biofundamentals 2.0

An introduction to molecular & evolutionary biology
preliminary edition

Michael W. Klymkowsky & Melanie M. Cooper

Molecular, Cellular & Developmental Biology, University of Colorado Boulder
Chemistry, Michigan State University

You know how it is.
You pick up a book, flip to the dedication, & find that, once again,
the author has dedicated a book to someone else & not to you.

Not this time.

Because we haven't yet met/have only a glancing acquaintance/are just crazy about each other/haven't seen each other in much too long/are in some way related/will never meet, but will, I trust, despite that, always think fondly of each other….

This one's for you.

(lifted from Neil Gaiman)

Preface: The biofundamentalist approach to teaching and learning basic biology

Our goal is to present the key observations and unifying concepts upon which modern biology is based. Once understood this is knowledge that will enable you to approach any biological process, from disease to kindness, from a scientific perspective. To truly understand biological systems we need to consider them from two complementary perspectives; how they came to be (the historic) and how their structures, traits, and behaviors are produced (the mechanistic).

We are biological entities, the products of complex developmental processes acting on inherited genetic information. We live in complex social arrangements with other humans and other organisms whose behaviors influence us in both subtle profound ways. As we alter our environment we inevitably alter ourselves. Science is a coherent strategy by which we seek to better understand the Universe and ourselves; how the physical world and its history shape and constrain what is and what is not possible. That said, science does not provide a prescription for how things should be. Science cannot tell us what is morally right and wrong, it can only attempt to explain what is and what might be. That said our scientific understanding of almost every topic, and particularly the remarkably complex behaviors of biological systems, is incomplete. It is not even clear that the Universe is necessarily coherent. The difficulties in producing a single theory that encompasses both the behavior of the vary large (gravity) and the very small (quantum mechanics) raises the possibility that a single theory of everything may not be possible or if possible may not be comprehensible to us.[1]

Scientific knowledge is a body of knowledge of varying degrees of certainty-some most unsure, some nearly sure, but none absolutely certain ... Now we scientists are used to this, and we take it for granted that it is perfectly consistent to be unsure, that it is possible to live and not know.
- Richard Feynman.

...it is always advisable to perceive clearly our ignorance.
– Charles Darwin.

Periodically a perspective known as scientism gains popularity in certain circles. It holds that science provides a complete and exclusive picture of the Universe, a picture that dictates how we should behave. We caution against this view, in part based on the lessons of history and in part because it violates our own deeply held (some might say, enlightenment) view that we are each unique individuals who are valuable in and of ourselves, deserving of respect, and not objects to be sacrificed to abstract ideals (that is, blown up or otherwise abused for scientific, political, religious, or economic reasons). A number of serious crimes against humanity and individuals have been justified based on purportedly unambiguously established "facts" or beliefs that later turned out to be untrue, seriously incomplete, tragically misapplied, or more or less irrelevant.[2] Crimes against

[1] Physics's pangolin: Trying to resolve the stubborn paradoxes of their field, physicists craft ever more mind-boggling visions of reality: http://aeon.co/magazine/science/margaret-wertheim-the-limits-of-physics/

[2] The Undergrowth of Science: http://www.salon.com/2000/11/30/gratzer/

people in the name of science are as unforgivable as crimes against people in the name of religion or political ideologies.

That said, scientific thinking is indispensable if we want to distinguish established, empirically supported observations from fantasies. Such fantasies can often be harmful, such as anti-vaccine campaigns that lead to an increase in deaths and avoidable diseases.[3] When we want to cure diseases, reduce our impact on the environment, or generate useful tools we are best served by adopting a dispassionate, empirically-based scientific approach to inform our decisions. Scientific studies help us decide between the possible and the impossible and to assess the costs and benefits of various interventions.

How biology differs from physics and chemistry

While it is true that biological systems, that is organisms, obey the laws of physics and chemistry they are more than highly complex chemical and physical systems. Why, you might well ask? Because each organism is a unique entity, distinguishable from others by the genetic information it carries and, at the molecular and cellular levels, by the stochastic events that have combined to influence its behavior. Even identical twins can be distinguished in terms of molecular and behavioral details. Moreover, each organism has a unique history that runs back in time for an unbroken period of ~3,500,000,000 years. To understand an organism's current shape, internal workings, and visible behaviors requires an appreciation of the general molecular, cellular, developmental, social, and ecological processes involved in producing these traits. These mechanistic processes are themselves the product of what the molecular biologist Francois Jacob (1920-2013) referred to as evolutionary tinkering, that is, the organism's evolutionary history.[4]

No organism, including us, was designed *de novo* (from the Latin meaning, anew). Rather each (including us) is the products of continuous evolutionary processes that occurred over long periods of time and involved a series of ancestors adapted to their own particular life styles (ecological niches), through a complex process that involved combinations of random (stochastic) and non-random events. These include mutational variation, various forms of genetic recombination, various types of selection, that arise through both internal processes and the organism's interactions with a changing environment. Because of these complex and interacting processes, one cannot readily deduce the details of an organism from physical first principles. Take for example the vertebrate eye, which behaves completely in accord with physical laws, nevertheless displays idiosyncrasies associated with its evolutionary history, idiosyncrasies that enable us to deduce that it arose independently from, for example the eyes of molluscs.[5] Evolutionary processes lead to the emergence of new traits and types of organisms and at the same time play a conservative role, maintaining organisms against the effects of molecular level noise in the form of deleterious mutations. The interactions between organisms and their environment produce evolutionary changes, albeit in often unpredictable ways. These processes can lead to the

---

[3] How vaccine denialism in the West is causing measles outbreaks in Brazil: http://www.theguardian.com/commentisfree/2014/apr/28/vaccine-denialism-measles-outbreaks-in-brazil and http://www.historyofvaccines.org/content/articles/history-anti-vaccination-movements

[4] Evolution and Tinkering: http://www.sciencemag.org/content/196/4295/1161.long and Tinkering: a conceptual and historical evaluation: http://www.ncbi.nlm.nih.gov/pubmed/17710845

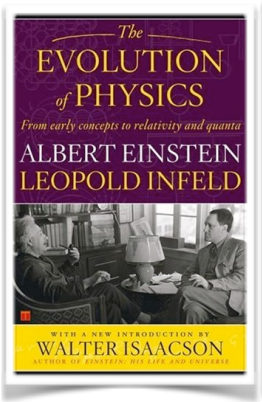
[5] How the Eye Evolved: http://www.nyas.org/publications/detail.aspx?cid=93b487b2-153a-4630-9fb2-5679a061fff7

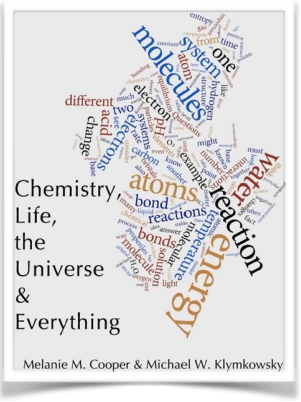
extinction of some organismic lineages as well as the appearance of new types of organisms from existing lineages and have led to the millions of different types of organisms currently in existence (in addition to those now extinct).

A second important difference between biological and physicochemical systems is that even the simplest of biological systems, an organism consisting of an individual cell (we will define what exactly a cell is in the next chapter) is more complex than the most complex physical system. Moreover, at the cellular and molecular levels there are often small numbers of specific molecules involved, so behaviors can be noisy and strictly deterministic behaviors are not always of primary importance. A bacterium, one of the simplest types of organisms in terms of molecular components, contains more than 3000 distinct genes, and hundreds to thousands of concurrent and interdependent chemical reactions, whose outcomes influence which genes are active (active genes are often said to be "expressed") and which are not, what ecological/environmental interactions are occurring, and how the bacterium responds to them. Nevertheless there are common themes that we will use and return to over and over again to make biological systems intelligible. We will rely on the fact that we can understand how molecules interact through collisions and binding interactions, how chemical reactions interact with one another, that is, how they are coupled through common intermediates, and how physical laws, in particular the laws of thermodynamics, constrain and shape biological behaviors.

The student's background and our teaching approach

While it is often the case that biology is taught early in a science sequence, this seems rather counterintuitive to us, since biological systems and processes are more complex that non-living physical or chemical systems even though biological systems are based on and constrained by physical and chemical principles. We recognize that it is unlikely that most students will enter the course completely comfortable with physical and chemical concepts, and we have written the text presuming very little. Where reference to physicochemical concepts is necessary, we have attempted to point them out explicitly and addressed them at a level that we believe should be adequate for students to be able to deal productively with the ideas presented. Given that biology students are a large fraction of the target cliental of introductory physics and chemistry courses, one can only hope that over time these courses will evolve to help life sciences students learn what they need to know. We suggest that students interested in learning more about the physical and chemical concepts that underlie biology might want to read Einstein and Infeld's "The Evolution of Physics" and our own "Chemistry, Life, the Universe, and Everything."





A Socratic, learning-centered approach to teaching: The complexity of biological systems can be daunting and all too often biology has been presented as a series of vocabulary terms, while little attention is paid to its underlying conceptual (sense-making) foundations. This emphasis on

memorization can be off-putting and, in fact, is not particularly valuable in helping the student to develop a working understanding of biological systems. Our driving premise is that while biological systems are complex, both historically and mechanistically, there is a small set of foundational observations and ideas that apply to all biological systems. Their complexity, and the incompleteness of our understanding, often make a perfect (complete and accurate) answer to biological questions difficult. Nevertheless, it is possible to approach biological questions in an informed, data-based (empirical) and logical manner. In general, we are less concerned with whether you can remember or reproduce the "correct" answer to a particular question and more with your ability to identify the facts and over-arching concepts relevant to a question and to then construct a plausible and logical answer.

