare the sizes of giant redwoods, blue whales, dogs, mushrooms, plankton, algae, amoeba, and bacteria. They occur over 8 orders of magnitude, from 1 µm (= 10^{-6} m) to 100 m (= 10^{2} m). However, their constituent cells are all <u>microscopic</u> and only range in size over 2 orders of magnitude, from 1-100 µm.

Note that the scale on the accompanying figure is not arithmetic, but <u>logarithmic</u> ($\log_{10}$) - this means that equal intervals along the axis represent *multiplications* of a fixed number (10 in this case) rather than *additions* of that number. There are many reasons for using log scales to examine data.

Log scales make it possible to plot widely <u>divergent data</u> onto the same graph. If we were to plot the sizes of a mouse (0.1 m) and a dog (1 m) on an arithmetic scale spanning the width of this page, adding the size of a killer whale (10 m) would require the page to be 200 times larger, whereas using a log scale would only involve doubling the size of the page.

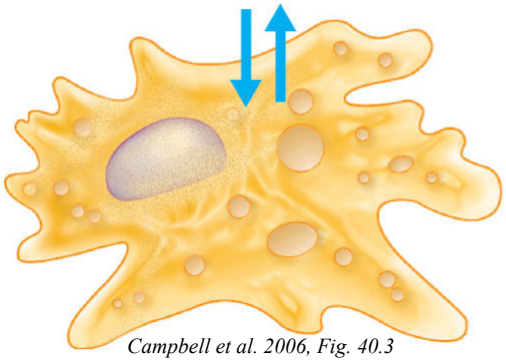
Log scales make it easier to <u>visualize</u> multiplicative factors between plots; e.g. chicken eggs are 10 times bigger than frog eggs.

Log scales often reveal <u>relationships</u> between variables not otherwise evident. Plots falling along curved lines using arithmetic scales may resolve to straight line relationships (<u>correlations</u>) using log scales. This is especially true for data involving powers. For instance, consider a cube to represent a cell and think about the consequences of enlarging it by doubling the length of its sides. While this would only involve a first order change in length (2^{1} = 2-fold change), it would involve a second order change in area (2^{2} = 4-fold change) and a third order change in volume (2^{3} = 8-fold). Similarly, a 10-fold change in length will involve a 100-fold change in area and a 1,000-fold change in volume.



Campbell et al. 2006, Fig. 6.2

Recognizing the impact of size variation on area and volume provides the key to understanding cell size constraints. Cells are limited to microscopic sizes in order to maintain high surface-to-volume ratios so that molecules can move throughout the whole cell. Even though molecular transport processes may involve diffusion (random movement down concentration gradient towards equilibrium), passive transport (facilitated diffusion through specific channels), or active transport (energy-dependent movement against concentration gradient using carrier protein/transporter/pump), they are only effective over small microscopic distances.

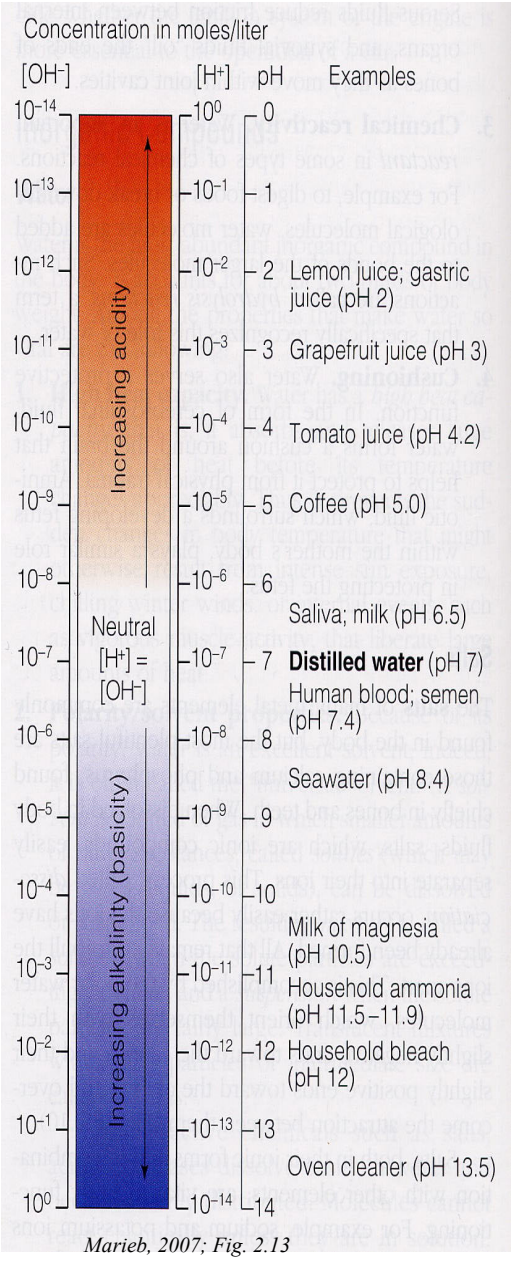


Campbell et al. 2006, Fig. 40.3

3. Why watery? Water is considered to be the fluid of life! Cells are composed of 70-95% water. The molecule, H_2O, has many unique properties. Due to its nonlinear shape, it has a polar charge that contributes to its cohesive (binding) and adhesive (wetting) properties. It exists in three physical states under prevailing climatic conditions, gas (water vapour), liquid (oceans, lakes, etc) and solid (ice). It has a high specific heat due to its kinetic energy and thus acts as a massive thermal bank which acts to stabilize global temperatures.

Water has remarkable chemical properties which allow it to function as a reactant (able to hydrolyse chemical reactions) and as a universal solvent (able to dissolve salts, sugars, and many proteins).

Atoms in water may occasionally lose or gain electrons, resulting in the dissociation of the molecule into positively-charged hydrogen ions (H^+) and negatively-charged hydroxide ions (OH^-); both are very reactive ions vital to the chemistry of life. The relative ratio of these ions contributes to the acid-base balance of a solution. In any aqueous solution, the product of H^+ and OH^- concentrations is constant at 10^{-14}, which can be written as the equation $[H^+]\,[OH^-] = 10^{-14}$. where square brackets indicate molar concentrations (moles per litre). In a neutral solution, both $[H^+]$ and $[OH^-]$ equals 10^{-7}, so the product = 10^{-14}. If acid is added in excess to increase $[H^+]$ to 10^{-6}, then $[OH^-]$ will decline proportionately to 10^{-8} (overall yielding $10^{-6} \times 10^{-8} = 10^{-14}$). Because ion concentrations can vary by a factor of one trillion or more, scientists used logarithms to compress this variation into the pH scale, defined as the negative logarithm of the hydrogen ion concentration, $pH = -\log[H^+]$. For a neutral solution, $[H^+]$ is 10^{-7} *M*, thus giving $pH = -\log 10^{-7} = -(-7) = 7$. Notice that pH declines as H^+ concentration increases, giving acids low pH and bases high pH values. Most biological fluids are in the range pH 6-8.



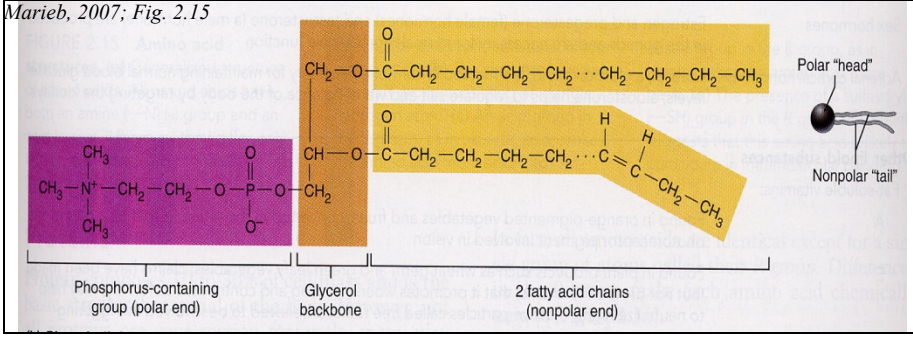
Marieb, 2007; Fig. 2.13

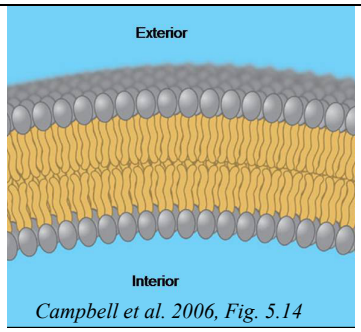
So far, we have established that life is composed of microscopic bags of water. We must now add the 4 essential building blocks of life, organic macromolecules (molecular mass > 1,000), comprising:

- lipids (fats/oils) (composed of fatty acids)
- carbohydrates (sugars) (composed of mono-, di-, poly-saccharides)
- proteins (composed of amino acids)
- nucleic acids (composed of nucleotides)

4. Why membrane-bound? Cells must be structurally bound by substances that are insoluble in water. Lipids (fats) provide those substances. Imagine trying to clean your greasy frypan in cold water. Most lipids are insoluble. They are composed of long chain fatty acids attached to a glycerol core. In our modern diet-conscious society, fats are perceived to be bad things, associated with obesity and chronic disease. Rightly so, but only when taken in excess. Lipids serve many functions: triglycerides and lipoproteins act as energy stores, cholesterol is the precursor of many steroid hormones; while phospholipids form membranes. They are essential building blocks, all cell membranes are composed of phospholipid bilayers. These polar molecules have hydrophilic heads and hydrophobic tails, which become assembled into bilayered sheets, with all heads oriented outwards and all tails inwards. These sheets form the core of all cell membranes, thus providing structural integrity.

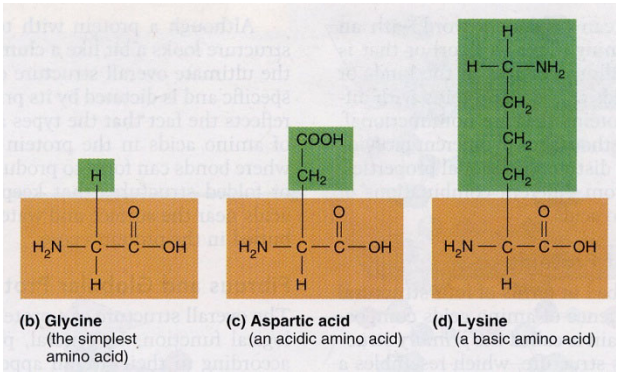
Marieb, 2007; Fig. 2.15



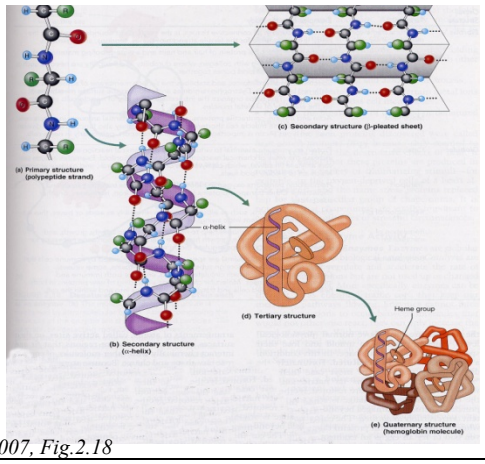


Campbell et al. 2006, Fig. 5.14

5. Why proteinaceous? Cells require many chemicals for metabolic processes, development and multiplication. The basic building blocks are proteins, polymeric molecules composed of chains of amino acids. While the numbers of proteins found in a cell may run into the hundreds of thousands, they are all formed from the same set of 20 amino acids. Proteins vary extensively in structure, each type having a unique three-dimensional shape due to four levels of conformational complexity: amino acid sequence (primary), coiling (secondary), folding (tertiary) and combination (quaternary).

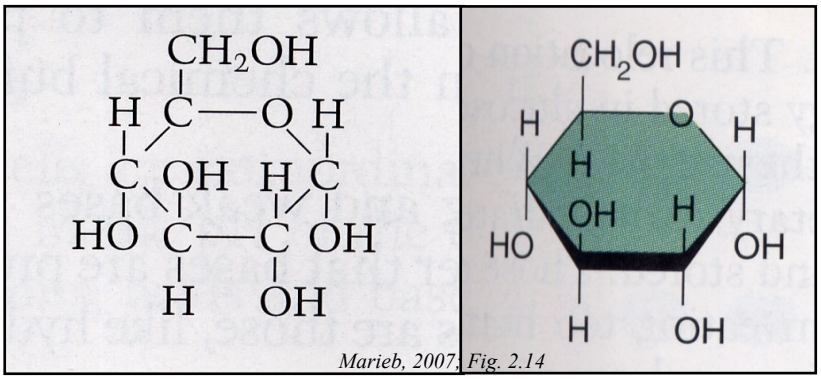


Marieb, 2007; Fig. 2.16



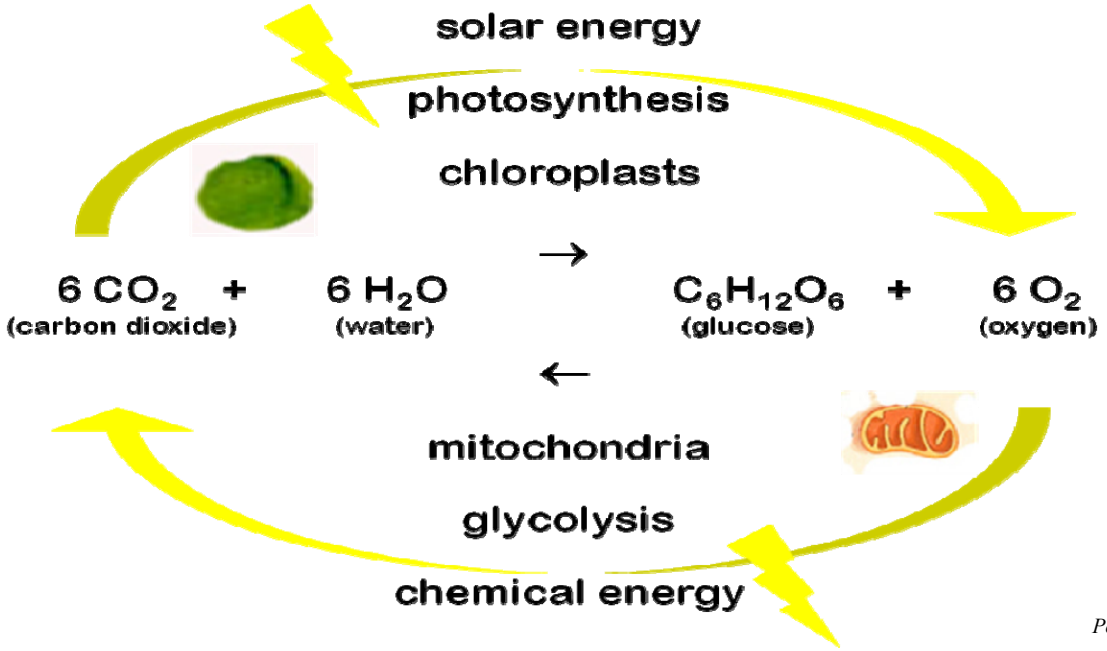
Marieb, 2007, Fig.2.18

6. Why sugar? Carbohydrates include monosaccharides (simple sugars), disaccharides (double sugars) and polysaccharides (polymers). They are all rich sources of chemical energy (stored in their molecular bonds) and their carbon skeletons serve as raw materials for the synthesis of other molecules, including amino acids (proteins) and fatty acids (lipids). Glucose ($C_6H_{12}O_6$) (shown adjacent) is the most common sugar involved in the chemistry of life. It is produced as an energy source by plants, and is ingested by animals as an energy source. It is a vital fuel for living organisms, and its cellular uptake is tightly regulated by various hormonal homeostatic mechanisms (insulin and glucagon in humans).



Marieb, 2007, Fig. 2.14

Carbohydrate metabolism is considered to be the most basic chemical reaction facilitating life on planet Earth. Atmospheric carbon dioxide is taken up by plants, where subcellular organelles (known as chloroplasts) use solar energy to bind it to water to create the sugar glucose, with oxygen being the by-product. The process is known as photosynthesis. In contrast, animals consume glucose (directly as herbivores or indirectly as carnivores) where different suncellular organelles (known as mitochondria) burn it in the presence of oxygen to liberate carbon dioxide and water thereby releasing chemical energy. This process is known as glycolysis (or oxidative phosphorylation). The plastids (chloroplasts and mitochondria) are thought to have evolved from bacteria taken up by early eukaryotes. The balance between plants and animals (producers and consumers) is vital for ecological harmony - it is estimated that plants should outnumber animals by at least ten to one to preserve current conditions (an indictment against further logging and land clearing which will tip the equation towards more greenhouse gases conducive to more global warming).



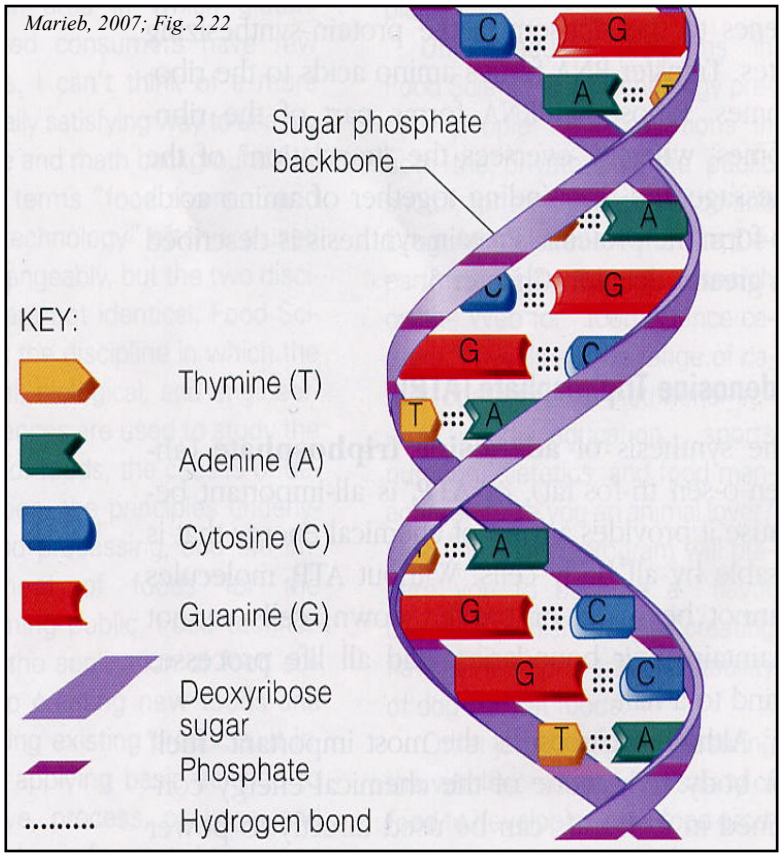
Pod

7. <u>Why self-replicating</u>? All life-forms have limited life-spans and ultimately die (their components effectively wear out). They must therefore replicate themselves in order for their species to survive.

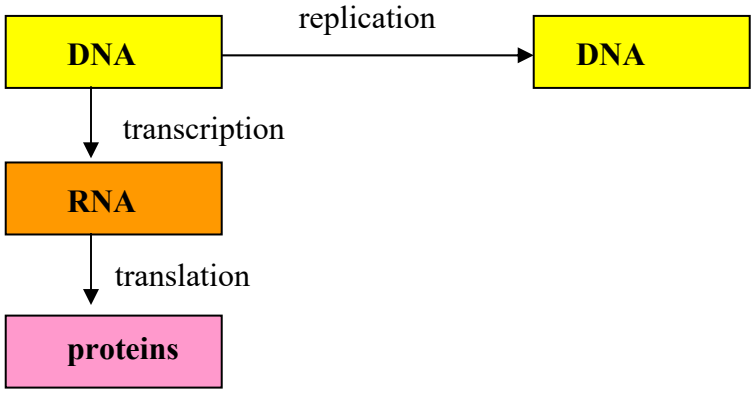
Whether cells <u>multiply</u> asexually (through fission (such as mitosis) or sexually (through meiosis), they essentially copy (replicate) their genetic codes spelt out by the nucleotide sequences of their <u>DNA</u> (deoxyribonucleic acid).

DNA is a linear polymer composed of four nucleotides: the purines, adenine (A) and guanine (G), and the pyrimidines, cytosine (C) and thymine (T) [T is substituted by uracil (U) in RNA]. Two strands of DNA are wound together in a <u>double helix</u>, such that only complementary bases are aligned (G will only align with C, and A only with T).

The DNA of prokaryotic cells (archaea and bacteria) lies in the cell cytoplasm, while that of eukaryotic cells (protista, plants, fungi, animals) is bound by an endomembrane to form the cell nucleus.

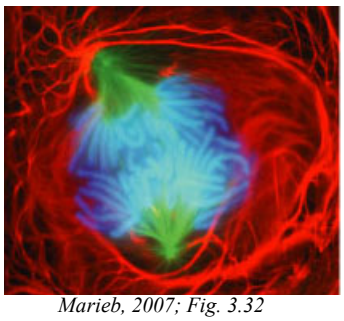


The <u>central dogma of life</u> is that of DNA <u>replication</u>, for it facilitates <u>inheritance</u> (through DNA replication) and <u>metabolism</u> (through DNA transcription to RNA and its translation to proteins).



The study of DNA and proteins forms the basis for modern molecular biology, which encompasses genomics, transcriptomics, proteomics and metabolomics. Later, we will examine the concepts behind sophisticated algorithmic analyses involved in bioinformatics, but for now, we need to consolidate our understanding of cell division and genetics.

Chapter 8.



Marieb, 2007; Fig. 3.32

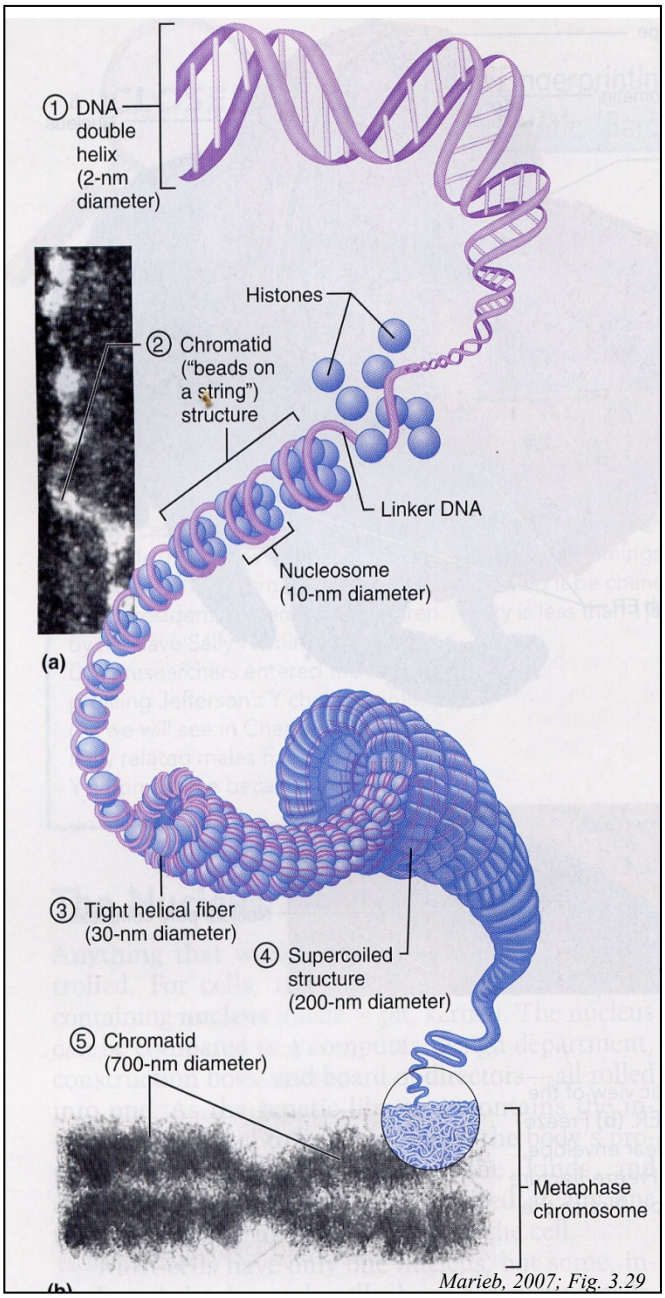
GENETICS

Vocabulary list: cell division, mitosis, meiosis, nucleotide, DNA, gene, chromosome, genetics, pedigree

Genetic material

Every living cell carries its genetic code of DNA sequences which can be differentially decoded throughout its life and which can be copied and passed onto the next generation of cells. Knowledge of the fundamental processes of cell division and genetic inheritance are mandatory for a proper understanding of contemporary molecular biology.

While we may read the genetic code as a linear array of four different nucleotides (A, T, G and C), the DNA molecule is certainly not linear. It is twisted into a double helix, which is coiled around histone proteins, which are wound into helical fibres, which are supercoiled into chromatin which is visible microscopically as the heterogeneous basophilic-staining nuclear material. Further condensation into individual chromosomes is only observed transiently during cell division.



Marieb, 2007; Fig. 3.29

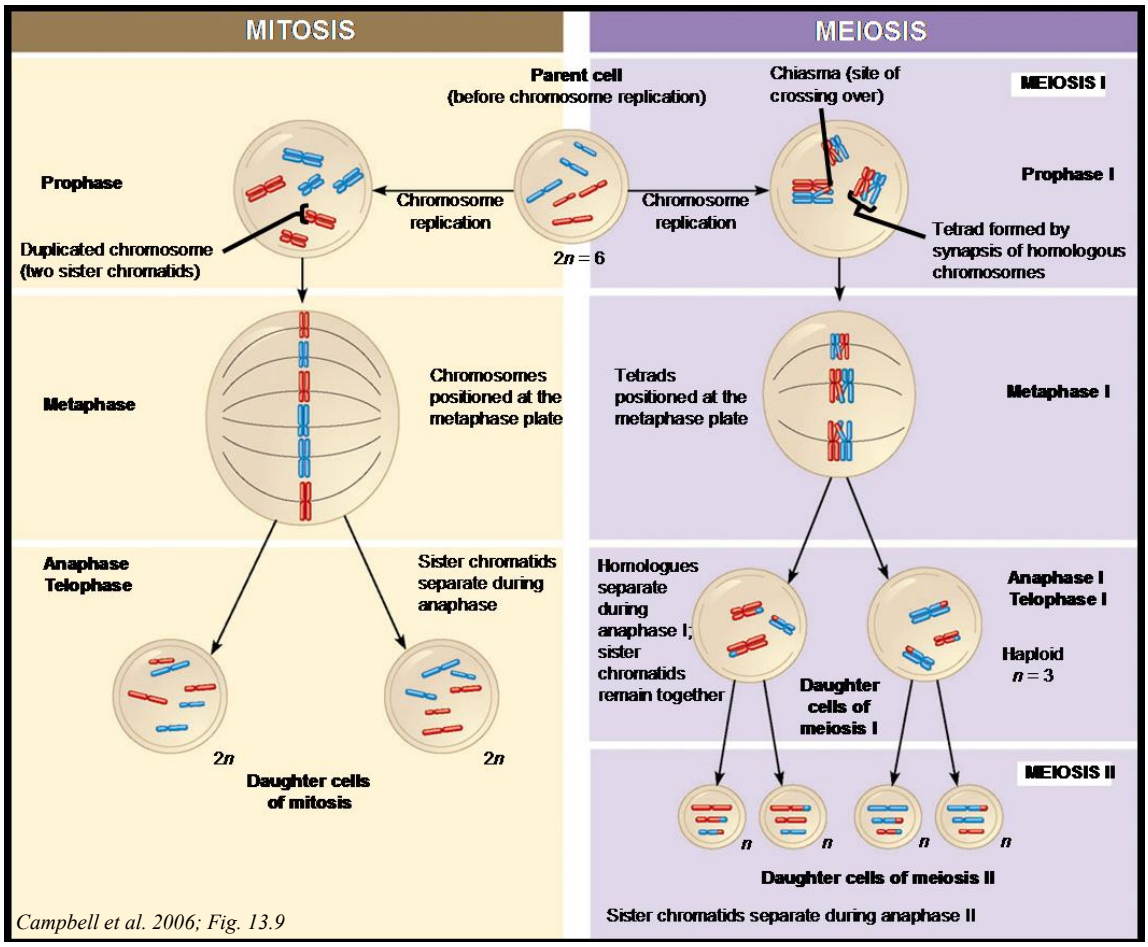
Cell cycle

All cells have a finite life span, usually ranging from minutes to months depending on the type of organism and cellular specialization. The predominant period of a cell's life is spent in interphase, characterized by a growth period (G_1), a synthesis period (S) where most DNA replication occurs, and another growth period (G_2), followed by cell division.

Cells use two processes to reproduce: mitosis (an asexual process where a sole parent cell 'splits' to form two identical daughter cells) and meiosis (a sexual process where two parents form gametes which combine to form a single novel daughter cell). Cells lost by an individual through senescence or injury are replaced by mitosis, and the daughter cells are faithful copies of the parent. In contrast, meiosis involves the creation a new generation of individuals, where the offspring are genetic combinations of two separate parents.

Cell division

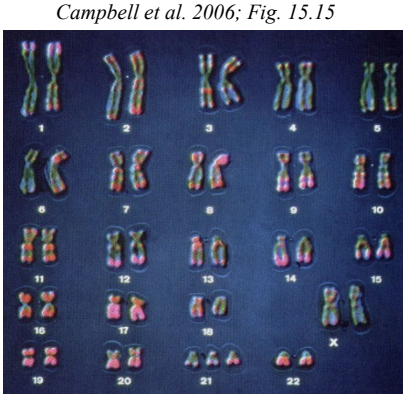
Cell divisional processes involve condensation of the DNA in parent cells into chromosomes which replicate and then separate (one separation in mitosis to form two daughter cells, two separations in meiosis to form four gametes). To complete the sexual reproduction process, a male gamete (sperm) from one parent must fertilize a female gamete (egg or ovum) from the other parent, thereby forming a zygote (embryo). Cells normally contain two copies of each chromosome, so they are said to be diploid (with 2n chromosomes), while gametes only contain one copy of each chromosome, and are haploid (with n chromosomes). When two haploid gametes combine, the resultant zygote is diploid (n + n = 2n), having inherited one maternal copy and one paternal copy of each chromosome. This concept of differential inheritance is central to classical genetics.



Campbell et al. 2006; Fig. 13.9

Chromosome number is constant within species, but varies between species, although not in accordance with the complexity of the organism. Humans have 23 pairs of chromosomes (2n = 46), dogs have 39 pairs, and mosquitoes 3 pairs. Various karyotyping techniques are used to screen for chromosomal abnormalities in cells harvested from embryos. The figure shows an abnormal human karyotype, known as trisomy-21 (three copies of chromosome 21). This genotype manifests in the phenotype known as Down syndrome, characterized by specific physical and intellectual features.

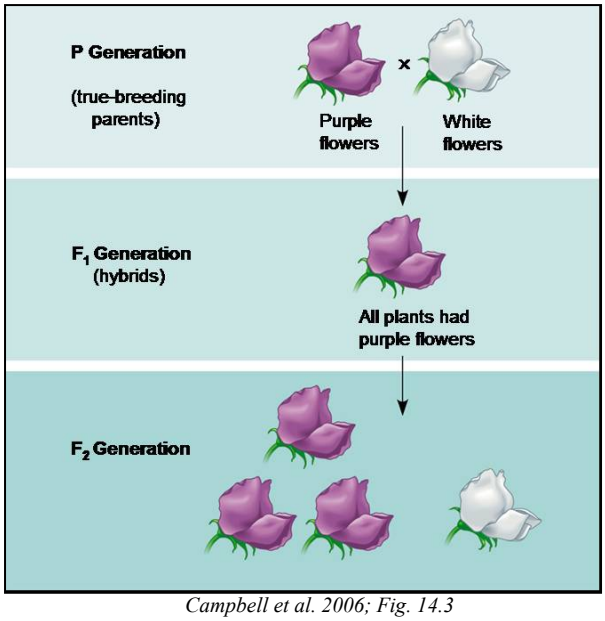
Campbell et al. 2006; Fig. 15.15



Mendelian (classical) genetics

An elegant series of experiments was performed by Mendel, an Austrian monk, who revealed patterns governing the inheritance of physical characteristics in garden peas.

He showed that some varieties always bred true to form, e.g. plants with purple flowers produced offspring with purple flowers. When he cross-pollinated plants with different characters, all the first-generation (F_1) offspring only displayed one trait, the other apparently having disappeared. However, when he cross-pollinated the F_1 hybrids, ~ ¼ of the second-generation (F_2) offspring displayed the missing trait.

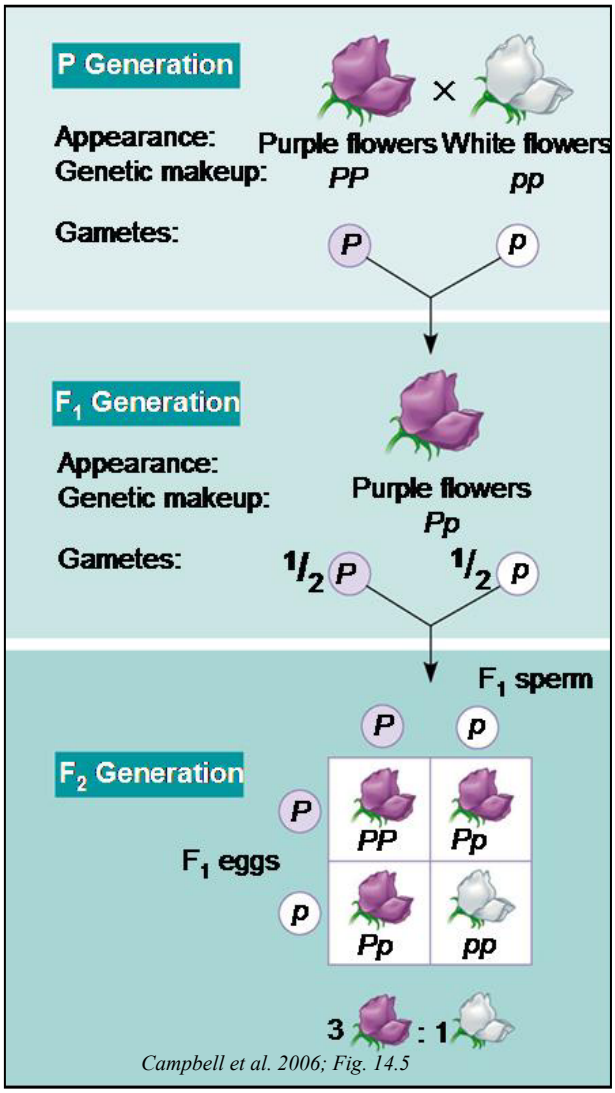


Campbell et al. 2006; Fig. 14.3

Mendel's observations led to the establishment of three rules of classical genetics: physical traits are passed from parents to offspring by units of inheritance (genes); offspring inherit two copies (alleles) of every gene (one from each parent); and some genes are dominant (expressed even when alleles differ = heterozygous) and some are recessive (only expressed when alleles are the same = homozygous).

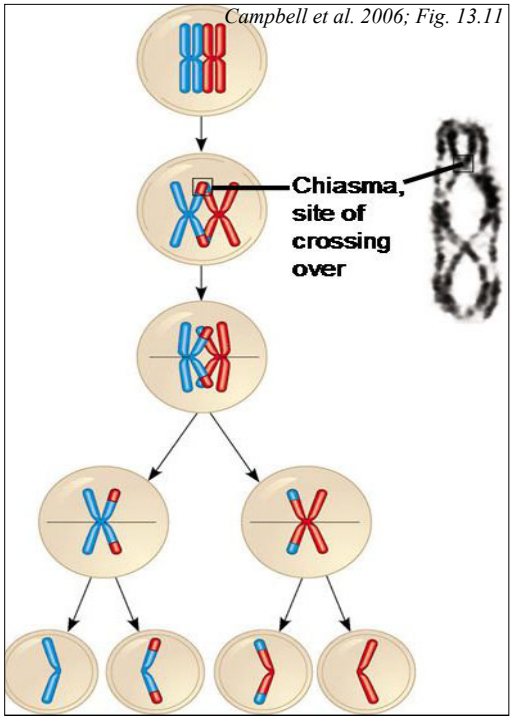
Knowing that gametes only carry one copy of each chromosome, and therefore only one allele for each gene, allows us to predict mating outcomes and the patterns of inheritance. By coding dominant alleles with capital letters (e.g. *P*) and recessive alleles with lower-case letters (e.g. *p*), we can construct a 2 x 2 matrix (called a Punnett square) to determine all possible combinations of alleles in the F_2 offspring. Random combination of gametes results in a 3:1 ratio of offspring expressing single dominant-gene traits.

The same principle can be applied to predicting the inheritance of multiple genes. Consider all possible outcomes from mating hybrids that are heterozygous for 2 genes (*Aa* and *Bb*). You should construct a 4 x 4 matrix to determine all 16 possible mating outcomes. If both genes are dominant, the ratio of phenotypes will be 9:3:3:1.



Campbell et al. 2006; Fig. 14.5

Non-Mendelian inheritance



Not all inheritance occurs according to strict Mendelian rules of allele segregation (in gametes) and random combination (in offspring). During the process of meiosis when paired chromosomes replicate, one arm from each pair can reciprocally exchange segments in a phenomenon known as chiasma, or crossing-over (shown in the attached diagram). This means that the parent cell has re-assorted its genetic material, so the resultant gametes (and subsequent offspring) will have novel combinations of genes. This is known as genetic recombination, the production of offspring with combinations of traits differing from those found in either parent. Crossing-over events particularly affect linked genes, that is, genes located close together on the same chromosome that tend to be inherited together. When a chiasma occurs between linked genes, the predicted breeding outcomes deviate from those expected from Mendel's laws of independent assortment.

Parental types (where no crossing-over has occurred) will occur with higher frequency than recombinant types (where crossing-over events have occurred, one event being more probable than two, and two more probable than three, etc). The frequency of recombination was found to be related to the distance between the genes on the chromosome: the shorter the distance, the lower the probability of a cross-over occurring between them and the lower the possibility of recombination. Geneticists can measure the percentage frequency of recombinant offspring and use the data to construct genetic (or linkage) maps, ordered lists of genetic loci along a particular chromosome. The distances between genes are expressed as map units (centimorgans), with one map unit equivalent to a 1% recombination frequency.

Population genetics

When considering population genetics, it should be apparent that genes and allele frequencies can change over time, resulting in evolutionary change. However, change must always be measured against a standard benchmark, in this case, a gene pool that is not evolving. The Hardy-Weinberg theorem states that the frequencies of alleles and genotypes in a population's gene pool remain constant from generation to generation, provided that the rules of Mendelian genetics apply. The population is then said to be in a state of Hardy-Weinberg equilibrium.

Consider the inheritance of a single gene in a population of organisms. The gametes are haploid, so they only carry one allele. Let the frequency of allele *A* = 0.8 and that of allele *a* = 0.2. Assuming random matings, we can calculate the frequencies of all three possible genotypes.