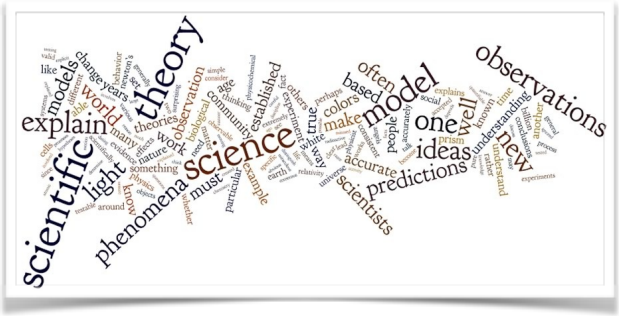
Going beyond memorization means that you will be expected to use and apply your understanding of key facts and overarching ideas to particular situations. This will require that you develop (through practice) the ability to analyze a biological situation or scenario; to identify what factors are critical (and recognize those that are secondary or irrelevant) and then apply your understanding to make predictions or critique conclusions. To this end we will repeatedly ask you to mentally dissect various situation to reach your own conclusions or solutions. To give you opportunities to practice, each section of the book includes a number of "questions to answer and ponder." You should be able to generate plausible answers to these questions, answers that we hope you will have an opportunity to present to, and analyze with, your instructor and fellow students. Where you do not understand how to proceed, you should storm into class able to articulate exactly why you are confused (something that often takes some serious thinking). You will need to actively search (and if you cannot find it, demand help in developing) a viable approach that enables you to answer those questions or to explain why the questions makes no sense. As part of this process, we have developed a number of interactive beSocratic activities, accessible through web links (BeSocratic.com) that are designed to develop your ability to construct models and explanations of important phenomena. In many cases, you will receive feedback within the context of the activity. That said, there is no substitute for discussions with other students and your instructors; that is, after all why one has experts in biology teaching biology courses. Ideas that you find obscure or that make no sense to you need to be addressed directly. Learning to critique or question an explanation will help you identify what is relevant, irrelevant, conceptually correct or logically absurd in your and your fellow students' thinking, so that by the time we reach the end of the course, you will have learned something substantial about biological systems.

We think the way we do because Socrates thought the way he did.
- Bettany Hughes

Revisions to the text: Because this is an introductory course and because the ideas presented are well established and foundational, we expect no need for dramatic revisions of content. That said, we have much to learn about how to help students master and apply complex biological ideas, so we are using student responses both from beSocratic activities, and from classroom interactions to identify effective activities and ineffective parts of the text so that they can be improved. New "editions" will incorporate these insights.

Chapter 1. Understanding science & thinking scientifically

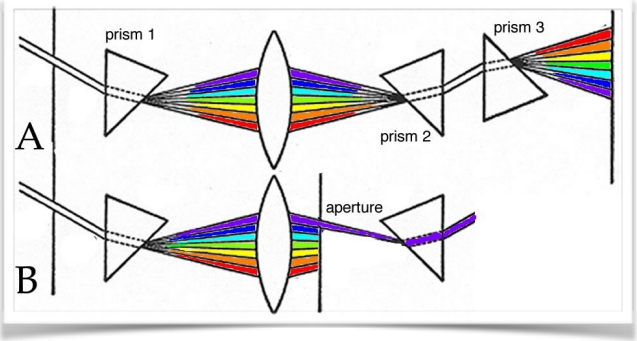
In which we consider what makes science a distinct, productive, and progressive way of understanding how the universe works that let us identify what is possible and what is impossible. We consider the "rules" that characterize a scientific approach to a particular problem.



A major feature of science, and one that distinguishes it from many other human activities, is its essential reliance upon shareable experiences rather than individual revelations. Thomas Paine (1737-1809), one of the intellectual parents of the American Revolution, made this point explicitly in his book *The Age of Reason*.[6] In science, we do not accept that an observation or a conclusion is true solely because another person claims it to be true. We do not accept the validity of revelation or what we might term "personal empiricism." What is critical is that, based on our description of a phenomena, an observation, or an experiment, others should in practice (or at the very least in theory) be able to repeat the observation or the experiment. Science is based on social (shared) knowledge rather than revealed truth.

> ***Revelation is necessarily limited to the first communication – after that it is only an account of something which that person says was a revelation made to him; and though he may find himself obliged to believe it, it can not be incumbent on me to believe it in the same manner; for it was not a revelation made to ME, and I have only his word for it that it was made to him.***
> ***– Thomas Paine, The Age of Reason.***

As an example, consider sunlight. It was originally held that white light was "pure" and that somehow, when it passed through a prism, the various colors of the spectrum, the colors we see in a rainbow, were created. In 1665, Isaac Newton (1642–1727) performed a series of experiments that he interpreted as demonstrating that white light was not pure but was, in fact, composed of light of different colors.[7] This conclusion was based on a number of distinct experimental observations. First, he noted that sunlight passed through a prism generated a spectrum of light of many different colors. He then used a lens to focus the spectrum emerging from the first prism so that passed through a second prism (Part A→); a beam of white light emerged from this second prism. One could then go on to show that the light emerging from the prism 1 lens prism 2 combination behaved the same as the original white light by passing it through a third prism, which again produced a spectrum. In the second type of experiment (Part B→), Newton used a screen with a hole in it, an aperture, and showed that light of a particular color was not altered when it passed through a second prism - no new colors were



[6] The Age of Reason: http://www.ushistory.org/paine/reason/singlehtml.htm

[7] Newton's Prism Experiments: http://micro.magnet.fsu.edu/primer/java/scienceopticsu/newton/ & http://youtu.be/R8VL4xm_3wk

produced. Based on these observations, Newton concluded that white light was not what it appeared to be – that is, a simple pure substance – but rather was composed (rather unexpectedly) of light of many distinct "pure" colors. The spectrum was produced because the different colors of light were "bent" or refracted by the prism to different extents. Why this occurred was not clear and neither was it clear what light is. Newton's experiments left these questions unresolved. This is typical: scientific answers are often extremely specific, elucidating a particular phenomena, rather than providing a universal explanation of reality.

Two basic features make Newton's observations and conclusions scientific. The first is reproducibility. Based on his description of his experiment others could, and were able to reproduce, confirm, and extend his observations. If you have access to glass prisms and lenses, you can repeat Newton's experiments yourself, and you would come to the same empirical conclusions; that is, you would observed the same phenomena that he did.[8] In 1800, William Herschel (1738-1822) did just that. He used Newton's experimental approach and discovered infrared (beyond red) light. Infrared light is light that is invisible to us but its presence can be revealed by the fact that when absorbed, say by a thermometer, it leads to an increase in temperature. In 1801, inspired by Herschel's discovery, Johann Ritter (1776 –1810) used the ability of light to initiate the chemical reaction: silver chloride + light → silver + chlorine to reveal the existence of another type of invisible light, which he called "chemical light" and which we now call ultraviolet light.[9] Subsequent researchers have established that visible light is just a small portion of a much wider spectrum of "electromagnetic radiation" that ranges from X-rays to radio waves. Studies on how light interacts with matter have led to a wide range of technologies, from X-ray imaging to an understanding of the history of the Universe. All these findings emerge, rather unexpectedly, from attempts to understand the rainbow.

The second scientific aspect of Newton's work was his clear articulation of the meaning and implications of his observations, the logic of his conclusions. These led to explicit predictions, such as that a particular color will prove to be homogenous, and not composed of other types of light. His view is that the different types of light, which we see as different colors, differ in the way they interact with matter. One way these differences are revealed is the extent to which they are bent when they enter a prism. Newton used some of these ideas when he chose to use mirrors rather than lenses to build his reflecting (or Newtonian) telescope. His design avoided the color distortions that arose when light passed through simple lenses.

The two features of Newton's approach make science, as a social and progressive enterprise, possible. We can reproduce a particular observation or experiment, and follow the investigator's explicit thinking. We can identify unappreciated factors that can influence the results observed and identify inconsistencies in logic or implications that can be tested. This becomes increasingly important when we consider how various scientific disciplines interact with one another.

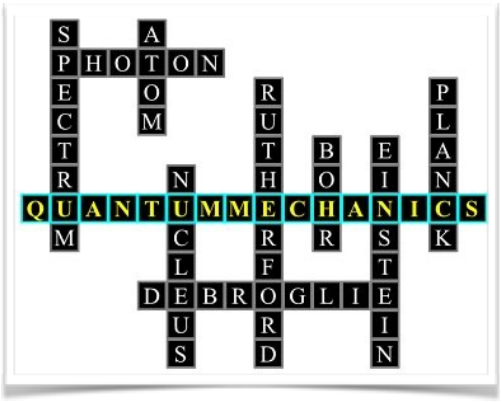
---

[8] Infrared astronomy: http://coolcosmos.ipac.caltech.edu/cosmic_classroom/ir_tutorial/discovery.html

[9] Ritter discovers ultraviolet light: http://coolcosmos.ipac.caltech.edu/cosmic_classroom/classroom_activities/ritter_bio.html

The interconnectedness of science

At one point in time, the study of biology, chemistry, physics, geology, and astronomy appeared to be distinct, but each has implications for the others, and they all deal with the real world. In particular, it is clear that biological systems obey the laws and rules established by physics and chemistry. As we will see, it was once thought that there were aspects of biological systems that somehow transcended physics and chemistry, a point of view known generically as vitalism. If vitalism had proven to be correct, it would have forced a major revision of chemistry and physics. As an analogy, the world of science is like an extremely complex crossword puzzle, where the answer to one question must be compatible with the answers to all of the others.[10] Alternatively, it can be that certain questions (and their answers) once thought to be meaningful can come to be recognized as irrelevant or meaningless. For example, how many angels can dance on the head of a pin is no longer considered a scientific question.



What has transpired over the years is that biological processes ranging from the metabolic to the conscious have been found to be consistent with physicochemical principles. What makes them distinctly different is that they are the product of evolutionary processes influenced by historical events that stretch back, in an uninterrupted “chain of being”, over billions of years. Moreover, biological systems in general are composed of many types of molecules, cells, and organisms that interact in complex ways. All this means is that while biological systems obey physicochemical rules, their behavior cannot be predicted based these rules. It may well be that life, as it exists on Earth, is unique. The only way we will know otherwise is to discover life on other planets, solar systems, galaxies, and universes (if such things exist), a seriously non-trivial but totally exciting possibility.

At the same time, it is possible that studies of biological phenomena could lead to a serious rethinking of physicochemical principles. There are in fact research efforts into proving that phenomena such as extrasensory perception, the continuing existence of the mind/soul after death, and the ability to see the future or remember the (long distant) past are real. At present, these all represent various forms of pseudoscience (and most likely, various forms of self-delusion and wishful thinking), but they would produce a scientific revolution if they could be shown to be real, that is, if they were reproducible and based on discernible mechanisms with explicit implications and testable predictions. This emphasizes a key feature of scientific explanations: they must produce logically consistent, explicit, testable, and potentially falsifiable predictions. Ideas that can explain any possible observation (something that some argue is the case for string theory in physics) are no longer science, whether or

[10] This analogy is taken from a talk by Alan Sokal: http://youtu.be/kuKmMyhnG94; graphic from http://scienceblogs.com/principles/2013/10/09/quantum-crosswords-my-tednyc-talk/

not they are “true” in some other sense.[11]

Models, hypotheses, and theories

Tentative scientific models are known as hypotheses. These are valuable in that they serve as a way to clearly articulate one’s assumptions. They form the logical basis for generating testable predictions about the phenomena they purport to explain. As scientific models become more sophisticated, their predictions can be expected to become more and more accurate or apply to areas that previous models could not handle. Let us assume that two models are equally good at explaining a particular observation. How might we judge between them? One way is the rule of thumb known as Occam's Razor (named after the medieval philosopher William of Occam, c. 1287–1347) or the Principle of Parsimony. This rule states that all other things being equal, the simplest explanation is the best. This is not to imply that an accurate scientific explanation will be simple, or that the simplest explanations are the correct ones, only that to be useful, a scientific model should not be more complex than necessary. Consider two models for a particular phenomena, one that involves angels and the other that does not. We need not seriously consider the model that invokes angels unless we can accurately monitor the presence of angels and if so, whether they are actively involved in the process to be explained. Why? Because angels, if they exist, clearly imply more complex factors that does a simple natural explanation. For example, we would have to explain what angels are made of, how they originated, and how they intervene in the natural world, that is, how they make matter do things. Do they obey the laws of thermodynamics or not? Under what conditions do they intervene? Are their interventions consistent or capricious? Assuming that an alternative, angel-less model is as or more accurate at describing the phenomena, the scientific choice would be the angel-less model. Parsimony (an extreme unwillingness to spend money or use resources) has the practical effect that it lets us restrict our thinking to the minimal model that is needed to explain specific phenomena. The surprising result, well illustrated by a TED talk by Murray Gell-Mann, is that simple, albeit often counter-intuitive rules, can explain much of the Universe with remarkable precision.[12] A model that fails to accurately describe and predict the observable world must be missing something and is either partially or completely wrong.

Scientific models are continually being modified, expanded, or replaced in order to explain more and more phenomena more and more accurately. It is an implicit assumption of many sciences that the Universe can be understood in scientific terms, and this presumption has been repeatedly confirmed but has by no means been proven.

A model that has been repeatedly confirmed and covers lots of observations is known as a theory – at least this is the meaning of the word in a scientific context. It is worth noting that the word theory is often misused, even by scientists who might be expected to know better. If there are multiple “theories” to explain a particular phenomena, it it more correct to say that i) these are not actually theories, in the scientific sense, but rather working models or simple speculations, and that ii) one or

---

[11] In this context, the lecture by Alan Sokol is worth listening to: http://www.guardian.co.uk/science/audio/2008/mar/03/alan.sokal.podcast. See also Farewell to Reality: http://www.math.columbia.edu/~woit/wordpress/?p=6002 ; http://www.math.columbia.edu/~woit/wordpress/ and http://www.scientificamerican.com/article/wronger-than-wrong/

[12] Beauty, truth and ... physics?: http://www.ted.com/talks/view/lang/en//id/194

more, and perhaps all of these models are incorrect or incomplete. A scientific theory is a very special set of ideas that explains, in a logically consistent, empirically supported, and predictive manner a broad range of phenomena. Moreover, it has been tested repeatedly by a number of critical and objective people – that is people who have no vested interest in the outcome – and found to provide accurate descriptions of the phenomena it purports to explain. It is not idle speculation. If you are curious, you might count how many times the word theory is misused, at least in the scientific sense, in your various classes.

That said, theories are not static. New or more accurate observations that a theory cannot explain will inevitably drive the revision or replacement of the theory. When this occurs, the new theory explains the new observations as well as everything explained by the older theory. Consider for example, gravity. Isaac Newton's law of gravity, describes how objects behave and it is possible to make extremely accurate predictions of how objects behave using its rules. However, Newton did not really have a theory of gravity, that is, an naturalistic explanation for why there is gravity and how it behaves the way it does. He relied on a supernatural explanation. When it was shown that Newton's law of gravity failed in specific situations, such as when an object is in close proximity of a massive object, like the sun, new rules and explanations were needed. Albert Einstein's Theory of General Relativity not only more accurately predicts the behavior of these systems, but also provided a naturalistic explanation for the origin of the gravitational force.[13] So is general relativity true? Not necessarily, which is why scientists continue to test its predictions in increasingly extreme situations.

Science is social

The social nature of science is something that we want to stress yet again. While science is often portrayed as an activity carried out by isolated individuals, the image of the mad scientist comes to mind, in fact science is an extremely social activity. It works only because it involves and depends upon an interactive community of scientists who keep each other (in the long run) honest.[14] Scientists present their observations, hypotheses, and conclusions are presented in the form of scientific papers, where their relevance and accuracy can be evaluated, more or less dispassionately, by others.

Over the long term, this process leads to an evidence-based, scientific consensus. Certain ideas and observations are so well-established that they can be reasonably accepted as universally valid, whereas others are extremely unlikely to be true, such as perpetual motion or "intelligent design creationism." These are ideas that can be safely ignored. As we will see, modern biology is based on a small set of theories[15] that include the Physicochemical Theory of Life, the Cell Theory and the Theory of Evolution. That said, as scientists we keep our minds open to exceptions and work to understand them. The openness of science means that a single person, taking a new observation or idea seriously, can

[13] A good video on General Relativity: http://www.bbc.co.uk/science/space/universe/questions_and_ideas/general_relativity#p009sgnl

[14] A good introduction of how science can be perverted is "The undergrowth of Science" by Walter Gatzer.

[15] Thinking about the conceptual foundations of the biological sciences: http://www.ncbi.nlm.nih.gov/pubmed/21123685

challenge and change accepted scientific understanding. That is not to say that it is easy to change the way scientists think. Most theories are based on large bodies of evidence and have been confirmed on multiple occasions. It generally turns out that most "revolutionary" observations are either mistaken, misinterpreted, or can be explained within the context of established theories. It is, however, worth keeping in mind that it is not at all clear that all phenomena can be put into a single "theory of everything." For example, it has certainly proven difficult to reconcile quantum physics with the general theory of relativity.

Gravity explains the motions of the planets, but it cannot explain who sets the planets in motion.
- Isaac Newton

A final point, mentioned before, is that the sciences are not independent of one another. Ideas about the behaviors of biological systems cannot contradict well established observations and theories in chemistry or physics. If they did, one or the other would have to be modified. For example, there is substantial evidence for the dating of rocks based on the behavior of radioactive isotopes of particular elements. There are also well established patterns of where rock layers (with specific ages) are found. When we consider the dating of fossils, we use rules and evidence established by geologists. We cannot change the age we assign to a fossil, making it inconsistent with the rocks that surround it, without challenging our understanding of the atomic nature of matter, the quantum mechanical principles involved in isotope stability, or geological mechanisms. A classic example of this situation arose when the physicist William Thompson (also known as Lord Kelvin)(1824-1907) estimated the age of the earth to be between 20 to 400 million years, based on the rate of heat dissipation of a once molten object, the earth. This was a time-span that seemed too short for various geological and biological processes, and greatly troubled Charles Darwin. Somebody was wrong, or better, this understanding was incomplete. The answer was with the assumptions that Kelvin had made; his calculations ignored the effects of radioactive decay (not surprising since radioactivity had yet to be discovered). These effects increased the calculated age of the earth by more than ten to one hundred fold, to about 5 billion years, an age compatible with both biological and geological processes.