- the probability of homozygous *AA* is 0.8 x 0.8 = p x p = p^2 = 0.64 (= 64%)
- the probability of homozygous *aa* is 0.2 x 0.2 = q x q = q^2 = 0.04 (= 4%)
- the probability of heterozygotes *Aa* and *aA* is 2(0.8 x 0.2) = 2(p x q) = 2pq = 0.32 (=32%)

The allele frequencies of the gametes (p + q for males, and p + q for females) have been used to calculate the genotype frequencies of the offspring according to the algebraic equation:

$$(p + q) \times (p + q) = p^2 + 2pq + q^2$$

The sum of genotype frequencies equals one, so the equation for Hardy-Weinberg equilibrium states that at a locus with two alleles, the three genotypes will appear in the proportions:

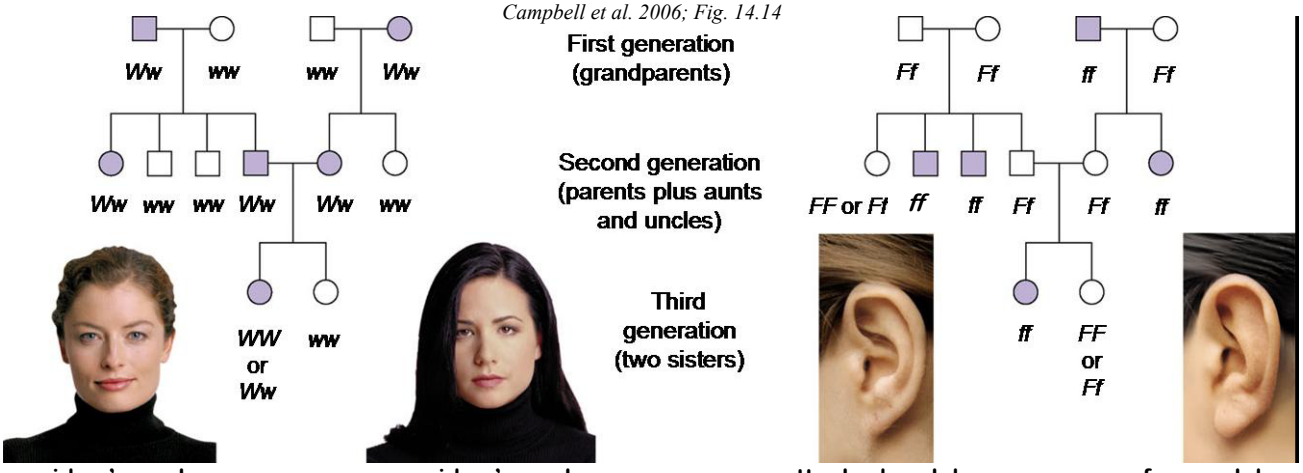
$$p^2 + 2pq + q^2 = 1$$

The allele and genotype frequencies in our hypothetical population will remain constant generation after generation, but only when five conditions are met (which occurs rarely in nature):

- random mating (disrupted when individuals select mates, especially related ones)
- large population size (small populations show chance fluctuations, known as genetic drift)
- no gene flow (must be a closed population, no allele transfer between populations)
- no mutations (gene insertions, deletions, substitutions modify the gene pool)
- no natural selection (differential survival and reproductive success of individuals changes allele frequencies)

Pedigrees

The inheritance patterns of particular traits can be traced and described using pedigrees. They have long been used to document births, deaths and marriages, particularly in royal families, but their contemporary use for genetic screening and counseling involves tracing the inheritance of alleles through several generations and predicting mating outcomes. Genes of particular interest are those associated with harmful genetic conditions, manifest by anatomical disorders or physiological, immunological or metabolic diseases. It is imperative that the pattern of inheritance of the condition be known: that is, whether the trait is dominant (expressed by homozygous dominant *DD* as well as heterozygotes *Dd*) or recessive (only expressed by homozygous recessive *dd*). An example of each pattern is illustrated below (circles represent females, squares represent males, shaded symbols indicate expression of trait, horizontal lines indicate matings, vertical lines show offspring).



The determination of relationships between living organisms has not only led to our understanding of genetics and inheritance, but it also provides the basis for the taxonomic classification of all life-forms and the study of their evolution (phylogeny).

Question:
Create a Punnett square to determine the frequency of the recessive trait albinism (lack of skin pigmentation) resulting from mating a heterozygous male with a heterozygous female.

Answer:

This is an example of dominant-recessive inheritance involving 2 alleles of 1 gene.
The Punnett square is 2x2:

		MALE	
		A	*a*
FEMALE	*A*	*AA*	*Aa*
	a	*Aa*	*aa*

The predicted frequency of normal phenotype (*AA* or *Aa*) = 3/4
The predicted frequency of albino phenotype (*aa*) = ¼

Question:
Although diploid organisms can only inherit 2 alleles for each gene, some genes have more than 2 alleles, leading to a phenomenon known as multiple-allele inheritance. For example, 3 alleles determine the ABO blood groups in humans (but each of us only inherits 2 alleles), I^A, I^B, and *i* (the former 2 are co-dominant and both are expressed when present, while the latter is recessive). If we assume the 3 alleles occur with equal frequency, construct a Punnett square to calculate the frequency of the blood groups A, B, AB, O.

Answer:

This is an example of dominant-recessive inheritance involving 3 alleles of 1 gene.
The Punnett square is 3x3:

	I^A	I^B	*i*
I^A	$I^A I^A$	$I^A I^B$	$I^A i$
I^B	$I^A I^B$	$I^B I^B$	$I^B i$
i	$I^A i$	$I^B i$	*i i*

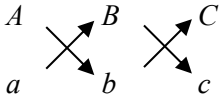
The predicted frequency of A blood group (I^AI^A, I^Ai) = 3/9
The predicted frequency of B blood group (I^BI^B, I^Bi) = 3/9
The predicted frequency of AB blood group (I^AI^B) = 2/9
The predicted frequency of O blood group (*ii*) = 1/9

Question: Three alleles (in random order *a, b, c*) are linked on a chromosome of a plant. An heterozygous hybrid [*AaBbCc*] was crossed with a recessive [*aabbcc*] and the types and numbers of gametes were recorded, as follows:

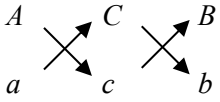
ABC	414	*aBc*	28	*abc*	386	*AbC*	20
aBC	80	*ABc*	1	*Abc*	70	*abC*	1

What is the order of these genes on the chromosome, *abc, acb* or *bac*?

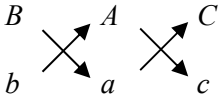
Answer: For each possible gene order, you need to consider the sequence of decreasing frequencies of gametes having 0, 1 or 2 cross-overs (frequent, infrequent and rare, respectively). This is best visualized by manually working out the results of 2 chiasmas for each combination:



yields *AbC* and *aBc*



AcB and *aCb*



BaC and *bAc*

Only the middle combination supports the very low frequencies of *Abc* (1) and *abC* (1),
so the correct order of the 3 genes on the chromosome is *acb*

Question: Using the data from the previous question, what is the correct distance between the three loci (in map units)?

	distance between *a* and *c*	distance between *b* and *c*
A.	2.1	2.4
B.	3.4	5.0
C.	5.0	15.2
D.	15.2	3.4
E.	15.2	5.0

Answer: To construct a linkage map based on recombination frequencies, you add the frequencies where gametes exhibit crossing-over events, and convert them to map units (1 unit = 1% recombination frequency).

In this case, the distance between *a* and *c* is equal to the frequency of gametes exhibiting cross-overs *Ac* (1 + 70) plus *aC* (1 + 80) which totals 152 (converts to 15.2 map units).
The distance between *b* and *c* is equal to the frequency of gametes exhibiting cross-overs *Bc* (1 + 28) plus *bC* (1 + 20) which totals 50 (converts to 5 map units).

The correct answer is therefore E (which can be mapped as follows):

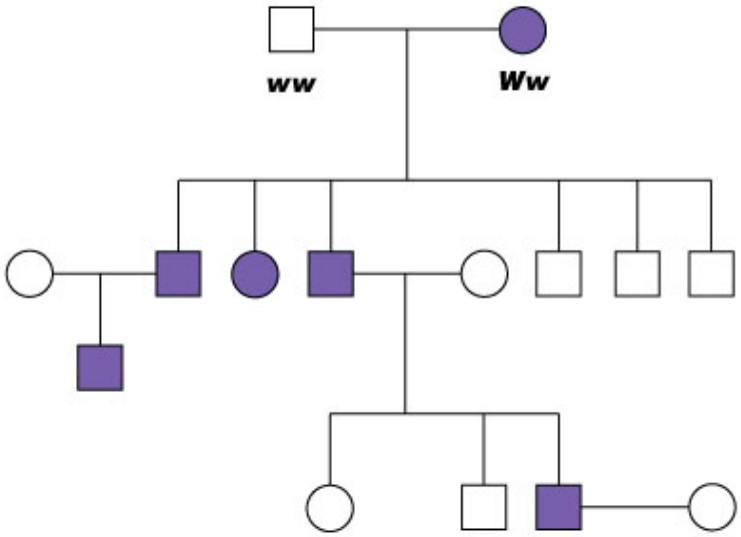
----------*a*----------*c*----------*b*----------
15.2 5.0

Question: Let the frequency of allele *A* = 0.6 and that of allele *a* = 0.4.
Assuming random matings, what are the frequencies of all three possible genotypes.

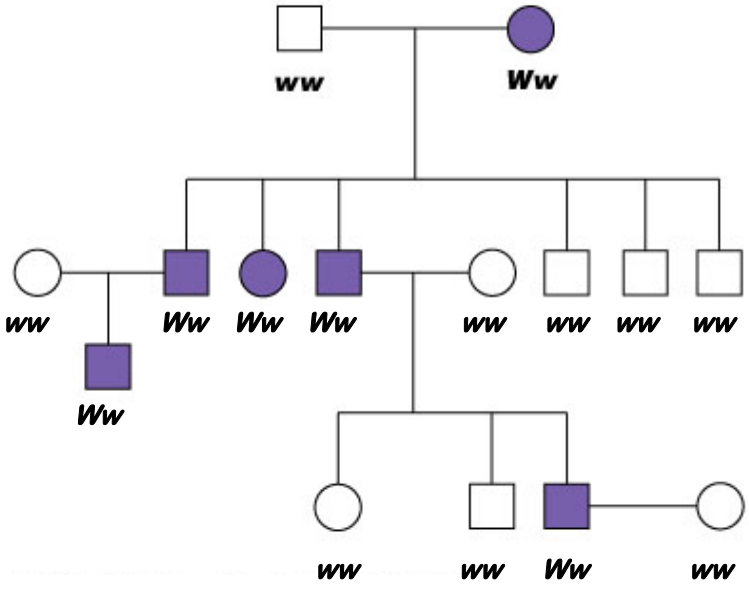
Answer: The Hardy-Weinberg equilibrium model predicts the genotype frequencies will conform to the quadratic equation $p^2 + 2pq + q^2 = 1$

the probability of homozygous *AA* is 0.6 x 0.6 = p x p = p^2 = 0.36 (= 36%).
the probability of homozygous *aa* is 0.4 x 0.4 = q x q = q^2 = 0.16 (= 16%).
the probability of heterozygotes *Aa* and *aA* is 2(0.6 x 0.4) = 2(p x q) = 2pq = 0.48 (= 48%)

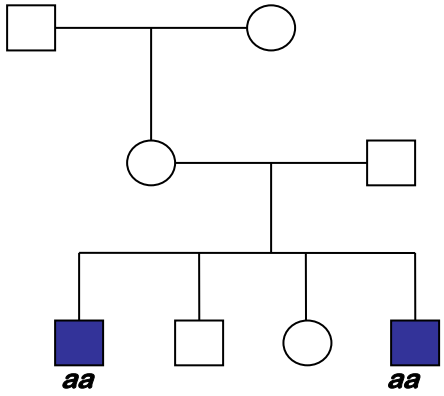
Question:
The following pedigree is for the dominant trait for woolly hair. Label the genotypes of all individuals on the pedigree chart.



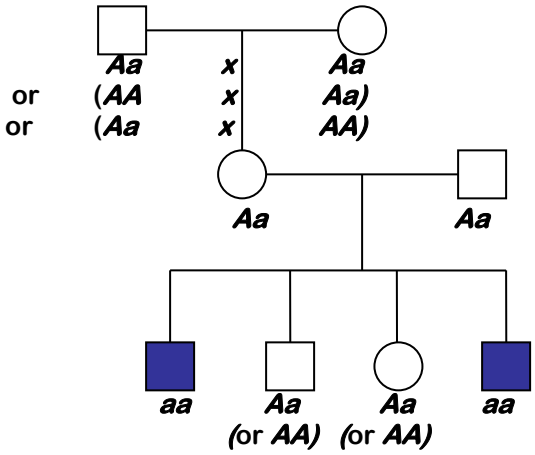
Answer:



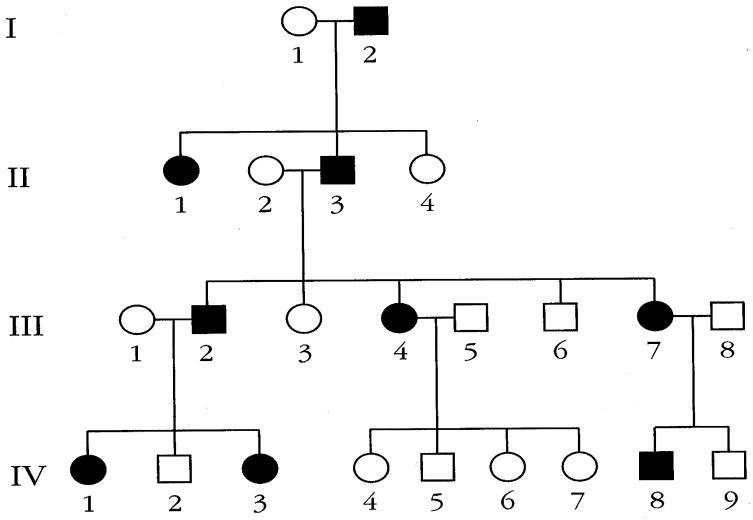
Question:
The following pedigree is for the recessive trait albinism. Label the genotypes of all individuals.



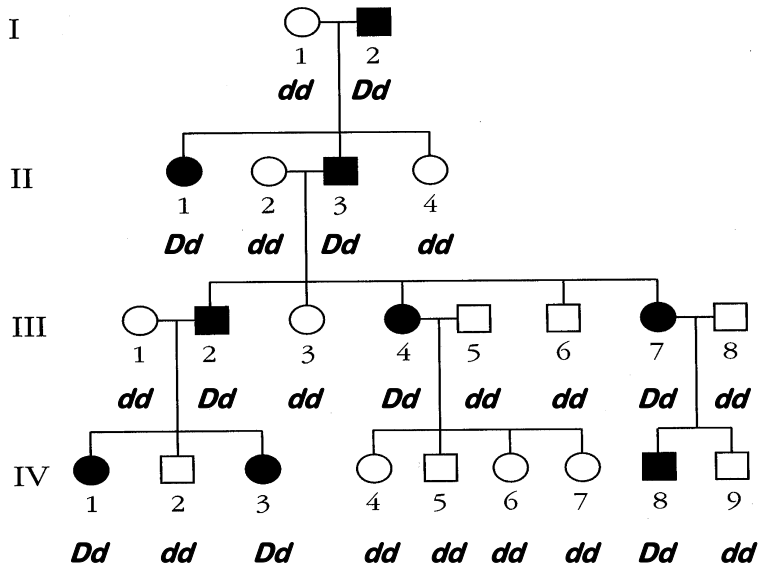
Answer:



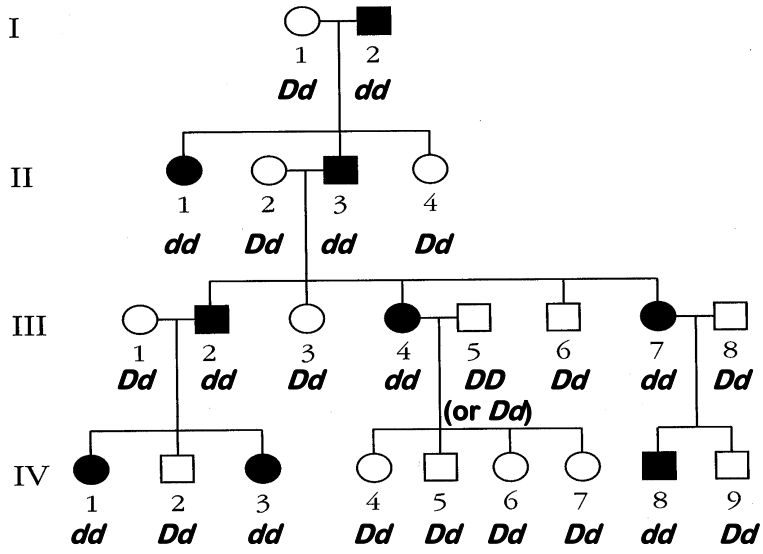
Question:
Work through the following pedigree to determine whether inheritance is dominant or recessive, and what the genotype (*DD*, *Dd* or *dd*) of individual III-5 would be in either case.



Answer:
If dominant:



If recessive:



The pedigree supports both patterns of inheritance.
(It cannot be used to differentiate between dominant or recessive inheritance).

If it is dominant, the genotype of III-5 is *dd*

If it is recessive, the genotype of III-5 is *DD*

Chapter 9.

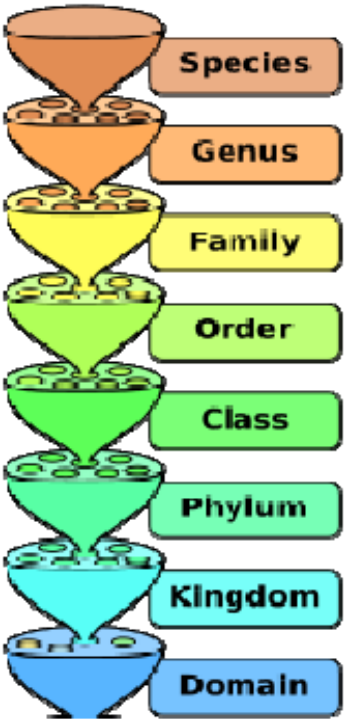
PHYLOGENY

Vocabulary list: taxonomy, classification, nomenclature, genus, species, biodiversity, species richness, systematics, homology, analogy, cladistics, monophyletic, paraphyletic, polyphyletic, phylogeny, phylogenetics, evolution, algorithm

Taxonomy

Taxonomy is the practice and science of cataloguing or classification, an important intellectual process which allows us to bring order and organization to store, sort and retrieve multiple items. Taxonomic schemes are composed of taxonomic units arranged in a hierarchical structure, typically related by subtype-supertype relationships.

Biological (Linnaean) classification is a method by which biologists group all living things according to their shared characteristics. Many characters have been used in classification systems, including elements of organismal structure, growth, development, ecology, behaviour, physiology, biochemistry and more recently, genotypic characters involving partial or complete gene sequences. Organisms may be known by many different common names but they all only have one scientific name, a two word name comprised of the genus name and species name. Humans are known scientifically as *Homo sapiens* (note the prescribed use of italics and capitals). This convention for naming species is referred to as binomial nomenclature, which is now regulated by International Codes in which 8 main taxonomic ranks are recognized: domain, kingdom, phylum, class, order, family, genus, species. Over 250 years, the number of kingdoms recognized by taxonomists has grown from two to six, mainly due to the recognition of new microbial assemblages. Domains are a relatively new grouping introduced in 1990. The majority of biologists accept the three-domain system, but often refer to 4-, 5- or 6-kingdom systems for historical or comparative purposes.



wikipedia, 2008 'species'

Linnaeus (1735)	Haeckel (1866)	Chatton (1937)	Copeland (1956)	Whittaker (1969)	Woese *et al.* (1977)	Woese *et al.* (1990)
2 kingdoms	3 kingdoms	2 empires	4 kingdoms	5 kingdoms	6 kingdoms	3 domains
-	Protista	Prokarota	Monera	Monera	Eubacteria	Bacteria
					Archaebacteria	Archaea
		Eukaryota	Protista	Protista	Protista	Eukarya
Vegetabilia	Plantae			Fungi	Fungi	
			Plantae	Plantae	Plantae	
Animalia	Animalia		Animalia	Animalia	Animalia	

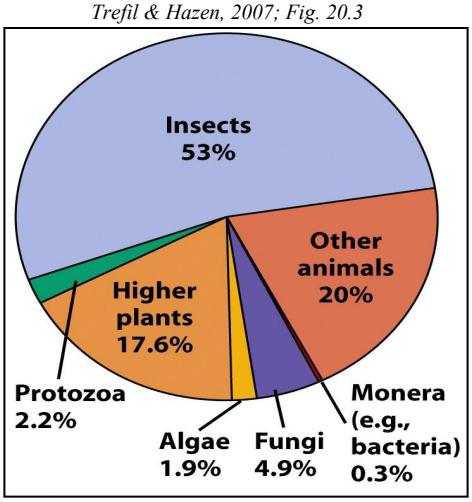
Species concept

The basic unit in biological classification is the species, commonly defined as an interbreeding population of individual organisms that produces fertile offspring. The biological species concept emphasizes reproductive isolation, the existence of biological factors (barriers) that impede two members from producing viable fertile hybrids. Such barriers are classified into pre-zygotic barriers which impede mating (habitat, temporal or behavioural isolation) or hinder fertilization (mechanical or gametic isolation) and post-zygotic barriers which prevent hybrid zygotes from maturing (reduced viability, fertility, survival). The biological species concept (based on separateness) is widely accepted in biology, but it does not apply to all organisms, e.g. there is no way to evaluate the reproductive isolation of fossils or asexual organisms such as prokaryotes and some lower eukaryotes. In these cases, other definitions are applied which emphasize the unity within a species, such as the morphological, paleontological, ecological and phylogenetic species concepts.

Biological diversity

Biodiversity has three main components or levels, genetic diversity, species diversity and ecosystem diversity. Genetic diversity encompasses the breadth of genetic variation occurring in all individuals within and between populations that are often associated with adaptations to local conditions. Species diversity refers to the variety of species (species richness) in an ecosystem or throughout the entire biosphere. Ecosystem diversity considers the variety of the biosphere's ecosystems where interactions among populations of different species can influence species and genetic diversity (e.g. the local extinction of a keystone predator can have a negative impact on the overall species richness of the community).

Trefil & Hazen, 2007; Fig. 20.3



Global biodiversity (the total number of species on Earth) has been estimated to range from 3-30 million species. While the numbers of large animal species are well known and only a few new species are discovered each decade, the situation for other fauna and flora are poorly known. Scientists frequently discover new species, particularly insects and other invertebrates, plants and microorganisms. Unfortunately, there is no central database to record information on biodiversity, so there are not even firm numbers on current extant species, with counts ranging from 1.5-1.9 million species. One recent study pronounced global biodiversity as N = 1,749,577 species.

Scientists have developed a number of methods for using current data to make estimates about global biodiversity. When the total number of known species in an ecosystem is plotted as a function of time (i.e. time of discovery), the resultant graph is an S-shaped (sigmoidal) curve. As time goes by and more species become known, the curve should flatten. If we assume that the discovery curve for insects, for example, will follow this same pattern, then by making a guess as to where we are on the curve, we can estimate what the final numbers will be. Another estimation technique is to do an exhaustive survey of organisms in a small geographic area, determine the ratio of known to unknown species present, and then assume that this ratio applies worldwide. Global estimates vary considerably and involve making some broad and generous assumptions, but it is imperative that we complete a census against which to assess future change, especially the impact of anthropogenic activities.

Species diversity

Species richness is the number of different species (*S*) in a given area. Estimates of *S* are typically used in conservation studies to score the biodiversity of specific ecosystems and predict their responses to change. Globally, there is a strong inverse correlation between species richness and latitude: the farther from the equator, the fewer species are found. The peak of species richness is not at the Equator, it is between 20-30°N. The gradient of species richness is also asymmetric about the Equator (increasing rapidly to the North but decreasing slowly to the South). The latitudinal gradient in species richness may be due to differences in available energy and resources (e.g. more solar radiation, land, water and minerals at lower latitudes). However, some studies indicate that areas of high primary productivity actually have lower species richness because a few species predominate and outcompete less productive species. Other variations in species richness have been attributed to altitudinal gradients, habitat diversity, biogeographic isolation and sampling anomalies (including sampling effort, area sampled, and numbers detected). The most common formula for calculating species diversity is the Simpson's diversity index:

$$D = \frac{\sum n(n-1)}{N(N-1)}$$

where D = diversity index
N = total number of organisms of all species found
n = number of individuals of a particular species

A high D value suggests a stable or ancient site, while a low D value suggests recent colonization or reduced diversity due to perturbation (e.g. through pollution or land development). In order to account for the probability of missing some species during sampling, the Jacknife estimate may be used:

$$S = n + \left(\frac{n-1}{n}\right)^k$$

where S = species richness
n = total number of species present in sample population
k = number of 'unique' species (of which only one organism was found in sample)

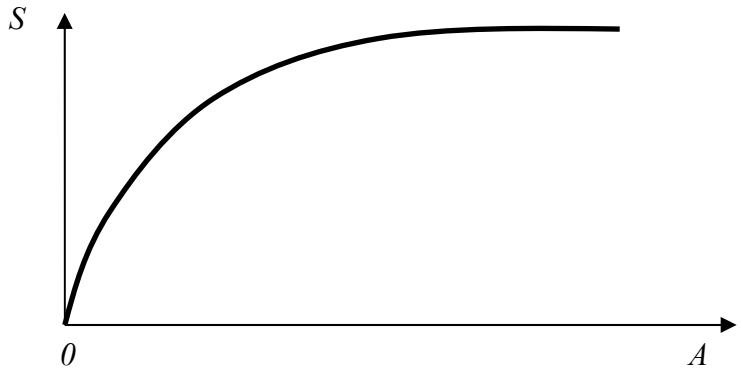
Species-area curves

The number of species found in a particular habitat is also dependent on the area of that habitat. The species-area relationship is very important in experimental ecology as it facilitates extrapolation of habitat biodiversity from data collected in smaller scale surveys, and also gives an estimate of the sampling effort required to produce meaningful results. The species-area relationship appears as a curve which can be modeled by the power function:

$$S = cA^z$$

where S = number of species
c = constant
A = area
z = constant (power)

When S and A are plotted on arithmetic axes, the resultant curve has a characteristic shape; increasing rapidly but then flattening to a plateau. When the plateau is reached, collecting samples from a greater area may result in the addition of only a few more species, so the additional sampling effort may not be warranted or justified. Estimation of the so-called 'minimal sampling area' is subjective, but many researchers define it as the area enclosing at least 95% of the total species found. The trouble with this estimation is that the total number of species is not known and the species-area curve does not usually approach an asymptote.

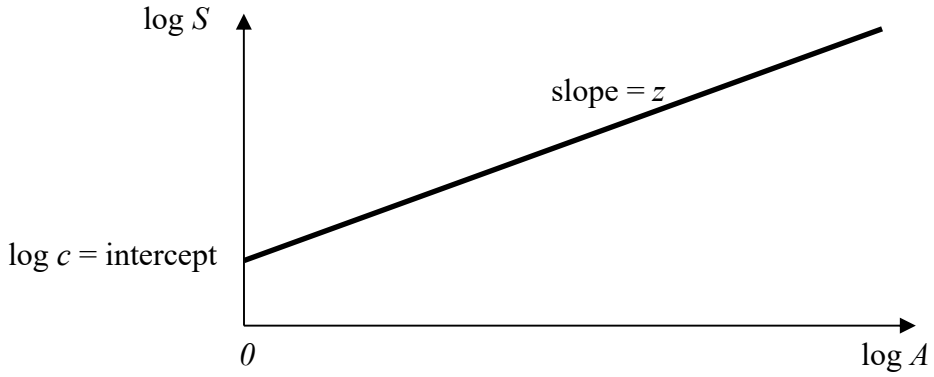


Researchers have used logarithmic transformations to simplify species-area relationships. When plotted on double log axes, the relationship between log S and log A appears linear. Logarithmic transformation turns the power function curve into a straight line with the equation:

$\log S = \log c + z \log A$

where $\log c$ = the y-intercept, and

z = the slope (gradient) of the straight line.



Ecologists can manipulate species-area model by changing values of not only of the variables but also the constants to predict what the impact of such changes will be on biodiversity. Many countries, municipalities, shires and councils now insist on environmental impact studies being conducted on specific habitats before approving rural or urban development projects. Power functions and logarithmic transformations of data play a leading role in such studies.

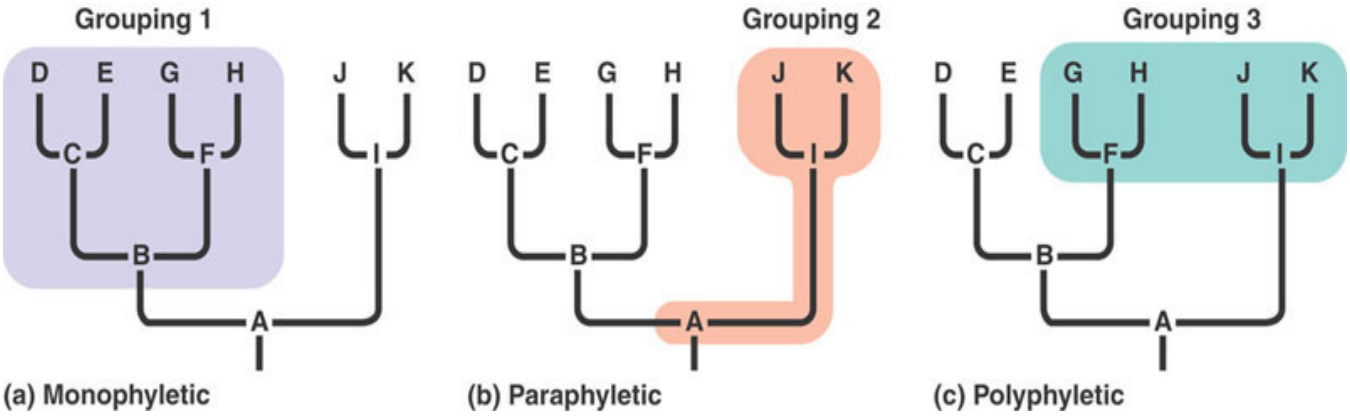
Appreciating biodiversity must not rely simply on species recognition, but should also involve assessing their genetic diversity, including species relatedness. Classification systems should therefore reflect kinship, that is, how organisms are linked through their genealogies.

Phylogeny

While early systems classified organisms for ease of identification, it is now generally accepted that classification should reflect the Darwinian principle of common descent. Species have an evolutionary history that explains the natural order of life. The most pertinent logic for classification is that of evolutionary relatedness. Phylogeny is the study of the evolutionary histories of groups of organisms. The processes of evolution are natural selection and other mechanisms that change the genetic composition of populations. Evolution is regarded as a branching process, whereby populations may speciate into separate branches, hybridize together, or terminate by extinction. Such processes are best visualized by the construction of phylogenetic trees, where branching orders can be shown and branch lengths can represent time. To reconstruct phylogeny, scientists use systematics, an analytical approach to understanding the diversity and relationships of living and extinct organisms. Evidence used to reconstruct phylogenies can be obtained from the fossil record, from morphological and biochemical similarities between organisms, and more recently, from comparisons of DNA and RNA nucleotide sequences within individual genes or even across entire genomes.

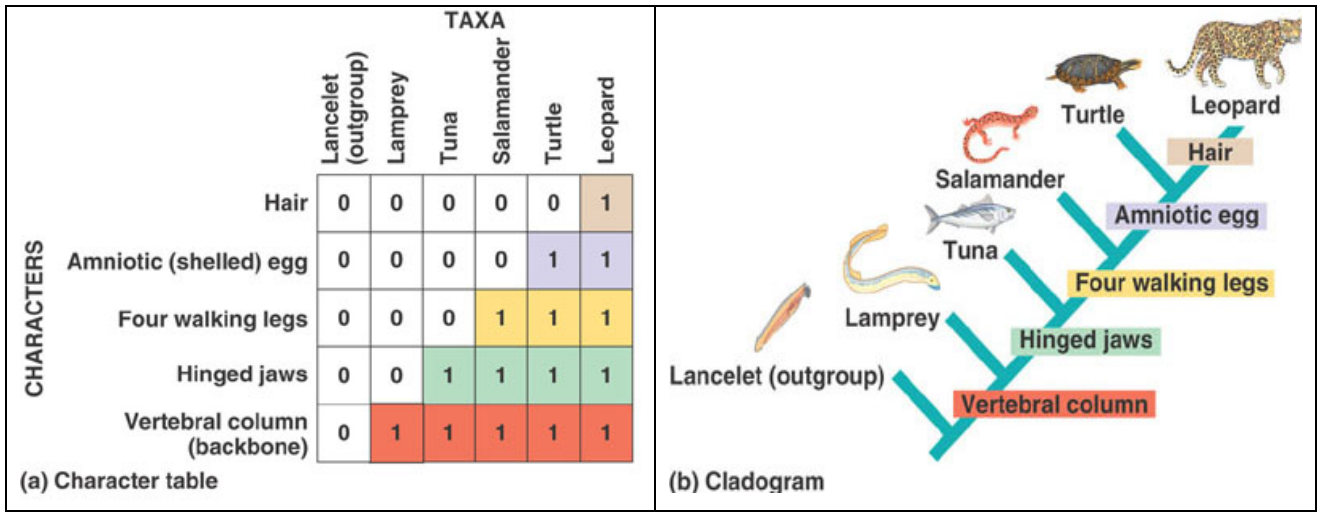
Similarity due to shared ancestry is called homology, while similarity due to convergent evolution is called analogy. When two organisms from different evolutionary lineages experience similar environmental pressures, natural selection may result in convergent evolution and analogous adaptations may arise. Distinguishing homology from analogy is critical in phylogenetic reconstruction. The more points of resemblance between two complex structures, the less likely it is that they evolved independently, e.g. the bones in human and chimpanzee skulls match almost perfectly, so it more likely they were inherited from a common ancestor, and less likely they had separate origins. The same principle applies to comparing gene sequences, if two organisms have closely similar sequences, it is highly likely that the genes are homologous.

Classification and phylogeny are linked. Since the 1960s, a trend called cladistic taxonomy (or cladistics) has emerged, arranging taxa in an evolutionary tree (called a cladogram). This system recognized clades as groups related by descent. If a clade included all the descendants of some common ancestral form, it was considered monophyletic, as opposed to paraphyletic (including some, but not all, descendants) or polyphyletic (derived from more than one ancestral taxon).



Campbell et al. 2006; Fig. 25.10

In addition to distinguishing between homology (shared ancestry) and analogy (convergent evolution), systematists must also differentiate homologous characters into shared derived characters (unique to a particular clade) and shared primitive characters (found in the clade being analysed as well as in older clades). A key step in cladistic analysis is outgroup comparison, which is used to differentiate them (only shared derived characters are useful for establishing relationships). Systematists need to identify an outgroup, a species closely related to those under study (ingroup), but not as much as they are to each other. It is assumed that any homologies shared by the ingroup and outgroup are primitive characters that were present in a common ancestor, while those present in some or all of the ingroup are derived characters which evolved later. For instance, to study the relationships among 5 vertebrates (leopard, turtle, salamander, tuna and lamprey), selecting a lancelet as the outgroup is a good choice as it lacks a backbone, unlike those in the ingroup. The presence of a notochord in lancelets and the embryos of the ingroup is a shared primitive character, and therefore not useful for sorting out ingroup relationships. The presence of jaws, absent in lampreys and present in the other ingroup taxa, helps to identify the earliest branch in the vertebrate cladogram. Analysing the taxonomic distribution of homologies enables us to identify the sequence in which derived characters evolved within the ingroup (cf. character table and cladogram).

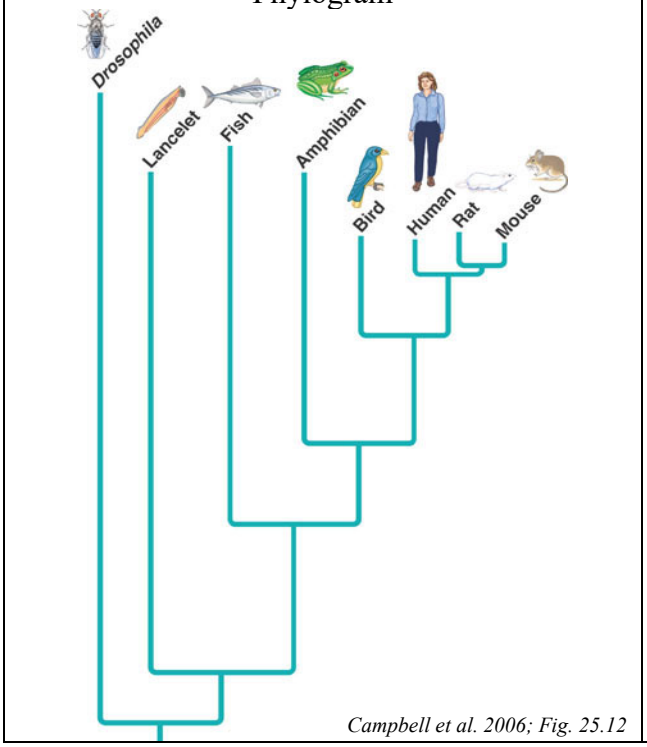


Campbell et al. 2006; Fig. 25.11

A cladogram is not a phylogenetic tree. To convert it into one, we need more information from sources such as the fossil record, which can indicate when and in which groups the characters first appeared. Any chronology represented by the branching pattern of a phylogenetic tree is relative (earlier versus later) rather than absolute (so many millions of years ago).

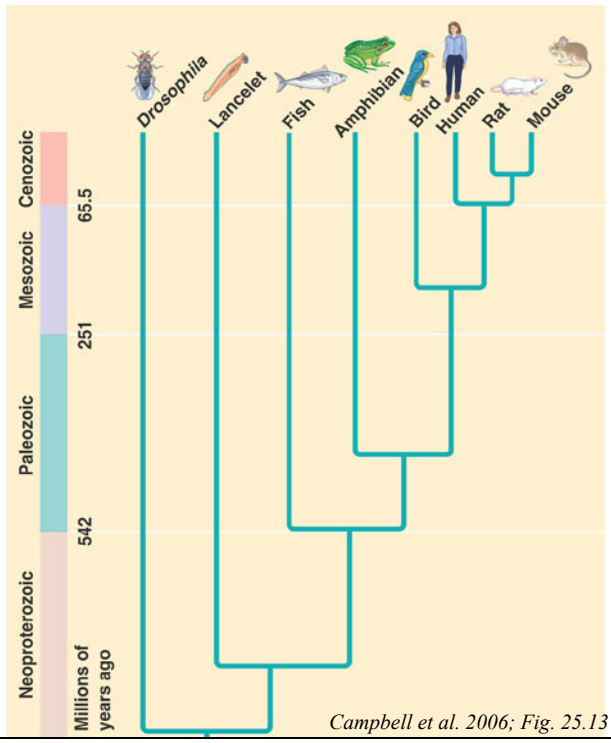
Some kinds of tree diagrams can be used to provide more specific information about timing. In a phylogram, the length of a branch reflects the number of genetic changes that have taken place in a particular DNA or RNA sequence in a lineage. Even though the branches in a phylogram may have different lengths, all the different lineages that descend from a common ancestor have survived for the same number of years. These equal amounts of chronological time are represented in an ultrametric tree, which has the same branching pattern as the phylogram, but all branches that can be traced from the common ancestor to the present are of equal lengths. Ultrametric trees do not contain the information about different evolutionary rates that can be found in phylograms. However, they draw on data from the fossil record to place certain branch points in the context of geological time.

Phylogram



Campbell et al. 2006; Fig. 25.12

Ultrametric tree



Campbell et al. 2006; Fig. 25.13

Tree building

Various mathematical procedures are used to create trees and test them for robustness. The principles of maximum parsimony and maximum likelihood help systematists reconstruct phylogeny. As data sets increase in size, it becomes more difficult to draw the phylogenetic tree that best describes evolutionary history. If you are analysing data for 4 species, there are 15 different ways to form a tree. If you are analysing 50 species, there are ~50^{45} (= 3×10^{76}) ways to form a tree. According to the principle of maximum parsimony, we look for the simplest explanation that is consistent with the facts. In the case of a tree based on morphological characters, the most parsimonious tree is the one that requires the fewest evolutionary events to have occurred in the form of shared derived characters. For phylograms based on DNA sequences, the most parsimonious tree requires the fewest base changes in DNA. The principle of maximum likelihood states that, given certain rules about how DNA changes over time, a tree should reflect the most likely sequence of evolutionary events.

Phenogram

Do classification systems reflect evolutionary relationships? Many are purported to do so, but it all depends on the characters analysed! Let's construct a phenogram (build a tree) using the average linkage cluster method (which effectively underpins all phylogenetic algorithms)

- Select any four people you know. Identify them individually as taxa A, B, C and D.
- Select four characters (traits) to differentiate them; such as gender, complexion, height, weight
- Identify 2 states for each character, gender (male, female); complexion (fair, dark); height (large, small); weight (beefy, winsome)

- Now create a 4x4 matrix of taxa versus characters

		TAXA			
		A	B	C	D
Character	Gender (m/f)				
	Complexion (fr/dk)				
	Height (lge/sm)				
	Weight (be/wi)				

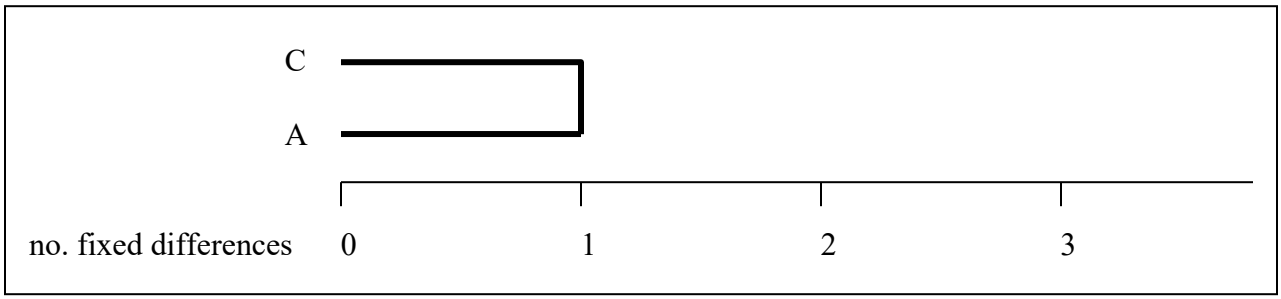
- Score the matrix by filling out the character states for each individual
  e.g.