Teaching and learning science

An important point to appreciate about science is that because of the communal way that it works, understanding builds by integrating one observation and idea into a network of others. As a result, science often arrives at conclusions that can be strange, counterintuitive, and sometimes disconcerting but nevertheless logically unavoidable. While it is now commonly accepted that the Earth rotates around its axis and revolves around the sun, which is itself moving around the center of the Milky Way galaxy, and that the Universe as a whole is expanding at what appears to be an ever increasing rate, none of these facts are immediately obvious and relatively few people who believe or accept them would be able to explain how we know them to accurately reflect the way the universe is organized. At the same time, when these ideas were first being developed they conflicted with the idea that the Earth was stationary, which, of course it appears to be, and located at the center of a static Universe, which also seems to be a reasonable presumption. Scientist's new ideas about the Earth's position in the Universe were often seen to pose a threat to the sociopolitical order and a number of people were threatened for holding "heretical" views on the topic. Most famously, these included the mystic Giordano Bruno (1548 –1600), who was burned at the stake for this and other ideas (some of

which are currently proposed by theoretical physicists) and Galileo Galilei (1564–1642), known as the father of modern physics. Interestingly the Roman Catholic Church placed Galileo's book, which proposed that the sun was the center of the solar system, on the list of forbidden books in 1616 and did not remove it until 1835. Galileo was arrested in 1633, tried by the Inquisition, forced to publicly recant his views on the relative position of the Sun and Earth, and spent the rest of his life under house arrest.[16]

The idea of us standing on the Earth which is rotating at ~1000 miles an hour and flying through space at about 67,000 miles per hour is difficult to reconcile with our everyday experience yet science has continued to generated even weirder ideas. Based on observations and logic, it appears that the Universe arose from "nothing" approximately 13.8 billion years ago.[17] Current thinking suggests that it will continue to expand forever at an increasingly rapid rate. Einstein's theory of general relativity implies that matter distorts space-time, which is really one rather than two discrete entities, and that this distortion produces the attraction of gravity

In the world of biology, it appears that all organisms are derived from a single type of ancestral cell that arose from non-living material between 3.5 to 3.8 billion years ago. There is an uninterrupted link between that cell and every cell in your body (and to the cells within every other living organism). You yourself are a staggeringly complex collection of cells. Your brain and its associated sensory organs, which generate consciousness and self-consciousness, contains approximately 86 billion (10^9) neurons as well as an equal number of non-neuronal (glial) cells. These cells are connected to one another through about 1.5 x 10^{14} connections, known as synapses.[18] How exactly such a system produces thoughts, ideas, dreams, feelings, and self-awareness remains quite obscure, but it is clear that these are all emergent behaviors that arise from this staggeringly complex natural system. Scientific ideas arise from the interactions between the physical world, our brains, and the social system of science that tests these ideas based on their ability to explain and predict the behavior of the observable universe.

One of the difficulties in understanding scientific ideas and their implications is that these ideas build upon a wide range of observations and are intertwined with one another. One cannot really understand biological systems without understanding the behavior of systems of chemical reactions, which requires an understanding of molecules, which rests upon an understanding of how atoms and energy behave and interact. To better grasp some of the challenges involved in teaching and learning science, we recommend that you watch a short video interview with the physicist Richard Feynman (1918-1988).[19] In it, he explains the complexity of understanding

[16]The History, Philosophy, and Impact of the Index of Prohibited Books: http://www.unc.edu/~dusto/dusto_prague_paper.pdf

[17] The Origin Of The Universe: From Nothing Everything?: http://www.npr.org/blogs/13.7/2013/03/26/175352714/the-origin-of-the-universe-from-nothing-everything

[18] Are There Really as Many Neurons in the Human Brain as Stars in the Milky Way? http://www.nature.com/scitable/blog/brain-metrics/are_there_really_as_many & http://onlinelibrary.wiley.com/doi/10.1002/cne.21974/abstract

[19] Feynman & magnets: http://www.youtube.com/watch?v=wMFPe-DwULM).

something as superficially (but not really) simple as how two magnets repel or attract one another.

It is our working premise that to understand a topic (or discipline), it is important to know some of the key observations and common rules upon which broader conclusions are based. To test one's understanding, it is necessary for you as a student to be able to approach a biological question, construct plausible claims for how (and why) the system behaves, based on various facts, observations, or explicit presumptions, which logically support your claim. You also need to present your model to others, knowledgeable in the topic, to get their feedback, to answer their questions and address their criticisms and concerns. Sometimes you will be wrong because your knowledge of the facts is incomplete, your understanding or application of general principles is inaccurate, or your logic is faulty. It is important to appreciate that generating coherent scientific explanations and arguments takes time and lots of practice. We hope to help you learn how to do, through useful coaching and practice. In the context of various questions, we (and your fellow students) will attempt to identify where you produce a coherent critique, explanation or prediction, and where you fall short. It will be the ability to produce coherent arguments, explanations, and/or predictions based on observations and concepts correctly applied in the context of modern biology, that we care about and hope to help you master in this course.

Questions to answer and ponder:

- A news story reports that spirit forces influence the weather. Produce a set of questions whose answers would enable you to decide whether the report was scientifically plausible.
- What features would make a scientific model ugly? See http://www.ted.com/talks/view/lang/en//id/194.
- How would you use Occam's razor to distinguish between two equally accurate models?
- Generate a general strategy that will enable you to classify various pronouncements as credible (that is, worth thinking about) or nonsense.
- Does the inability to measure something unambiguously make it unreal? Explain what is real.
- How should we, as a society, deal with the tentative nature of scientific knowledge matter?
- If "science" concludes that free will is an illusion, would you accept it and behave like a robot?

Chapter 2: Life's diversity and origins

In which we consider what biology is all about, namely organisms and their diversity. We discover that organisms are built of one or more, sometimes many cells. We consider the origins of organisms, their basic properties, and their relationships to one another.

Biology is the science of organisms, how they function, behave, interact, and, as populations, have and can evolve. As we will see, organisms are discrete, highly organized, bounded but open, non-equilibrium, physicochemical systems. Now that is a lot of words, so the question is what do they mean? How is a rock different from a mushroom that looks like a rock? What exactly, for example, is a bounded, non-equilibrium system? The answer is not simple, it assumes a knowledge of thermodynamics, a topic that we will address more directly in Chapter 5. For the moment, when we talk about a non-equilibrium system, we mean a system that can do various forms of work. Of course that means we have to define what we mean by work. For simplicity, we will start by defining work as something that takes the input of energy. In the context of biological systems, work involves generating and maintaining molecular gradients, driving unfavorable, that is energy-requiring, reactions, such as the synthesis of various biomolecules including nucleic acids, proteins, lipids, and carbohydrates required for growth and reproduction, and the generation of movement, and so on. Much of this involves the concept of energy, which is itself quite abstract and difficult to master. For our purposes, we will focus on what is known as free energy, which is what enables things to happen. When a system is at equilibrium its free energy is 0, which means that there are no macroscopic (visible) or net changes occurring. The system is essentially static, even though at the molecular level there are still movements due to the presence of heat. Organisms maintain their non-equilibrium state (their free energy is much greater than zero) by importing energy in various forms form the external world. They are different from other such systems in that they contain a genetic (heritable) component. For example, while non-equilibrium systems occur in nature – hurricanes and tornados are non-equilibrium systems – they differ from organisms in that they are transient. They arise *de novo* and when they dissipate they leave no offspring. In contrast, each organism alive today arose from one or more pre-existing organisms (its parent) and each organism, with some special exceptions, has the ability to produce offspring. As we see, the available evidence indicates that each and every organism, past, present, and future, has (or will have) an uninterrupted history stretching back billions of years. This is a remarkable conclusion, given the obvious fragility of life.

Biology has only a few over arching theories. One of these, the Cell Theory of Life, explains the historic continuity of organisms, while the Theory of Evolution by Natural Selection (and other processes), explains how populations of organisms have changed over time. Finally, the Physicochemical Theory of Life explains how it is that organisms can display their remarkable properties without violating the laws that govern physical and chemical systems.

What is life, exactly?

Clearly, if we are going to talk about biology, and organisms and cells and such, we have to define exactly what we mean by life. This raises a problem peculiar to biology as a science. We cannot define life generically because we know of only one type of life. We do not know whether this type of life is the only type of life possible or whether radically different forms of life exist elsewhere in the universe or even on Earth, in as yet to be recognized forms.

While you might think that we know of many different types of life, from mushrooms to whales, from humans to the bacterial communities growing on the surfaces of our teeth (that is what dental plaque is, after all), we will see that the closer we look the more these different "types of life" are in fact simply versions of a common underlying motif, they are one type of life. Based on their common chemistry, molecular composition, cellular structure, and the way that they encode hereditary information in the form of molecules of deoxyribonucleic acid (DNA), all topics we will consider in depth later on, there is little reasonable doubt that all organisms are related, that is they are descended from a common ancestor.