		TAXA			
		A	B	C	D
Character	Gender (m/f)	m	m	m	f
	Complexion (fr/dk)	fr	dk	dk	fr
	Height (lge/sm)	lg	lg	lg	sm
	Weight (be/wi)	wi	be	wi	be

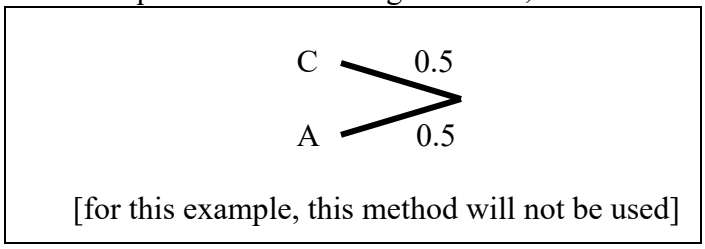
- This data is now converted into a 4 x 4 matrix (taxa versus taxa) showing the numbers of fixed differences between individuals, e.g. A and B differ at 2 character states, etc

Fixed differences		TAXA			
		A	B	C	D
Taxa	A	0	2	**1**	3
	B		0	1	3
	C			0	4
	D				0

- Select the two taxa with the fewest differences, in this case, start with A and C which have one (1) fixed difference. Use this data to start tree-building. Plot taxa against distance.



Alternatively, if the fixed differences are considered to be genetic distances, the taxa can be plotted with equidistant bifurcating branches, i.e. 0.5 units each

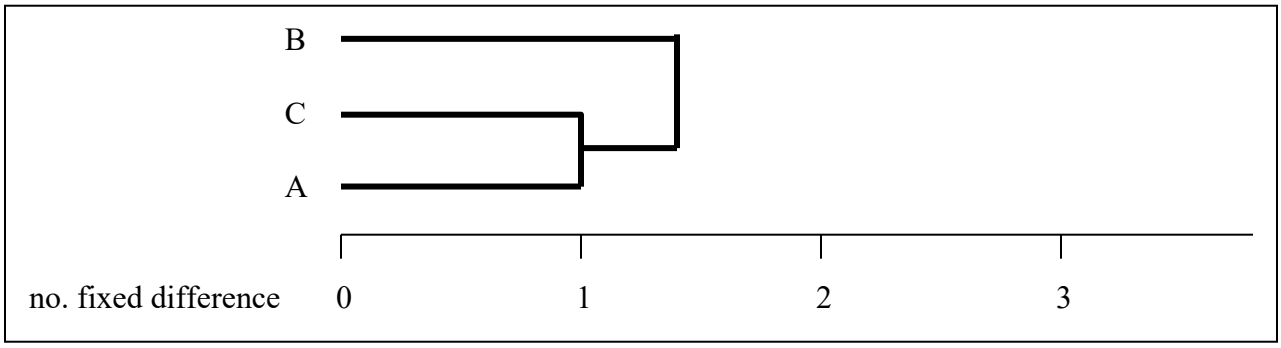


[for this example, this method will not be used]

- Now collapse the 4x4 matrix into a 3x3 matrix by averaging differences between AC and B and D e.g. 2 differences between A & B, plus 1 difference between C & B, gives (2+1) differences, so average difference between AC & B = (2+1)/2 = 1.5

Fixed differences		TAXA		
		AC	B	D
Taxa	AC	0	**(2+1)/2=1.5**	(3+4)/2=3.5
	B		0	3
	D			0

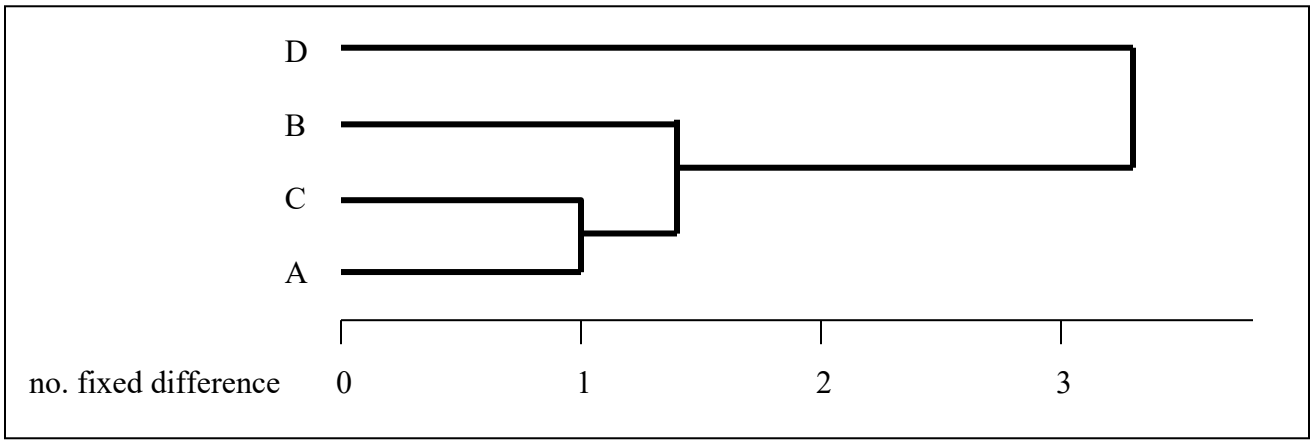
- Find the next least difference, in this case it is between AC & B (so plot taxon B)



- Now collapse the 3x3 matrix into a 2x2 matrix by averaging differences between ACB and D

Fixed differences		TAXA	
		ACB	D
Taxa	ACB	0	(3.5+3)/2=3.25
	D		0

- Plot final taxon D



The resultant tree shows the relationships between the individuals inferred from the characters analysed. The tree can be variously interpreted e.g. A, B, C are closely-related (shows how common males are!), D is a distant relative (shows how far removed females are!), D is ancestral (shows how essential females are/were?), etc. Do the relationships revealed have any actual biological or genetic basis?

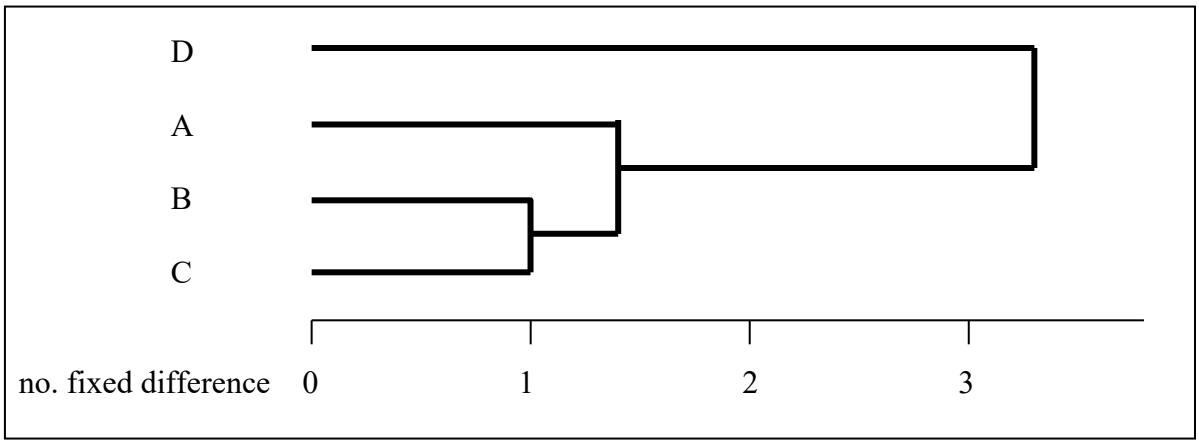
From our simple little example, we should draw TWO main conclusions:

1. Not all characters are phylogenetically informative!

   Garbage in, garbage out! The relationships revealed in a tree are only as good as the assumptions made in character selection. The above example made the assumptions that gender, complexion, height and weight were phylogenetically informative. While they may have a genetic basis (sex chromosomes, complexion genes, etc), the four characters examined do not reveal any evolutionary relationships. The inclusion of significantly more characters (hundreds) in modern phylogenetic analyses, including cladistic and molecular analyses of phenotypic and genotypic characters, still make the same assumptions. Whether we look at 50 biological characters (structure, function, development, ecology, etc) or 1,000 molecular characters (single nucleotides in the genetic code), we must interpret any phylogenetic inferences cautiously as not all characters may be informative: some may be divergent; some may be convergent; and many may be misleading. Indeed, many molecular phylogenies differ depending on the genes and gene sequences examined.

2. Multiple trees are often generated.

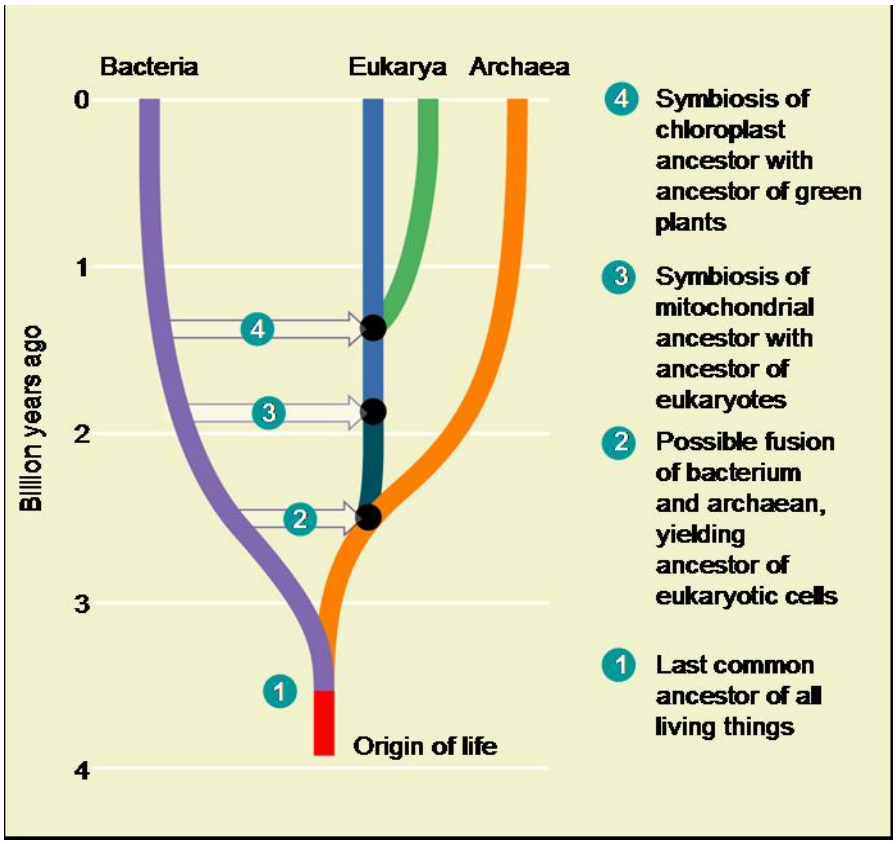
   Which tree is best? Depending on the data, many trees differing in topography (branching order) can often be generated. In the example above, an alternative tree can be generated from the data - start building from a different point e.g. the difference between B & C instead of A & C (both give only one fixed difference). This would generate the following tree:



The occurrence of trees of differing topographies is an important concept in phylogenetic reconstruction. Numerous alternate trees can be generated from large data sets. We must then look for consensus trees, those supported by good boot-strap replicate scores, using other algorithms.

Gene transfer

Organisms can generally inherit genes in two ways: from parent to offspring (vertical gene transfer), or by horizontal or lateral gene transfer, in which genes jump between unrelated organisms, a common phenomenon in prokaryotes. Lateral gene transfer has complicated the determination of phylogenies of organisms since inconsistencies have been reported depending on the gene chosen. Carl Woese came up with the three-domain theory of life (eubacteria, archaea and eukaryotes) based on his discovery that the genes encoding ribosomal RNA are ancient and distributed over all lineages of life with little or no lateral gene transfer. Therefore rRNA are commonly recommended as molecular clocks for reconstructing phylogenies. This has been particularly useful for the phylogeny of microorganisms, to which the species concept does not apply and which are too morphologically simple to be classified based on phenotypic traits.



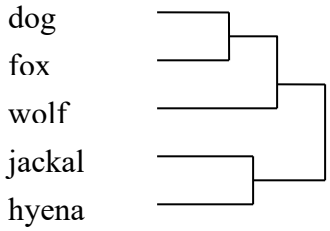
Molecular systematics helps us to understand phylogenetic relationships that cannot be measured by comparative anatomy and other non-molecular methods. Gene sequence analyses allow us to compare the variation occurring between widely divergent taxa (e.g. mammals and bacteria) as well as within individual species (e.g. microbial strain variation). The ability of molecular trees to encompass both short and long periods of time is based on the fact that different genes evolve at different rates. For example, the DNA that codes for ribosomal RNA (rRNA) changes relatively slowly, so comparisons of DNA sequences in these genes can be used to sort out relationships between taxa that diverged hundreds of millions of years ago. In contrast, mitochondrial DNA (mtDNA) evolved relatively recently and can be used to explore recent evolutionary events, such as relationships between groups within a species. The search for phylogenetic trees that are parsimonious and likely has been facilitated largely by computers and the newly emergent field of bioinformatics.

Question. Dogs, foxes, jackals, wolves and hyenas were all scored for the presence (+) or absence (-) of seven phenotypic characters (I-VII). The results are shown in the following Table.

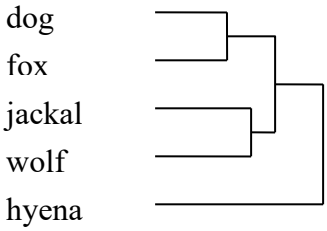
Character	I	II	III	IV	V	VI	VII
Dog	+	+	+	+	+	+	+
Fox	-	+	+	-	+	+	-
Jackal	+	-	+	-	+	-	-
Wolf	+	+	+	-	+	+	+
Hyena	-	-	-	+	-	-	-

Which phenogram indicates the phylogenetic relationships between these five animal groups based on the data given?

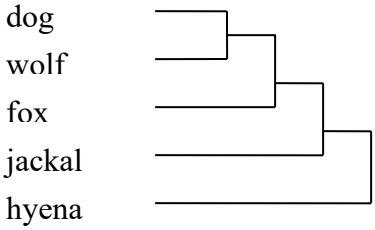
A.



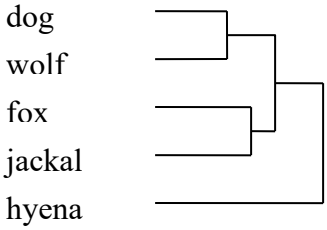
B.



C.



D.

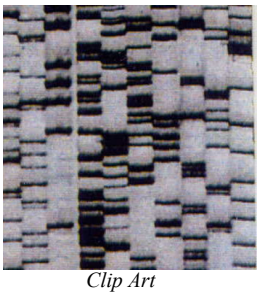


Answer: Construct matrix of fixed differences in characters states between taxa:

	Dog	Fox	Jackal	Wolf	Hyena
Dog	0	3	4	1	6
Fox		0	3	2	5
Jackal			0	3	3
Wolf				0	7
Hyena					0

Find smallest fixed difference; in this case, it is between dog and wolf.
Continue tree-building to reveal answer (i.e. phenogram shown in C)

Chapter 10.



Clip Art

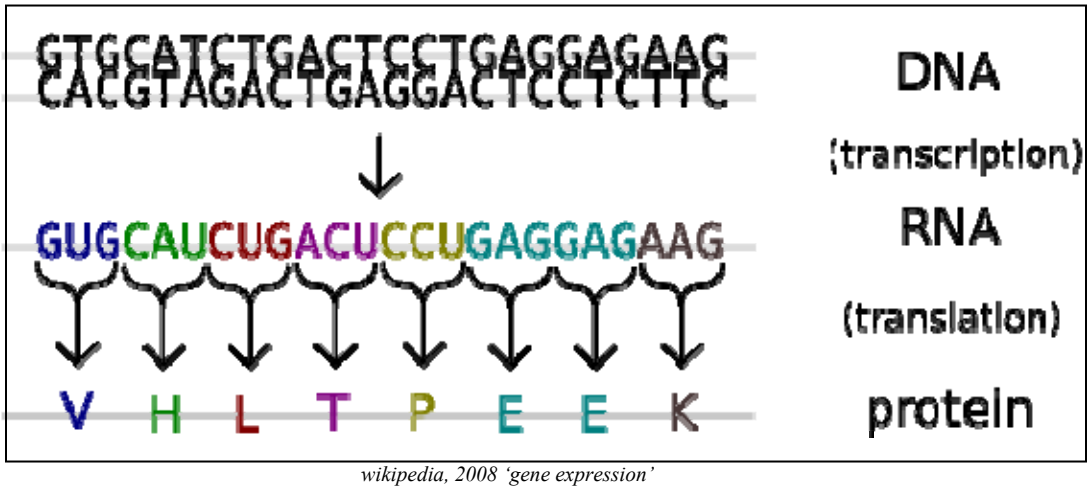
BIOINFORMATICS

Vocabulary list: bioinformatics, computational biology, sequences, nucleotides, DNA, RNA, transcription, protein, translation, strings, searches, insertions, deletions, mutations

Bioinformatics and computational biology (sometimes called systems biology) involve the use or development of techniques, including applied mathematics, informatics, statistics, computer science, artificial intelligence, chemistry, and biochemistry to solve biological problems, usually on the molecular level. Major research efforts include sequence alignment, gene finding, genome assembly, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, and the modeling of evolution.

Sequence data

The central dogma of molecular biology postulates a linear information flow between genetic material (DNA), RNA and protein molecules, which are responsible for all major biological functions within a cell. The information contained in sections of a DNA molecule (as sequences of G, C, A, T nucleotides) is first re-coded into messenger RNA molecules by a process called transcription. RNA is chemically similar to DNA and is also composed of four nucleotides. Important differences are that the nucleotide thymine in DNA is replaced by uracil (U) in RNA, and the deoxyribose in DNA is replaced by ribose in the backbone of RNA (making the RNA molecule less stable and subject to degradation with time). The RNA molecules are then translated by ribosomes into proteins, which are biopolymers composed of chains of amino acids. There are 20 different types of amino acids which are encoded by various combinations of nucleotide triplets (codons).



wikipedia, 2008 'gene expression'

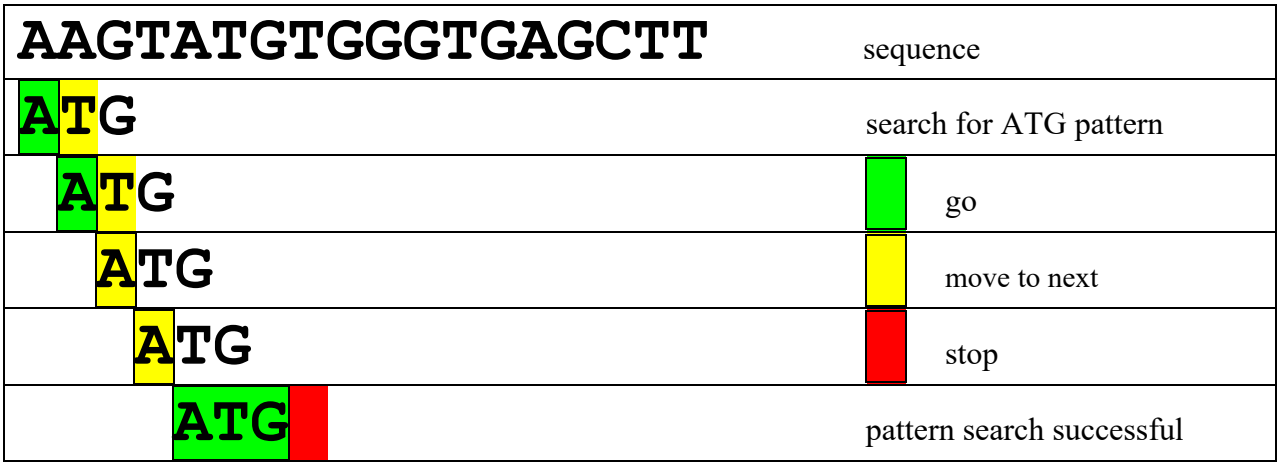
Millions of DNA sequences have been decoded over the last decade from hundreds of organisms, primarily using the so-called shotgun sequencing technique which does not give a sequential list of nucleotides, but rather the sequences of thousands of small DNA fragments (each about 600-800 nucleotides long). The ends of these fragments overlap and, when aligned in the right way, make up the complete genome. Shotgun sequencing yields sequence data quickly, but the task of assembling the fragments can be quite complicated for larger genomes. The only practical way to search the databases to facilitate sequence comparisons is to use sophisticated computer programs.

Software tools for bioinformatics range from simple command-line tools to more complex graphical programs available from various institutions. The National Center for Biotechnology Information (NCBI), US National Institutes of Health, provides a popular web-based implementation that searches their databases. The computational tool best-known among biologists is BLAST, an algorithm for determining the similarity of arbitrary sequences against other sequences, generally from curated databases of protein or DNA sequences. These programs compensate for mutations in DNA sequences, in order to identify strings of sequences that are related, but not necessarily identical.

Strings

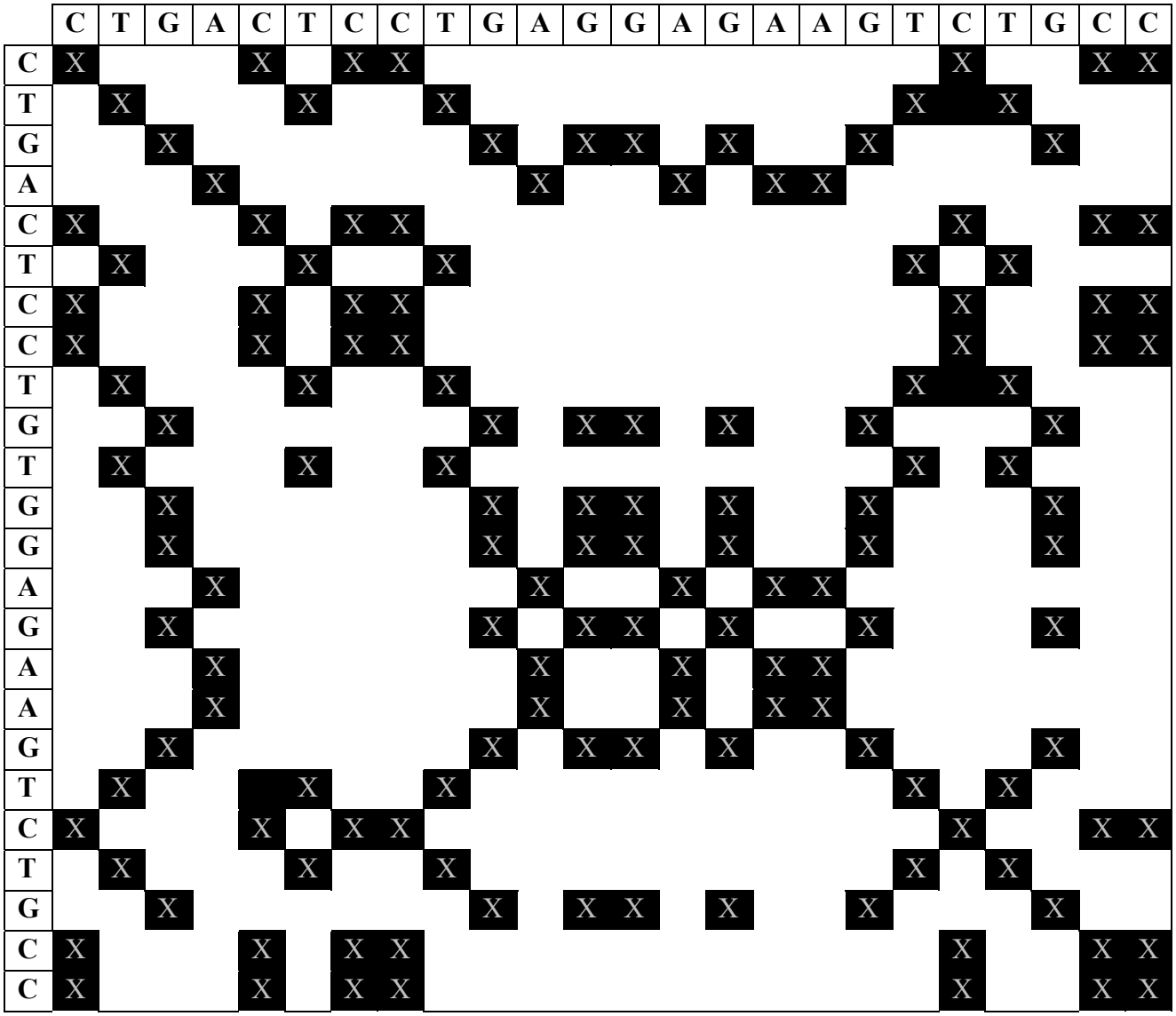
Strings of characters are the most basic data structure to store information. We use the alphabet to store text character strings in the form of words, phrases, sentences, paragraphs, chapters, books, etc. We use strings of binary digits (= bits) to store data in computers in quantities measured in kilobytes, megabytes, gigabytes, etc. In biology, strings of nucleotide sequences in our DNA can now be deciphered and stored in computerized databases. While vast amounts of information can be archived, it is advantageous for us to be able to search those archives, retrieve information and make comparisons (quickly!). We are familiar with web archives (such as Wikipedia) and web search engines (such as Google) which give rapid access to desired information. A key concept in science is the comparison of two objects, in this case, we compare two strings encoding the objects. A simple analogy is to compare the spelling of a word (string of alphabetic characters) with that given in a dictionary. In biology, we compare strings of DNA sequences between different organisms. For example, the sequence of a DNA fragment from the rare Babelfish was ACTTGAGCCAT while the sequence of a comparable fragment from a human was ACTTTAGCCAT. While comparing short strings can be simple and straightforward, identifying similarities and differences between long strings contained in large databases can be daunting.

Pattern searches: Various search algorithms have been developed to determine whether a pattern (short string of characters) is contained within a larger string, and if so, at which location. Pattern searches are used to find start (ATG) and stop (TAA, TGA, TGG) codons, cleavage sites for restriction enzymes, and primer matches for polymerase chain reaction (PCR) amplification. The simplest, but also least efficient, approach is to check whether the pattern matches the string at every possible position in the string. This is called the 'brute-force' or 'naive string' search algorithm. The computational complexity (hence run-time) of the algorithm depends on both the length of the pattern and the length of the string. If the string is *m* characters long and the pattern is *n* characters long, it takes (*m* x *n*) computations to analyse the complete data set.



Visual comparison: Humans are very good at identifying patterns visually. One of the first ways to compare biological sequences was to generate identity matrices and visualise them in so-called dot plots which show the similarities between two sequences in a comparison matrix. Each element in the matrix compares the character in one sequence with the character of the other sequence. Identical character pairings are visualised by a black pixel while different characters are coloured white.

The matrix below (dot plot) shows the results of comparing DNA fragments encoding for haemoglobin from a person with sickle cell disease and a healthy person.



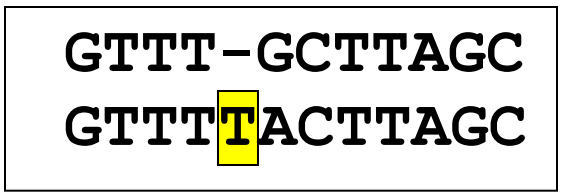
Sequence analyses:

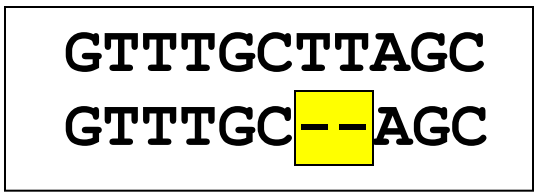
Biologists use bioinformatics not only to find identical matching sequences, but also sequences that are most similar to those of interest. Biological sequences may not be perfectly identical because of experimental errors (artefacts that may have occurred in the determination of the sequence) or natural variation (mutations in sequences in the course of evolution).

Four different types of sequence mutations are recognized:

1. Point mutations (substitutions) where one base is replaced by another (e.g. G for A at position 5). The most common type are transitions that exchange a purine for a purine (A ↔ G) or a pyrimidine for a pyrimidine, (C ↔ T), while less common types are transversions that exchange a purine for a pyrimidine or a pyrimidine for a purine (C/T ↔ A/G).

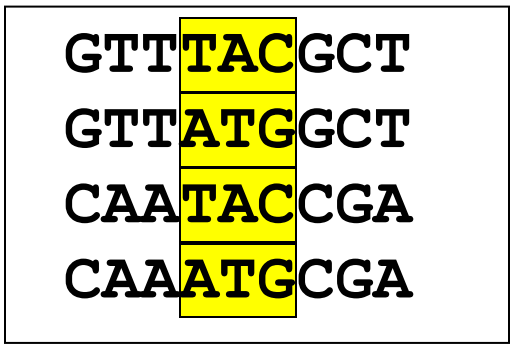
2. "Indel" mutations where bases are inserted (T at position 5) or deleted (TT at positions 7&8).





3. Chromosomal mutations involving large scale changes, through:
   - gene amplification (tandem repeats or duplications),
   - gene loss (deletions of large regions),
   - jumping genes (insertions or deletions of transposable elements), and
   - fusion genes (juxtapositioning of previously separate regions).

4. Recombination mutations involving cross-over events (chiasmas) and gene conversion (similar to cut-and-paste editing) (e.g. cross over of TAC and ATG at positions 4-6)



The ultimate effect of sequence variation may have negligible, severe or fatal consequences. Point mutations that occur within the protein coding region of a gene may be classified as silent mutations (still code for the same amino acid), mis-sense mutations (code for a different amino acid) or non-sense mutations (code for a stop thereby truncating the protein). Insertions and deletions in the coding region of a gene may alter splicing of the mRNA (splice site mutation), or cause a shift in the reading frame (frame-shift mutation), both of which can significantly alter the gene product.

Bioinformatics has found multiple uses in the comparative characterization of genes (genomics), transcriptional products (transcriptomics) and translational products (proteomics).

Genome annotation: The deciphering of an organism's genome involves identifying and marking genes and other biological features within the DNA sequences. Several software systems have been developed to find genes encoding for whole or partial proteins (both known and impuned), transfer RNA and other genetic elements.

Comparative genomics: Various comparative techniques are available to look for correspondence between genes (orthology analyses) or other genomic features in different organisms. Inter-genomic maps make it possible to trace the evolutionary processes responsible for the divergence of two genomes (through point mutations, insertions, deletions and recombination events).

Gene expression: The expression of many genes can be determined by measuring mRNA levels using a variety of techniques (such as microarrays, expressed sequence tags, parallel signature sequencing, multiplexed *in situ* hybridization). These techniques are subject to measurement bias and background noise so various computational tools have been developed to separate signal from noise in high-throughput studies.

Gene regulation: Bioinformatics techniques are now being used to examine the complex series of events involved in differential gene regulation, including signal pathways, promoter-suppressor analyses, up-regulated or down-regulated transcription and expression, and alterations in protein activity. Cells can be examined at different stages of development and under different conditions of stress (heat shock, starvation, etc.) to determine co-expression relationships.

Evolutionary biology: Informatics has facilitated numerous detailed and intensive studies on the origin and descent of species, as well as their change over time. It has enabled researchers to characterize species using multiple genotypic characters, explore correlations between phenotypic and genotypic classification systems, infer phylogenetic relationships between diverse taxa, examine complex evolutionary events such as gene duplication and lateral gene transfer, and build complex population models to predict change over time.

Biodiversity assessment: The total genomic complement can be harvested from all the species present in a particular environment, whether it be a biofilm in pipe, a puddle of water, a scoop of soil, or the entire biosphere of the planet Earth. Databases are used to collect the species names, descriptions, distributions, genetic information, status and size of populations, habitat needs, and how each organism interacts with other species.

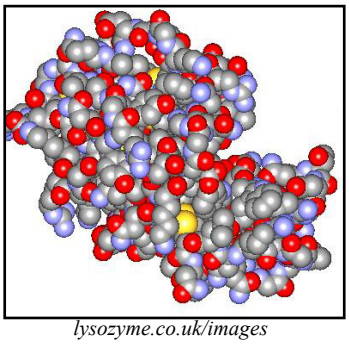
Analysis of protein expression: Microarrays and mass spectrometry are used to examine proteins present in biological samples. Bioinformatics helps match data against predicted masses from protein sequence databases, even when multiple, but incomplete, peptides from each protein are detected.

Protein structure: The primary structure of a protein (amino acid sequence) depends on the codon sequences of the gene that encodes it. Thereafter, the secondary, tertiary and quaternary structure of the protein depends on molecular interactions resulting in complex, yet often predictable, coiling and folding patterns. In the structural branch of bioinformatics, homology is used to determine which parts of a protein are important in structure formation and interaction with other proteins.

Chapter 11.

ENZYMES

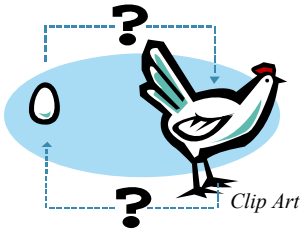
Vocabulary list: RNA World theory, metabolism, catalyst, enzyme, heredity, nucleotides, DNA, RNA, replication, transcription, translation, protein synthesis, amino acid, bioenergetics, free energy, activation energy, substrate, enzyme kinetics, Michaelis-Menten constant



lysozyme.co.uk/images

Conundrum?

All life on Earth has a cellular basis; be it prokaryotic (cells without true nuclei, such as archaea and bacteria) or eukaryotic (cells with membrane-bound nuclei, such as protista, fungi, plants and animals). Even the viruses, which could be considered acellular, still need host cells. All these cells need to survive and replicate. Any theories developed on the origins of life must account for both fundamental traits: metabolism and heredity.



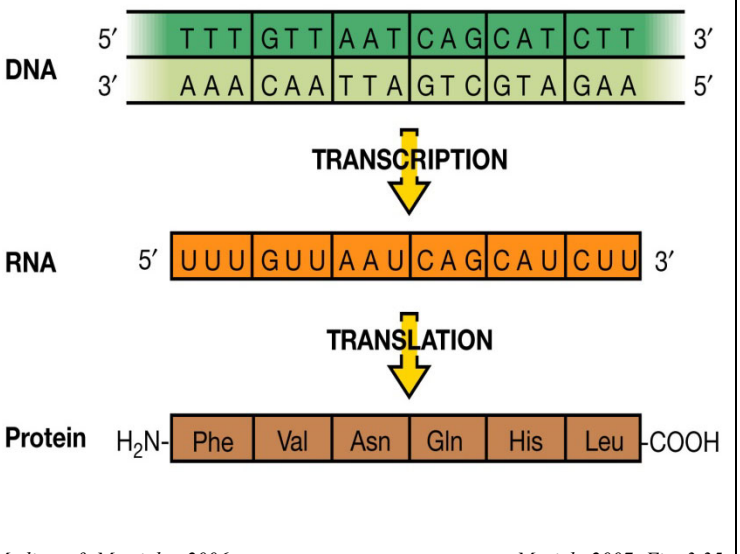
Clip Art

Reflecting on this issue produces a 'chicken-and-egg' conundrum.
Which came first?

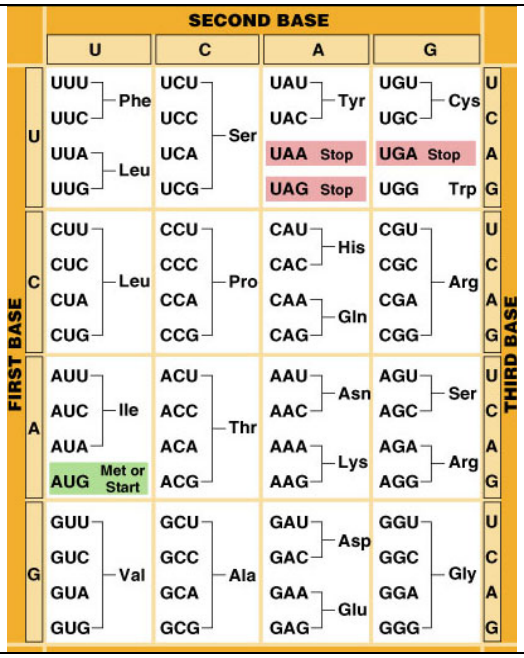
Metabolism: Cells survive by deriving energy and building materials through controlled biochemical reactions. These metabolic processes essentially involve catabolic decomposition ($AB \rightarrow A + B$) and anabolic synthesis ($A + B \rightarrow AB$). If the reactants relied upon random chance encounters (through Brownian motion), reactions would occur at an extremely low frequency, if at all. Fortunately, reactions may be substantially amplified by catalysts, chemicals that speed up reactions but are not consumed by them. Metabolic processes rely on catalysts in the form of enzymes, protein molecules whose unique three-dimensional (3D) configuration acts to bring reactants together with greater frequency. These enzymes are essential for life processes, but where did the first enzymes come from?

Heredity: Life goes on because cells replicate. They produce offspring through asexual and sexual processes. The daughter cells are usually faithful copies (or composites) of the parents, ensuring the transfer or inheritance of traits essential for cell survival and future replication. Changes in copy stringency form the basis for the evolution of species through the natural selection of characters favouring survival and reproduction. The base molecular units of replication are DNA, sequences of which form genes arrayed on chromosomes. Strands of DNA pair up to form a double helix, with complementary base pairing rules providing vital proof-correction. DNA is the great self-replicating molecule of life, but where did the first DNA come from?

The conundrum arises through the mutual dependencies of DNA and enzymes. DNA provides the genetic code required for protein synthesis (through transcription and translation) and enzymes are required for DNA replication. But which came first?



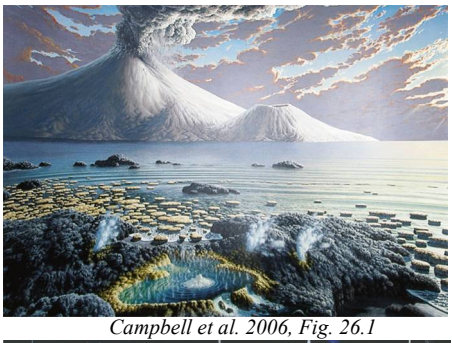
Madigan & Martinko, 2006



Marieb, 2007; Fig. 3.35

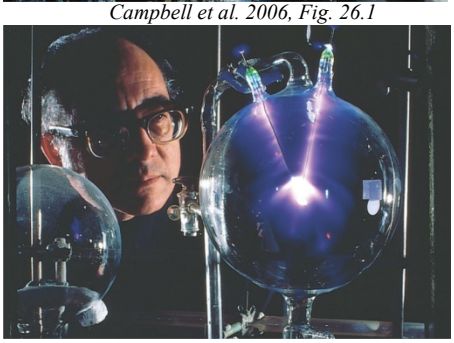
There are currently several fascinating theories on the origins of life, with current consensus favouring a middle-ground option, the RNA World theory, based on the properties of the different molecules. DNA is an excellent replicator but a poor catalyst, while protein enzymes are excellent catalysts but poor replicators. In comparison, RNA exhibits good replicator and good catalyst properties. It is similar to DNA but has a 3D shape instead of a double helix due to partial self-pairing and hairpin bends. Various RNA molecules have been found to be self-replicating enzymes; including transfer RNA (tRNA) which acts as an enzyme facilitating the reading of messenger RNA (mRNA) and protein synthesis in ribosomes, and ribozymes which are enzyme-like RNA molecules that catalyze reactions during RNA splicing.

Scientists had long reasoned that for life to arise, the atmosphere of pre-biotic Earth would have to be carbon-rich (in the form of compound gases such as carbon dioxide and methane), nitrogen-rich (as it is today, but originally compounded with hydrogen as ammonia) and reducing (lacking free oxygen, unlike today where free oxygen is produced by plants and most organisms rely on it for oxidation).



Campbell et al. 2006, Fig. 26.1

Some elegant experiments were conducted over 50 years ago by Miller and Urey in the USA whereby they simulated primordial environmental conditions in glass flasks. They added water for an ocean to an atmosphere of methane, ammonia, hydrogen and later carbon dioxide, and mimicked conditions of volcanic heat and lightning sparks. Within a week, they observed the spontaneous synthesis and enrichment of organic compounds within the primordial soup, including amino acids and sugars (the basic components of DNA and RNA).



Campbell et al. 2006, Fig. 4.2

It is interesting to consider the dynamics of anabolic reactions (A + B → AB). If the reaction relies on random chance encounters of components A and B, few would ever come into contact, and even fewer would react to produce AB. The reaction rate depends on the frequency of fruitful contacts. However, when an enzyme (ABase) is introduced into the system, it catalyses the reaction by capturing components A and B and orienting them to form AB. The product is released and the enzyme is then free to repeat the process. The production of AB is substantially up-regulated, typically in the order of a million to a trillion times. The situation is even better if the system is auto-catalytic, that is, the enzyme ABase and the product AB are the same. Under the right conditions, the production of AB would be exponential. The self-replication and catalytic properties of AB would theoretically provide the right ingredients for the origin of life. It is mooted that AB may be RNA.

Recent studies have supported this theory. Rebek in the USA developed an auto-catalytic system whereby reaction products catalyses their own formation. More recently, Spiegelman in the USA succeeded in getting RNA from a bacteriophage (a virus normally infecting a bacterium) to self-replicate in test tubes containing only precursor molecules. He even observed evolution in action, in the form of spontaneous mutation and natural selection over many generations (to the point the virus became adapted to living only in test tubes and no longer in bacteria). The contention that life began in warm sunny ponds has also been challenged by the recent discovery of life in hot deep rocks. Numerous thermophilic bacteria have been found around undersea geothermal (volcanic) vents and geological core samples are revealing their huge biomass in deep rocks and fissures far removed from sunlight and at crushing pressures. Whatever the primaeval conditions were for the origins of life, all extant life-forms utilize similar strategies for metabolic survival and heritable replication.

Bioenergetics

Energy is defined as the ability to do work. In biochemical reactions, it is calculated in units of kilojoules (kJ), a measure of heat energy. Chemical reactions involve changes in energy. The free energy (G) of a chemical reaction is defined as the energy released that is available to do useful work. The change in free energy during a reaction is expressed as ΔG^{0}' where the Greek symbol delta Δ denotes change, and the superscripts ^{0} and ' mean standard conditions (pH 7, 25^{0}C, 1 atmosphere, all reactants and products at 1M).

Consider the reaction A + B → C + D.
If ΔG^{0}' is negative, the reaction will proceed with the release of energy (exergonic reaction).
If ΔG^{0}' is positive, the reaction requires energy to proceed (endergonic).

To calculate ΔG^{0}', you need to know the free energy of the reactants and products. This is the free energy of formation (G_f^{0}), the energy released or required during the formation of a given molecule from its elements. For most compounds, G_f^{0} is negative, indicating that compounds tend to form spontaneously (with energy being released). Using G_f^{0} values, it is possible to calculate ΔG^{0}' by subtracting the sum of the free energies of formation of the reactants (A and B) from that of the products (C and D), given by the formula:

$$\Delta G^{0}' = G_f^{0}\,[C + D] - G_f^{0}\,[A + B]$$

Free energy calculations using standard conditions are only approximations of the free energy changes that actually occur when a reaction takes place in nature. What is most relevant in bioenergetic calculations is not ΔG^{0}', but ΔG, the free energy change that occurs under the actual

conditions of the reaction. ΔG takes into account the actual concentrations of the reactants and products, and is calculated as:

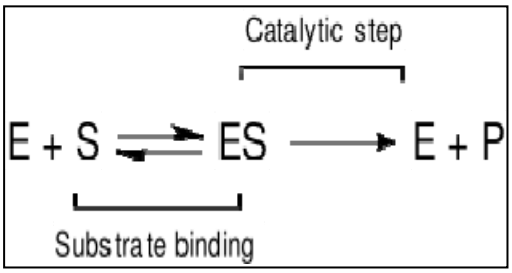
$$\Delta G = \Delta G^{0\prime} + RT \ln K$$

where R = a constant (8.29 J/mol/Kelvin), T = temperature (Kelvin), and
K = the equilibrium constant for the reaction.

Free energy calculations reveal only whether energy is released or required for a given reaction. They do not provide information about the rate of the reaction. Reactions may not proceed spontaneously even though energy may be released, because the reactants must first be activated. Chemical bonds in the reactants must first be broken and this requires energy, called the activation energy (Ea or $\Delta G^{\ddagger}$). For many reactions, the activation energy barrier is virtually insurmountable in the absence of catalysis, but in the presence of a proper catalyst, it is much less formidable.

Enzymes

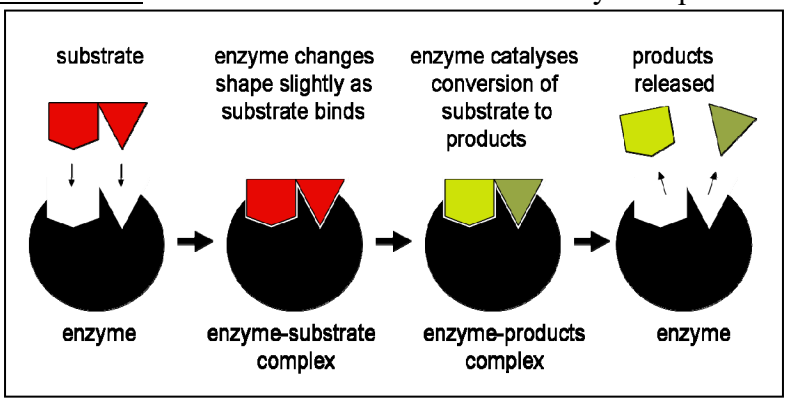
Enzymes are proteins that accelerate (catalyse) chemical reactions, binding to substrates and turning them into products, without being consumed themselves. Enzyme reactions occur in two stages: the first involving the enzyme (E) binding reversibly to the substrate (S) forming the enzyme-substrate complex (ES); and the second involving the catalysis of the complex releasing the product (P) as well as the enzyme itself (unaltered and ready to go again).



wikipedia, 2008 'enzyme'

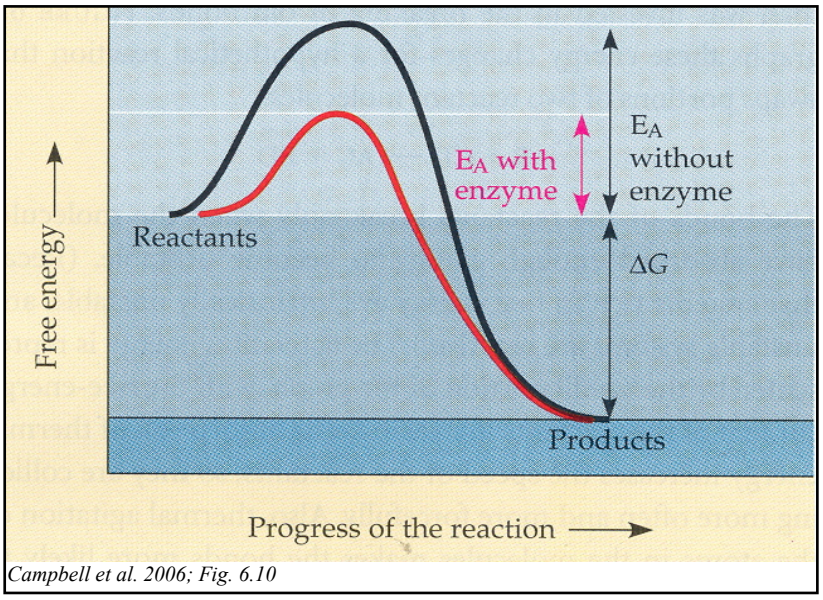
Like all proteins, enzymes are made as long, linear chains of amino acids that fold to produce a three-dimensional product, which determines their activity. Enzymes are generally globular proteins and range from just 62 to over 2,500 amino acid residues in size. Most enzymes are much larger than the substrates they act on, and only a small portion of the enzyme (the active site around 3–4 amino acids) is directly involved in catalysis. Most enzymes can be denatured (unfolded and inactivated) by heating, which destroys the three-dimensional structure of the protein. Depending on the enzyme, denaturation may be reversible or irreversible.

Enzymes are usually very specific as to which reactions they catalyze and the substrates that are involved in these reactions. Complementary shape, charge and hydrophilic/hydrophobic characteristics of enzymes and substrates are responsible for this specificity. The complementary geometric shapes of enzymes and substrates, whereby they fit exactly into one another, gave rise to the "lock and key" model to explain enzyme specificity, but it failed to explain the stabilization of the transition state that enzymes achieve. The "induced fit" model is now the most currently accepted hypothesis of enzyme action. Since enzymes are rather flexible structures, the active site is continually reshaped by interactions with the substrate. As a result, the amino acid side chains which make up the active site are moulded into the precise positions that enable the enzyme to perform its catalytic function. The active site continues to change until the substrate is completely bound, at which point the final shape and charge is determined.



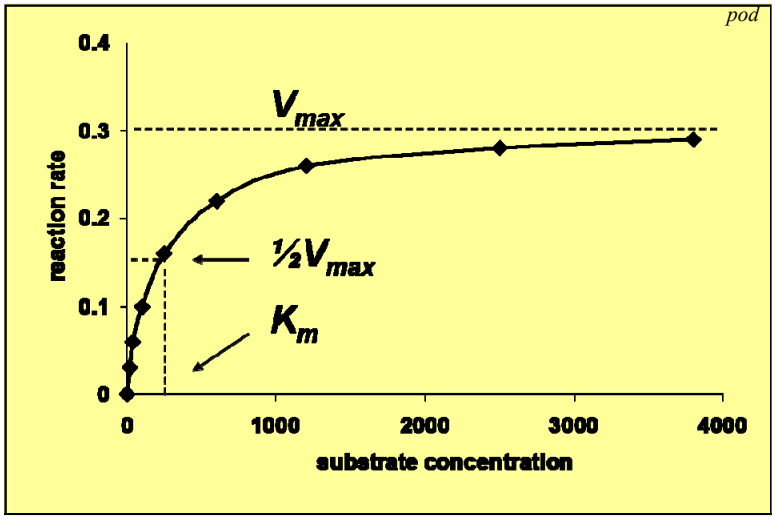
pod

Enzymes are not consumed by the reactions they catalyze, nor do they alter the equilibrium of those reactions. Like all catalysts, enzymes work by providing an alternative mechanism involving a different transition state and lowering the activation energy (*Ea* or $\Delta G^{\ddagger}$) for a reaction. This has the effect that more molecular collisions have the energy needed to reach the transition state. Hence, enzymes perform reactions much faster, more specific or at lower temperatures. Substrates usually need a large amount of energy to reach the transition state, which then decays into the end product. An enzyme stabilizes the transition state, reducing the energy needed for its formation and thus reducing the energy required to form products. Enzymes reduce the amount of energy needed to start the chemical reaction. However, they cannot make energetically unfavourable reactions possible, they have no effect of the equilibrium of the reaction because the rate of both the forward and reverse reactions are equally affected. The net free energy change of a reaction is the same whether an enzyme is used or not, the enzyme just makes it easier to activate.



Campbell et al. 2006; Fig. 6.10

Enzymes can catalyze up to several million reactions per second. To find the maximum speed of an enzymatic reaction, the substrate concentration is increased until a constant rate of product formation is seen (flattening of saturation curve). As substrate concentration increases, saturation occurs because more of the free enzyme is converted into the substrate-bound ES form. At the maximum velocity (*Vmax*) of the enzyme, all the enzyme active sites are bound to substrate, and the amount of ES complex is the same as the total amount of enzyme. However, *Vmax* is only one kinetic constant of enzymes. The amount of substrate needed to achieve a given rate of reaction is given by the Michaelis-Menten constant (*Km*), which is the substrate concentration required for an enzyme to reach ½ *Vmax*. Each enzyme has a characteristic *Km* for a given substrate, and this can show how tight the binding of the substrate is to the enzyme. Another useful constant is *kcat*, which is the number of substrate molecules handled by one active site per second. The efficiency of an enzyme can then be expressed in terms of *kcat/Km* (called the specificity constant) which incorporates the rate constants for all steps in the reaction. Because the specificity constant reflects both affinity and catalytic ability, it is useful for comparing different enzymes against each other, or the same enzyme with different substrates.



Enzyme kinetics

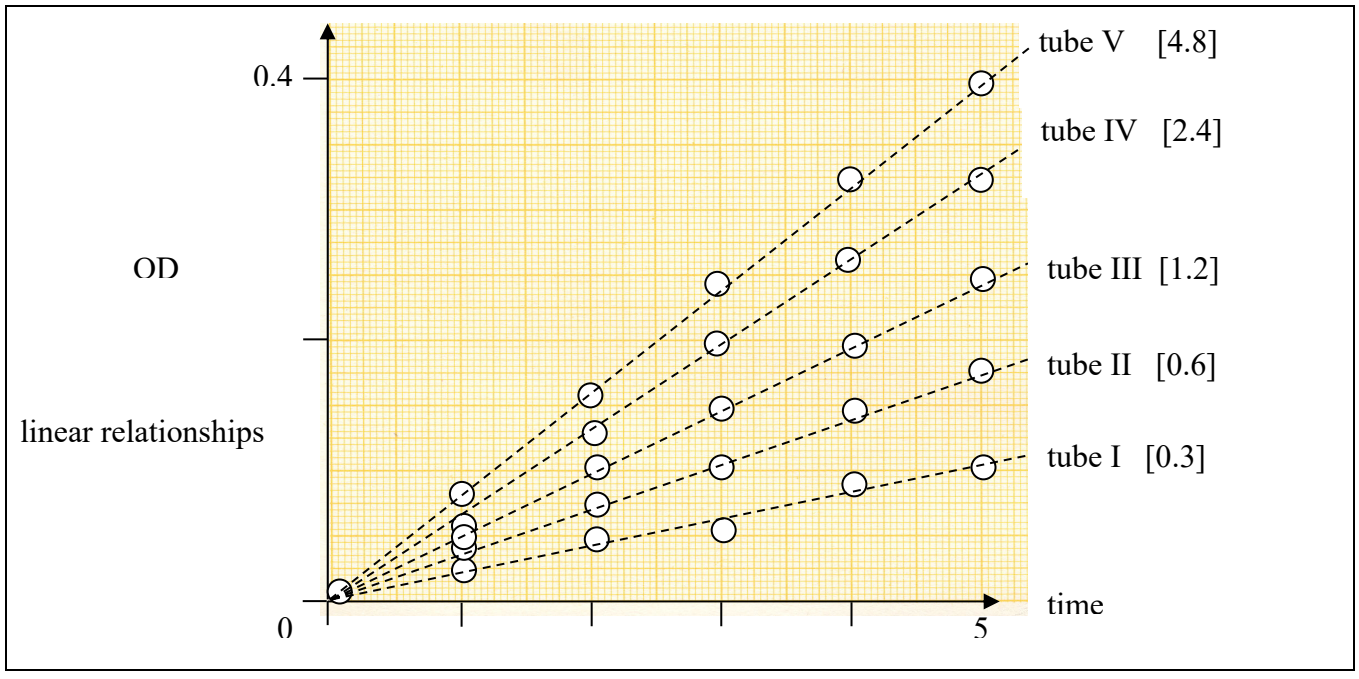
Enzyme reactivity (catalytic power) can be quantitated (and visualized) using a 3-step process:

- by measuring enzyme reactions against different substrate concentrations [*S*] (plot against time);
- by calculating the average rates of reaction (= enzyme velocity, V_i) (plot V_i against [*S*]); and
- by determining the Michaelis-Menten constant (K_m), which is the substrate concentration required for the enzyme to reach one-half of its maximum velocity ($\frac{1}{2}V_{max}$) (using a Lineweaver-Burk double-reciprocal plot of $1/V_i$ against $1/[S]$)

Consider the process by which slices of apple turn brown. This is because the enzyme o-diphenol oxidase catalyzes the oxidation of phenols in the apple to form dark-coloured products. We can determine the enzyme kinetics by reacting some freshly-squeezed apple juice against different concentrations of a defined substrate: in this case, catechol, which is converted to o-quinine which is further oxidized to dark products. We set up 5 test tubes containing double dilutions of catechol, giving concentrations of 0.3, 0.6, 1.2, 2.4 and 4.8 mM. We add a fixed amount of apple juice to each tube and then use a spectrophotometer to measure the optical density (OD) at 540 nm wavelength of the solution in each tube each minute for 5 minutes; yielding the following results:

Tube	I	II	III	IV	V
Substrate concentration [S] mM	0.3	0.6	1.2	2.4	4.8
Optical density at time zero OD_0	0	0	0	0	0
Optical density at 1 minute OD_1	0.021	0.038	0.050	0.065	0.082
Optical density at 2 minutes OD_2	0.040	0.070	0.099	0.128	0.163
Optical density at 3 minutes OD_3	0.059	0.104	0.148	0.195	0.240
Optical density at 4 minutes OD_4	0.081	0.140	0.195	0.260	0.325
Optical density at 5 minutes OD_5	0.100	0.175	0.240	0.320	0.405

What type of relationship exists between OD and time? Graphical plots reveal them to be linear.

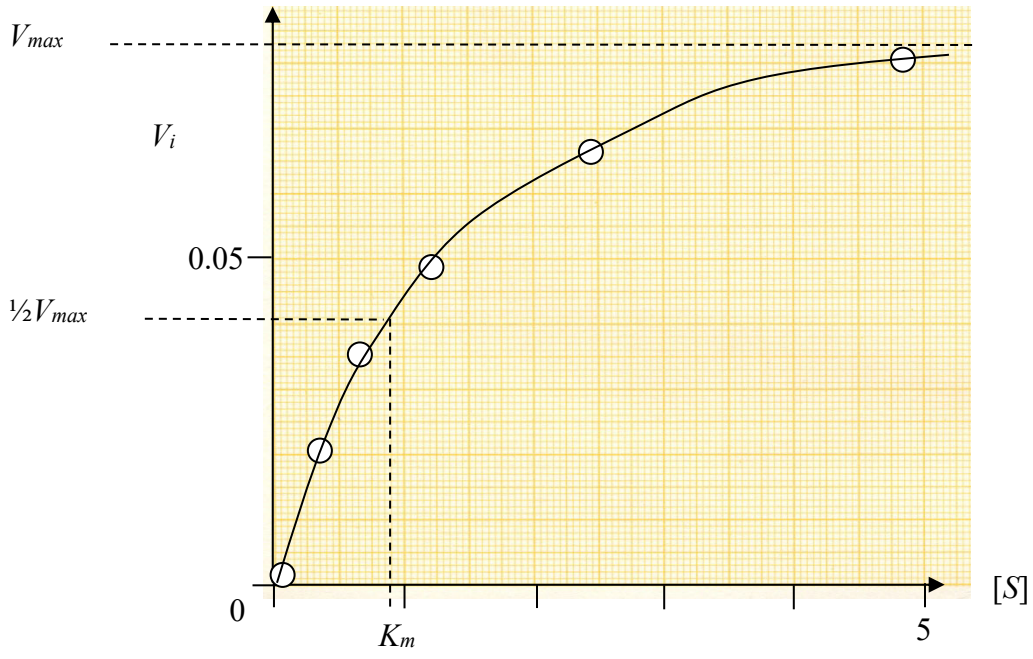


Early reactions increase linearly with time, but later plateau off as the substrate depletes. If your graphs of OD against time were almost linear, calculate the initial reaction velocity (V_i) for each tube. [velocity = gradient of line = rise over run = average change in OD per minute]

Tube	I	II	III	IV	V
Substrate concentration [*S*] mM	0.3	0.6	1.2	2.4	4.8
Initial velocity *Vi* (= ΔOD/min)	(0.100-0)/5 = 0.020	0.175/5 = 0.035	0.240/5 = 0.048	0.320/5 = 0.064	0.405/5 = 0.081

When the amount of substrate is in substantial excess to the amount of enzyme, the early reaction rate (initial velocity, *Vi*) rises almost linearly with increasing [*S*]. However, as [*S*] increases, the velocity levels off (forming a rectangular hyperbola). The asymptote represents the maximum velocity of the reaction, V_{max}. The substrate concentration that produces one-half V_{max} is called the Michaelis-Menten constant K_m.

(i) Plot V_i against [*S*]. What type of relationship is evident?
(ii) Roughly estimate maximum velocity V_{max} (draw on graph as asymptote, estimated plateau level).
(iii) Determine ½V_{max} on graph and then read corresponding [*S*] = Michaelis-Menten constant (K_m)



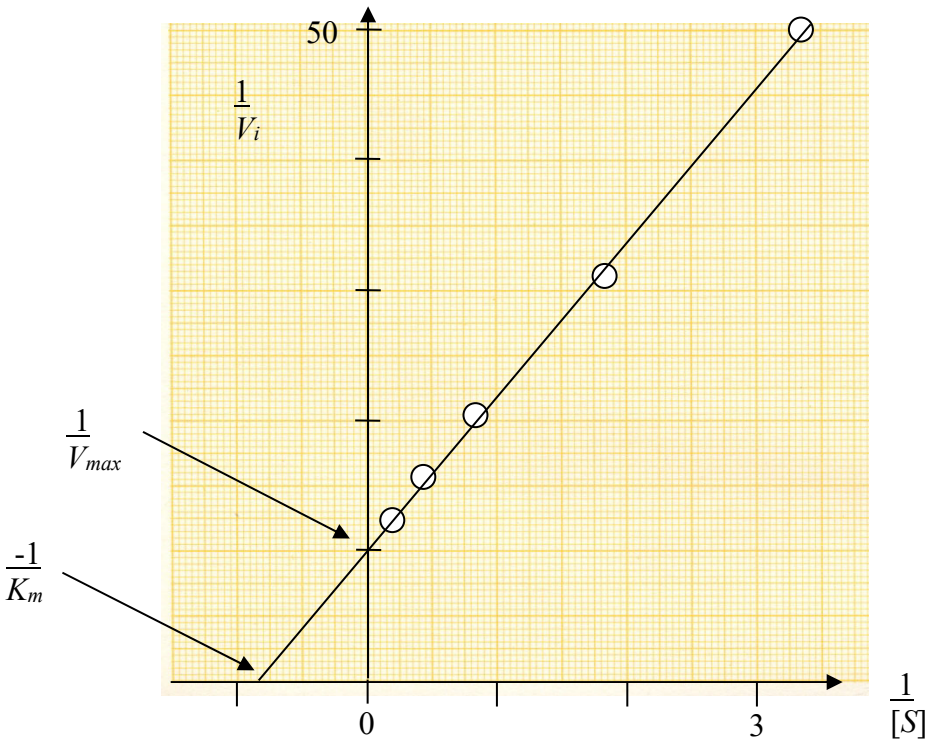
(i) Relationship is like a power function
(ii) V_{max} on graph ~ 0.085
(iii) ½V_{max} on graph ~0.043,
so corresponding [S] = Michaelis-Menten constant (K_m) ~0.9

K_m is an inverse measure of the affinity or strength of binding between the enzyme and its substrate. The lower the K_m, the greater the affinity (so the lower the [*S*] needed to achieve a given rate).

A more precise way to determine V_{max} and K_m is to make a double-reciprocal plot of V_i and [S] (that is, plot $1/V_i$ against $1/[S]$) (this is called a Lineweaver-Burk plot). Calculate the reciprocals of your data in the following table.

Tube	I	II	III	IV	V
Substrate concentration [S] mM	0.3	0.6	1.2	2.4	4.8
Reciprocal [S] = $1/[S]$	3.33	1.67	0.83	0.42	0.21
Initial velocity V_i	0.020	0.035	0.048	0.064	0.081
Reciprocal V_i = $1/V_i$	50	31.7	20.8	15.6	12.3

(i) Make a Lineweaver-Burk plot of the data. What type of relationship exists between $1/V_i$ and $1/[S]$?
(ii) The y-intercept = $1/V_{max}$. Calculate V_{max}
(iii) The x-intercept = $-1/K_m$ Calculate K_m



(i) There is a linear relationship between $1/V_i$ and $1/[S]$

(ii) The intercept on the y-axis ($1/V_i$) corresponds to $1/V_{max}$
in this case, $1/V_{max}$ = 10, so V_{max} = 0.100

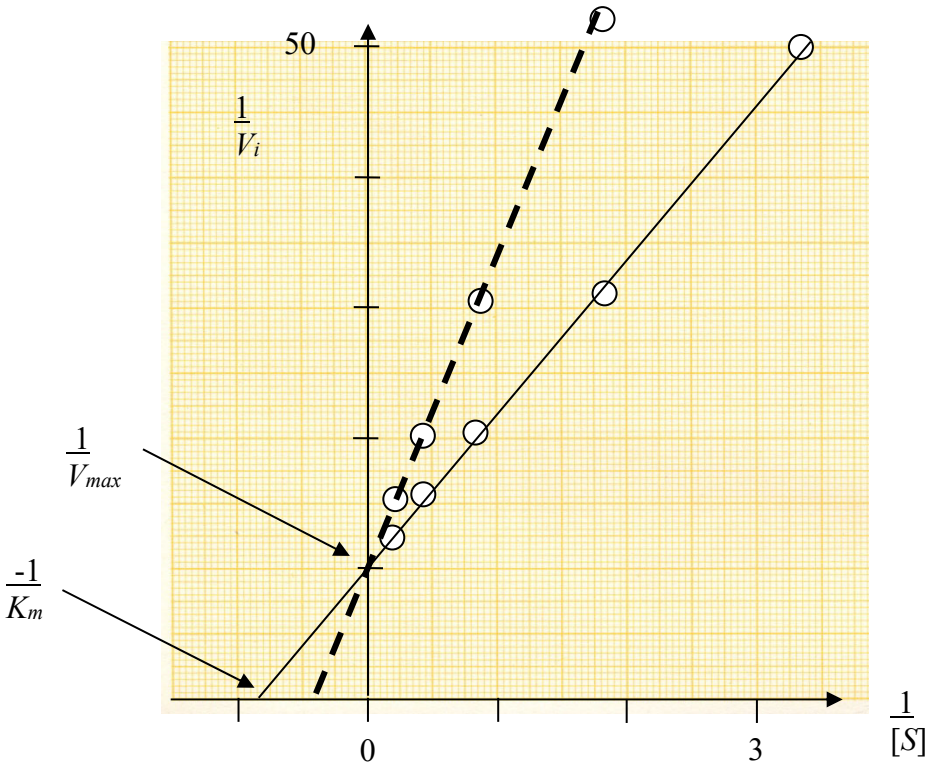
(iii) The intercept on the x-axis ($1/[S]$) corresponds to $-1/K_m$
in this case, $-1/K_m$ = -0.8, so K_m = 1.25 mM

It takes a substrate concentration of 1.25 mM to achieve one-half the maximum enzyme velocity

You decide to run the whole series again, but this time you also add a competitive inhibitor to each tube. You use para-hydroxybenzoic acid (PHBA) which binds to the same site as catechol but is not acted upon. You obtain the following results.

Tube	I	II	III	IV	V
Substrate concentration [*S*] mM	0.3	0.6	1.2	2.4	4.8
Reciprocal [*S*] = 1/[*S*]	3.33	1.67	0.83	0.42	0.21
ΔOD = initial velocity V_i	0.011	0.019	0.032	0.048	0.060
Reciprocal V_i = 1/ V_i	90.9	52.6	31.3	20.8	16.7

(i) Make a Lineweaver-Burk plot of the data. What do you notice about the graphs?
(ii) The y-intercept = $1/V_{max}$. Calculate V_{max}
(iii) The x-intercept = $-1/K_m$ Calculate K_m



(i) The graphs intersect at the same point on the y-axis

(ii) They have the same y-intercept = $1/V_{max} = 10$, so $V_{max} = 0.100$

(iii) However, the x-intercept has changed, now giving $-1/K_m = -0.4$, so $K_m = 2.50$ mM

It now takes a substrate concentration of 2.50 mM to achieve one-half the maximum enzyme velocity due to competition for active sites with the PHBA inhibitor.

Terminology

An enzyme's name is often derived from its substrate or the chemical reaction it catalyzes, with the word ending in -ase (e.g. lactase, alcohol dehydrogenase). This convention may result in different enzymes (called isoenzymes) with the same function having the same basic name. Isoenzymes differ in their amino acid sequences and thus exhibit differential structural and/or functional characteristics.

A standard nomenclature has therefore been developed, where all enzymes are allocated an EC number. The top-level classification involves 6 groups:

- EC 1 oxidoreductases (catalyze oxidation/reduction reactions)
- EC 2 transferases (transfer functional groups, e.g. methyl or phosphate groups)
- EC 3 hydrolases (catalyze the hydrolysis of various bonds)
- EC 4 lyases (cleave various bonds by means other than hydrolysis and oxidation)
- EC 5 isomerases (catalyze isomerization changes within a single molecule)
- EC 6 ligases (join two molecules with covalent bonds)

The SI derived unit for measuring catalytic activity is the katal, which is moles per second. The degree of activity of a catalyst can also be described by the 'turn over number' (TON) and the catalytic efficiency by the 'turn over frequency' (TOF). The biochemical equivalent is the enzyme unit.

Enzymes serve a wide variety of functions inside living organisms. They are indispensable for signal transduction and cell regulation, often via kinases and phosphatases. They also generate movement, with myosin hydrolysing ATP to generate muscle contraction and also moving cargo around the cell as part of the cytoskeleton. Other ATPases in the cell membrane are ion pumps involved in active transport. Enzymes often work together in a specific order, creating metabolic pathways, where one enzyme takes the product of another as a substrate. Indeed, a metabolic pathway such as glycolysis could not exist independently of enzymes.

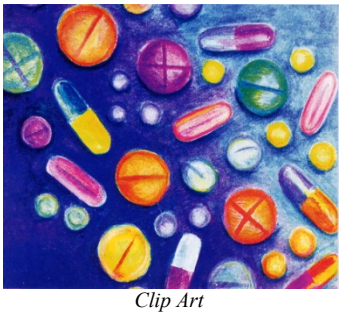
Because enzymes are such efficient catalysts, they are subject to exquisite control, often through negative and positive feedback. A cell can control enzyme activity by:

- altering production (induce/inhibit transcription and translation of enzyme genes)
- compartmentalization (confine them to specific cellular compartments)
- changing activity (activate/inhibit precursors e.g. through feedback mechanisms)
- modifying molecules (post-translational modification, e.g. phosphorylation, glycosylation)
- imposing activation requirements (reducing/oxidising environment, pH change, etc)

Due to their wide spectrum of activity, their specificity for particular reactions, and their impressive catalytic properties, enzymes have found extensive use in various industries.

baking industry	amylases proteases	breakdown starch in flour to sugar breakdown proteins in flour for biscuit manufacture
brewing industry	amylases glucanases pullulanases proteases	split polysaccharides in malt improve beer filtration adjust fermentation remove cloudiness
dairy industry	rennin lipases lactases	hydrolyse protein to manufacture cheese to ripen some cheeses breakdown lactose to glucose
meat industry	trypsin papain	predigest baby foods meat tenderizer
paper industry	amylases xylanases cellulases ligninases	breakdown starches (aids sizing and coating paper) reduce bleach (used to decolourize paper) breakdown fibres (smooths paper) remove lignin (softens paper)
biofuel industry	cellulases ligninases	breakdown cellulose for sugar fermentation breakdown lignin waste
cleaning industries	amylases lipases cellulases	detergents to remove starches detergents to remove fats fabric conditioners
molecular biology	DNA polymerase restriction enzymes	produce DNA (esp. via polymerase chain reaction) cut DNA

Chapter 12.



Clip Art

PHARMACOLOGY

Vocabulary list: pharmacology, drug, treatment, therapy, prevention, prophylaxis, pharmacodynamics, dose-response curve, pharmacokinetics, concentration-time profile, absorption, distribution, metabolism, excretion

Drugs

The Oxford dictionary defines 'pharmacology' as 'the science of action of drugs on body', and 'drug' as a 'medicinal substance, organic or inorganic, used alone or as an ingredient'. Throughout history, various substances have been endowed with medicinal qualities and folklore herbal remedies have long been used by humans (and even by some animal populations). Many drugs come from plants, particularly those species that have developed chemical defenses against pests, pathogens and disease. Even today, bioscreening companies are still finding new active compounds produced by unicellular and multicellular flora and fauna. Indeed, finding novel medicinal compounds is often cited as another compelling reason for conserving Earth's biodiversity.

Scientists are continuing to identify the active ingredients in herbal remedies and they are constantly attempting to create synthetic analogues to overcome problems in supply and demand. This spawned an incredibly successful pharmaceutical industry which has had an extraordinary impact on health and welfare (medical, veterinary and agricultural).

Mehlhorn, 2001; Fig. 3

ambulant Renaissance pharmacist

A huge range of prescription and nonprescription drugs are produced by multinational companies under a staggering number of names (one international non-proprietary name, but with several regional non-proprietary names and many proprietary brand names). Drugs come in a vast array of topical preparations (liniments, lotions, ointments, dips, shampoos, washes, pour-ons, spot-ons, collars, creams, sprays, powders, aerosols), oral formulations (tablets, pills, capsules, bolus, liquids, emulsions) or parenteral formulations (liquids in ampoules/vials, or solid implants, for subcutaneous, intramuscular, intravenous or intraperitoneal administration).

In human society, drug use has two different connotations. Most drugs are perceived to be beneficial, they are taken within guidelines (with or without prescription) for the treatment (therapy) or prevention (prophylaxis) of infection, illness and disease. Other drugs (including prescription drugs) are detrimental when taken in excess, ultimately they do harm, but are nonetheless taken for recreational purposes because they provide transient pleasures. Some are, or have been, regarded as socially acceptable (such as alcohol and tobacco), while others are highly illegal in many societies due to their addictive and destructive effects (such as heroin and amphetamines).

Quantitating drugs is the accurate and reliable measurement of biologically active chemicals using sophisticated analytical instrumentation. Successful trace detection methods are underpinned by an understanding and application of mathematics - simple algebra, basic calculus, exponential and logarithmic functions. As a complement to statistics, physics, chemistry, earth sciences and biology, a knowledge of how to quantitate drugs has many and increasingly important applications of global significance. These include: personal and national security; forensic methodology; law enforcement; quality control; sporting performance; drug discovery, design and development; pharmaceutical clinical trials; clinical medicine and basic scientific research. Drug quantitation methodologies are performed in industry and government institutions to demanding international standards. Real life applications of drug quantitation draw upon contested philosophical, ethical and legal stances. The study of drugs involves two complementary areas:

- pharmacodynamics, the study of what a drug does to the body; and
- pharmacokinetics, the study of what the body does to a drug.

Pharmacodynamics (PD)

PD is the study of the biochemical and physiological effects of drugs on the body (or on infective microorganisms within the body) and the mechanisms of drug action and the relationship between drug concentration and effect.

The majority of drugs either mimic or inhibit normal physiological/biochemical processes or inhibit pathological processes. Drugs may act as stimulants, depressants, toxins or substitutes in their chemical interactions with enzymes, structural proteins, carrier proteins, ion channels, hormones, neuromodulators or neurotransmitters. For instance, some general anaesthetics work by disordering neural membranes and altering Na^+ influx. Enzyme-substrate binding alters the production or metabolism of key endogenous chemicals (e.g. aspirin inhibits cyclooxygenase preventing inflammation). Many drugs act as ligands which bind to receptors influencing cellular processes, either resulting in enhanced action (agonist), blocked action (antagonist), or even opposite action (inverse agonist). For drugs to work, they must reach specific target concentrations. Many factors affect drug concentrations; such as patient size, age, genetic disposition, physiology, metabolism, etc. There may also be undesirable side-effects including: cell mutation (carcinogenic activity); metabolic disturbances, physiological damage and abnormalities. The therapeutic window is the amount of drug between the amount that gives an effect (effective dose) and the amount that gives more adverse effects than desired.

The rates of drug (ligand) binding to receptors can be used to determine the equilibrium concentration of bound receptors. For the reaction $L + R \leftrightarrow L.R$, the equilibrium dissociation constant (*Kd*) is defined as:

$$Kd = [L]\,[R] / [L.R]$$

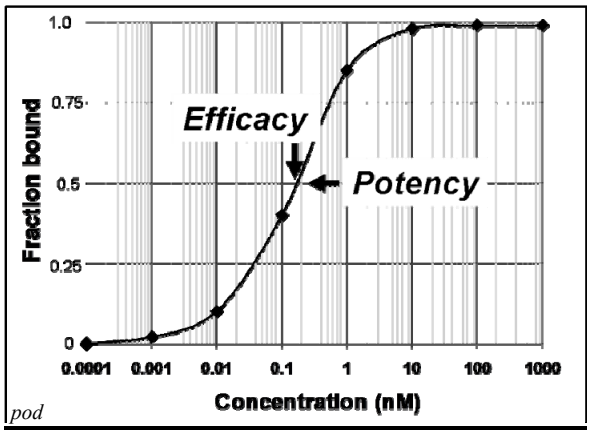
where L = ligand, R = receptor, and square brackets [ ] denote concentration.

The fraction of bound receptors (occupancy) is found as $(1 + [R]/[L \cdot R])^{-1}$, which can be expressed as:

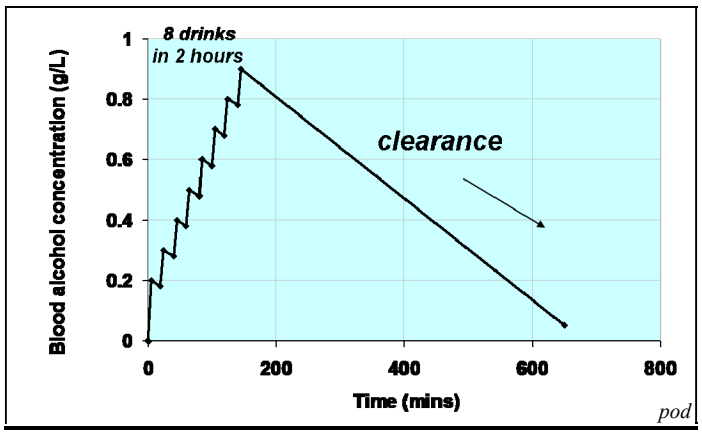
$$\text{fraction bound} = \frac{1}{1 + (Kd/[L])}$$

The relationship between occupancy and pharmacological response is usually non-linear. Often the response is determined as a function of log[L] to consider many orders of magnitude of concentration. It is useful to note that 50% of the receptors are bound when [L]=*Kd*.

Pharmacodynamics: dose-response curve



Pharmacokinetics: concentration-time curve



Pharmacokinetics (PK)

PK is a branch of pharmacology dedicated to the determination of the fate of substances administered externally to a living organism (applies mainly to drugs, but in principle concerns other substances such as nutrients, metabolites, hormones, toxins, etc). PK considers four main areas: the extent and rate of absorption, distribution, metabolism and excretion (ADME scheme).

Pharmacokinetic analyses are performed by noncompartmental or compartmental methods. Noncompartmental methods estimate the exposure to a drug by estimating the area under the curve of a concentration-time graph, with the trapezoidal rule (numerical differential equations) the most common area estimation method. Due to the dependence of the length of '*x*' in the trapezoidal rule, the area estimation is highly dependent on the blood/plasma sampling schedule. That is, the closer your time points are, the closer the trapezoids are to the actual shape of the concentration-time curve. Compartmental methods use kinetic models to describe and predict the concentration-time curve. PK compartmental models are often similar to kinetic models used in other scientific disciplines such as chemistry. The advantage of compartmental to noncompartmental analysis is the ability to predict the concentration at any time. The disadvantage is the difficulty in developing a proper model.

Bioanalytical methods are necessary to construct a concentration-time profile. Chemical techniques are employed to measure the concentration of drugs in biological matrix, most often plasma or urine. Mass spectrometry is often used due to the complex nature of the matrix and the need for high sensitivity to observe low dose and long time point data. Standard curves and internal standards are used for the quantitation of pharmaceuticals in samples taken at different time points. Blank and time-zero samples are important in determining background levels and insuring data integrity.

It is very important that concentrations of drugs can be accurately determined (and predicted) in order to maximize therapeutic benefits and minimize potential side-effects. Continuous monitoring of drug concentrations in the body is often not possible, but blood samples can be taken at regular intervals and tested. From these discrete samples, a continuous model can often be assumed and defined by differential equations.