We cannot currently answer the question of whether the origin of life is a simple, likely, and predictable event given the conditions that existed on the Earth when life first arose, or whether it is an extremely rare and unlikely event. In the absence of empirical data, one can question whether scientists are acting scientifically or more as lobbyists for their own pet projects when they talk about doing astrobiology or speculating on when we will discover alien life forms.[20] That said, asking seemingly silly questions, provided that empirically-based answers can be generated, has often been the critical driver of scientific progress. Consider, for example, current searches for life on Earth, almost all of which are based on what we already know about life. Specifically, the methods used rely on the fact that all known organisms use DNA to encode their genetic information; they would not recognize types of life that are dramatically different. In particular, they would not detect organisms that used a different method (not DNA) to encode genetic information. But if we could generate, de novo, living systems in the laboratory we would have a better understanding of what functions are necessary for life and how to look for such "non-standard" organisms in new ways. It might even lead to the discovery of alternative forms of life right here on Earth, assuming they exist.[21] That said, until someone manages to create or identify such non-standard forms of life, it seems quite reasonable to concentrate on the characteristics of life as we know them.

So, let us start again in trying to produce a good definition, or given the fact that we know only of one version of life, a useful description of what we mean by life. First, the core units of life are organisms, which are individual living objects. From a structural and thermodynamic perspective, each organism is a bounded, non-equilibrium system that persists over time and, from a practical point of view, can produce one or more copies of itself. Even though organisms are composed of one or more cells, it is the organism that is the basic unit of life. It is the organism that reproduces new organisms.[22]

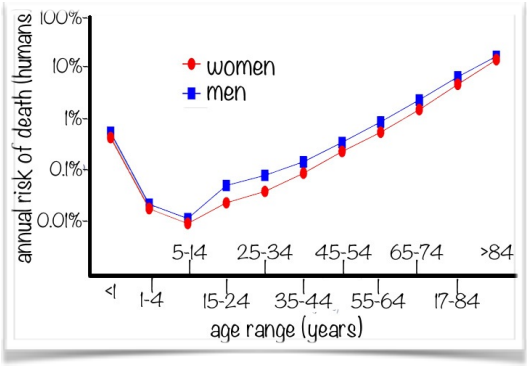
---

[20] The possibility of alternative microbial life on Earth: http://www.ncbi.nlm.nih.gov/pubmed/18053938

[21] Signatures of a shadow biosphere: http://www.ncbi.nlm.nih.gov/pubmed/19292603

[22] In Chapter 4, we will consider how multicellular and social organisms come to be.

Why the requirement for and emphasis on reproduction? This is basically a pragmatic criterion. Assume that a non-reproducing form of life was possible. A system that could not reproduce runs the risk of death (or perhaps better put, dissolution) by accident. Over time, the probability of death for a single individual will approach one, that is certainty.[23] In contrast, a system that can reproduce makes multiple copies of itself and so minimizes, although by no means eliminates, the chance of accidental extinction (the death of all descendants). We see the value of this strategy when we consider the history of life. Even though there have been a number of mass extinction events over the course of life's history, organisms descended from a single common ancestor that appeared billions of years ago continue to survive and flourish.



So what does the open nature of biological systems mean? Basically, organisms are able to import, in a controlled manner, energy and matter from outside themselves, to export waste products into their environment.[24] This implies that there is a distinct boundary between the organism and the rest of the world. All organisms have such a barrier (boundary) layer, as we will see, and the basic barrier appears to be a homologous structure of organisms - that is, it was present in and inherited from the common ancestor. What is important about this barrier is that it is selective, it allows the capture or entry of energy and matter. As we will see, the importation of energy, specifically energy that can be used to drive various cellular processes, is what enables the organism to maintain its non-equilibrium nature and its dynamic structure. The boundary must be able to retain the valuable structures generated, while at the same time allow waste products to leave. This ability to import matter and export waste enables the organism to grow and to reproduce. We assume that you have at least a basic understanding of the laws of thermodynamics, but we will review the basic ideas captured in these laws later, in Chapter 5.

We see evidence of the non-equilibrium nature of organisms most obviously in the ability of organisms to move, but it is important for all aspects of the living state. In particular, organisms use energy, captured from their environment, to drive various chemical reactions and mechanical processes that by themselves are thermodynamically unfavorable. To do this, they use networks of thermodynamically favorable reactions coupled to thermodynamically unfavorable reactions. An organism that reaches thermodynamic or chemical equilibrium is dead.

There are examples of non-living, non-equilibrium systems that can "self-organize" or appear de novo. Hurricanes and tornados form spontaneously and then disperse. They use energy from their environment, which is then dispersed back into their environment (a process associated with increased entropy). They differ from organisms in that they cannot produce offspring - they are the result of specific atmospheric conditions. They are individual entities, unrelated to one another, which do not and cannot evolve. Tornados and hurricanes that formed billions or millions of years ago would (if we could

[23] image modified from "risk of death" graph: http://www.medicine.ox.ac.uk/bandolier/booth/Risk/dyingage.html

[24] In fact, this is how they manage to organize themselves, by exporting entropy. So be careful when people (or companies) claim to have a zero-waste policy, which is an impossibility according to the laws of thermodynamics that all systems obey.

observe them) be similar to those that form today. Since we understand (more or less) the conditions that produce them, we can predict fairly reliably the conditions that will lead to their formation and how they will behave once they form. In contrast, organisms present in the past were different from those that are alive today. The further in the past we go, the more different they appear. Some ancient organisms became extinct, some gave rise to the ancestors of current organisms. In contrast, all tornados and hurricanes originate anew, they are not derived from parental storms.

Question to answer and ponder:

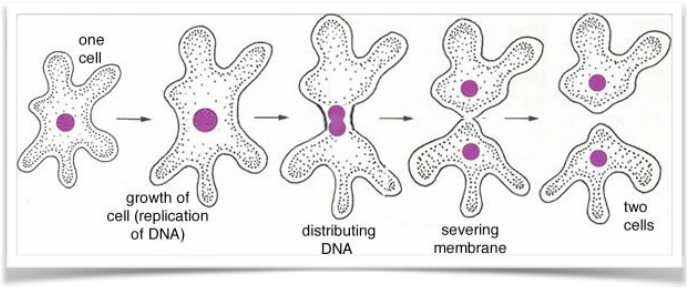
- Using the graph on risk of death as a function of age in humans, provide a plausible model for the shape of the graph.
- Why are the points connected? Wouldn't it make more sense to draw a smooth line between them? which better captures the reality of the situation?
- Extrapolate when the probability of death reaches 1 and explain why it is never 0.
- What factors would influence the shape of the curve? How might the curve differ for different types of organisms?
- Make a model of what properties a biological boundary layer needs to possess. Using your current knowledge, how would you build such a boundary layer?

The cell theory and the continuity of life

Observations using microscopes revealed that all organisms examined contained structurally similar "cells." Based on such observations, a rather sweeping conclusion were drawn by naturalists toward the end of the 1800's. Known as the Cell Theory, it has two parts. The first is that every organism is composed of one or more cells (in some cases billions of cells) and non-cellular products produced by cells, such as bone, hair, scales, and slime. The cells that the Cell Theory deals with are defined as bounded, open, non-equilibrium physicochemical systems (a definition very much like that for life itself). The second is that cells arise only from pre-existing cells. The implication is that organisms (and the cells that they are composed of) arise in this way and no other way. We now know (and will consider in great detail as we proceed) that in addition to their basic non-equilibrium nature, cells also contain a unique material that encodes hereditary information in a physical and relatively stable form, namely molecules of double-stranded deoxyribonucleic acid (DNA). Based on a wide range of data, the Cell Theory implies that all organisms currently in existence (and the cells from which they are composed) are related through an unbroken series of cell division events that stretch back in time. Other studies, based on comparing the information present in DNA molecules, as well as careful comparisons of how cells are constructed, at the molecular level, suggests that there was a single common ancestor that lived between 3.5 to 3.8 billion years ago. This is a remarkable conclusion, given the (apparent) fragility of life - it implies that each cell in your body has a multibillion year old history. What the cell theory does not address is the processes that lead to the origin of the first organisms (cells).

The earliest events in the origin of life, that is, exactly how the first cells originated and what they looked like are unknown, although there is plenty of speculation to go around. Our confusion arises in large measure from the fact that the available evidence indicates that all organisms that have ever lived on Earth share a single common ancestor, and that that ancestor, likely to be a singled-cell organism, was already quite complex. We will discuss how we came to these conclusions, and their

implications, later on in this chapter. One rather weird point to keep in mind is that the "birth" of a new cell involves a continuous process by which one cell becomes two. Each cell is defined, in part, by the presence of a distinct surface barrier, known as the cell or plasma membrane. The new cell is formed when that original membrane pinches off to form two distinct cells (FIG→). The important point is that there is no discontinuity, the new cell does not "spring into life" but rather emerges from the preexisting cell. This continuity of cell from cell extends back in time back billions of years. We often define the start of a new life with the completion of cell division, or in the case of humans and other sexually reproducing multicellular organisms, a fusion event, specifically the merger of an egg cell and a sperm cell. But again there is no discontinuity, both egg cell and sperm cell are derived from other cells and when they fuse, the result is also a cell. In the modern world, all cells, and the organisms they form, emerge from preexisting cells and inherit from those cells both their cellular structure, the basis for the non-equilibrium living system, and their genetic material, their DNA. When we talk about cell or organismic structures, we are in fact talking about information, stored in the structure, information that is lost if the cell/organism dies. The information stored in DNA molecules (known as an organism's genotype) is more stable, it can survive the death of the organism, at least for a while. In fact, information-containing DNA molecules can move between unrelated cells or from the environment into a cell, a process known as horizontal gene transfer (which we will consider in detail toward the end of the book).



The organization of organisms

Some organisms consist of a single cell, others are composed of many cells, often many distinct types of cells. These cells vary in a number of ways and can be extremely specialized (particularly within the context of multicellular organisms), yet they are all clearly related to one another, sharing many molecular and structural details. So why do we consider the organism rather than the cell to be the basic unit of life? The distinction may seem trivial or arbitrary, but it is not. It is a matter of reality versus abstractions. It is organisms, whether single or multicellular, that produce new organisms. As we will discuss in detail when we consider the origins of multicellular organisms, a cell within a multicellular organism normally can neither survive outside the organism nor produce a new organism - it depends upon cooperation with the other cells of the organism to reproduce. In fact, each multicellular organism is an example of a cooperative, highly integrated social system. The cells of a typical multicellular organism are part of a social system in which most cells have given up their ability to reproduce a new organism; their future depends upon the reproductive success of the organism of which they are a part. It is the organism's success in generating new organisms that underlie evolution's selective mechanisms. Within the organism, the cells that give rise to the next generation of organism are known as germ cells, those that do not (and die with the organism) are known as somatic cells.[25] All organisms in the modern world, and for apparently the last ~3.5 billion years, arise from a pre-existing organism or,

[25] If we use words that we do not define and that you do not understand, look them up!

in the case of sexually reproducing organisms, from the cooperation of two organisms, another example of social evolution which we will consider in greater detail in Chapter 4. We will also see that breakdowns in such social systems can lead to the death of the organism or disruption of the social system. Cancer is the most obvious example of an anti-social and evolutionarily short-sighted behavior of cells within a multicellular organism.

Spontaneous generation and the origin of life

The ubiquity of organisms raises obvious questions: how did life start and what led to all these different types of organisms? At one point, people believed that these two questions had a single answer, but we now recognize that they are really two quite distinct questions and their answers involve distinct mechanisms. An early commonly held view (by those who thought about such things) was that supernatural processes produced life in general and human beings in particular. The articulation of the Cell Theory and the Theory of Evolution by Natural Selection, which we will discuss in detail in the next chapter, concluded quite persuasively that life had a single successful origin and that various natural evolutionary processes generated the diversity of life.

But how did life itself originate? It used to be widely accepted that various types of organisms, such as flies, frogs, and even mice, could arise spontaneously, from non-living matter.[26] Flies, for example, were thought to appear from rotting flesh and mice from wheat. If true, on-going spontaneous generation would have profound implications for our understanding of biological systems. For example, if spontaneous generation based on natural processes was common, there must be a rather simple process at work, a process that (presumably) can produce remarkably complex outcomes (all bets are off if the process is supernatural). Also, if each organism arose independently, we might expect that the molecular level details of each would be unique, since they presumably arose independently from different stuff and under different conditions compared to other organisms of the same type. However, we know this is not the case, since all organisms are clearly related and can be traced back to a single ancestor (a conclusion to which we return, repeatedly.)

A key event in the conceptual development of modern biology was the publication of Francesco Redi's (1626 –1697) paper entitled "Experiments on the Generation of Insects" in 1668. He hypothesized that spontaneous generation did not occur. His hypothesis was that the organisms that appeared had developed from "seeds" deposited by adults. His hypothesis led to a number of clear predictions. One was that if adult flies were kept away from rotting meat, for example, maggots (the larval form of flies) would never appear no matter how long one waited. Similarly, the type of organism that appeared would depend not on the type of rotting meat, but rather on the type of adult fly that had access to the meat. To test his hypothesis Redi set up two sets of flasks - both contained meat. One set of flasks were exposed directly to the air and so to flies, the other was sealed with paper or

He who experiments increases knowledge. He who merely speculates piles error upon error.
- Arabic epigraph quoted by Francisco Redi.

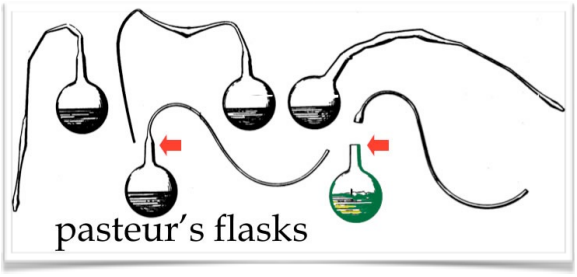
[26]Farley, J., The spontaneous generation controversy (1700-1860): The origin of parasitic worms. J. Hist. Biol., 1972. **5**: 95-125 (http://link.springer.com/article/10.1007%2FBF02113487) and The spontaneous generation controversy (1859-1880): British and German reactions to the problem of abiogenesis. J. Hist. Biol., 1972. 5: 285-319 (http://www.jstor.org/stable/4330578)

cloth. Maggots appeared only in the flasks open to the air. Redi concluded that organisms as complex as insects (and too large to pass through the cloth) could arise only from other insects, or rather eggs laid by those insects – that life was continuous.

The invention of the light microscope and its use to look at biological materials by Antony van Leeuwenhoek (1632-1723) and Robert Hooke (1635-1703) led to the discovery of a completely new and totally unexpected world of microbes or microscopic organisms. We now know these as the bacteria, archaea, protozoa, unicellular algae, and microscopic fungi, such as yeasts. Although it was relatively easy to generate compelling evidence that macroscopic (that is, big) organisms, such as flies, mice, and people could not arise spontaneously, it seemed plausible that microscopic and presumably much simpler organisms could form spontaneously.

The discovery of microbes led a number of scientists to explore their origin and reproduction. Lazzaro Spallazani (1729-1799) showed that after a broth was boiled it remained sterile (that is, without life) as long as it was isolated from contact with fresh air. He concluded that microbes, like larger organisms, could not arise spontaneously but were descended from other microbes, many of which were floating in the air. Think about possible criticisms to this experiment – perhaps you can come up with ones that we do not mention!

One criticism was that it could be that boiling the broth destroyed one or more key components that were necessary for the spontaneous formation of life. Alternatively, perhaps fresh air was the "vital" ingredient. In either case, boiling and isolation would have produced an artifact that obscured rather than revealed the true process. In 1862 (note the late date, this was after Charles Darwin had published *On the Origin of Species* in 1859), Louis Pasteur (1822-1895) carried out a particularly convincing set of experiments to address both of these concerns. He sterilized broths by boiling them in special "swan-necked" flasks. What was unique about his experimental design was the shape of the flask neck; it allowed air but not airborne microorganisms to reach the broth. Microbes in the air were trapped in the bended region of the flask'a neck. This design enabled Pasteur to address a criticism of previous experiments, namely that access to air was necessary for spontaneous generation to occur. He found that the liquid, even with access to air, remained sterile for months. However, when the neck of the flask was broken the broth was quickly overrun with microbial growth. He interpreted this observation to indicate that air, by itself, was not necessary for spontaneous generation, but rather was normally contaminated by microbes. On the other hand, the fact that the broth could support microbial growth after the neck was broken indicated that the heating of the broth had not destroyed some vital element needed for spontaneous generation or standard growth to occur. In the language of modern scientific experimentation, breaking the flask served as a positive control – it showed that the boiled media could have supported spontaneous generation if such a process were possible. Of course, not all (in fact, probably not any) experiment is perfect. For example, how would one argue against the objection that the process of spontaneous generation normally takes tens to thousands of years to occur? If true, this would invalidate Pasteur's conclusion. Clearly an experiment to address that possibility has its own practical issues. Nevertheless, the results of various experiments on spontaneous generation led to the



conclusion that neither microscopic nor macroscopic organisms could arise spontaneously, at least not in the modern world. The problem, at least in this form, became uninteresting to working scientists.

Does this mean that the origin of life is a supernatural event? Not necessarily. Consider the fact that living systems are complex chemical reaction networks. In the modern world, there are many organisms around who are actively eating complex molecules to maintain their non-equilibrium state and to grow and reproduce. If life were to arise by a spontaneous but natural process, it is possible that it could take thousands to hundreds of millions of years to occur. We can put some limits on the maximum time it could take from geological data using the time when the Earth's surface solidified from its early molten state to the first fossil evidence for life (about 100 to 500 million years). Given the tendency of organisms to eat one another, one might argue (as did Darwin) that once organisms had appeared in a particular environment they would have suppress any subsequent spontaneous generation events - they would have eaten the molecules needed for the process. But, as we will see, evolutionary processes have led to the presence of organisms essentially everywhere on Earth that life can survive - there are basically no welcoming and sterile places left within the modern world. Here we see the importance of history. According to the current scientific view, life could arise *de novo* only in the absence of life; once life had arisen, the conditions had changed. The presence of life is expected to suppress the origin of new forms of life.

It is often said that all the conditions for the first production of living organisms are now present. But if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc. present, that a proteine compound was formed, ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed.
- Charles Darwin (1887).