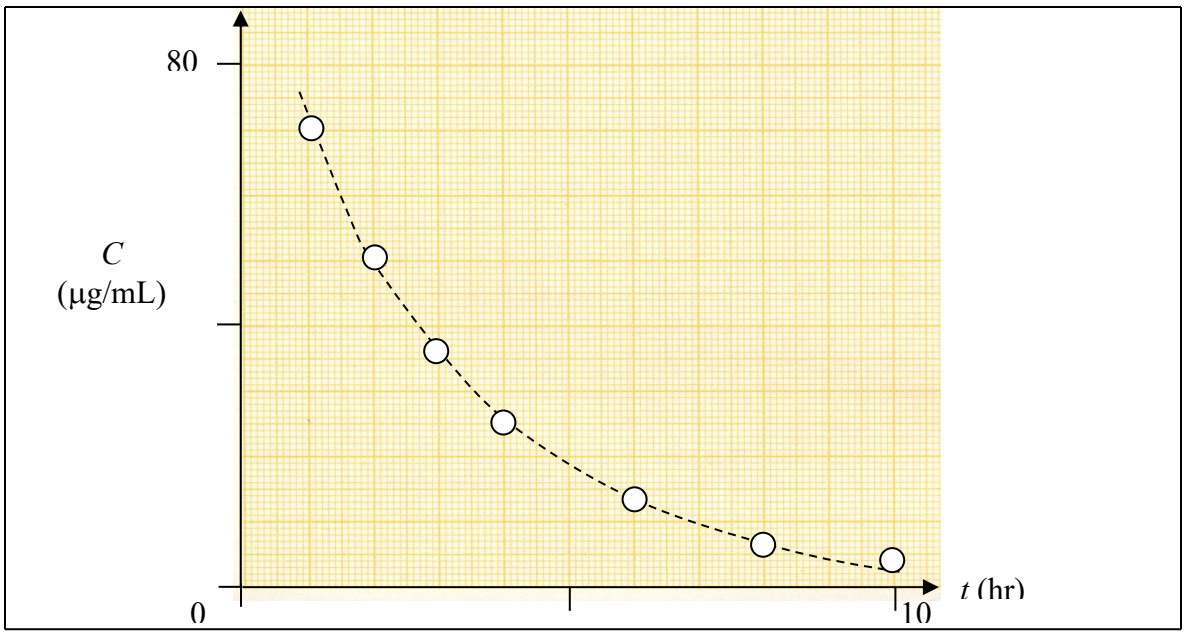
Concentration-time curves

Work through the following example of a one-compartment (whole body) linear (first order) kinetic model for a drug taken as a single bolus (one-dose).

Sorry Susan is suffering from a massive migraine and has gulped down an un-measured dose of an analgesic drug. Her husband, Pharmacologist Pete, is concerned and regularly monitors the drug concentration in her blood, with the following results:

Time, *t* (hrs)	1	2	3	4	6	8	10
Concentration in plasma, *C* (µg/mL)	70	50	35	25	12.5	6.2	3.1

Plot the concentration-time curve



Having plotted the drug concentration-time curve, we can derive a relevant equation for the changing concentration and then calculate relevant PK parameters; including:

- peak concentration (C_0),
- elimination rate constant (*kel*),
- initial dose (*DOSE*),
- apparent volume of distribution (V),
- area under curve (*AUC*) and
- half-life of elimination ($t_{½}$).

Peter wants to roughly estimate the relationship between concentration and time, so he calculates the **average rate of change in concentration** over time (= slope or tangent of the curve between each sampling time) using the formula: slope = rise over run = $\Delta C / \Delta t$
Complete his calculations:

Time, t (hrs)	1	2	3	4	6	8	10
Concentration in plasma, C (µg/mL)	70	50	35	25	12.5	6.2	3.1

$\Delta C / \Delta t = [C(t_2) - C(t_1)] / (t_2 - t_1)$ (50-70)/(2-1) = -20

$\Delta C / \Delta t = [C(t_3) - C(t_2)] / (t_3 - t_2)$ -15

etc -10

-6.2

-3.15

-1.55

Average C (crude estimation of plot point, not necessarily midpoint) (70+50)/2 = 60

(it would be more appropriate to calculate the slope of the tangent at each of the sampling points, but we only want a rough estimation) 42.5

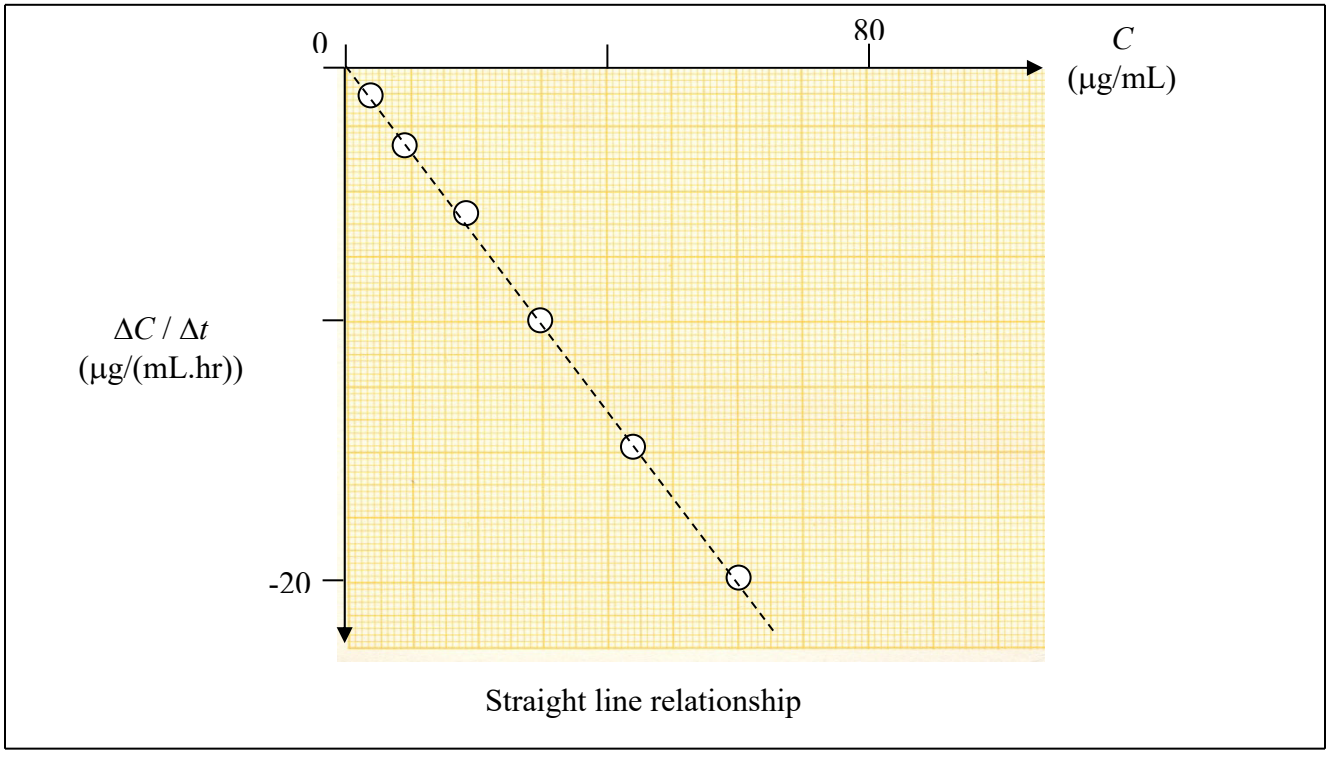
30

18.75

9.35

4.65

Plot the rate of change versus plasma concentration. What type of relationship did you get?



Straight line relationship

When first-order kinetics are obeyed (as in this model), the crude plot will yield a straight line relationship. This behaviour can be expressed mathematically as a differential equation:

rate of change of concentration versus time $= -\Delta C / \Delta t = -dC / dt = kel \,.\, C$

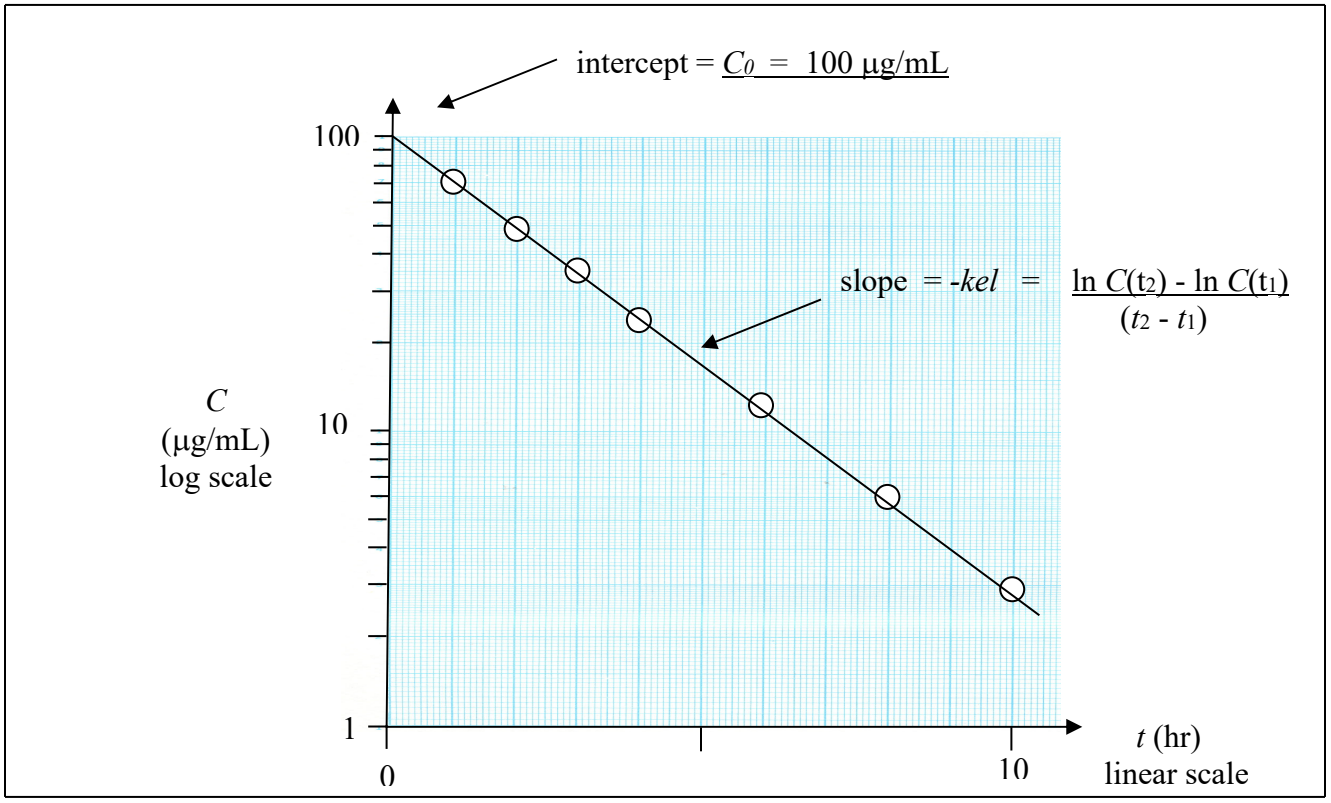
where ***kel*** = a proportionality constant called the "**elimination rate constant**"

By integrating the equation, we can get rid of the differential term and the equation resolves to a single exponential equation (whereby the fall in plasma concentration follows monoexponential decay):

$$C_t = C_0 \,.\, e^{-kel.t}$$

Make a semi-log plot of C versus t to determine the **initial (peak) plasma concentration C_0**
(the plot should give a straight line with an intercept of C_0).

Time, t (hrs)	1	2	3	4	6	8	10
Concentration in plasma, C (μg/mL)	70	50	35	25	12.5	6.2	3.1



Calculate the **elimination rate constant (*kel*)** by determining the slope of the line (= -*kel*) from $t = 1$ to $t = 6$.

slope $= -kel = [\ln C(6) - \ln C(1)] / [t(6) - t(1)]$
$= [\ln 12.5 - \ln 70] / [6 - 1]$
$= [2.5257 - 4.2485] / 5$
$= -0.345$
so $kel = 0.345\ hr^{-1}$

This model assumes that the dose of drug taken was quickly absorbed and uniformly mixed throughout the whole volume of distribution. In pharmacology, this is defined as the **apparent volume (*V*)** which equals the amount of drug in the body divided by the plasma concentration; that is,

$V = DOSE / C_0$

Calculate the initial dose take by Susan if $V = 10$ L

$DOSE = C_0 . V$
$= 100\ \mu g/mL \times 10\ L$
$= 100\ mg/L \times 10\ L$
$= 1{,}000\ mg$
$= 1\ g$

The **area under the plasma concentration-time curve (*AUC*)** is very useful for calculating the relative efficacy of different drug products. Using the trapezoidal rule and integration, we can resolve:

$AUC = C_0 / kel$ [which is equivalent to $DOSE / [V . kel]$

Calculate *AUC* in our model.

$AUC = C_0 / kel$
$= 100 / 0.345 = 300\ \mu g/mL\ hr$

An important property of first-order kinetics is the **half-life of elimination ($t_{½}$).** This is the time taken for the plasma concentration to fall to half its original value, where:

$\ln C/2 = \ln C - kel . t_{½}$
so $\ln [(C/2).(1/C)] = - kel . t_{½}$
$\ln 2 = kel . t_{½}$
thus $t_{½} = \ln 2 / kel = 0.693 / kel$

Calculate $t_{½}$ in our model

$t_{½} = \ln 2 / kel$
$= 0.693 / 0.345 = 2\ hr$

<u>Mastery Question</u>

In medicine (and forensics), we can use established values of *kel* for different drugs and *V* for average patients to calculate the dosage required to yield a plasma concentration at a given time. In this model where $kel = 0.345\ hr^{-1}$ and $V = 10$ L, what *DOSE* would Susan have taken to have a plasma concentration of 12.5 μg/mL after 6 hours?

We are given $t = 6$ hr, $C_6 = 12.5$ μg/mL (= 12.5 mg/L), $V = 10$ L, and $kel = 0.345\ hr^{-1}$

We substitute $C_0 = DOSE / V$ and $t = 6$ into our exponential equation $C_t = C_0 . e^{-kel.t}$
to give $C_6 = [DOSE / V] . e^{-kel \times 6}$
yielding $12.5 = [DOSE / 10] . e^{-0.345 \times 6}$
which solves to give $DOSE = 990.5$ mg
that is, ~ 1 g (which is similar to the *DOSE* calculated previously)

Exemplar: ALCOHOL

The word 'alcohol' has a broad general meaning, often refering to ethanol (grain alcohol or wine spirits) or to any alcoholic beverage. In chemical terms, an alcohol is any organic compound in which a hydroxyl group (-OH) is bound to a carbon atom of an alkyl or substituted alkyl group. The general formula for a simple alcohol is $C_nH_{2n+1}OH$. Ethanol (C_2H_5OH) is a colorless, volatile liquid obtained by the fermentation of sugars. It is the most widely used and abused agent throughout the world. Acute alcohol abuse can lead to intoxication characterized by mild to profound sensory, motor and cognitive impairment. Chronic abuse can lead to psychological dependence and produce a variety of physical diseases significantly influencing morbidity and mortality.

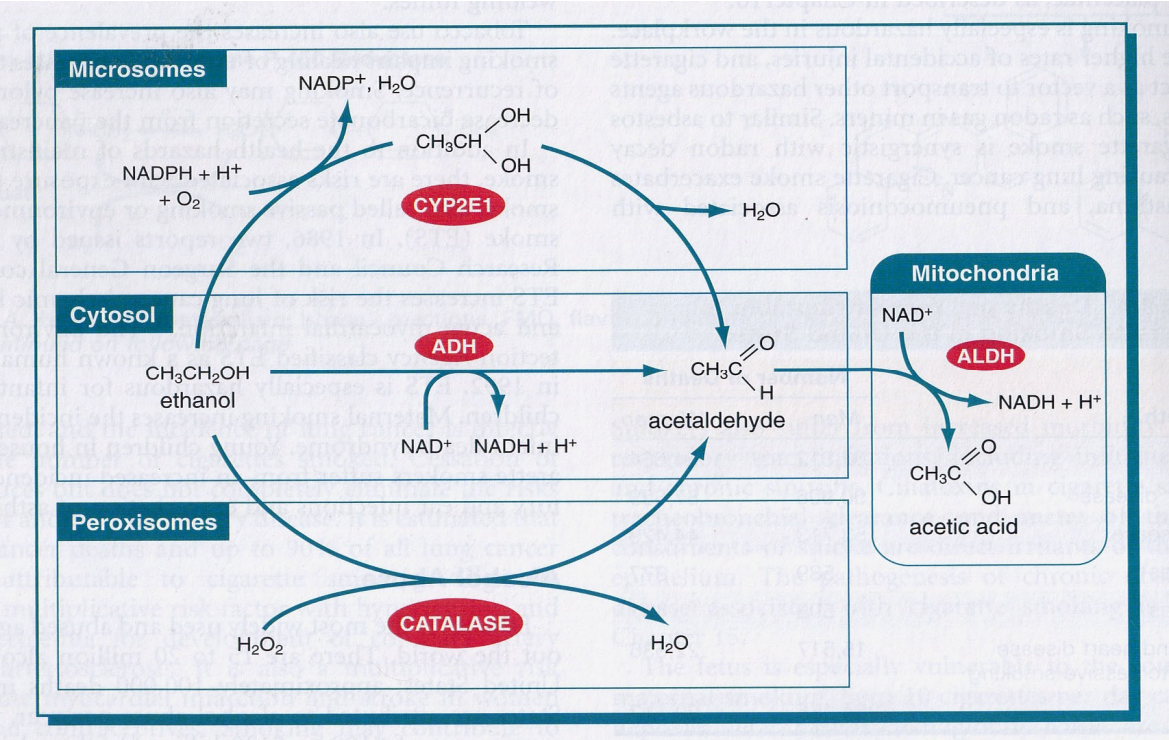
Pharmacodynamics (PD):

Ethanol is a central nervous system (CNS) depressant, but its actual mode of action is unknown (despite some resemblances to anaesthetics). It is thought that ethanol affects fluidization of membrane phospholipids and alters signal transduction in the brain (involving inhibition of GABA (γ-aminobutyric acid) neurotransmitters, voltage-gated calcium channels and NMDA (N-methyl-D-aspartate) receptors). Acute ethanol intoxication include slurred speech, motor incoordination, increased self-confidence and generally acting stupid. Chronic use can cause a wide range of systemic effects, involving the liver (fat deposits, inflammation, scar formation), CNS (cerebellar degeneration, peripheral neuropathy), cardiovascular system (cardiomyopathy, hypertension), gastro-intestinal tract (gastritis, pancreatitis), and the reproductive system (testicular atrophy, abortion, foetal alcohol syndrome). To date, no specific receptor for ethanol has been identified, but chronic use results in psychologic and physical dependence.

Pharmacokinetics (PK):

Alcohol is readily absorbed from the gastro-intestinal tract, and is one of only a few substances that can be absorbed from the stomach (20-25% from stomach, 75-80% from duodenum). It fits somewhere between a food and a drug. It supplies calories but cannot be broken down and stored as protein, fat or carbohydrate (it yields 7.1 kcal/g). Peak absorption rates range from 0.5-2 hours after ingestion, depending on the fasting or fed state of the individual and their metabolic rate, age, sex, and weight as well as the quantity and frequency of ingestion. Many people drink socially and consume several to many drinks over a single evening (with binge drinking being popular amongst teenagers). Once absorbed through the gut, alcohol is distributed rapidly throughout the body in the bloodstream where it is removed by a combination of metabolism, excretion and evaporation. The relative proportions vary from person to person, but typically most alcohol is metabolized by the liver (~90%), some evaporates through the breath (1-5%), some is excreted in urine (1-3%) and a very small proportion is excreted in sweat, tears, etc (<0.5%). Excretion into urine typically begins after about 40 minutes, whereas metabolisation commences as soon as the alcohol is absorbed. Alcohol metabolism proceeds simultaneously in hepatocytes by three pathways:

- the alcohol dehydrogenase (ADH) system in the cytoplasm;
- the microsomal ethanol-oxidizing system (MEOS) in the endoplasmic reticulum; and
- the catalase system in peroxisome organelles.



All three pathways result in the production of acetaldehyde, a very toxic metabolite which causes facial flushing, increased heart rate and nausea. Acetaldehyde is subsequently metabolized to acetic acid by the enzyme aldehyde dehydrogenase (ALDH). Alcohol metabolism requires a cofactor, nicotinamide adenine dinucleotide (NAD), that is also necessary for other metabolic processes, such as the metabolism of pyruvates, urates and fatty acids. Because alcohol competes for use of NAD, it disrupts liver metabolism, resulting in metabolic acidosis, impaired glycolysis, enhanced lipid synthesis (fat accumulation), ketogenesis (ketone production) and fibrogenesis (fibre production). These disturbances lead to progressive liver disease, characterized by three stages:

- fatty liver disease (hepatocytes become filled with lipid vacuoles);
- alcoholic hepatitis (inflammation or swelling of the liver), and finally
- cirrhosis (fibrotic scar formation).

Many physiologically active materials are removed from the bloodstream (by metabolism or excretion) at a rate proportional to their concentration, so they exhibit exponential decay with a characteristic half-life. However, this is not true for alcohol. Typical doses of alcohol actually saturate the enzymes' capacity, so that alcohol is removed from the bloodstream at an approximately constant rate. Saturation kinetics, also known as zero-order kinetics and simple kinetics, means that the time course of disappearance of the drug from the plasma is linear (the drug is removed at a constant rate that is independent of plasma concentration).

The ethanol clearance rate varies considerably between individuals; experienced male drinkers with a high body mass may process up to 30 grams per hour (but a more typical figure is 10 grams per hour). Persons below the age of 25, women, persons of certain ethnicities, and persons with liver disease may process alcohol more slowly. There are currently no known drugs or other ingestible agents which will accelerate alcohol metabolism. Alcohol ingestion can be slowed by ingesting alcohol on a full stomach. Spreading the total absorption of alcohol over a greater period of time decreases the maximum alcohol level, decreasing the hangover effect.

Blood alcohol concentrations (BAC):

Driving motor vehicles under the influence of alcohol (DUI) is illegal in most countries due to the greater frequency of involvement in crashes resulting in injury and death. Strict legal limits have been established by law enforcement agencies (0.05-0.10%) and they are rigidly monitored in systematic and random testing programs. Individuals can be compelled to provide samples (breath or blood) for alcohol content determination. BAC is usually measured as mass per volume. For example, a BAC of 0.05% means 0.5‰ (permille) or 0.5 grams of alcohol per 1000 millilitres of blood. BAC is measured in many different units throughout the world, such as:

1 g/kg = 1 g/L = 100mg/dL = 1 mg/cc = 100 mg% = 1 decigrams% = 0.1 g% = 0.1% = 1 ‰

A BAC reading of 0.2% represents very serious intoxication (most first-time drinkers would be unconscious by about 0.15%), and 0.35% represents potentially fatal alcohol poisoning. The accepted LD_{50} (or lethal dose for 50% of adult humans) is 0.4%. For habitual heavy drinkers, those numbers can at least double. In extreme cases, individuals have survived BACs as high as 0.9%.

The number of drinks consumed is a very poor measure of intoxication largely because of variation in physiology and individual alcohol tolerance. However, it is generally accepted that the consumption from sober of two standard drinks (containing a total of 20 grams) of alcohol will increase the average person's BAC roughly 0.05% (a single standard drink consumed each hour after the first two will keep the BAC at approximately 0.05%), but there is much variation according to body weight, sex, and body fat percentage. Furthermore, neither BAC nor the number of drinks consumed are necessarily accurate indicators of the level of impairment. Tolerance to alcohol varies from one person to another, and can be affected by such factors as genetics, adaptation to chronic alcohol use, and synergistic effects of other drugs.

BAC can be roughly estimated using a mathematical approach (not as accurate as instrument analyses) by measuring the total amount of alcohol consumed (in grams) divided by the total amount of water in the body (in litres), effectively giving alcohol per volume. The total water weight of an individual can be calculated by multiplying his or her body weight by their percent water. Gender plays an important role in the total amount of water that a person has. In general, men have a higher percent of water (~58%) than women (~49%), thereby lending credence to the generalization that men require more alcohol than women to achieve the same BAC level. Men are also generally heavier than women. The more water a person has, the more alcohol is required to achieve the same BAC level.

Most calculations of alcohol to body mass simply use the weight of the individual, and not specifically their water content. Nonetheless, consider a 68 kg woman, she would have a total amount of water of 68 x 0.49 = 33.3 kg. This is equivalent to 33,300 mL of water (1 L of water weighs 1 kg, and 1 L = 1000 mL). If she had consumed 2 standard drinks (total of 20 g alcohol) within a short period of time, her BAC would be 20 g per 33,300 mL = 0.06%.

Over 70 years ago, a mathematical model was developed by Widmark to predict BAC. It is still widely used in forensic science. The model is an open 1-compartment model with zero-order elimination process - it is assumed that the alcohol after consumption is quickly taken into the body and spread over the total body water, i.e. it is distributed rapidly into the bloodstream from the stomach and small intestines, and further into the watery fluids in and around somatic cells. Thereafter, the alcohol is eliminated at a constant rate.

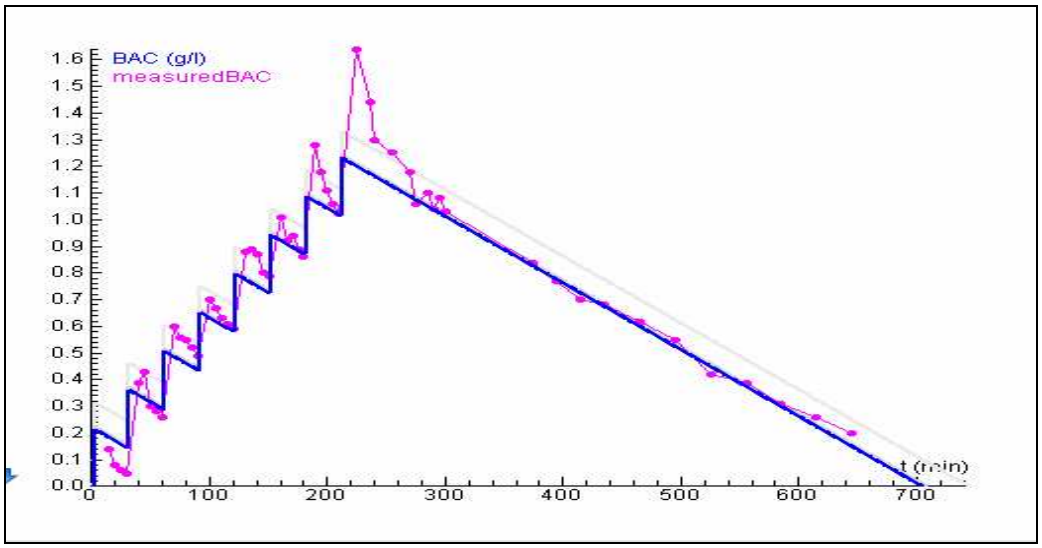
After absorption, the blood alcohol level is represented in the Widmark model by the formula:

$$BAC = [D / (r.W)] - \beta.t$$

where : D = the amount of alcohol consumed (in grams),
r = the so-called Widmark factor,
W = the body weight (in kg),
β = the rate of metabolism (clearance rate in g/L/h), and
t = the time (in hours) after consuming alcohol.

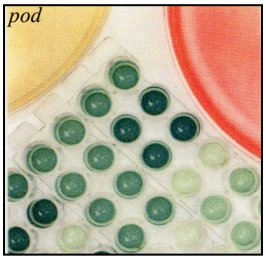
The rate of alcohol metabolism is individual, it differs for men and women, it varies with age, and it depends on circumstances (before or after a meal). Generally, it varies from 0.10-0.20 g/L/h. The Widmark factor is also individual and depends mainly on body composition. Mean values are 0.68 for men and 0.55 for women (women generally have a higher percentage of body fat and therefore less body water than men). The product $r.W$ is equal to the volume of distribution (V_d), the theoretical volume of total body water compartment into which the alcohol is distributed. Various methods can be found in the literature to estimate r (or V_d) from variables such as height (H, in cm), weight and age:

e.g. $r_{(men)} = 0.3161 - 0.004821.W + 0.004632.H$
$r_{(women)} = 0.3122 - 0.006446.W + 0.004466.H$



The graph shows the BAC in an individual who has consumed 8 standard drinks over 3½ hours. While most clearance graphs show trend lines as smooth curves or straight lines, in reality BAC can fluctuate wildly over very short time intervals, an effect known as steepling (giving a zig-zag concentration-time profile). Some steepling may be due to variable test conditions and heterogeneous reagents/samples, but most is due to physiological processes, including food consumption, periodic evacuation of stomach contents, variations in peripheral blood flow, hypertension, activity levels and other stressors. Despite many attempts by defendants accused of DUI offences to cast doubt on BAC readings, the technologies approved by judicial systems as the basis for prosecution have been validated by large cohort studies and include internal and external standards. Do the crime, pay the fine!

Chapter 13.



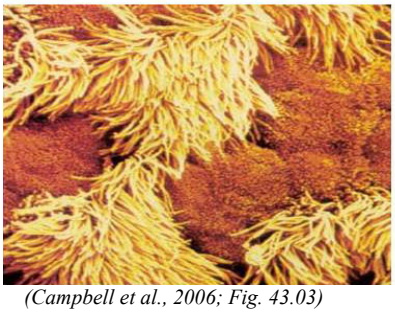
IMMUNO-SEROLOGY

Vocabulary list: immunity, innate, acquired, adaptive, lymphocytes, T cells, B cells, serum, antibody, antigen, immunoserology, dilution, titre, prevalence, sensitivity, specificity, false positive, false negative

The immune system

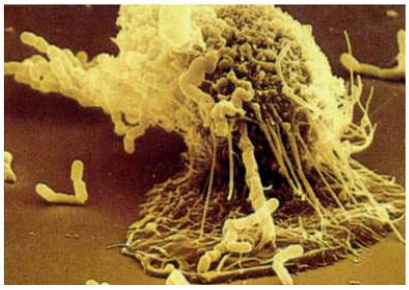
The human body must defend itself against invasion by pathogenic organisms (viruses, bacteria, fungi, protozoa, helminths, arthropods) as well as the uncontrolled growth of abnormal cells (tumours). The immune system provides protection against both, but its various elements must be constantly regulated to maintain an appropriate level of protection. Heightened responses may cause auto-immune diseases, allergic reactions and hypersensitivity, while deficient responses leave us susceptible to infections and tumour (cancer) development. Cells of the immune system are produced in special organs and tissues and move around the body in the circulatory and lymphatic systems. They communicate and interact using numerous molecular signals (chemokines, cytokines, lymphokines). Collectively, there are three lines of immunological defense: barriers formed by membranes; innate immunity provided by phagocytes; and acquired (adaptive) immunity reliant on lymphocytes.

membrane barrier

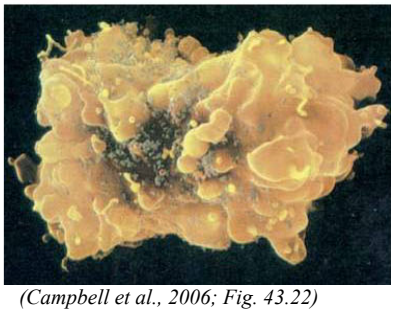


(Campbell et al., 2006; Fig. 43.03)

phagocyte (macrophage)



lymphocyte (T cell)



(Campbell et al., 2006; Fig. 43.22)

Cells and organs

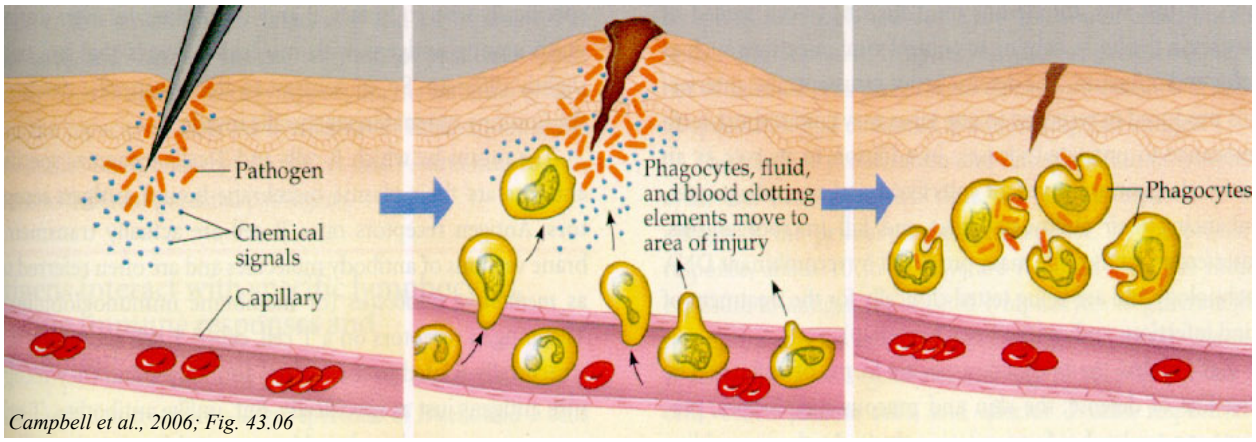
All cells of the immune system are formed by pluripotent stem cells in the bone marrow. Myeliod precursors give rise to phagocytic and dendritic cells, while lymphoid precursors give rise to lymphocytes. Cells of the immune system are evident in the blood as leucocytes (white blood cells) [both granulocytes (neutrophils, eosinophils and basophils) and agranulocytes (lymphocytes)]; in tissues as macrophages and antigen-presenting cells; and in lymphoid organs and tissues. Primary lymphoid organs (bone marrow and thymus) are said to be 'productive'; they 'produce' effector cells. In particular, they are the sites where lymphocytes proliferate, differentiate, acquire specific antigen receptors, learn to distinguish between self and non-self, and mature into functional cells (**B** cells in the **B**one marrow, and **T** cells in the **T**hymus). Secondary lymphoid organs are said to be 'reactive'; they are the sites where lymphocytes 'react' with foreign material (antigens) initiating specific immune responses. These organs include lymph nodes (which filter extracellular fluid, = lymph), the spleen (which filters the blood), and mucosal-associated lymphoid tissue (MALT)(which protects parts of the gastrointestinal, respiratory and genitourinary tracts).

Barrier defenses (nonspecific)

The first line of defense comprises physical and chemical barriers to infection; membranes and their secretions that foreign invaders first encounter. These barriers are nonspecific; that is, they are presented to all invaders, irrespective of type. The skin presents a formidable external barrier, while mucous membranes line all internal tubular organs producing mucus and various other chemical secretions (particularly the enzyme lysozyme). If invaders penetrate barrier defenses, they then meet hungry killer cells.

Innate defenses (nonspecific)

The second line of immunological defense involves phagocytic cells that seek out invaders in host tissues. The phagocytic cells may reside within specific tissues (fixed macrophages) or circulate throughout the body in the blood (neutrophils and monocytes). They engulf and digest foreign material irrespective of type: thus the immunity is said to be innate and nonspecific. It has long been known that infections elicit inflammation, innate responses characterized by four cardinal signs: redness (rubor), heat (calor), swelling (tumour) and pain (dolor).



Campbell et al., 2006; Fig. 43.06

Localized inflammation occurs when injured cells release chemical signals (chemokines) causing dilation and increased permeability of blood vessels, allowing fluid and large numbers of phagocytes to enter the tissues to seek out and destroy invaders. Inflammation of organs is signified by adding the postfix '*–itis'* to the organ name (hepatitis indicating inflammation of the liver, encephalitis that of the brain, etc). Innate immune responses are particularly effective against bacterial invaders, whereas other larger organisms have become quite adept at avoiding death by phagocytosis. Survivors are not safe, however, as they then trigger a multitude of other defensive mechanisms.

Acquired defenses (specific)

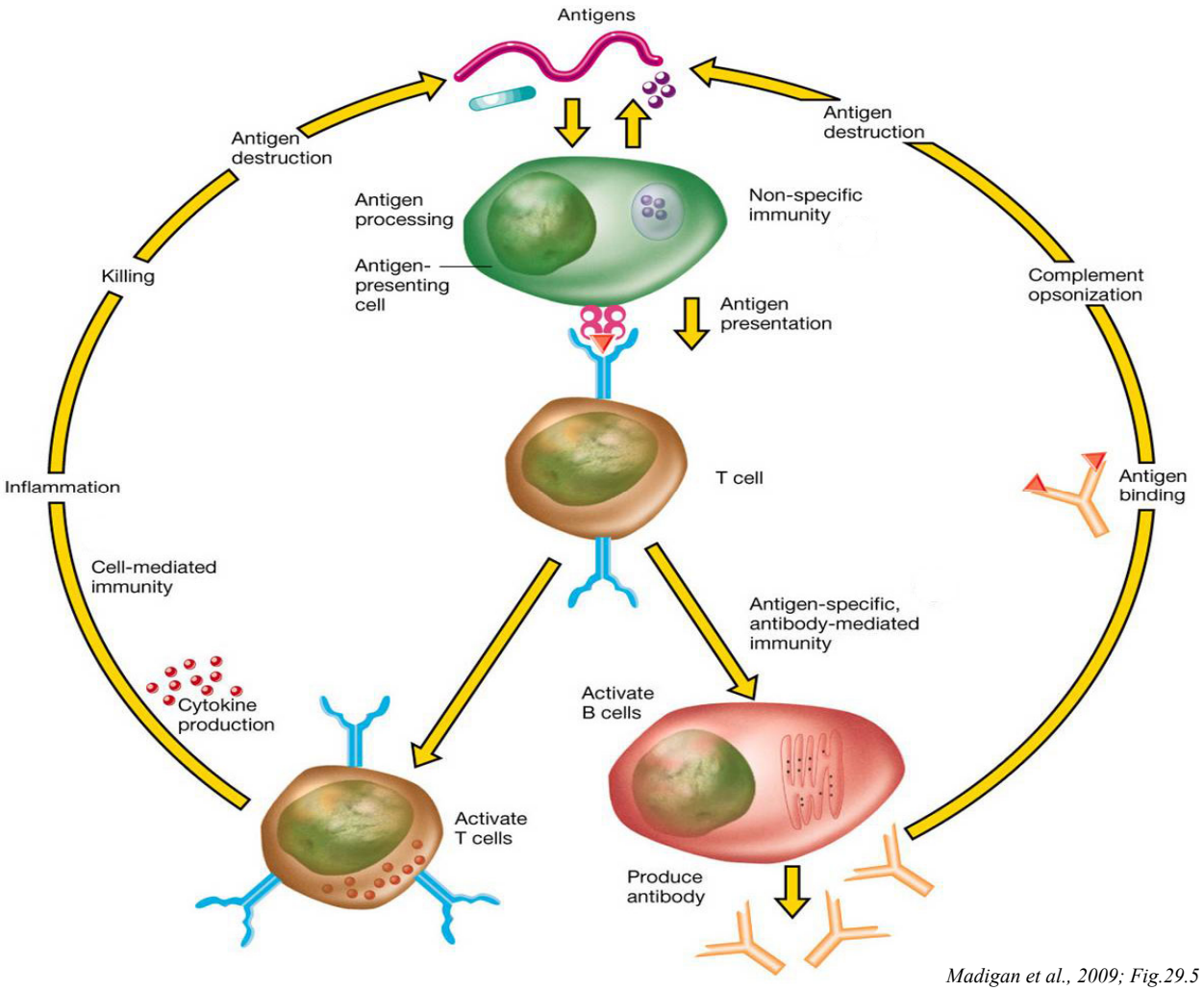
The third line of defense (adaptive or acquired immunity) is highly specific and is reliant on lymphocytes (providing specificity, diversity, memory and self-tolerance). T cells invoke cell-mediated responses while B cells invoke humoral responses (involving antibodies). Both T and B cells recognize antigens via specific surface antigen receptors. Self tolerance develops as cells bearing receptors for native molecules are destroyed or rendered nonresponsive. Major histocompatibility complex (MHC) molecules are crucial to antigen presentation. Class I MHC molecules located on all nucleated cells present antigens to cytotoxic T cells, while class II MHC on macrophages and B cells present antigens to helper T cells.

When an antigen binds to a particular lymphocyte, it is activated to produce numerous identical copies (clonal selection). The primary immune response (first exposure to an antigen) results in clones of short-lived infection-fighting effector cells as well as clones of long-lived memory cells. Subsequent exposure to the same antigen activates the memory cells and the resultant secondary immune responses are faster, stronger and often protective (providing the basis for vaccination).