The death of vitalism

Naturalists originally thought that life itself was a type of supernatural process, too complex to obey or be understood through the laws of chemistry and physics.[27] In this vitalistic view, organisms were thought to obey different laws from those acting in the non-living world. For example, it was assumed that molecules found only in living organisms, and therefore known as organic molecules, could not be synthesized outside of an organism; they had to be made by a living organism. In 1828, Friedrich Wöhler (1800 –1882) challenged this view by synthesizing urea in the laboratory. Urea is a simple organic molecule, $O=C(NH_2)_2$ found naturally in the waste derived from living organisms. Urine contains lots of urea. Wöhler's *in vitro* or "in glass" (as opposed to *in vivo* or in life) synthesis of urea was simple. In an attempt to synthesize ammonium cyanate (NH_4NCO), he mixed the inorganic compounds ammonium chloride (NH_4Cl) and silver cyanate (AgNCO). Analysis of the product of this reaction revealed the presence of urea. What actually happened was this reaction:

$$AgNCO + NH_4Cl \rightarrow NH_4NCO + AgCl \rightarrow O=C(NH_2)_2 + AgCl.$$

Please do not memorize the reaction, what is of importance here is to recognize that this is just another chemical reaction, not exactly what the reaction is.

[27] In a sense this is true since many physicists at least do not seem to understand biology.

While simple, the *in vitro* synthesis of urea had a profound impact on the way scientists viewed so called organic processes. It suggested that there was nothing supernatural involved in the synthesis of urea; it obeyed the laws of chemistry. Based on this and similar observations on the *in vitro* synthesis of other, more complex organic compounds, we (that is, scientists) are now comfortable with the idea that all molecules found within cells can, in theory at least, be synthesized outside of cells, using appropriate procedures. Organic chemistry has been transformed from the study of molecules found in organisms to the study of molecules containing carbon atoms, although a huge amount of time and effort is now devoted to the industrial synthesis of a broad range of organic molecules.

Questions to answer & to ponder:

- General a scheme that you could use to determine whether something was living or not.
- Why does the continuity of cytoplasm from generation to generation matter? What (exactly) is transferred?
- Why did the discovery of bacteria reopen the debate on spontaneous generation?
- How is the idea of vitalism similar to and different from intelligent design creationism?
- Is spontaneous generation unscientific? Explain your answer.

Thinking about life's origins

There are at least three possible approaches to the study of life's origins. A religious (i.e. non-scientific) approach would likely postulate that life was created by a supernatural being. Different religious traditions differ as to the details of this event, but since the process is supernatural it cannot, by definition, be studied scientifically. Nevertheless, intelligent design creationists often claim that we can identify those aspects of life that could not possibly have been produced by natural processes, by which they mean various evolutionary and molecular mechanisms, which we will discuss in the next chapter. It is important to consider whether these claims would, if true, force us to abandon a scientific approach to the world around us in general, and the origin and evolution of life in particular. Given the previously noted interconnectedness of the sciences, one might well ask whether a supernatural biology would not also call into question the validity of all scientific disciplines. For example the dating of fossils is based on geological and astrophysical (cosmological) evidence for the age of the Earth and the Universe, which themselves are based on physical and chemical observations and principles. A non-scientific biology would be incompatible with a scientific physics and chemistry. The lesson of history, however, is different. Predictions as to what is beyond the ability of science to explain have routinely been demonstrated by scientists to be wrong, often only a few years after such predictions were made!

Another type of explanation for the appearance of life on Earth, termed panspermia, assumes that advanced aliens brought (or left) life on Earth. Perhaps we owe our origins to casually discarded litter from these alien visitors. Unfortunately, the principles of general relativity, one of the best confirmed of all scientific theories, limit the speed of travel and given the size of the Universe, travelers from beyond the solar system seem unlikely, if not totally impossible. Moreover panspermia simply postpones but does not answer the question of how life began. Our alien visitors must have come from somewhere and panspermia does not explain where they came from. Given our current models for the

history of the Universe and the Earth, understanding the origin of alien life is really no simpler than understanding the origin of life on Earth. On the other hand, if there is life on other planets and moons in our solar system, and we retrieve and analyze it, it would be extremely informative, particularly if it could be shown that it originated independently rather than being splashed from the Earth through various astronomical impact events.[28]

Experimental studies on the origins of life

One strategy to understanding how life might have arisen involves experiments to generate plausible precursors of living systems in the laboratory. The experimental studies carried out by Stanley Miller (1930-2007) and Harold Urey (1893-1981) were early and influential example of this approach.[29] These two scientists made an educated, although now apparently incorrect, guess as to the composition of Earth's early atmosphere. They assumed the presence of oceans and lightning. They set up an apparatus to mimic these conditions and then passed electrical sparks through their experimental atmosphere. After days they found that a complex mix of compounds had formed. Included in this mix were many of the amino acids found in modern organisms, as well as lots of other organic molecules. Similar experiments have been repeated with combinations of compounds more likely to represent the environment of early Earth, with similar results: various biologically important organic molecules accumulate rapidly.[30] Quite complex organic molecules have been detected in interstellar dust clouds, and certain types of meteorites have been found to contain complex organic molecules. During the period of the heavy bombardment of Earth, between about 4.1 and 3.9 billion years ago, meteorite impacts could have supplied substantial amounts of organic molecules.[31] It therefore appears likely that early Earth was rich in organic molecules, the building blocks of life.

Given that the potential building blocks were present, the question becomes what set of conditions were necessary and what steps led to the formation of the first living systems? Assuming that these early systems were relatively simple compared to modern organisms (or the common ancestor of life for that matter), we hypothesize that the earliest proto-biotic systems were molecular communities of chemical reactions isolated in some way from the rest of the outside" world. This isolation or selective boundary was necessary to keep the system from dissolving away or dissipating. One possible model is that such systems were originally tightly associated with the surface of specific minerals and that these mineral surfaces served as catalysts, speeding up important reactions (we will return to the role of catalysts in biological systems later on). Over time, these pre-living systems acquired more sophisticated boundary structures (membranes) and were able to exist free of the mineral surface, perhaps taking small pieces of the mineral with them.

The generation of an isolated but open system, which we might call a protocell was a critical step in the origin of life. Such an isolated system has important properties that are likely to have

---

[28] Top 5 Bets for Extraterrestrial Life in the Solar System: http://www.wired.com/wiredscience/2009/01/et-life/

[29] The Miller-Urey experiment:http://www.ucsd.tv/miller-urey/ and http://en.wikipedia.org/wiki/Miller–Urey_experiment

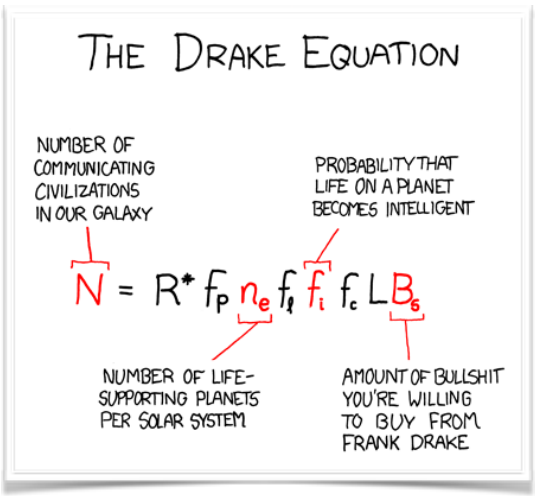
[30] A reassessment of prebiotic organic synthesis in neutral planetary atmospheres: http://www.ncbi.nlm.nih.gov/pubmed/18204914

[31] A time-line of life's evolution: http://exploringorigins.org/timeline.html

facilitated the further development of life. For example, because of the membrane boundary, changes that occur within one such structure will not be shared with neighboring systems. Rather, they can accumulate and favor the survival of one system over its neighbors. Such systems can also reproduce in a crude way by fragmentation. If changes within one such system improved its stability, its ability to accumulate resources, or its ability to survive and reproduce, that system, and its progeny, would be likely to become more common. As these changes accumulate and are passed from parent to offspring, the organisms will inevitably evolve (as we will see in detail in the next chapter.)

Questions to answer & to ponder:

- If we assume that spontaneous generation occurred in the distant past, why is it not occurring today? How could you tell if it were?
- In 1961, Frank Drake, a radio astronomer, proposed an equation to estimate the number of technological civilizations that exist within the observable Universe (N).[32] The equation is $N = R^* \times f_p \times n_e \times f_l \times f_i \times f_c \times L$ where



R^*=The rate of formation of stars suitable for the development of intelligent life.
f_o = The fraction of those stars with planetary systems.
n_e = Tthe number planets, per solar system, with an environment suitable for life.
f_l = The fraction of suitable plants on which life actually appears.
fi = The fraction of life-bearing planets on which intelligent life emerges.
f_c = The fraction of civilization that develop a technology that releases detectable signs of their existence into space.
L = The length of time such civilizations release detectable signals into space.
Identify those parts of the Drake equation that can be established (at present) empirically and that cannot, and explain your reasoning.

Mapping the history of life on earth

Assuming that life arose spontaneously on early Earth, we can now look at what we know about the history of Earth and the fossil record to better understand the appearance and diversification of life. This is probably best done by starting with what we know about where the Universe and Earth came from. The current scientific model for the origin of the universe is known as the Big Bang. It arose from efforts to answer the question of whether the fuzzy nebulae identified by astronomers were located within or outside of our galaxy. This required some way to determine how far these nebula were from Earth. Edwin Hubble (1889-1953) and his co-workers were the first to realize that nebula were in fact galaxies in their own right, each very much like our own Milky Way and each composed of many billions of stars. This was a surprising result, since it made Earth, sitting on the edge of one among many, many galaxies seem less important. It is a change in cosmological perspective similar to that associated with the idea that the sun, rather than Earth, was the center of the solar system (and the Universe).