Cell-mediated responses (defense against intracellular pathogens)

Helper T cells interact with antigen-class II MHC complexes on macrophages through T cell receptors and the cell surface protein CD4. Contact stimulates the helper T cells to grow and divide, as does interleukin-1 secreted by macrophages. The helper T cells secrete interluekin-2 which activates other cells in the immune system; including cytotoxic T cells (stimulated to become active killer cells), B cells (stimulated to become antibody-producing plasma cells) as well as additional helper T cells (stimulated to divide more rapidly and increase cytokine production).

Viral, bacterial and tumour antigens are gathered in nucleated cells of the body and are presented on the surface by class I MHC proteins. They are recognized by cytotoxic T cell receptors and the cell surface protein CD8. Cytotoxic T cells are stimulated by contact and by interleukin-2 to release the protein perforin which punctures the target cell, allowing water and ions to rush in causing swelling and lysis.



Madigan et al., 2009; Fig.29.5

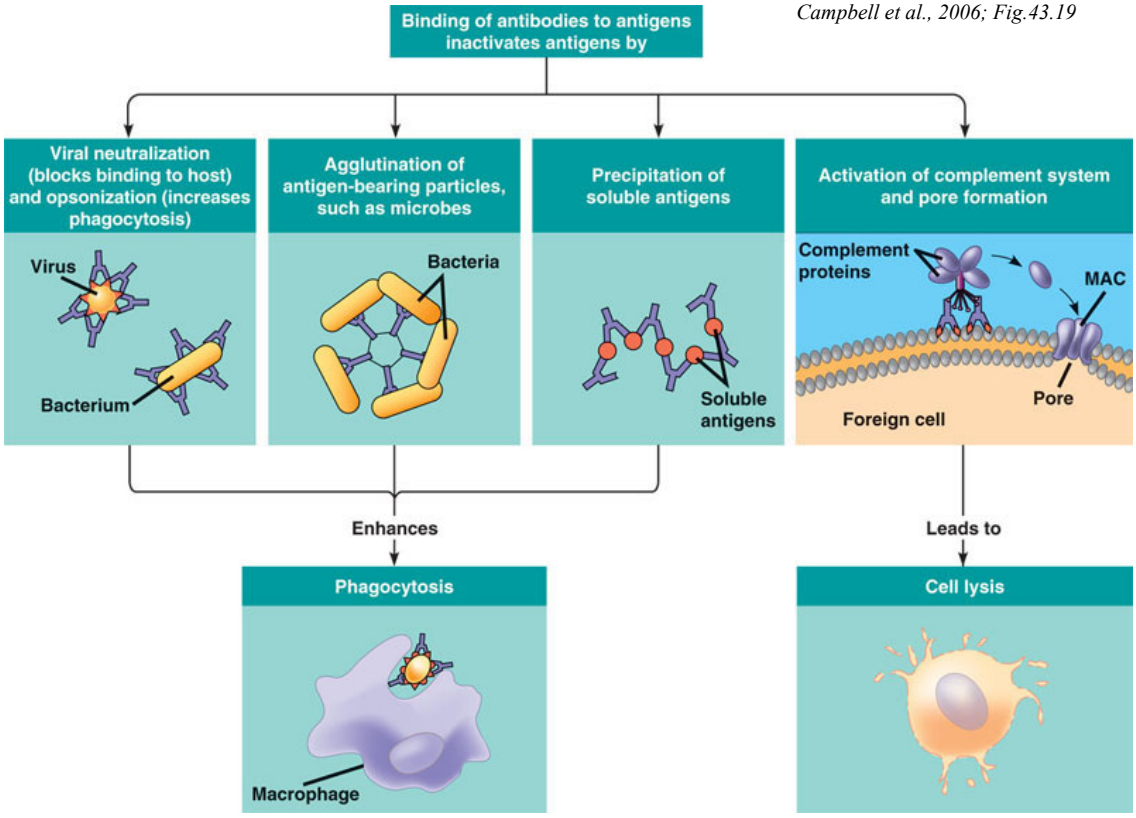
Humoral responses (defense against extracellular pathogens)

B cells have antibody receptors in their plasma membranes specific for particular antigens. Successful binding stimulates B cell proliferation producing clones of plasma cells and memory cells. Plasma cells secrete antibodies in increasing amounts peaking 10-17 days after activation. Antibodies flow through the body's fluids and tag foreign cells and molecules for destruction.

Helper T cells may also contribute to humoral responses involving T-dependent antigens. B cells may internalize antigens by endocytosis where they bind with class II MHC proteins and move to the cell surface. The B cell now presents the same antigen that fits in the receptor of the helper T cell. When contact occurs, the helper T cells secrete interleukin-2 and other cytokines that activate B cells to divide into plasma and memory cells.

The binding of antibodies to antigens tags foreign cells and molecules for destruction via neutralization, agglutination, precipitation and complement fixation. Neutralization of bacteria and viruses occurs when antibodies bind to and block the activity of the antigen. Neutralized viruses cannot invade cells and microbes and soluble antigens coated with antibodies are phagocytosed by macrophages (enhanced attachment called opsonization). Agglutination is the process whereby bacteria or viruses are effectively neutralized by clumping. Antibodies have at least two binding sites allowing them to crosslink microbes which are then phagocytosed by macrophages (opsonization). Precipitation is the cross-linking of soluble antigens to form immobile precipitates which are then phagocytosed by macrophages (opsonization). Complement fixation is the activation of a series of serum proteins leading to lysis of a microbe. First, complement proteins bind to IgM or IgG antibodies that have tagged a foreign cell. Some of the proteins are activated to form a membrane attack complex that creates a pore in the target cell, allowing water and ions to rush in causing the cell to swell and lyse.

Campbell et al., 2006; Fig.43.19



Hypersensitivity

Hypersensitivity is an adaptive immune response that occurs in an inappropriate or exaggerated way, resulting in tissue damage or some other detrimental response. There are four major categories:

- Type I (immediate) hypersensitivity may occur within minutes and is generally provoked by environmental allergens and involves the production of IgE and degranulation of mast cells. Examples include anaphylactic shock, asthma, hives, and drug allergies.

- Type II (antibody-dependent cytotoxic) hypersensitivity occurs within hours when antibody binds to self or foreign antigens resulting in cell death from phagocytosis, killer cell activity or complement-mediated lysis. Examples include haemolytic anaemia and Rh blood group incompatibility.

- Type III (immune complex) hypersensitivity occurs within days and is caused by the deposition of antigen-antibody complexes in blood vessels and tissues. Examples include serum sickness and glomerulonephritis.

- Type IV (delayed type) hypersensitivity occurs over days and is mediated by T cells and macrophages producing skin reactions (erythema, itching, eczema, necrosis) on second exposure to simple antigens. Examples include contact dermatitis due to nickel, poison ivy, soaps, cosmetics, etc.

Immune pertubations

Autoimmune diseases (such as rheumatoid arthritis and insulin-dependent diabetes) occur when the immune system loses tolerance for self and turns against certain native molecules.

Immunodeficiencies may be congenital (such as agammaglobulinaemia and SCID), acquired through disease (cancers such as Hodgkin's disease) or infection (such as HIV-AIDS) and possibly influenced by physiological condition (reduced due to stress or depression). Individuals are rendered susceptible to a variety of opportunistic infections and pathological disorders.

Immunological recognition determines the success of blood transfusions, tissue grafts and organ transplantations. Incompatibility between donor and recipient can have severe and even fatal consequences. Autografts (involving the same individual) and isografts (involving genetically identical individuals) do not induce immune responses, whereas allografts (involving non-identical individuals) and xenografts (involving different donor and recipient species) do provoke strong immune responses leading to graft rejection (unless strong immunosuppressive drugs are used).

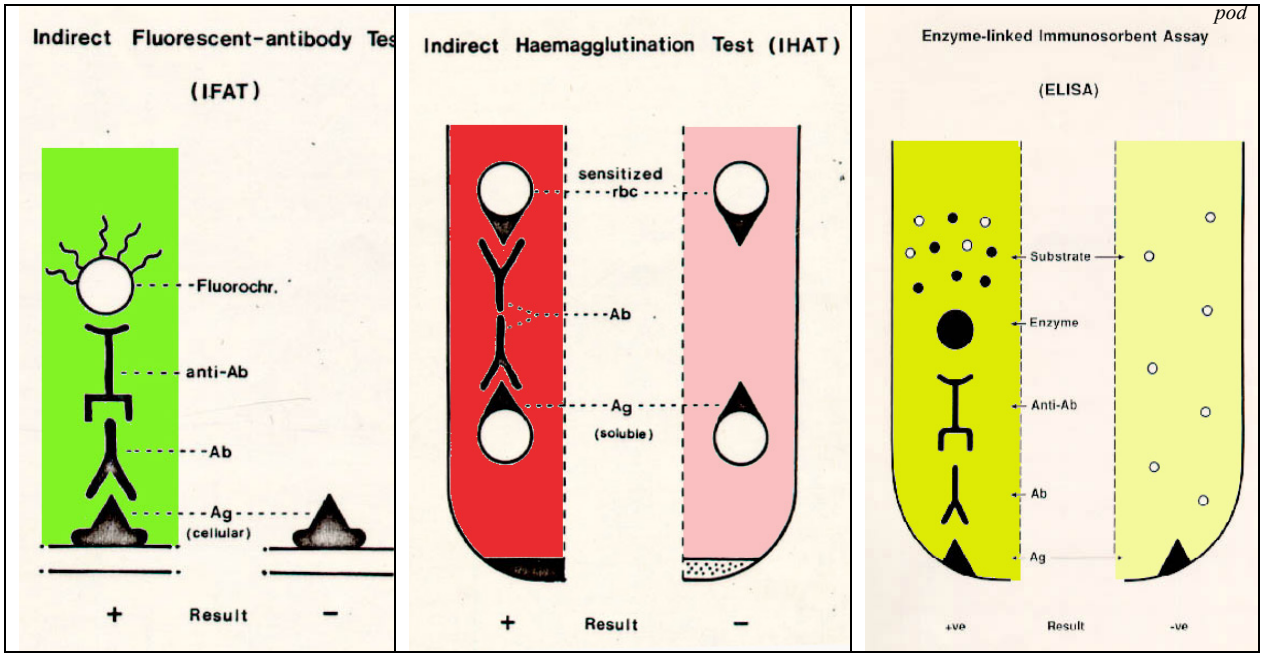
Active protective immunity against disease can be generated by previous exposure (natural infection) or by immunization (vaccination) with inactivated, attenuated or subunit vaccines. Consider your own vaccination history.

Passive immunity involves the transfer of antibodies between individuals, either naturally (such as transplacentally or via maternal colostrum) or artificially (immunoprophylaxis/immunotherapy).

Antibody-antigen tests

Animals respond to most infectious diseases by forming antibodies against the infecting pathogen as part of their immunological defenses. Antibodies (also called gamma-globulins or immunoglobulins) are produced by plasma cells (transformed B lymphocytes) and are secreted into the blood stream to circulate through the body. When they come into contact with the relevant antigen, they bind to it and tag it for destruction.

The presence of specific antibodies is frequently used as an indicator of infection, particularly for diseases which have nonspecific symptoms or clinical signs (fever, diarrhoea). A variety of immuno-serological tests have been developed to detect host antibodies against microbial antigens; including fluorescent-antibody labelling, haemagglutination and enzyme immunoassays.



All tests operate on the same principle. Host serum is incubated in test vessels which have been coated with microbial antigens. If antibodies are present, they bind to the antigens and are subsequently revealed by chemical indicator systems (fluorescence, haemolysis, enzyme-substrate colour change).

It is particularly important to quantitate the amount of antibody present as this provides an indication of the severity of infection and the immunocompetence of the host. The concentration of many other chemicals present in blood (hormones, electrolytes, drugs, etc) can be measured in absolute terms and expressed in specific units (µg/ml, etc). However, measuring the concentration of antibodies is much more difficult due to their variable specificity, cross-reactivity, highly reactive nature, and strong binding affinity to detection systems.

Mathematics came to the rescue! An accurate indication of the amount of antibody present can be obtained by serially diluting the blood to the point it no longer tests positive. The last positive dilution is then called the end-point titre and this indicates the quantity of antibody present. For example, an end-point titre of 1/100 indicates there was enough specific antibody present to still elicit a positive reaction when the blood was diluted 100 times. It is common for antibody titres during active acute infections to be in the vicinity of 1/500 to 1/1000.

Note that this reciprocal notation represents a fraction not a ratio; 1/10 means 1 in 10 (while 1:10 means 1 as to 10). A 1/10 dilution is therefore made by adding 1 part blood (usually serum) to 9 parts diluent (usually physiological saline) to give a 1 in 10 dilution. The most commonly used series are doubling dilutions beginning with a 1/2 dilution (i.e. 1/2, 1/4, 1/8, 1/16, 1/32, 1/64, etc). Serial dilutions are usually made on 96-well plastic microtitre plates (wells arranged in 8 rows and 12 columns).

The results of serological tests are best read objectively and quantitatively; that is, without subjective interpretation by individual operators and with measurement of a related parameter, such as spectrophotometric absorbance (optical density), intensity of fluorescence, or degree of haemolysis. Various instruments have been developed to measure test results, compare and adjust them to standards and reference controls, and then calculate and present the results.

Regrettably, the test results can be influenced by many factors which affect the integrity of the relationships between parameters (such as edge effects, detectable levels, accuracy, interference, competition, cross-reactivity). Nonetheless, the end-point titre (= last positive dilution) of any particular sample can be given with a high degree of confidence.

Interpreting the significance of the test results requires thorough knowledge of the kinetics (onset and duration) and dynamics (intensity) of the host response to infection. Longitudinal samples are obtained to determine whether antibody titres in a particular individual remain stable or whether they are increasing or decreasing (plot titre over time). The results provide strong presumptive evidence on the status of infection within that individual, thus allowing appropriate therapy.

Question:
The cattle on Peter's ranch have become sick, so you collect serum samples to test for antibodies against various microbial pathogens. You prepare serial two-fold titrations of serum in 96-well microtitre plates. Beginning with a 1 in 2 dilution in the first well (top-left corner), what are the highest serial titrations that can be made running down a single column, and along a single row? What mathematical function is involved and what formula applies?

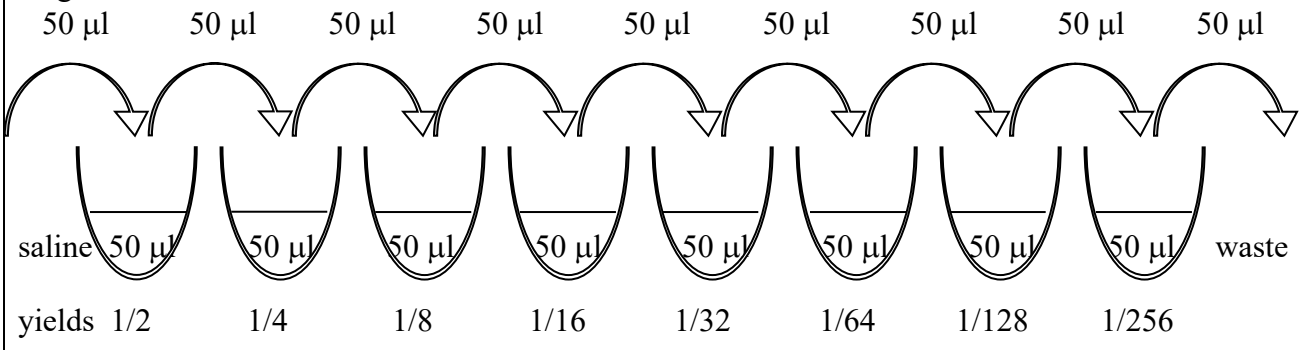
Answer:
Columns - each column has 8 wells so you can make 8 dilutions
(1/2, 1/4, 1/8, 1/16, 1/32, 1/64, 1/128, 1/256)
Rows - each row has 12 wells so you can make 12 dilutions
(1/2, 1/4, 1/8, 1/16, 1/32, 1/64, 1/128, 1/256, 1/512, 1/1024, 1/2048, 1/4096)

Mathematically, this is exponential decay (to the base 2) in the form $f(x) = Ca^{kx} = 1/2^x$
When x = 8, f(x) = 1/256 (this is a titre, not a fraction, so we do not resolve to 0.00390625)
When x = 12, f(x) = 1/4096 (this is a titre, not a fraction, so we do not resolve to 0.00024414

Question:
The serological test you are using requires 50 µl of each titration in each test well. How would you physically make a doubling dilution series from 1/2 to 1/256 (use a diagram to illustrate your answer)? What is the minimum amount of blood and diluent needed to make this series?

Answer:
Serial titration method (need only use one 50 µl micropipette and two tips)
- add 50 µl of diluent to each of 8 consecutive wells
- then add 50 µl of whole blood to first well (giving 100 µl of 1/2 dilution), mix
- then take 50 µl of 1/2 dilution and add to next well (giving 100 µl of 1/4 dilution)
- repeat another 6 times

diagram



In total, need 8 x 50 µl diluent = 400 µl (= 0.4 mL), and 50 µl of blood

Question:
You conduct an enzyme immunoassay on eight dilutions of serum from your prize bull and measure the intensity of colour (absorbance) in each test well:

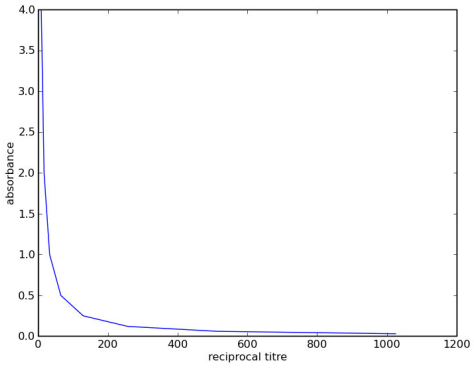
titre	1/8	1/16	1/32	1/64	1/128	1/256	1/512	1/1024
absorbance	4.0	2.0	1.0	0.5	0.25	0.12	0.06	0.03

Assuming the central concept of serial titration holds true (namely, that every double dilution halves the antibodies present, and therefore should halve the absorbance reading), what would the shape of the curve be if you plotted reciprocal titre (x) against absorbance (y) on graph paper with:

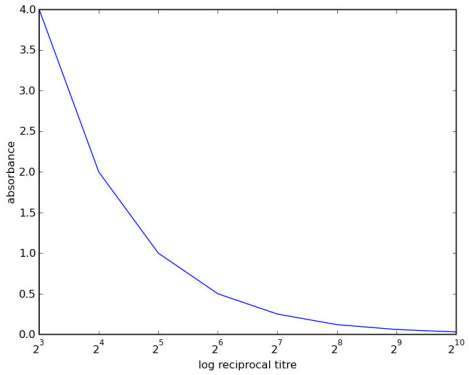
(i) arithmetic scales for both x and y
(ii) logarithmic scale for x and arithmetic scale for y
(iii) arithmetic scale for x and logarithmic scale for y
(iv) logarithmic scales for both x and y

Answers:

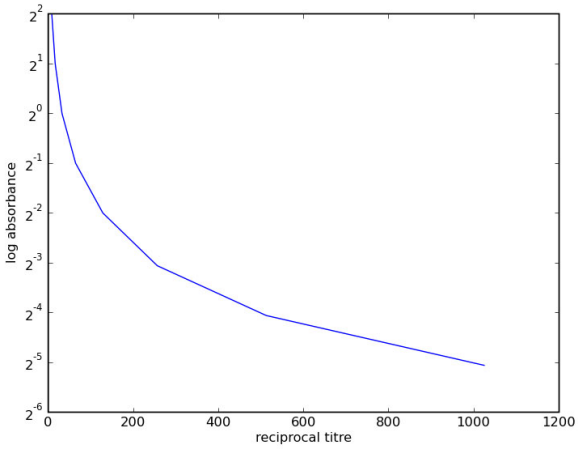
i) arithmetic plot → exponential curve



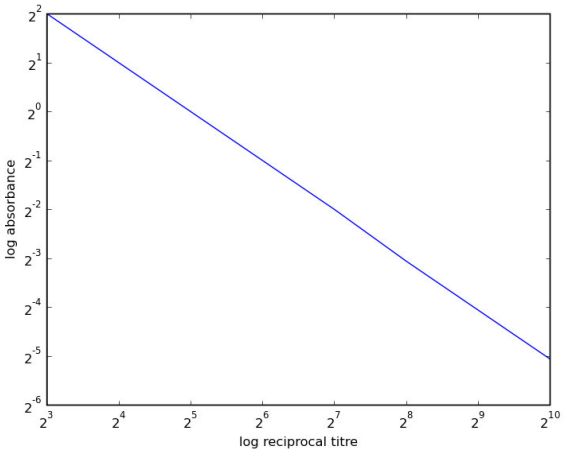
ii) semi-log (x) plot → exponential curve



iii) semi-log (y) plot → exponential curve



iv) double log plot → linear (straight line)



Test Sensitivity and Specificity

All serological tests provide indirect presumptive evidence of infection by demonstrating the presence and amount of antibody against a particular pathogen. These tests are not 100% perfect for a variety of reasons; such as:

- acute infection (host recently infected and antibodies not yet formed)
- chronic infections (host infected and cured years ago but antibodies persist)
- low dose infections (host immune response so low it cannot be measured)
- cross-reactivity (antibodies cross-react against other antigens)

It is therefore important that users of any particular test know how good it is (as determined by objective quantitative assessment). Measurement of test efficacy is done by comparing it against another more conventional test held to be very accurate (representing the 'gold standard'). Test sensitivity and specificity are then calculated as shown in the following table. Regrettably, gold standard tests are not always available so serological test efficacy may be poorer than reported as they were compared against semi-flawed tests. Nonetheless, these parameters (usually expressed as percentages) should be included in the product information accompanying all commercial test kits.

INFECTION STATUS

		INFECTED	NOT INFECTED	
TEST RESULT	POSITIVE	true positive (A)	false positive (B)	(A+B)
	NEGATIVE	false negative (C)	true negative (D)	(C+D)
		(A+C)	(B+D)	(A+B+C+D)(=N)

Prevalence: True prevalence of infection = (A+C)/N
Test prevalence (seroprevalence) = (A+B)/N

Accuracy: Accuracy = (A+D) / N

Sensitivity is the probability of a positive test in an infected individual = A/(A+C)
[clinical pneumonic = SnNout = with a sensitive test, a negative result rules the diagnosis out]
High sensitivity means there were few or no false negatives, while low sensitivity means they were frequent. The consequences of returning a false negative result (when the patient is actually infected) involves failure to treat, worsening of the disease and possible fatality.

Specificity is the probability of a negative test in a non-infected individual = D/(D+B)
[clinical pneumonic = SpPin = with a specific test, a positive result rules the diagnosis in]
High specificity means there were few or no false positives, while low specificity means they were frequent. The consequences of returning a false positive result (when the patient was actually not infected) include unwarranted treatment, waste of time and money, and potential harm to the patient through invasive treatment or toxic side-effects.

Question:
You compared the results of your antibody test with clinical observations on the disease status of each cow; revealing positive antibody responses in 30 diseased cows and 20 non-diseased cows, and negative antibody responses in 10 diseased and 40 non-diseased cows. Construct a result table, and calculate the '**sensitivity**' and '**specificity**' of the antibody test.
cf. Good website for definitions: www.cebm.utoronto.ca/glossary/spsn.htm

Answer:

Construct 2x2 result table

	Diseased	Non-diseased
Antibody positive	true positive (A) = 30	false positive (B) = 20
Antibody-negative	false negative (C) = 10	true negative (D) = 40

Sensitivity is the probability of a positive test in a diseased individual
= A/(A+C) = 30/40 = 0.75 (or 75%)

Specificity is the probability of a negative test in a non-diseased individual
= D/(D+B) = 40/60 = 0.67 (or 67%)

Chapter 14.

ECOLOGY

Vocabulary list: Functions, graphics, matrices, logarithms, exponentials, ecology, biosphere, ecosystems, communities, populations, life tables, survivorship curves, exponential growth, logistic growth

Clip Art

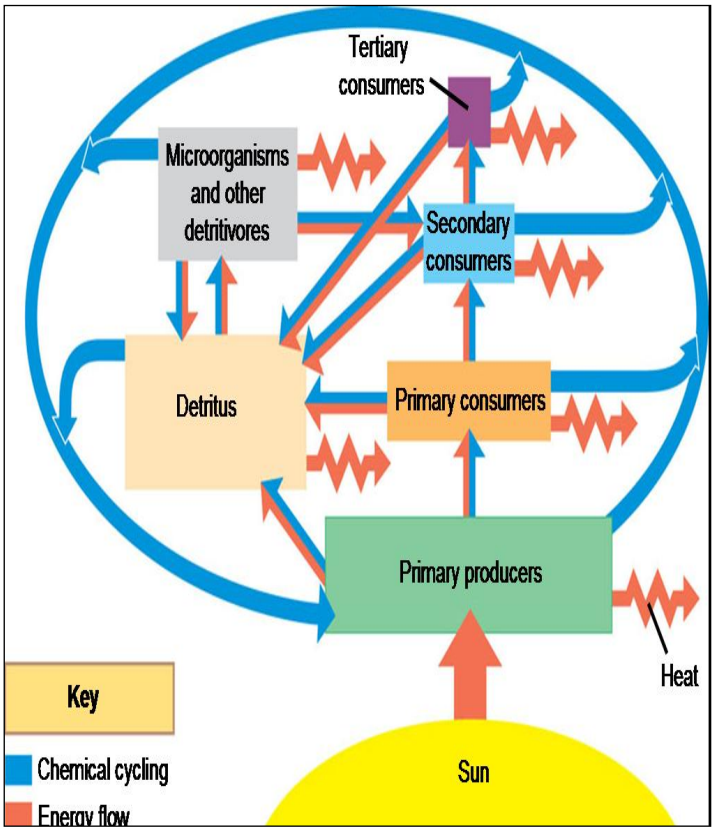
Co-existence

Scientists recognize many different levels of biological organization, from the miniscule to the majestic. We have examined molecules (the building blocks of matter) and cells (the basic units of life). We know living things range from single-celled organisms (simple, but by no means primitive) to multicellular organisms (with cellular specialization to form complex tissues and organs). We also recognize several levels of collective co-existence, where organisms live together in:

- populations (all the individuals of a species within a given area);
- communities (all species of living organisms within a given area);
- ecosystems (all living things within a given area, together with all the non-living components in that area with which life interacts); and the
- biosphere (all the environments on Earth inhabited by life).

As you can see, the definitions of these collective concepts have elastic boundaries, in that the area of study must be given or specified by the scientist. Accordingly, ecosystems can vary in size from an aquarium to a lake, meadow, mountain range, or continent. The dynamics of all ecosystems includes two major processes:

- nutrient cycling; and
- energy flow.

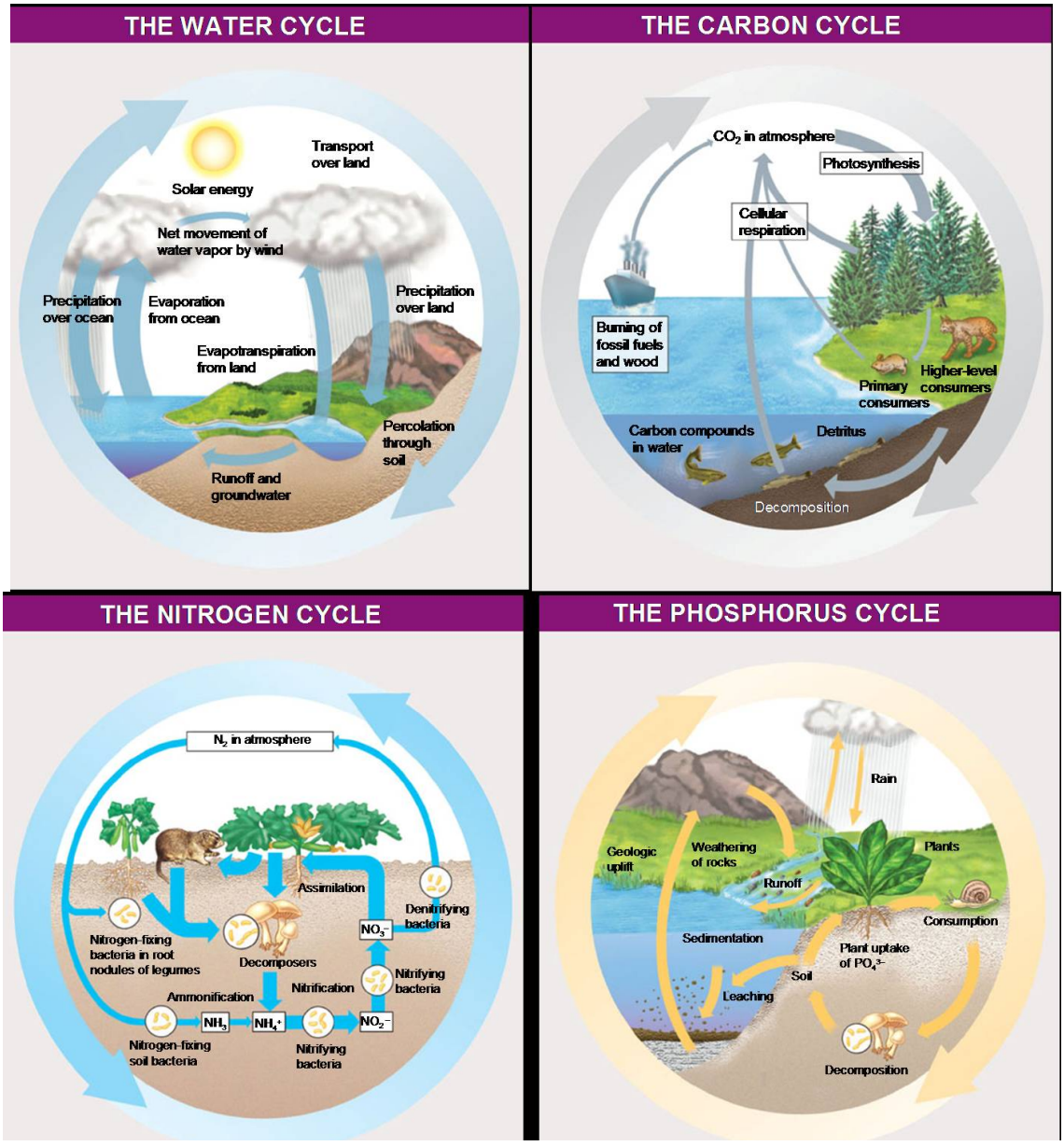


Campbell et al. 2006; Fig. 54.2

Nutrients are constantly recycled within ecosystems, they are used to build organic materials, which subsequently degrade releasing them back into the system. All chemical elements (such as carbon and nitrogen) pass through complex cycles that incorporate both living and nonliving parts of an ecosystem. In contrast, energy constantly flows into an ecosystem (usually as sunlight), where it is converted to chemical energy by producers (usually photosynthetic organisms) and utilized by consumers (herbivores and carnivores) and decomposers (microbes).

Ecosystems are therefore said to recycle matter while energy flows through.

Ecology is the study of interactions between organisms and their environments; a holistic science involving many disciplines. Ecologists seek to understand organismal biodiversity, distribution and abundance (species richness, temporal and spatial variation) with respect to biotic and abiotic (environmental) influences; particularly when human activities can have profound effects, whether accidental (like oil spills), deliberate (urban development) or unintended (acid rain, holes in the ozone layer, global warming). Analyses of abiotic (physicochemical) parameters invariably include the three major cycles on Earth, involving the atmosphere, oceans and soil. The atmospheric cycle involves the movement of gases and water vapour in weather and climatic cycles. The hydrologic cycle describes the movement of water between the oceans, land masses and atmosphere. The rock cycle concerns the processes by which rocks and soils are formed, altered and destroyed over millennia.



Ecosystem ecology emphasizes energy flow and chemical cycling among the various biotic and abiotic components. Community ecology deals with the interactions between the whole array of species in a community, including competition, predation, herbivory, symbiosis, and disease. Population ecology concentrates mainly on factors that affect how many individuals of a particular species live in an area.

Population ecology

Population ecology is the study of populations in relation to the environment, including environmental influences on population density, distribution, age structure, and size. However, measuring population size can be difficult. In many instances, it is impossible to count the total number of individuals, so population numbers are estimated. Two methods of doing so are:

- quadrat counts: count the number of individuals in a series of randomly located plots, calculate the average density, and extrapolate to estimate the population size.
- mark-recapture method: capture a number of individuals, mark and release them. Some time later, capture another sample of individuals (which will include some marked ones). The proportion of marked recaptured individuals in the second sample is considered equal to the proportion originally sampled from the population.

Question: A marine biologist, wanting to calculate the number of fish (N) that lived on a reef, captured a sample of individuals (numbering S1), tagged them and released them. One month later, he collected another sample (numbering S2) and found several marked individuals amongst them (numbering S3). Which formula can be used to calculate N?

A. N = (S2 x S3) / S1
B. N = (S1 x S2 x S3)
C. N = (S1 x S3) / S2
D. N = (S1 x S2) / S3
E. N = (S2 + S1) / S3

Answer: D [the following proportions are considered equal, S1/N = S3/S2]

Demography is the study of the vital statistics of populations and how they change over time. Populations wax and wane in size as individuals join (births and immigration) and leave (deaths and emigration). The population growth rate ordinarily refers to the change in population over a specific time period expressed as a percentage of the number of individuals in the population at the beginning of that period; given by the formula:

Growth rate = [(births + immigration) - (deaths + emigration)] / population

Question. The world factbook [www.cia.gov/library/publications/the-world-factbook/] gives the following figures for Australia (as of July 2007): current population = 20,434,176; net migration rate = 3.78 per 1000; birth rate = 12.02 per 1,000; death rate = 7.56 per 1,000.
(i) What is the current annual percentage growth rate of the Australian population?
(ii) Assuming the current growth rate remains constant, write an expression for Australia's population *t* years from July 2007.
(iii) In what year does the model predict that the population will exceed 30 million for the first time?

Answers:
(i) The formula for growth rate = [(births + immigration) - (deaths + emigration)] / population
This can be restructured according to the data given to yield:
annual % growth rate = annual % birth rate + net annual % migration rate - annual % death rate
= 1.202 + 0.378 - 0.756 = 0.824 %

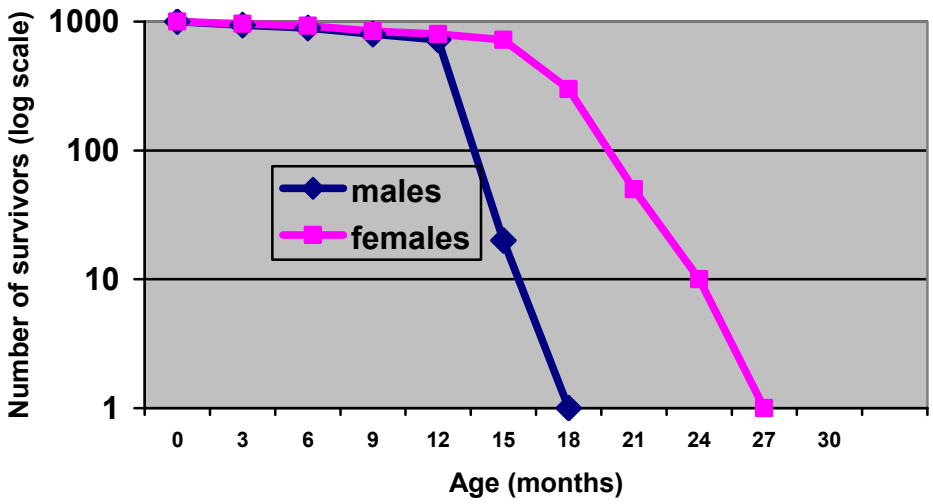
(ii) The equation is $P(t) = 20434176e^{0.00824t}$

(iii) We have $30000000 = 20434176e^{0.00824t}$ so $1.468 = e^{0.00824t}$ so $\ln(1.468) = 0.00824t$
so t = 46.6 years. Thus the population will first exceed 30000000 during the year 2053.

The survival pattern of a population can be summarized in a life table, best constructed by following the fate of a specific cohort (a group of individuals of the same age) from birth until death.

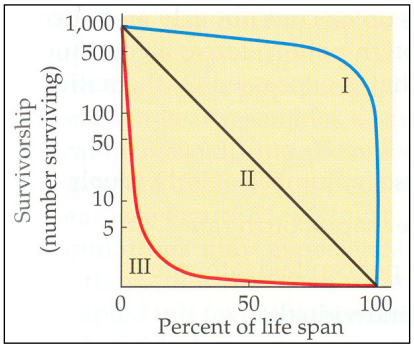
Life table for Australian marsupial mouse *Antechinus stuartii*									
Season	Age (months)	Males				Females			
		No. alive at start of season	Proportion alive	No. deaths during season	Death rate	No. alive at start of season	Proportion alive	No. deaths during season	Death rate
winter	0-3	100	1.00	6	0.060	100	1.00	4	0.040
spring	4-6	94	0.94	5	0.053	96	0.96	3	0.031
summer	7-9	89	0.89	9	0.101	93	0.93	9	0.097
autumn	10-12	80	0.80	8	0.100	84	0.84	4	0.048
winter	13-15	72	0.72	70	0.972	80	0.80	8	0.100
spring	16-18	2	0.02	2	1.000	72	0.72	42	0.583
summer	19-21	0	0			30	0.30	25	0.833
autumn	22-24	0	0			5	0.05	4	0.90
winter	25-27	0	0			1	0.01	1	1.00

A graphic way of presenting the data in a life table is by constructing a **survivorship curve**, that is, plotting the numbers of individuals in a cohort of 1,000 individuals still alive at each age.



Several patterns of survivorship are exhibited by natural populations:

- Type I curves (flat to begin, reflecting low death rate in early and middle life, then dropping steeply as death rates increase among aged groups)
- Type II curves (intermediate, almost linear on log scale, reflecting constant mortality over life span)
- Type III curves (drops quickly at start, reflecting high death rates early in life, then flattening out as death rates decline for survivors, usually organisms producing large numbers of offspring but providing little parental care)



Purves et al. 1997; Fig. 51.7

Stage-structured diagrams can also be used to depict the life-cycles of organisms, with the arrows connecting stages showing the proportion of the population transitioning between stages within any one time interval. Simultaneous equations can be developed to model population structures, which allow matrix operations to be used to solve equations for any given time interval.

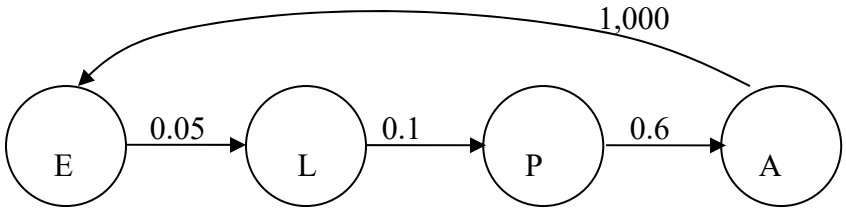
Question. The following life table applies to fleas.

Time period	Developmental stage	Number alive
1	Eggs	10,000
2	Larvae	500
3	Pupae	50
4	Adults*#	30

*The male:female ratio of adult fleas is 1:1
#Adult female fleas lay a total of 2,000 eggs

(a) Draw a stage-structured model of the life-cycle showing transitions between stages.
(b) Write a series of simultaneous equations to calculate the numbers of individuals within each stage from time t to time $t+1$.
(c) What is the transition matrix for this stage-classified population model?

Answers:
(a) Stage-structured model



[the sequential development of fleas is evident from the diagram, and the proportions transitioning between stages are shown numerically above the arrows]

(b) Simultaneous equations calculate numbers of individuals within each stage from time t to time $t+1$.

$$
\begin{aligned}
E_{t+1} &= 0\,E_t + 0\,L_t + 0\,P_t + 1{,}000\,A_t &&= 1{,}000\,A_t \\
L_{t+1} &= 0.05\,E_t + 0\,L_t + 0\,P_t + 0\,A_t &&= 0.05\,E_t \\
P_{t+1} &= 0\,E_t + 0.1\,L_t + 0\,P_t + 0\,A_t &&= 0.1\,L_t \\
A_{t+1} &= 0\,E_t + 0\,L_t + 0.6\,P_t + 0\,A_t &&= 0.6\,P_t
\end{aligned}
$$

[it is conceptually easier to include all stages in sequence (scoring zero for those not involved in transitions, and numeric values for those that do transition between stages)]

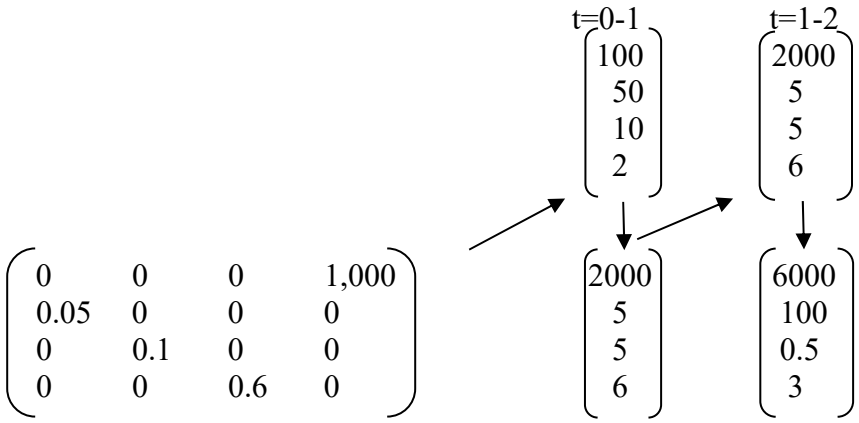
(c) Transition matrix (revealed by deleting letters but retaining numbers in simultaneous equations)

$$
\begin{pmatrix}
0 & 0 & 0 & 1{,}000 \\
0.05 & 0 & 0 & 0 \\
0 & 0.1 & 0 & 0 \\
0 & 0 & 0.6 & 0
\end{pmatrix}
$$

Question: You bought a new bed for your dog from friends but it contained 100 eggs, 50 larvae, 10 pupae and 2 adult fleas. Use the transition matrix to determine how many adult fleas will you have after two time periods have passed (begin with t = 0)

.

Answer: (calculated by two repetitions of matrix multiplication)



There will be 3 adult fleas after 2 time periods.

Question: Another ectoparasite of dogs only has two life-cycle stages (juveniles and adults) which cycle monthly. The transition matrix for this parasite is: $\begin{bmatrix} 0.4 & 2 \\ 0.5 & 0.2 \end{bmatrix}$

In July, there were 44 juveniles and 9 adults on your dog. How many were there in June (i.e. in the <u>previous</u> month)?

Answer: (calculated by determining the inverse of the matrix, and then solving the matrix equation)

$A = \begin{bmatrix} 0.4 & 2 \\ 0.5 & 0.2 \end{bmatrix}$, Population in June = $P_i = \begin{bmatrix} x \\ y \end{bmatrix}$, and Population in July = $P_{i+1} = \begin{bmatrix} 44 \\ 9 \end{bmatrix}$

When $A P_i = P_{i+1}$, then $P_i = A^{-1} P_{i+1}$

Calculate $A^{-1} = \frac{1}{ad - bc} \times \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{-0.92} \times \begin{bmatrix} 0.2 & -2 \\ -0.5 & 0.4 \end{bmatrix}$

Matrix multiply to get $P_i = \frac{1}{-0.92} \times \begin{bmatrix} 0.2 & -2 \\ -0.5 & 0.4 \end{bmatrix} \text{ x } \begin{bmatrix} 44 \\ 9 \end{bmatrix} = \frac{1}{-0.92} \times \begin{bmatrix} -9.2 \\ -18.4 \end{bmatrix} = \begin{bmatrix} 10 \\ 20 \end{bmatrix}$

So there were 10 juveniles and 20 adults in June!

Various models have been developed to analyse population growth, but all require making assumptions about the population being examined, either under ideal or constrained conditions.

Exponential model of population growth

This model describes population growth in an idealized, unlimited environment. While this may not frequently occur naturally, it can be used to examine the capacity for a species to increase in population size and the conditions for that to occur. We have seen that population growth rate is based on influx (births and immigration) and efflux (deaths and emigration). The population increases in size when birth rates exceed death rates, and when immigrations exceed emigrations. If we ignore migrations for simplicity sake, we can express the change in population size mathematically as:

$$\Delta N/\Delta t = B - D$$

where ΔN = change in population size; Δt = time interval; B = births; and D = deaths.

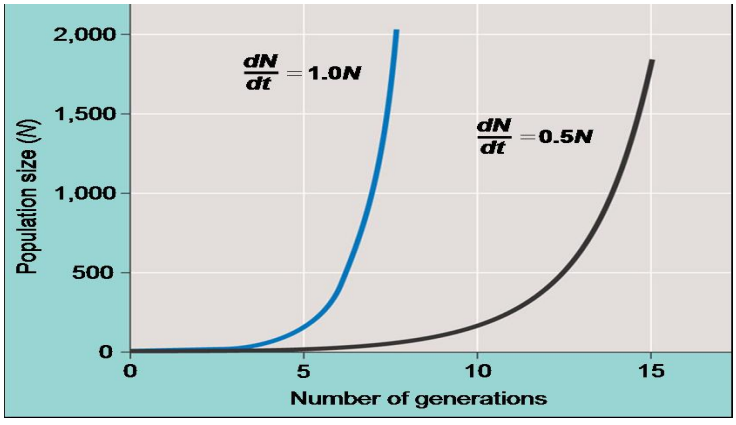
Births and deaths can also be expressed as "per capita rates", by calculating the average number occurring per individual during the specified time period. For example, if there are 34 births per year in a population of 1,000 individuals, the annual per capita birth rate, b = 34/1000 (= 0.034). We can use this rate to calculate the expected number of births per year in a population of any size by using the formula $B = bN$. Similarly, if there are 24 deaths in the population over the same period, the annual per capita death (mortality) rate, m = 24/1000 (= 0.024), and $D = mN$. We can now simplify our equation for population change over time as:

$$\Delta N/\Delta t = bN - mN$$

Taking this one step further, we can amalgamate the birth and death rates into the "per capita rate of increase (r)" [= (b - m)]. Obviously, the population is growing when $r > 0$, declining when $r < 0$, and stable when $r = 0$ (called ZPG = zero population growth). The equation for change in population size is therefore further simplified to :

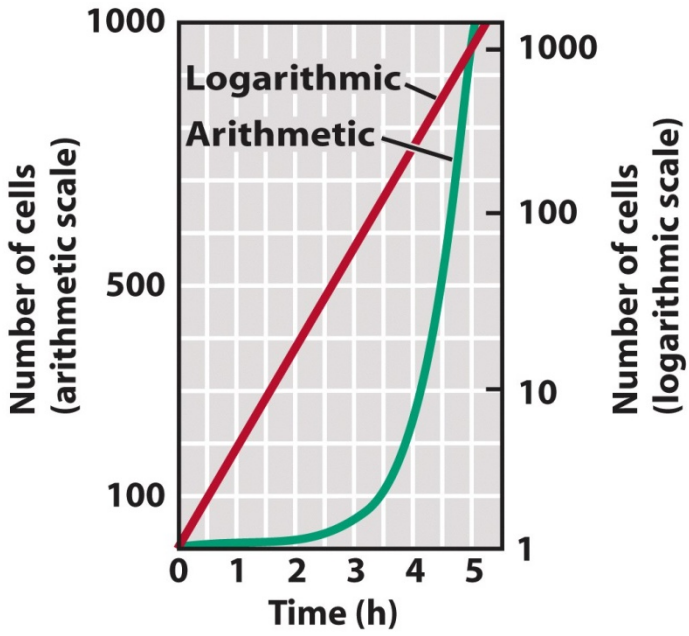
$$\Delta N/\Delta t = rN \qquad \text{[expressed as } dN/dt = rN \text{ using differential calculus]}$$

Under ideal conditions, population growth can become exponential. Births exceed deaths, so the population grows and accumulates more new individuals per unit of time when it is large. When population size is plotted over time, exponential growth is seen as a characteristic J-curve (increase becoming steeper with time). J-shaped curves are characteristic of populations that are introduced into new or unfilled environments (especially in the microscopic world of microbes) or whose numbers have been drastically reduced by a catastrophic event and are rebounding.

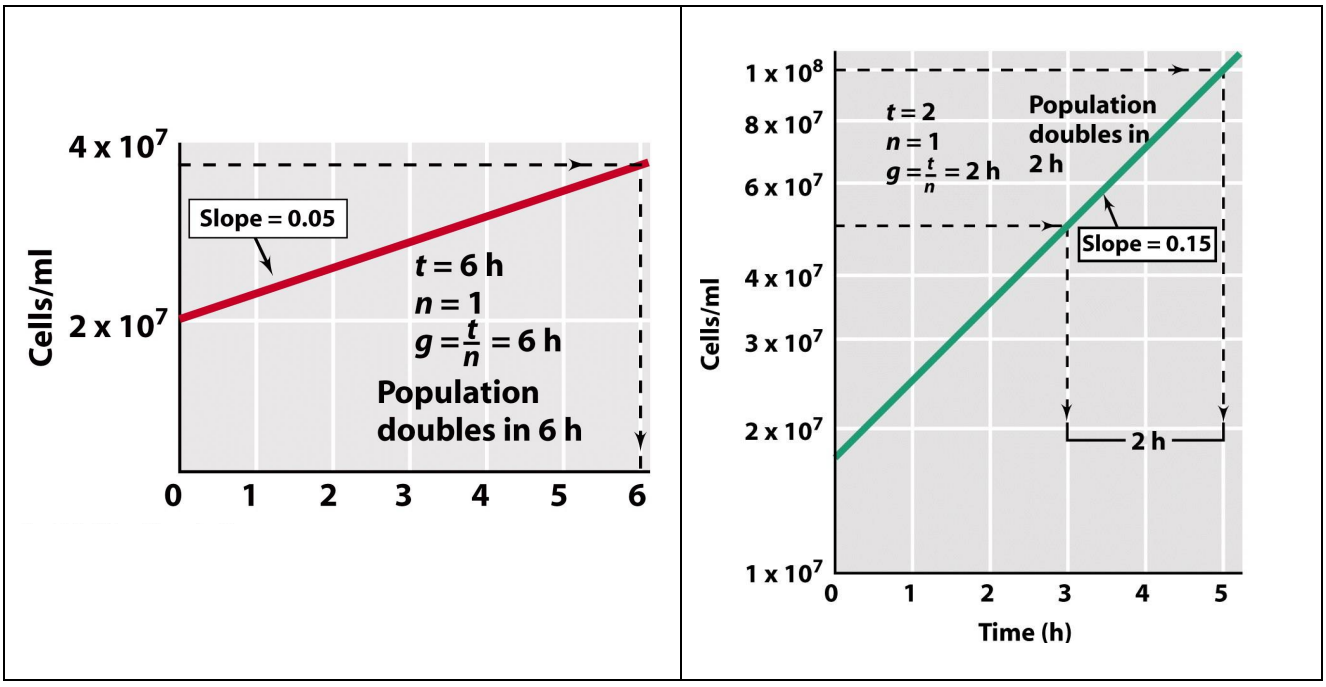


By way of an example, let us consider bacterial growth. Bacterial cells divide by doubling themselves (through mitosis, budding, fission), that is, one parent cell split to form two daughter cells. The time required for this to occur is called the generation time (or doubling time). Depending on the organism and the available conditions, bacterial generation times range from minutes to months. Consider the population size arising from a single bacterium with a doubling time of 30 minutes.

Time (h)	Total number of cells
0	1
0.5	2
1	4
1.5	8
2	16
2.5	32
3	64
3.5	128
4	256 (2^8)
4.5	512 (2^9)
5	1,024 (2^{10})
5.5	2,048 (2^{11})
6	4,096 (2^{12})
.	.
.	.
10	1,048,576 (2^{19})



When you plot this exponential growth against time using arithmetic scales, you derive the typical J-curve with a continuously increasing slope. However, when cell number is plotted on a logarithmic ($\log_{10}$) scale against time, the plots fall along a straight line. These semilogarithmic graphs are convenient to use for estimating generation times from a set of growth data, based on the relationship where generation time (g) equals time (t) over number of generations (n): that is: $g = t/n$.



The exponential increase in bacterial cell number is a geometric progression of the number 2. The cell division process can be expressed as: 2^0 (= 1) → 2^1 (= 2) → 2^2 (= 4) → 2^3 (= 8) etc. The mathematical relationship between the number of cells present initially (N_0) and the number present after a period of exponential growth (N) is given by:

$N = N_0 2^n$ where n = number of generations

Using logs yields $\log N = \log N_0 + n \log 2$

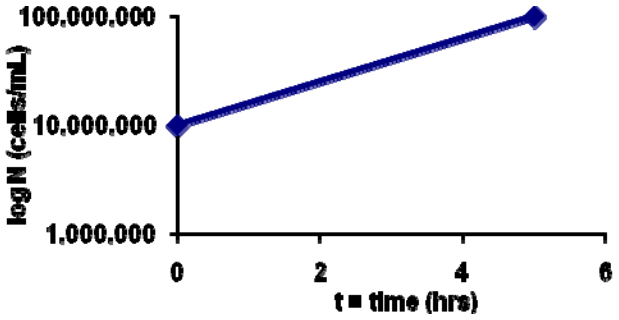
so $n = [\log N - \log N_0] / \log 2$

$= [\log N - \log N_0] / 0.301$

$= 3.3 [\log N - \log N_0]$

The generation time g can also be calculated from the slope of the line obtained in a semilogarithmic plot of exponential growth. The slope is equal to $\log 2n/t$ (= $0.301n/t$). This term is called the specific growth rate (k). Another index of growth is the reciprocal of the generation time, called the division rate (v). The division rate = $1/g$ and has units of reciprocal hours (h^{-1}). While g is a measure of the time taken for a population to double in number, v is a measure of the number of generations that occur per unit time. The slope of the line relating cell number to time is equal to $v/3.3$. Knowing n and t, you can calculate g, k and v for microbes growing under different culture conditions (and test treatments).

Question: A microbiologist took a swab sample from the hand of a mathematician and grew some extraordinary bacteria on a culture plate. He plotted cell concentration on a logarithmic scale against time, and generated the following graph.



$N_5 = 10^8$ when $t = 5$

$N_0 = 10^7$ when $t = 0$,

(i) The formula for calculating cell concentration (N) as a function of the number of generations (n) is $N = N_0 2^n$. Find the number of generations between t=0 and t=5.
(ii) Calculate the generation time, g.
(iii) Calculate the slope (= specific growth rate, k).

Answers:

(i) $N = N_0 2^n$ transforms to $\log N = \log N_0 + n \log 2$

thereby giving number of generations $n = 3.3(\log N - \log N_0)$

$= 3.3 (\log 10^8 - \log 10^7)$

$= 3.3 (8-7)$

$= 3.3$

(ii) generation time $g = t/n = 5/3.3 = 1.52$ hours

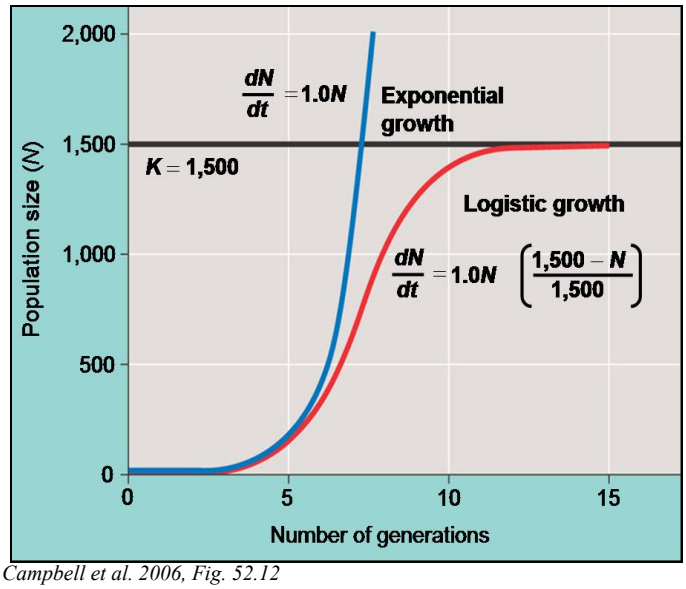
(iii) specific growth rate $k = \log 2 \times n / t = 0.301n/t = (0.301 \times 3.3) /5 = 0.20$

Logistic model of population growth

This model incorporates the concept of carrying capacity, because resources are usually limited. As population density increases, each individual has access to a smaller share of the total resources available. Ultimately, there is an upper limit to population size; the 'carrying capacity' defined as the maximum stable population size that a particular environment can support. Carrying capacity is not fixed but varies over time and space. Energy limitation often determines carrying capacity, although other factors, such as shelters, refuges from predators, soil nutrients, water and suitable breeding sites, can be limiting. If individuals cannot obtain sufficient resources to reproduce, the per capita birth rate b will decline. If they cannot find and consume enough energy to maintain themselves, the per capita death rate m may increase. A decrease in b or an increase in m results in a lower per capita rate of increase r. In the logistic population growth model, r declines as carrying capacity is reached.

Mathematically, we start with the equation for exponential growth, adding an expression that reduces the rate of increase as N increases. If the carrying capacity is K, then $(K - N)$ is the number of additional individuals the environment can accommodate, and $(K - N)/K$ is the fraction of K that is still available for population growth. Multiplying r by this fraction gives the growth rate of the population as: $dN/dt = rN[(K - N)/K]$

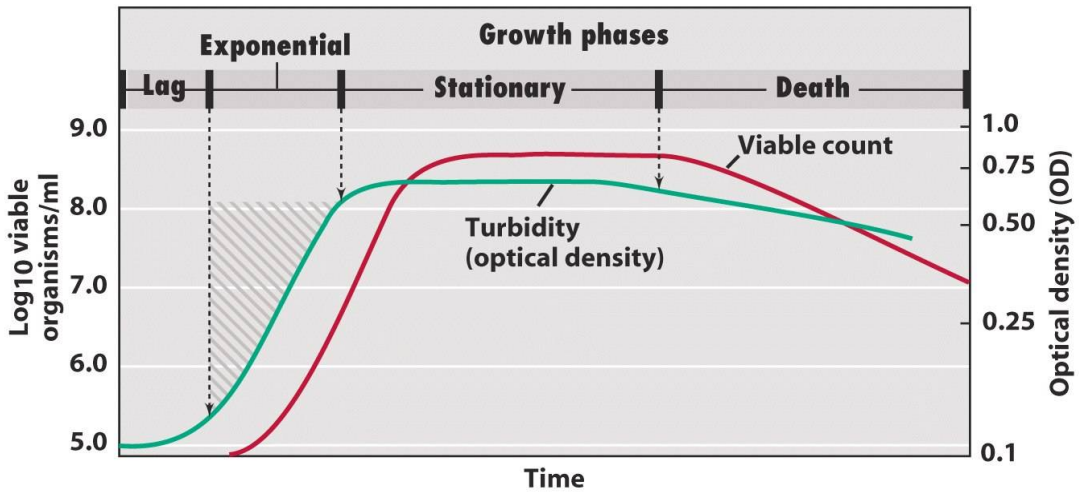
When N is small compared to K, the term $[(K - N)/K]$ is large and the rate of increase is steep. But when N is large and approaches K as resources become limited, the term $[(K - N)/K]$ is small and so is the rate of population growth. When N is plotted over time, the logistic model of population growth produces an S-shaped (sigmoid) curve. Population growth rate slows dramatically as N approaches K.



Campbell et al. 2006, Fig. 52.12

Some of the assumptions built into the logistic model do not apply to all populations. Most populations do not adjust instantaneously and approach their carrying capacity smoothly. There is often a time-lag before the negative effects of increasing population are realized, while other populations overshoot their carrying capacity before settling down. Some populations fluctuate greatly, making it difficult to define the carrying capacity. Not all individuals will have the same effect on population growth. Some populations show an Allee effect, in which individuals may have a more difficult time surviving or reproducing if the population is too small (such as animals being unable to find mates).

Remember our bacterial population. When resources are finite, population growth occurs in four distinct phases. An initial lag phase is followed by a characteristic exponential (log) growth phase. However, as essential nutrients are depleted and/or toxic products build up, the population enters a stationary phase which persists until diminishing resources can no longer support the whole population and cells begin to die (the death phase).



The logistic population growth model is useful in conservation biology for estimating how rapidly a particular population might increase in numbers after it has been reduced to a small size, or for estimating sustainable harvest rates for fish or wildlife populations. Note that logistic growth equations have been developed to model populations exhibiting continuous growth (differential equation incorporating any value for time) or discrete growth (arithmetic equation incorporating values for time which are integers only – such as months, years, or even generations).

Question: Pastoral Pete has just bought a new property in the middle of Queensland. The farm is 5,000 hectares in size and he has been told he can run cattle at a stocking density of 0.2 cattle/hectare. He buys 200 head of cattle, and the stock agent optimistically guarantees him an annual birth rate of 80%. His neighbours tell him to expect a 5% mortality rate due to disease and predators. Pete does not have faith in his mathematical prowess so he hires you to calculate how many years it will take for the cattle population to exceed 900? (consider his date of purchase as year 0)?

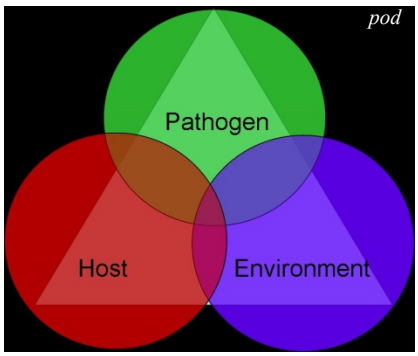
Answer:
You must first calculate the carrying capacity (K = 0.2 cattle/hectare x 5,000 hectares = 1,000 cattle) and the growth rate (r = births - deaths = 0.80 - 0.05 = 0.75)
For continuous growth, the logistic growth equation is a differential equation. However, for discrete growth (when time intervals are integers only), the logistic growth equation is:

$$L_{i+1} = L_i + r\,L_i\,(1 - L_i/K)$$

so $$L_{i+1} = L_i + 0.75\,L_i\,(1 - L_i/1{,}000)$$

in year 0, the population size L_i = 200
in year 1, the population size = 320
in year 2, the population size = 483
in year 3, the population size = 670
in year 4, the population size = 836
in year 5, the population size = 939 (exceeds 900)

Chapter 15.

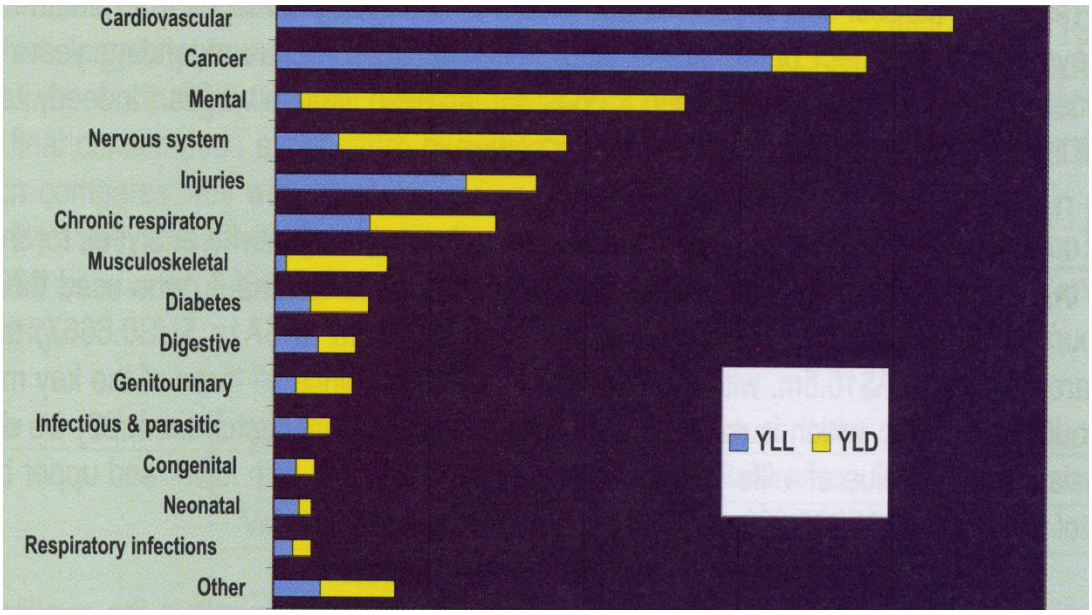


EPIDEMIOLOGY

Vocabulary list: epidemiology, disease, communicable, morbidity, mortality, prevalence, incidence, intensity, case control studies. odds ratio, cohort studies, relative risk, outbreak, mathematical model, basic reproduction number, epidemic, differential equations, SIR model, susceptible, infected, resistant, contact number

Epidemiology literally translates from Greek terms to mean "the study of what is upon the people", suggesting that it applies only to human populations. However, the term is widely used in zoology (sometimes supplanted by the term epizootiology), botany and microbiology. Epidemiology is the study of factors affecting the health and illness of populations. It provides a quantitative foundation for evidence-based medicine for identifying risk factors for disease and determining optimal treatment approaches. Studies may be conducted on communicable (infectious) diseases (caused by viruses, bacteria, fungi, protozoa, helminths, and arthropods) and non-communicable diseases (usually classified as inflammatory, cardiovascular, neoplastic, genetic, developmental, endocrine, nutritional, autoimmune, traumatic, senescence, iatrogenic, or idiopathic in origin). Diseases may be evident in a population as sporadic (occasional), endemic (established/persistent), epidemic (outbreak) or pandemic (global in distribution).

To study the occurrence, spread and control of diseases, epidemiologists employ a range of study designs generally categorized as descriptive (observational), analytic (testing for significant relationships) and experimental (testing treatments and other interventions). Generally speaking, epidemiologists study relationships between cause (exposure) and effect (morbidity and mortality). Morbidity is defined as the incidence of illness, and mortality as the incidence of death. Public health agencies often combine these measurements into a single equation: DALY = YLD + YLL, where DALY = Disability Adjusted Life Years (combined burden); YLD = Years Life lost due to Disability (morbidity burden); and YLL = Years Life Lost due to premature death (mortality burden).



NHMRC, 2002. Impact of diseases.

Assessments of the impact of disease are based on measurements of its distribution and abundance within a population (incorporating spatial and temporal variation); including:

- prevalence (cross-sectional study of population at selected point or period)
  [if 70 people were infected from a population of 100 examined in July, prevalence = 70%]
- incidence (longitudinal study of population over specified time)
  [if prevalence is 70% in July and 80% in October, incidence increased by 10% over 3 months]
  [Note, can have negative incidence, when prevalence decreases over time]
- intensity/severity (different categories, often qualitative, sometimes quantitative)
  [based on range of symptoms/signs, often supported by clinical parameters, or, for infectious diseases, estimations of the numbers of organisms per host]

These parameters may exhibit longitudinal fluctuations (periodic variation over days, weeks, months or years) due to a wide variety of factors (demographic, socioeconomic, behavioural, geographic, and climatic). For example, climatic changes associated with global warming are predicted to cause many diseases to increase in incidence, distribution and severity (due to host translocations, greater susceptibility/diminished resistance, expanded ranges and enhanced survival of pathogens and vectors, ineffective/inappropriate treatment/control). Establishing cause and effect can be problematic, so several guidelines (non-prescriptive criteria) have been proposed for assessing evidence of causation (the Bradford-Hill criteria): involving:

- Strength (a small association does not mean that there is not a causal effect),
- Consistency (replication by different persons in different places with different samples),
- Specificity (the more specific the association, the greater the probability),
- Temporality (the effect must occur after the cause, and include any expected delays),
- Biological gradient (higher exposure should generally lead to higher incidence),
- Plausibility (there should be a plausible mechanism to explain the relationship), and
- Coherence (correlation between field and laboratory findings).

There are 4 main types of epidemiological studies: case series, case controls, cohorts and outbreaks.

1. Case series describe the experience of a single patient, or a group of patients with a similar diagnosis. They are purely descriptive and often rely on an astute clinician identifying an unusual feature of a patient's disease and formulating a hypothesis. Analytical studies are then done to investigate possible causal factors.

2. Case control studies involve matching cases with disease to comparable controls without disease (from the same local population), and then looking back through time at potential exposures both populations may have encountered. A 2 x 2 matrix is constructed, displaying exposed cases (A), exposed controls (B), unexposed cases (C) and unexposed controls (D).

	CASES	CONTROLS
EXPOSED	A	B
UNEXPOSED	C	D

The parameter used to measure association is the odds ratio (OR), which is the ratio of the odds of exposure in the cases (A/C) to the odds of exposure in the controls (B/D), that is,

$$OR = (AD) / (BC)$$

If OR >> 1, then the conclusion is "those with the disease are more likely to have been exposed". If OR ~ 1, then exposure and disease were probably not associated, and if OR << 1, this suggests that exposure may have been protective in the causation of disease. Case control studies are usually faster and more cost effective than cohort studies, but are sensitive to bias (such as recall and selection bias).

Question: Case control studies were used to show an association between smoking cigarettes and lung cancer. The following table presents information about potential risk factors in 1,000 people with lung cancer and 1,000 matched controls:

Potential risk factor		Patients with lung cancer	Age-matched negative controls (without lung cancer)
History of smoking	Smoker	898	245
	Non-smoker	102	755
Proximity to industrial pollution (smog)	Urban dwelling	447	399
	Rural dwelling	553	601
Other respiratory condition	Asthmatic	42	104
	Non-asthmatic	958	896

Calculate the odds ratio for each factor, and interpret the results.

Answer:

Potential risk factor	(AD) / (BC)	OR	Significance
History of smoking	(898x755) / (245x102)	27.1	lung cancer associated with smoking
Proximity to industrial pollution (smog)	(447x601) / (399x553)	1.2	lung cancer not associated with smog
Other respiratory condition	(42x896) / (104x958)	0.38	asthma may be protective

3. Cohort studies select subjects based on their exposure status, and then follows them forward through time to assess their outcome status. The results are tabulated in a 2x2 matrix:

	CASES	CONTROLS
EXPOSED	A	B
UNEXPOSED	C	D

The parameter generated to measure association is the relative risk (RR) which is the ratio of the incidence of disease in the exposed group (A/(A+B)) to that in the unexposed group (C/(C+D)), that is,

$$RR = (A/(A+B)) / (C/(C+D))$$

If RR >> 1, then the conclusion is "those with the exposure were more likely to develop disease." If RR ~ 1, then there was no association, and if RR << 1, then exposure may have been protective.

Question: Cohort studies are currently being used to examine whether brain tumour formation may be associated with mobile phone use, infections of the brain by the sporozoan parasite *Toxoplasma* which is transmitted by cats, or by wearing hats. The following table presents information about tumour formation in 1,000 people:

Potential risk factor		Patients developing brain tumours	People not developing brain tumours
Mobile phone use	phone use	49	699
	no phone use	6	246
Dwelling with cats	cats as pets	8	142
	no cats	47	803
Habitual hat use	wears hats	1	51
	no hats	54	894

Calculate the relative risk for each factor, and interpret the results.

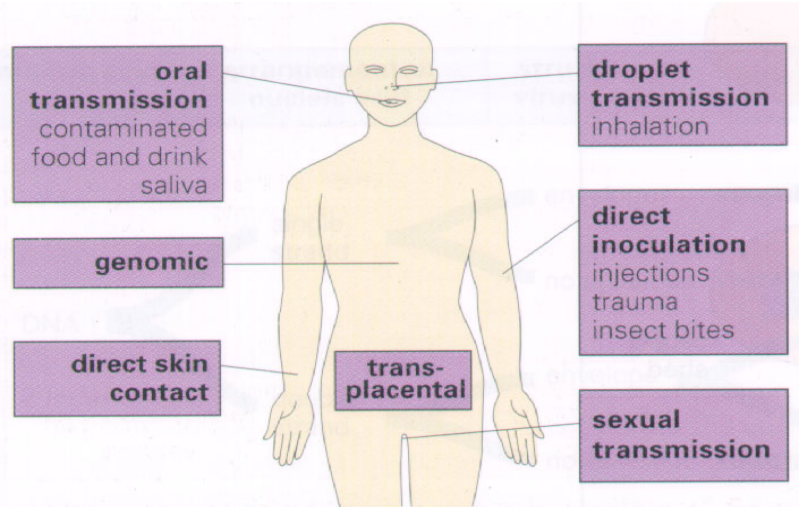
Answer:

Potential risk factor	$\frac{A}{A+B}$ / $\frac{C}{C+D}$	RR	Significance
Mobile phone use	$\frac{49}{49+699}$ / $\frac{6}{6+246}$	2.75	tumours may be associated with mobile phone use
Dwelling with cats	$\frac{8}{8+142}$ / $\frac{47}{47+803}$	0.96	tumours not associated with living with cats
Habitual hat use	$\frac{1}{1+51}$ / $\frac{54}{54+894}$	0.34	wearing hats may be protective

Prospective cohort studies have many benefits over retrospective case control studies. The RR is a more powerful effect measure than the OR, as the OR is just an estimation of the RR, since true incidence cannot be calculated in a case control study where subjects are selected based on disease status. Temporality can be established in a prospective study, and confounders are more easily controlled for. However, they are more costly, and there is a greater chance of losing subjects to follow-up based on the long time period over which the cohort is followed.

4. Outbreak investigations. A disease outbreak is a classification used in epidemiology to describe the sudden appearance, or sudden increase in incidence, of a disease in a specified group or population. Local outbreaks may be confined to a small area or village, while epidemics are outbreaks affecting regions or countries, and pandemics are global disease outbreaks.

Outbreaks may originate from single point sources or from multiple common sources. Exposure to infections may be irregular, periodic or continuous and may involve contact with environmental sources (water, food or air-borne), zoonotic (animal-to-human) and anthropogenic (human-to-human) sources. Several modes of transmission are commonly recognized for infectious diseases: venereal, faecal-oral, aerosol, predator-prey, and vector-borne transmission.



When investigating outbreaks, epidemiologists have developed a number of widely accepted steps; these include:

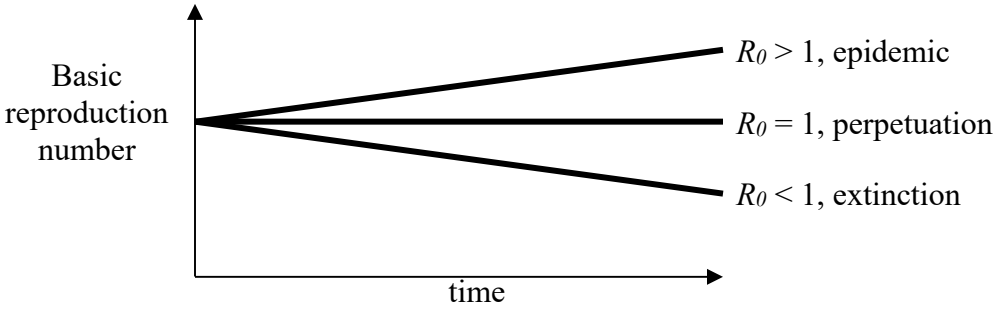
- verify the diagnosis,
- confirm the existence of the outbreak above background levels,
- define inclusion criteria (what constitutes a case),
- conduct descriptive assessment (time, place, people),
- develop a hypothesis (causes?),
- test hypothesis (collect and analyse data),
- refine hypothesis as appropriate and carry out further study,
- develop and implement control and prevention systems, and
- release findings to greater community.

Mathematical epidemiological models

Biological systems are inherently complex, with many interactions occurring between their component parts, some conspicuous but most cryptic. Nonetheless, many attempts have been made to develop mathematical models to reliably monitor and predict disease distribution and abundance. Simulation models are vital to governments and public health agencies to prepare and respond to disease outbreaks through resource allocation, infrastructure development, preparedness training and public education. While such models have frequently been used in sciences such as physics and engineering, they have only relatively recently been applied to infectious diseases. This has necessitated the development of a specific vocabulary to define parameters as well as adopting a de-constructist approach to identify integral components and mathematical relationships.

Infectious disease models distinguish between 'microparasites' and 'macroparasites' *sensu latu*. Microparasites are infectious disease agents, such as viruses, bacteria and some protozoans, which reproduce directly, often at very high rates, within the host. They are generally small, have short generation times, and usually produce long-lasting immunity against re-infection. The duration of infection is usually short relative to the life span of the host, so infections are typically transient. Macroparasites, such as helminths and arthropods, have no direct reproduction within the definitive host, producing transmission stages that pass from the host to complete their life-cycle. They are typically large and have longer generation times, which can often be a significant fraction of the host's life expectancy. Immunity tends to be of relatively short duration once the parasites are removed, and infections are often persistent, with hosts being continually re-infected.

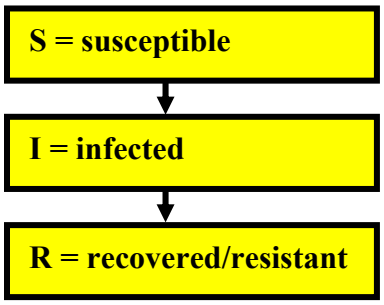
A concept from ecology that is central to infectious disease models is the basic reproduction number, R_0, which is defined as the average number of secondary infections resulting from a single primary infection within a completely susceptible population. Pathogens must reproduce themselves to survive.



If $R_0 = 1$, the pathogen is stable within the population and infections will perpetuate.
If $R_0 < 1$, the pathogen will slowly die out as there will not be enough new infections to sustain it.
If $R_0 > 1$, then the number of infections will increase with time and there may be an epidemic.

However, populations are never completely susceptible, as some individuals are already infected, and others have been infected and are now immune (resistant to secondary infection). The effective reproductive rate, $R = R_0.s$, where s = the fraction of the population that is susceptible. At equilibrium (i.e. when $R = 1$), $R_0 = 1/s$. It is possible to estimate s in a population by conducting serological surveys (for the presence of host antibodies which are taken as indicative of prior exposure and resistance). If s is low (e.g. 20% seronegative), R_0 must be high (1/0.2) to maintain equilibrium, whereas if s is high (e.g. 80% seronegative), R_0 must be low (1/0.8) to maintain equilibrium. The objective of mass vaccination and other control programs is to reduce s (fraction susceptible) in order to reduce R_0 (and thereby reduce the spread of infections).

Disease models need to account for variability in host susceptibility, particularly over time. The SIR model has been developed for disease transmission within populations that exhibit three states: susceptible (S), infected (I) and resistant (R). The model uses differential equations (i.e. algebraic equations that include a derivative or slope) to describe the movement of individuals between states with respect to time.



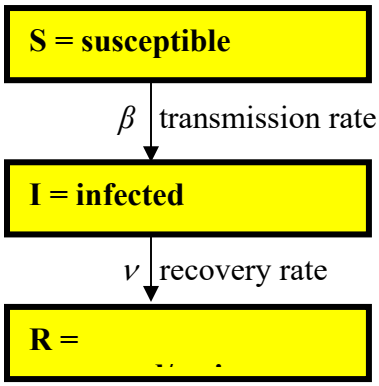
dS/dt = change in number of susceptibles (S) over time

dI/dt = change in number of infected (I) over time

dR/dt = change in number of resistants (R) over time

The movement of individuals between states is predictable, those who leave one state progress to the next. Entries are added, departures are subtracted, and the total population number N = S+I+R remains constant. As the first step in the differential equation modelling process, we identify the independent variable as time (t) and the dependent variables as the fractions of the population in each state. Let s = the susceptible fraction of the population (= S/N), i = the infected fraction of the population (= I/N), and r = the resistant fraction of the population (= R/N). [Although it may seem more natural to work with numbers, some of the calculations are much simpler when we use fractions instead (remember that numbers and fractions are proportional to each other, so they will give the same information about the spread of disease)].

We must now make some assumptions about the rates of change of our dependent variables. In the simplest possible model, we are ignoring births so no one is added to the susceptible group. The only way an individual leaves the susceptible group is by becoming infected. This will depend on the amount of contact between susceptibles and infecteds. Suppose that each infected individual has a fixed number (β) of contacts per unit time that are sufficient to spread the disease. For example, if each infected person makes one potential-infecting contact every two days, then $\beta = ½$. But not all these contacts are with susceptible individuals. Assuming a homogeneous mixing of the population and an equal likelihood of encounter, the fraction of contacts with susceptibles is s, thus on average, each infected individual generates $\beta.s$ new infected individuals per unit time. We also assume that a fixed fraction (ν) of the infected group will recover per unit time. For example, if the average duration of infection is three days, then, on average, one-third of the currently infected population will recover each day (i.e. $\nu = 1/3$). We use these parameters in differential equations to obtain derivatives for our dependent variables.



$ds/dt = -\beta.s.i$ = susceptible equation

$di/dt = \beta.s.i - \nu.i$ = infected equation

$dr/dt = \nu.i$ = resistant equation

The spread of disease depends on the parameters β and ν, measures of transmission and recovery. These parameters must be estimated to generate numerical solutions of the differential equations. The fraction ν of infecteds recovering in a given day can be estimated from observation of infected individuals. Specifically, ν is roughly the reciprocal of the patent or infective period (i.e. the number of days an individual is sick enough to infect others). There is no direct way to estimate β, but there is an indirect way. Consider the ratio of β to ν. This can be expressed as:

$\beta/\nu = \beta \times 1/\nu$ = number of close contacts per day per infected x number of days infected
= number of close contacts per infected.

This ratio is called the **contact number**, $c = \beta/\nu$. It is a combined characteristic of the population and of the disease. In similar populations, it measures the relative contagiousness of the disease, because it tells us indirectly how many of the contacts are close enough to actually spread the disease. Calculus can be used to estimate c after the epidemic has run its course, then β can be calculated as $c\nu$. Consider the differential equations for s and i (and substitute β/c for ν in the second equation):

$ds/dt = -\beta.s.i$
$di/dt = \beta.s.i. - \nu.i$
$= \beta.s.i - (\beta/c).i$
$= \beta.i.[s - (1/c)]$

The complicated terms in both equations cancel out leaving something simpler when we divide the second equation by the first . The Chain Rule can be used to reconfigure the equations to:

$[di/dt]/[ds/dt] = \beta.i.[s - (1/c)] / -\beta.s.i$
$di/ds = [s - (1/c)] / - s$
$= -1 + (1/cs)$

The resultant differential equation determines the infected fraction i as a function of the susceptible fraction s. Note three things: the only parameter that appears is c (which we are trying to determine); the equation is independent of time (which means that the relationship between i and s holds true for the entire duration of the epidemic); and the right-hand side is an explicit function of s (which is now the independent variable). Solving the differential equation (finding the anti-derivative) gives: $i = -s + (1/c) \ln s + q$ where q is a constant

We can obtain two known values for i and s by using the two extremes of time of the outbreak. At the beginning of the outbreak (when $t = 0$), $i = 0$ and $s = 1$. At the end of the outbreak (when $t = \infty$), $i = 0$ (again) and s will have settled to its steady state. Because $i = 0$ in both cases, we can combine the equations for $t = 0$ and $t = \infty$:

As $i_0 = i_\infty$, we obtain	$-s_0 + (\ln s_0)/c + q = -s_\infty + (\ln s_\infty)/c + q$
which simplifies to:	$-s_0 + (\ln s_0)/c = -s_\infty + (\ln s_\infty)/c$
given $s_0 = 1$, we obtain	$-1 + (\ln 1)/c = -s_\infty + (\ln s_\infty)/c$
knowing $\ln 1 = 0$, we obtain	$-1 = -s_\infty + (\ln s_\infty)/c$
which rearranges to	$c = \ln s_\infty / (s_\infty - 1)$

The steady state achieved at s_∞ is directly observable after the epidemic as the fraction of the population that did not get the disease. By determining this value, we can calculate c, then β ($= c.\nu$).

In simple models, the contact number provides a rough estimate of the basic reproduction number, that is, c (or β/ν) ~ R_0, as both parameters assess pathogen transmission potential and the spread of infections within a population, but after making different assumptions (R_0 assumes a completely susceptible population, while c accounts for a fraction becoming resistant). Nonetheless, it is obvious that R_0 and c will increase (indicating enhanced spread) when β is high and ν is low, and R_0 and c will decrease (limited spread) when β is low and ν is high.

Example: Let us model a particular disease to show the relationships between the basic reproductive number of the pathogen (R_0), the transmission rate (β), the infectious period (IP), the recovery rate (ν), and the susceptible fraction of the population (s). The values used in this example are actually quite close to those applicable to measles infections in humans.

Let the infectious period for the disease be two weeks (IP = 2 wk)
and assume that one infected person will infect a further five people ($R_0 = 5$).

The transmission rate $\beta = R_0 / IP = 5 / 2 = 2.5$ wk^{-1} [$\Rightarrow$ IP = R_0 / β]
The recovery rate $\nu = 1 / IP = 1 / 2 = 0.5$ wk^{-1} [$\Rightarrow$ IP = $1 / \nu$]
This gives IP = $R_0 / \beta = 1 / \nu$
$\Rightarrow R_0 = \beta / \nu$ [transmission rate over recovery rate] [$= 2.5 / 5 = 5$]

At equlibrium, the effective reproduction rate R will equal 1,
so $R_0 = 1 / s$ [the reciprocal of the susceptible fraction]
$\Rightarrow s = 1 / R_0 = \nu / \beta$ [recovery rate over transmission rate]
[$= 0.5 / 2.5 = 0.2 = 20\%$ population threshold]

The aim is to curb the epidemic by $\uparrow \nu$ [recovery rate]
or $\downarrow \beta$ [transmission rate]
or both to markedly $\uparrow s$ ($= \nu / \beta$)

Parasite populations

Host-parasite interactions differ from those of predator-prey models because parasitic diseases do not necessarily kill their hosts, and hosts may recover and develop immunity to re-infection. Basically, two types of models have been developed for host-parasite interactions:

- Microparasite models (usually involving acute, transient, multiplicative infections)
- Macroparasite models (usually involving chronic, prolonged, cumulative infections)

Working excamples of both models are available as free downloads
from the Populus website at www.cbs.umn.edu/populus/

Infectious Microparasitic Diseases

Microparasites, in the broadest sense, include viral, bacterial and protozoan pathogens. These micro-organisms reproduce quickly, reach very high intensities within individual hosts, cause acute transient infections (short duration compared to host life span), are often limited by host immune responses, and recovered individuals may develop protective immunity against re-infection. Population dynamics of microparasites are driven largely by transmission between hosts. The SIR model classifies the host population as susceptible (S), infected (I) or recovered/resistant (R), without accounting for within-host abundance/intensity.

Populations are quantitated as:

N = total host population density
S = susceptible host population density
I = infected host population density
R = recovered (immune) host population density

Changes in SIR populations depend on host variables, namely:

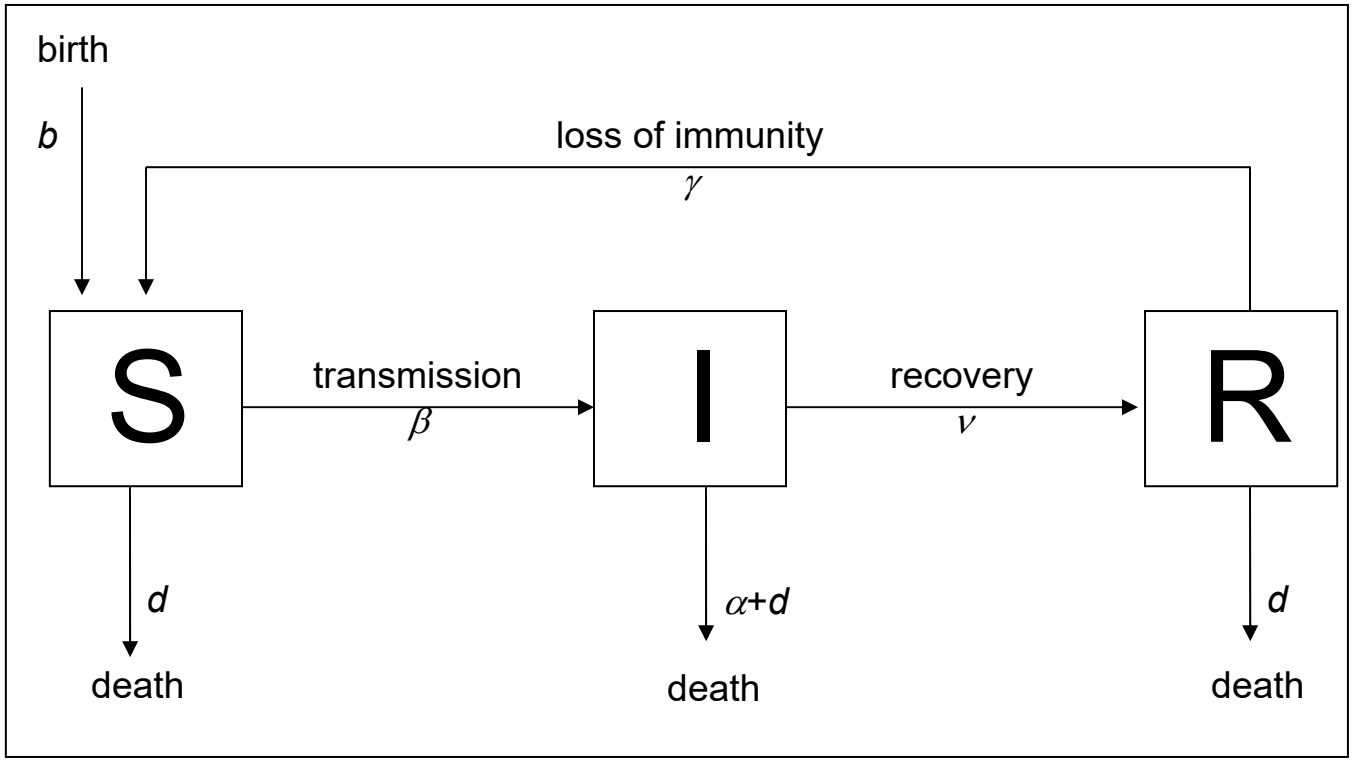
b = host birth rate
d = host death rate (natural mortality)

Changes also depend on parasite variables, including:

α = disease-induced mortality rate
β = between-host transmission rate
ν = recovery rate
γ = rate of immunity loss

Movement between SIR states are modelled by a series of differential equations, whereby numbers change over time, that is, dS/dt, dI/dt, dR/dt. Such changes are said to be instantaneous (not discrete) rates of change. When there is no change in rate (for instance, $dS/dt = 0$), that population is stable; but when the rate is > 0 or < 0, the population increases or decreases, respectively.

The differential equations differ according to whether transmission is density-dependant (DD) or frequency-dependant (FD). DD transmission considers that infections occur in direct proportion to the number of encounters between susceptible and infected individuals, which is simply the product of their densities (SI). FD transmission recognizes that mixing of the population is not homogeneous and that many individuals have limited numbers of contacts. The frequency of susceptible hosts (S/N) therefore determines the number of transmissions rather than the absolute density (S). A concept map of the SIR model is shown below.



The differential equations for the **SIR-DD** model are:

$dS/dt = b(S+I+R) - dS - \beta SI + \gamma R$ [in the SIR-FD model, βSI is replaced with $(\beta SI)/N$]

$dI/dt = \beta SI - (\alpha+d+\nu)I$

$dR/dt = \nu I - (d+\gamma)R$

The equations can be integrated to obtain the net reproductive rate of disease

$R_0 = (\beta S)/(\alpha+d+\nu)$

For disease to persist, R_0 must exceed 1,
so we can set $R_0 = 1$ to get the threshold susceptible-host population density to sustain the parasite, that is

$S_T = (\alpha+d+\nu)/\beta$

The implication is that the disease will go extinct unless $S > S_T$

Macroparasitic Infections

Macroparasites generally include helminths and arthropods which usually cause chronic and persistent infections. Disease severity depends on the number of parasites present which in turn depends on exposure to free-living infective stages. Infections are often over-dispersed, as a small fraction of the host population may harbour most of the parasites. It is not sufficient to divide the host population into S, I or R classes, as macroparasite models must track the intensity of infection in individual hosts (often by assuming a probability distribution for parasite loads). Models take into account host population size (H) as well as populations of adult parasites (P) (inside hosts) and free-living stages (W) (outside hosts).

The Anderson & May (A&M) model characterizes between-host variation in parasite burdens with a negative binomial distribution. The decoupled A&M model is a simplified version whereby infective stages are assumed to be short-lived and at equilibrium.

The Dobson & Hudson (D&H) model includes a hypobiotic stage of arrested parasite development following infection, before maturation of adult parasites affect host vitality.

Parameters used in the models include:

H = total host population size (numbers)
P = total parasite population size (total numbers of adults)
W = population of free-living infective stages (eggs and larvae)

Host parameters include:

b = host birth rate
d = host death rate (natural mortality)

Parasite parameters involve:

α = disease-induced mortality rate (/parasite/time)
β = transmission rate per host contact (/host/time)
λ = birth rate of parasite eggs or larvae
κ = negative binomial aggregation parameter (a dimensionless constant inversely proportional to assumed parasite aggregation among hosts)
μ_P = natural mortality rate of adult parasites
γ = infective stage mortality rate

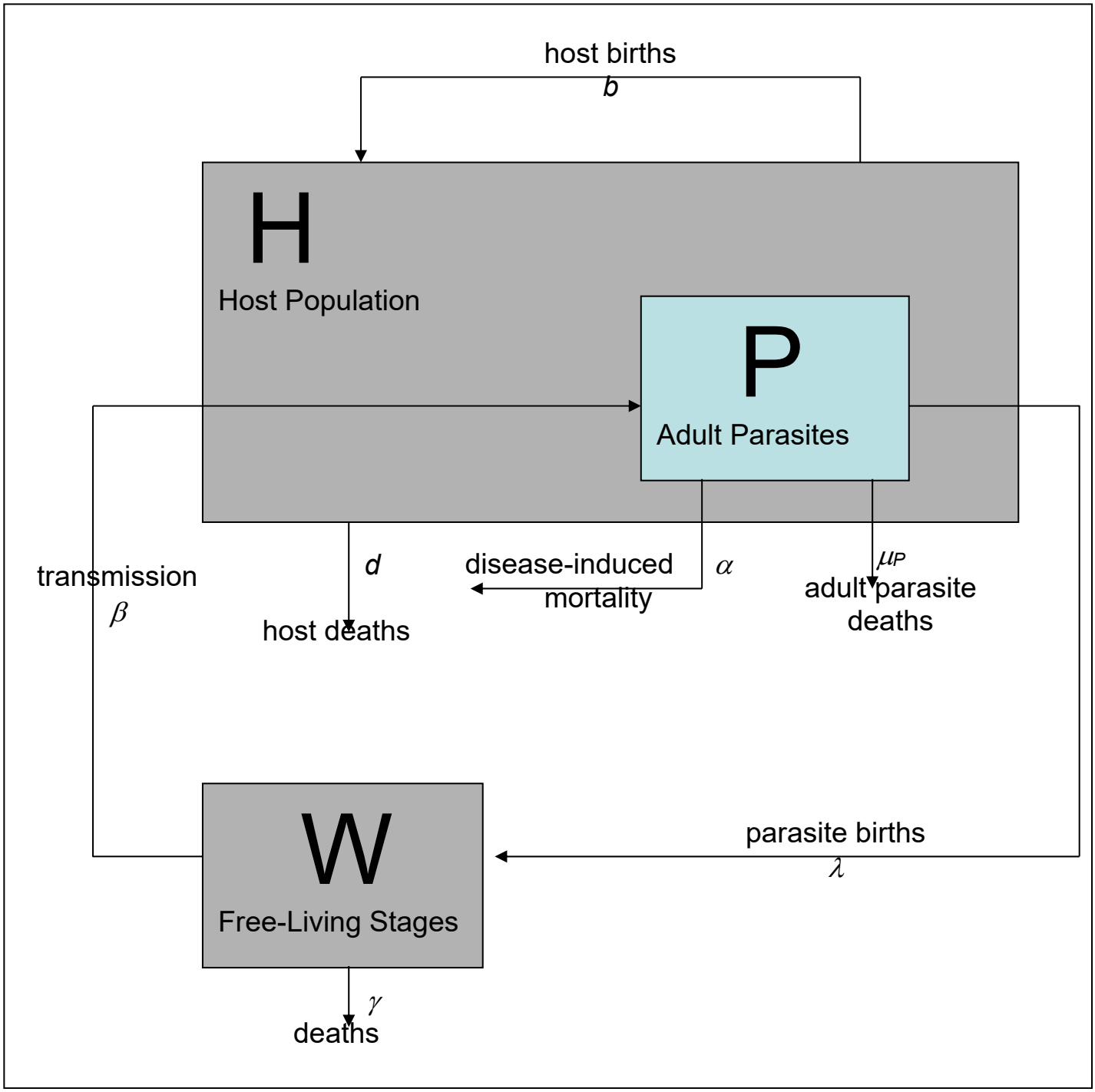
Movement between HPW states is modelled by the following differential equations:

$$\mathrm{d}H/\mathrm{d}t = (b-d)H - \alpha P$$

$$\mathrm{d}P/\mathrm{d}t = \beta WH - (\mu_P+d+\alpha)P - \alpha P^2/H[(k+1)/k]$$

$$\mathrm{d}W/\mathrm{d}t = \lambda P - \gamma W - \beta WH$$

The concept map for the A&M macroparasite model with direct transmission is shown below:



Host and parasite populations may reach equilibrium if $\lambda - (\mu_P + d + \alpha) > (b-d).[(\kappa+1)/\kappa]$

If so, the parasite is capable of regulating host populations to an equilibrium where $P/H = (b-d)/\alpha$
If not, but $\lambda - (\mu_P + d + \alpha) > 0$, then hosts grow at an exponential rate (< disease-free rate).
If not, but $\lambda - (\mu_P + d + \alpha) < 0$, then the parasite cannot be maintained.

Chapter 16.

Clip Art

ENGINEERING

Vocabulary list: engineering, technology, materials science, stress, strain, compressive strength, tensile strength, shear strength

Engineering is the application of science and technology to the design, analysis, and/or construction of useful products and services for the community within economic, environmental and resource constraints. The history of engineering stems from the earliest times when humans began to make clever inventions, such as the pulley, lever, and wheel. An engineer is literally someone concerned with engines; derived from the Latin *ingenium* and meaning "innate quality, especially mental power, hence a clever invention". An engineer is therefore someone who makes useful or practical inventions. Much of science involves the analysis of complex systems, a way of thinking common to engineers. Creating function by building a complex system, and getting it to work provides compelling proof that the scientist understands the components and how they work in synchrony. Scientists can gain an understanding of a system through a constructionist approach (build-up) or a deconstructionist approach (pull-apart, reverse engineering). Engineering is a broad discipline involving many branches, historically categorized as:

- mechanical engineering (design of physical or mechanical systems, such as engines, power trains, kinematic chains and vibration isolation equipment)
- civil engineering (design and construction of public and private buildings, bridges, etc);
- chemical engineering (conversion of raw materials into usable commodities);
- electrical engineering (design of electrical systems and goods, such as transformers); and
- aerospace engineering (design of aircraft, spacecraft and related topics).

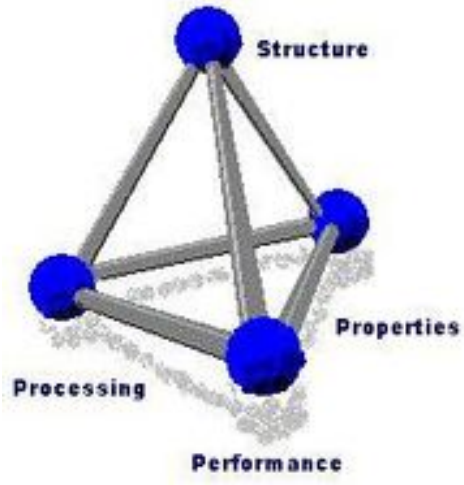
With the rapid advancement of technology, many new fields are gaining prominence and new branches are developing, such as environmental engineering, computer engineering, software engineering, biomedical engineering, nanotechnology, molecular engineering, mechatronics, etc. All areas of endeavour require a detailed and integrated knowledge of the cognate scientific disciplines, particularly mathematics, physics and chemistry. Engineering principles are also finding application in many biological fields; e.g. mechanical concepts applied to human movement studies (biomechanics), fluid dynamics to cardiovascular blood flow, electrical signalling to nerve conduction, etc.

The inventions people use, perhaps more than any other facet, defines the technological sophistication of a society. Humans have progressed through the Stone Age, the Bronze Age, the Iron Age and the Industrial Revolution, and are now experiencing the technological marvels of the Silicon Age (computer microchips). Technology is a broad concept that deals with a species usage and knowledge of tools and crafts, and how it affects their ability to control and adapt their environment. In human society, it is a consequence of science and engineering. Technology can refer to material objects of use to humanity, such as machines, hardware or utensils, but can also encompass broader themes, including systems, methods of organization, and techniques. People's use of technology began with the conversion of natural resources into simple tools. The prehistoric discovery of the ability to control fire increased the available sources of food and the invention of the wheel helped humans in travelling in and controlling their environment. Recent technological developments, including the printing press, the telephone, and the Internet, have lessened physical barriers to communication and allowed humans to interact on a global scale. However, not all technology has been used for peaceful purposes; the development of weapons of ever-increasing destructive power has progressed throughout history, from clubs to nuclear weapons.

Technology has affected society and its surroundings in a number of ways. In many societies, technology has helped develop more advanced economies (including today's global economy) and has allowed the rise of a leisure class. Many technological processes deplete natural resources and produce unwanted by-products (pollution) to the detriment of the Earth. Various implementations of technology influence the values of a society and new technology often raises new ethical questions. Philosophical debates have arisen over the use of technology in society, with disagreements over whether technology improves the human condition or worsens it. Neo-Luddism, anarcho-primitivism, and similar movements criticise the pervasiveness of technology in the modern world, claiming that it harms the environment and alienates people; proponents of ideologies such as transhumanism and techno-progressivism view continued technological progress as beneficial to society and the human condition. Irrespective of philosophical opinion, most human populations are surrounded by hundreds of materials in everyday life.

Materials science

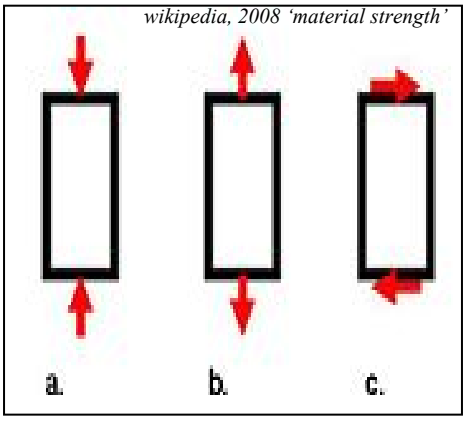
wikipedia, 2008 'materials science'



Look around your immediate environment and you will quickly recognize many examples of natural materials (wood, stone, clay, skins, fibres) and manufactured materials (concrete, bitumen, steel, aluminium, rubber, synthetic fibres, plastics, etc). Materials science is an interdisciplinary field involving the structure and properties of matter and their application (processing and performance) to various areas of science and engineering. Industrial applications include materials design (planning, development, economics), processing (casting, rolling, welding, ion implantation, crystal growth, thin-film deposition, sintering, glassblowing, etc.), and analytical techniques (characterization through electron microscopy, X-ray diffraction, calorimetry, nuclear microscopy, Rutherford backscattering, neutron diffraction, etc.). Material scientists also deal with the extraction of materials and their conversion into useful forms (ingot casting, foundry techniques, blast furnace extraction, electrolytic extraction, etc). The study of metal alloys is a significant part of materials science. The alloys of iron (steel, stainless steel, cast iron, tool steel, alloy steels) make up a large proportion both in quantity and commercial value. The alloys of aluminium and titanium are renowned for their high strength-to-weight ratios, and those of magnesium for their ability to provide electromagnetic shielding. Polymers and ceramics are also an important part of materials science. Polymers are the raw materials (resins) used to make plastics, and include polyethylene, polypropylene, polyvinyl-chloride, polystyrene, polyurethane, and polycarbonates.

Materials display many different properties, involving strength, hardness, flexibility, density, solubility, texture, lustre, colour, melting point, etc. These properties are dependent on three essential features: the kinds of atoms involved; the arrangement of those atoms; and how the atoms are bonded together. These, taken together and related through the laws of thermodynamics, govern a material's microstructure, and thus its properties. Strength is the ability of a solid to resist changes in shape. This bears a direct relationship to the chemical bonds involved; e.g. van der Waals forces (such as in graphite) are the weakest, ionic bonds (such as in glass and ceramics) are stronger, and covalent bonds (such as in plastics and Kevlar) are the strongest. All materials are held together by the bonds between atoms. When an outside force is applied, the atoms must shift their positions in response - bonds stretch and compress. Equal and opposite forces are generated within the material to resist the outside forces, in accordance with Newton's third law of motion.

Material strength



Material strength is not a single property, because there are different ways of placing an object under stress. Scientists and engineers recognize three very different kinds of strength:
(a) compressive strength (ability to withstand crushing);
(b) tensile strength (ability to withstand pulling apart); and
(c) shear strength (ability to withstand twisting).
Materials differ markedly in the kinds of strength they possess. Some exhibit high tensile strength but poor shear strength, some have great compressive strength but poor tensile strength. Composite materials combine the properties of two or materials, the strength of one offsetting the weakness of the other, e.g. reinforced concrete where steel rods, with great tensile strength, are embedded in concrete, which has great compressive strength.

Uniaxial stress is defined as force over area, given by the formula $\sigma = F/A$, where F = force (newtons) and A = area (m^2). The area can be the undeformed area or the deformed area, depending on whether engineering stress or true stress is used. The effects of dynamic loading are probably the most important practical part of the strength of materials, especially the problem of fatigue. Repeated loading often initiates brittle cracks, which grow slowly until failure occurs. Fatigue strength is a measure of the strength of a material or a component under cyclic loading, and is given as stress amplitude or stress range ($\Delta\sigma = \sigma_{max} - \sigma_{min}$), along with the number of cycles to failure.

The point at which material stops resisting external forces and begins to bend, break or tear is called its elastic limit. Elasticity is the ability of a material to return to its previous shape after stress is released. In many materials, the relation between applied stress and the resulting strain is directly proportional (up to a certain limit), giving a straight line when graphed. The slope of the line is known as Young's modulus and can be used to determine stress-strain relationships. Plasticity is the opposite of elasticity and is defined as unrecoverable strain; plastic deformation is retained even after the relaxation of the applied stress. Most elastic materials are also capable of plastic deformation, while brittle materials, like ceramics, do not experience any plastic deformation and will fracture under relatively low stress. Materials such as metals usually experience a small amount of plastic deformation before failure while soft or ductile polymers will deform much more.

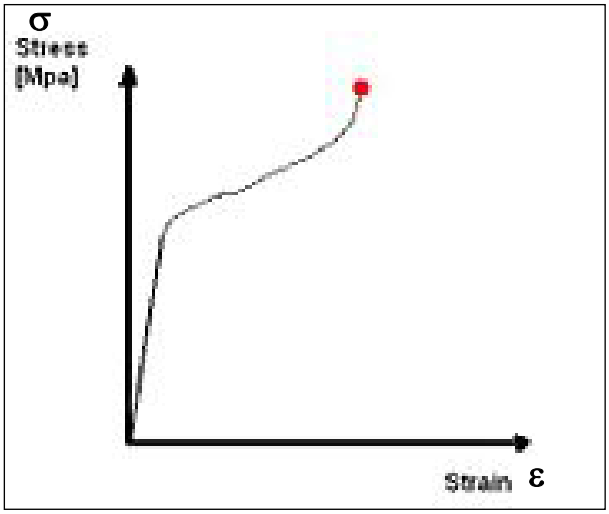
The ultimate strength of a material is often quoted as force per unit of cross section area (N/m^2), e.g. the ultimate tensile strength (UTS) of AISI 1018 steel is 440 MN/m^2. The SI unit of stress is the pascal, where 1 Pa = 1 N/m^2. In Imperial units, the unit of stress is given as pounds-force per square inch (psi). The 'Factor of Safety' (FS) is a design constraint that an engineered component or structure must achieve, and is given by the formula $FS = UTS / R$, where UTS = ultimate force (or stress), and R = applied stress. For example, to achieve a factor of safety of 4, the allowable stress in an AISI 1018 steel component is calculated as $R = UTS / FS = 440/4 = 110$ MPa ($= 110 \times 10^6$ N/m^2).

Compressive strength

The compressive strength of a material is its capacity to withstand uniaxially directed pushing forces, compressive stress, before it fails completely and is crushed. Concrete can be made to have high compressive strength, e.g. many concrete structures have compressive strengths in excess of 50 MPa, whereas a material such as soft sandstone may have a compressive strength as low as 5 or 10

MPa. On an atomic level, the molecules or atoms are forced together when in compression whereas they are forced apart when in tension. Since atoms in solids always try to find an equilibrium position and distance between other atoms, forces arise throughout the entire material which oppose both compression and tension.

Compressive strength is measured experimentally by means of a compressive test, where a specimen (usually cylindrical) is placed between two compression plates. As force is applied, the specimen becomes shortened as well as spreading laterally. A stress (σ)–strain (ε) plot by the instrument would look similar to that shown in the diagram. The compressive strength of the material corresponds to the stress at the terminal end-point. The initial region of the plot shows a linear relationship, where the material follows Hooke's Law, according to the formula $\sigma = E\varepsilon$, where E = Young's modulus (in this case, for compression). The linear region terminates at the yield point, above which the material behaves plastically and will not return to its original length once the load is removed.



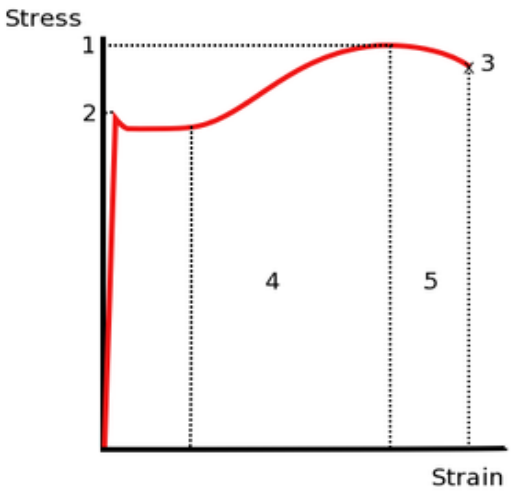
There is a difference between the engineering stress and the true stress. By definition, the uniaxial stress $\sigma = F/A$. The area of the specimen, however, varies on compression. It is some function of the applied load, that is, $A = f(F)$. The force divided by area at the start of the experiment is known as the engineering stress and is defined by $\sigma_e = F/A_0$, where A_0 = original specimen area. The corresponding engineering strain is defined by: $\varepsilon_e = (l - l_0) / l_0$, where l = current specimen length [m] and l_0 = original specimen length [m]. The compressive stress would therefore correspond to the point on the engineering stress-strain curve (ε_e, σ_e) defined by:

$$\sigma_e = F^* / A_0 \text{ and}$$
$$\varepsilon_e = (l^* - l_0) / l_0$$

where F^* = load applied just before crushing and l^* = specimen length just before crushing.

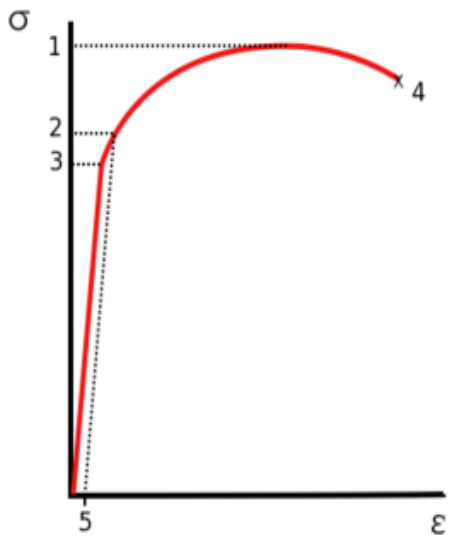
Tensile strength

The tensile strength of a material is the amount of stress required to pull something apart, such as stretching a rope, wire or structural beam to breaking point. There are actually three definitions of tensile strength: yield strength (the stress at which material strain changes from elastic deformation to plastic deformation); ultimate strength (the maximum stress a material can withstand) and breaking strength (the stress coordinate on the stress-strain curve at the point of rupture). These various definitions are shown in the stress-strain graph for low-carbon steel (where 1 = ultimate strength, 2 = yield strength, 3 = rupture, 4 = strain hardening region, and 5 = necking region).



Metals including steel have a linear stress-strain relationship up to the yield point. In some steels, the stress decreases after the yield point due to the interaction of carbon atoms and dislocations in the stressed steel. For most metals, the yield point is not sharply defined. All deformation below the yield strength is recoverable (elastic), but all deformation above the yield point is not (plastic). After the yield point, steel and many other ductile metals undergo a period of strain hardening, in which the stress increases with strain up to the ultimate strength. After a metal has been loaded to its yield strength, it begins to "neck" as the cross-sectional area of the specimen decreases due to plastic flow. When necking becomes substantial, it may cause a reversal of the engineering stress-strain curve, where decreasing stress correlates to increasing strain because of geometric effects. This is because the engineering stress and engineering strain are calculated assuming the original cross-sectional area before necking. If the graph is plotted in terms of true stress and true strain the curve will always slope upwards and never reverse, as true stress is corrected for the decrease in cross-sectional area. The peak stress on the engineering stress-strain curve is known the ultimate tensile strength. After a period of necking, the material will rupture and the stored elastic energy is released as noise and heat. The stress on the material at the time of rupture is known as the breaking stress.

Ductile metals do not have a well defined yield point. The yield strength is typically defined by the "0.2% offset strain", determined by finding the intersection of the stress-strain curve with a line parallel to the initial slope of the curve and which intercepts the abscissa at 0.002. A stress-strain curve typical of aluminum along with the 0.2% offset line is shown adjacent (where 1 = ultimate strength, 2 = yeld strength, 3 = proportional limit stress, 4 = rupture, and 5 = offset strain, typically 0.002). Brittle materials such as concrete and carbon fiber do not have a yield point, and do not strain-harden which means that the ultimate strength and breaking strength are the same. Typical brittle materials do not show any plastic deformation but fail while the deformation is elastic. One of the characteristics of a brittle failure is that the two broken parts can be reassembled to produce the same shape as the original component. A typical stress-strain curve for a brittle material will be linear.



The breaking strength of a rope is often specified in units of force, such as newtons, without specifying the cross-sectional area of the rope. This is often loosely called tensile strength, but this is not a strictly correct use of the term. In brittle materials such as rock, concrete, cast iron, or soil, tensile strength is negligible compared to the compressive strength and it is assumed zero for many engineering applications.

Tensile strength can be defined for liquids as well as solids. For example, when a tree draws water from its roots to its upper leaves by transpiration, the column of water is pulled upwards by capillary action, and this force is transmitted down the column by its tensile strength. Air pressure from below also plays a small part in a tree's ability to draw up water, but this alone would only be sufficient to push the column of water to a height of about ten metres, and trees can grow much higher than that.

Shear strength

The shear strength of a material is the amount of stress required for it to be distorted or broken by opposing twisting forces which shear successive layers within the material, i.e. the layers shift laterally over each other. In engineering, shear strength is important when designing the dimensions and composition of structural components to resist or yield to shear forces, for example, steel rivets reinforcing metal constructions or the safety bolts in aircraft ejector seats. Because opposing forces are involved, shear stress (τ) is measured as: $\tau = (\sigma_1 - \sigma_2) / 2$, where σ_1 = major principal stress, and σ_2 = minor principal stress. In general, ductile materials fail in shear (e.g. aluminium) whereas brittle materials (e.g. cast iron) fail in tension.

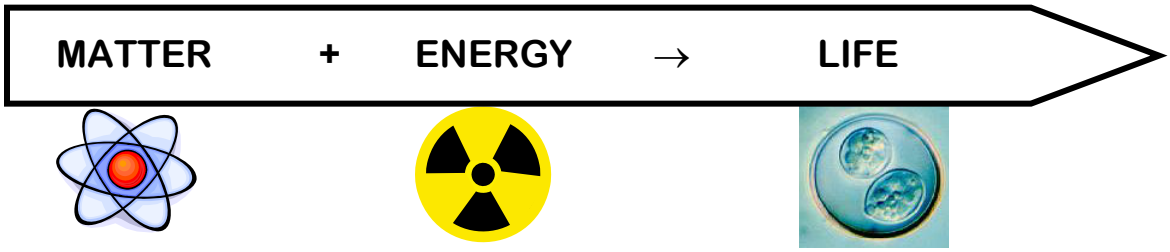
Question: What force would be required to shear a 5 cm diameter bolt made of steel with a shear stress rating of 110 MPa?
Answer: We are given $\tau = 110$ MPa $= 110 \times 10^6$ and we can calculate bolt area $A = \pi r^2 = \pi\ 0.025^2 = 1.96 \times 10^{-3}$ m^2 We can rearrange the formula $\tau = F/A$ to give $F = \tau A$ $= (110 \times 10^6) \times (1.96 \times 10^{-3})$ N.m^{-2} m^2 $= 2.156 \times 10^5$ N

Engineering has dramatically changed the face of our planet. We have applied our knowledge to the construction of advanced civilizations, sophisticated technologies, monumental structures and other tangible evidence of our existence on Earth. Regrettably, our activities have also led to marked chemical and industrial pollution, radioactive waste, global warming, diminishing biodiversity, etc. Paradoxically, while science and engineering has created many environmental problems, we have little recourse other than to turn to science and engineering to find solutions to those problems.

17. Concluding remarks

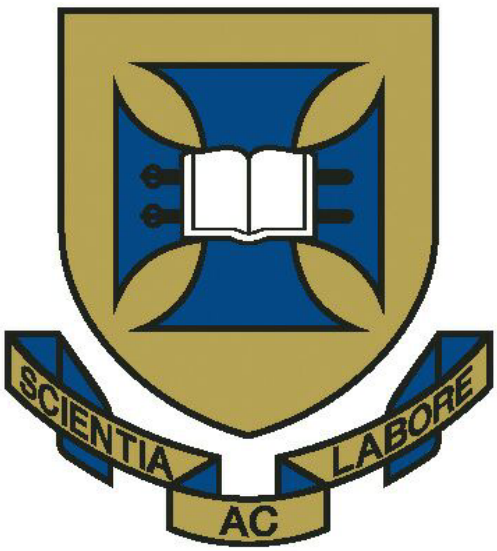
These course notes have attempted to compress hundreds of years of accumulated knowledge into a short overview. The notes are by no means complete but should distill the essence of various disciplines of relevance to any serious contemporary scientist.

Personally, I like learning using illustrations, anecdotes, exemplars, analogies, metaphors, similes, parables, models, etc. They give me a framework upon which to hang information, and provide clues for my overtaxed memory. I like thinking about science as though I were creating Frankenstein's monster by following a recipe: take some matter, charge it with energy, generate life.



mass	potential	cells
space	kinetic	organisms
atoms	electromagnetic	populations
molecules	chemical	communities
reactions	nuclear	ecosystems

I constantly use this triad of thought to check that I am seeing the big picture in my scientific endeavours. You will probably develop your own quirky mannerisms throughout your career. I hope you enjoy your journey as a scientist as much as I have mine. It is enormously satisfying professionally and personally, but can only be accomplished (as stated in The University of Queensland crest and motto) "*scientia ac labore*" ("by means of knowledge and hard work").



18. References:

Blackman A, Bottle SE, Schmid S, Mocerino M, Wille U. 2008. Chemistry. John Wiley & Sons [ISBN 9-78047081-0866]

Campbell NA, Reece JB & Meyers N. 2006. Biology. 7th edition, Pearson Education [ISBN 1741033861]

Dawkins R. 2004. The Ancestor's Tale. Phoenix/Orion Books, London. [ISBN 0753819961]

Flannery T. 2006. The Weather Makers. Text Publishing, Melbourne. [ISBN 1-920885-84-6]

Haefner JW. 2005. Modeling biological systems: principles and applications. Springer [ISBN 0387250115]

Klein DR 2005. General Chemistry. John Wiley & Sons [ISBN 0-471-71662-6]

Knight RD. 2008. Physics for scientists and engineers. 2nd ed. Pearson Addison Wesley [ISBN 9780805327366]

Kumar V., Abbas AK & Fausto N. 2005. Robbins and Cotran - Pathologic basis of disease. 7th edition, Elsevier Saunders [ISBN 0721601871]

Lecointre G & Le Guyader H. 2006. The tree of life: a phylogenetic classification. Harvard University Press [ISBN 0674021835]

Madigan MT & Martinko JM 2006. Brock. Biology of microorganisms. 11th ed. Pearson Education Int. [ISBN 0131968939]

Madigan MT, Martinko JM, Dunlap PV & Clark DP. 2009. Brock. Biology of microorganisms. 12th ed. Pearson Education Int. [ISBN 0321536150]

Marieb EN & Hoehn K. 2007. Human anatomy and physiology. 7th ed. Pearson Education. [ISBN 0805359095]

Mehlhorn H. 2001. Encyclopedic reference of parasitology, 2nd ed. Springer Pub. [ISBN 3540668195]

Purves WK, Orians GH, Heller HC & Sadava D. 1997. Life: the science of biology. WH Freeman & Co. [ISBN 0716728699]

Trefil J & Hazen RM. 2007. The sciences: an integrated approach. John Wiley & Sons [ISBN 100471769924]

Walker JS. 2007. Physics. 3rd ed. Pearson Prentice Hall [ISBN 0131536311]