[32] The Drake equation: http://www.seti.org/drakeequation and cartoon: http://xkcd.com/384/

To measure the movement of galaxies with respect to Earth Hubble and colleagues used the Doppler shift, which is the effect on the wavelength of sound or light by an object's velocity relative to an observer. In the case of light emitted from an object moving toward the observer, the wavelength will be shortened, that is, shifted to the blue end of the spectrum. Light emitted from an object moving away from the observer will be lengthened, that is, shifted to the red end of the spectrum. Based on the observed Doppler shifts in the wavelengths of light coming from stars in galaxies and the observation that the further a galaxy appears to be from Earth, the greater that shift is toward the red, Hubble concluded that galaxies, outside of our local group, were all moving away from one another. Running time backward, he concluded that at one point in the past, all of the matter and energy in the universe must have been concentrated in a single point. A prediction of this Big Bang model is that the Universe is estimated to be ~13.8 +/- 0.2 billion (10^9) years old. This is a length of time well beyond human comprehension; it is sometimes referred to as deep time - you can get some perspective on deep time using the Here is Today website (http://hereistoday.com). Other types of data have been used to estimate the age of Earth and the other planets in the solar system as ~4.5 x 10^9 years.

After Earth first formed, a heavy bombardment of extraterrestrial materials, such as comets and asteroids, collided with it. This bombardment began to subside around 3.8 to 3.9 billion years ago and reached its current level by about 3.5 billion years ago.[33] It is not clear whether life arose multiple times and was repeatedly destroyed during the early history of Earth (4.5 to 3.6 billion years ago) or if the origin of life was a one-time event, taking hundreds of millions of years before it succeeded, which then managed to survive and expand around 3.8 to 3.5 billion years ago.

Fossils evidence for the history of life on earth

The earliest period in Earth's history is known as the Hadean, after Hades, the Greek god of the dead. The Hadean is defined as the period between the origin of Earth up to the first appearance of life. Fossils provide our only direct evidence for when life appeared on Earth. They are found in sedimentary rock, that is rock formed when fine particles of mud, sand, or dust entombed an organism before it can be eaten by other organisms. Hunters of fossils (paleontologists) do not search for fossils randomly but use geological information to identify outcroppings of sedimentary rocks of the specific age they are studying in order to direct their explorations.

Early in the history of geology, and before Darwin proposed the modern theory of evolution, geologists was recognized that fossils of specific types were associated with rocks of specific ages. This correlation was so robust that rocks could be accurately dated based on the types of fossils they contained without exception. At the same time, particularly in a world that contains young earth creationists who claim that Earth was formed less than 10,000 years ago, it is worth remembering both the interconnectedness of the sciences and that geologists do not rely solely on fossils to date rocks. This is in part because many types of rocks do not contain fossils. The non-fossil approach to dating rocks is based on the physics of isotopes and the chemistry of atomic interactions. It uses the radioactive decay of elements with isotopes with long half-lives, such as ^{235}Ur which decays into ^{207}Pb with a half-life of ~704 million years and ^{238}Ur which decays into ^{206}Pb with a half life of ~4.47 billion

[33] The violent environment of the origin of life:http://www.sciencedirect.com/science/article/pii/0016703793905436

years. Since these two Pb isotopes appear to be formed only through the decay of Ur, the ratios of Ur and Pb isotopes can be used to estimate the age of the rock.

To use isotope abundance to date rocks, it is critical that all of the atoms in a mineral measured stay there, that none wash in or away. Since Ur and Pb have different chemical properties, this can be a problem in some types of minerals. That said, with care, and using rocks that contain chemically inert minerals, like zircons, this method can be used to measure the age of rocks to an accuracy of within 1% or better. These and other types of evidence support James Hutton's (1726-1797) famous dictum that Earth is ancient, with "no vestige of a beginning, no prospect of an end."[34] We know now, however, that this statement is not accurate; while very very old, Earth coalesced around 5 billion years ago and will disappear when the sun expands and engulfs it in about 5.5 billion years from now.[35]

But, back to fossils. There are many types of fossils. Chemical fossils are molecules that, as far as we know, are naturally produced only through biological processes.[36] Their presence in ancient rock implies that living organisms were present at the time the rock formed. These first appear in rocks that are between 3.8 to 3.5 x 10^9 years old. What makes chemical fossils problematic is that there may be non-biological but currently undiscovered or unrecognized mechanisms that could have produced them, so we have to be cautious in our conclusions.

Moving from the molecular to the physical, are trace fossils. These can be subtle or obvious. Organisms can settle on mud or sand and make impressions. Burrowing and slithering animals make tunnels or disrupt surface layers. Leaves and immotile organisms can leave impressions. Walking animals can leave footprints in sand, mud, or ash. How does this occur? If the ground is covered, compressed, and converted to rock, these various types of impressions can become fossils. Later erosion can then reveal these fossils. For example, if you live near Morrison, Colorado, you can visit the rock outcrop known as Dinosaur Ridge and see trace fossil dinosaur footprints; there may be similar examples near where you live.

We can learn a lot from trace fossils, impressions can reveal the general shape of an organism or its ability to move or to move in a particular way. To move, it must have some kind of muscle or alternative mobility system and probably some kind of nervous system that can integrate information and produce coordinated movements. Movement also suggests that the organisms that made the trace had something like a head and a tail. Tunneling organisms are likely to have had a month to ingest sediment, much like today's earthworms - they were predators, eating the microbe they found in mud.

In addition to trace fossils, there are also the type of fossils that most people think about, which are known as structural fossils, namely the mineralized remains of the hard parts of organisms such as teeth, scales, shells, or bones. As organisms developed hard parts, fossilization, particularly of organisms living in environments where they could be buried within sediment before being dismembered and destroyed by predators or microbes, became more likely.

---

[34] http://www.talkorigins.org/faqs/geohist.html

[35] http://www.youtube.com/watch?v=iauIP8swfBY

[36] Although as Wohler pointed out, they can be generated in the laboratory.

Unfortunately for us (as scientists), many and perhaps most types of organisms leave no trace when they die, in part because they live in places where fossilization is rare or impossible. Animals that live in wood lands, for example, rarely leave fossils. The absence of fossils for a particular type of organisms does not imply that these types of organisms do not have a long history; rather it means that the conditions where they lived and died or their body structure is not conducive to fossilization. Many types of living organisms have no fossil record at all, even though, as we will see, there is molecular evidence that they arose tens to hundreds of millions of years ago.

Life's impact on the earth

Based on fossil evidence, the current model for life on Earth is that for a period of ~2 x 10^9 (billion) years the only forms of life on Earth were microscopic. While the exact nature of these organisms remains unclear, it seems likely that they were closely related to prokaryotes, that is, bacteria and archaea. While the earliest organisms probably used chemical energy, relatively soon organisms appeared that could capture the energy in light and use it to drive various thermodynamically unfavorable reactions. A major class of such reactions involves combining CO_2 (carbon dioxide), H_2O (water), and other small molecules to form carbohydrates (sugars), and other important biological molecules such as lipids, proteins, and nucleic acids. At some point during the early history of life on Earth, organisms appeared that released molecular oxygen (O_2) as a waste product of such light-driven reactions, known generically as oxygenic photosynthesis. These oxygen-releasing organisms became so numerous that they began to change Earth's surface chemistry - they represent the first life-driven ecological catastrophe.

The level of atmospheric O_2 represents a balance between its production, primarily by organisms carrying out oxygenic photosynthesis, and its removal through various chemical reactions. Early on as O_2 appeared, it reacted with iron to form deposits of water insoluble Fe (III) oxide - that is, rust. This rust reaction removed large amounts of O_2 from the atmosphere, keeping its levels low. The rusting of iron in the oceans is thought to be largely responsible for the massive banded iron deposits found around the world.[37] O_2 also reacts with organic matter, as in the burning of wood, so when large amounts of organic matter are buried before they can react, as occurs with the formation of coal, more O_2 accumulates in the atmosphere. Although it was probably being generated and released earlier, by ~2 billion years ago, atmospheric O_2 had appeared in detectable amounts, and by ~850 million years ago it had risen to significant levels. Atmospheric O_2 levels have changed significantly since then, based on the relative rates of its synthesis and destruction. Around 300 million years ago, atmospheric O_2 levels had reached ~35%, almost twice the current level. It has been suggested that it was these high levels of atmospheric O_2 that made possible the evolution of giant insects.[38]

Although we tend to think of O_2 as a natural and benign substance, it is in fact a highly reactive and potentially toxic compound and its appearance posed challenges and provided opportunities to

[37] Paleoecological Significance of the Banded Iron-Formation: http://econgeol.geoscienceworld.org/content/68/7/1135.abstract

[38] see Atmospheric oxygen, giant Paleozoic insects and the evolution of aerial locomotor performance: http://jeb.biologists.org/content/201/8/1043.full.pdf

many organisms. As we will see later on O_2 can be "detoxified" through reactions that lead to the formation of water and this type of reaction appears to have been co-opted for other purposes. For example, through coupled reactions O_2 can be used to capture the maximum amount of energy from food, leading to the generation of CO_2 and H_2O, both of which are very stable.

Around the time that O_2 levels were first rising, that is about 10^9 years ago, the first trace fossil burrows appear in the fossil record. These were likely to have been produced by simple worm-like, macroscopic multicellular organisms, known as metazoans, capable of moving along and through the mud on the ocean floor. About 0.6 x 10^9 years ago, new, more complex structural fossils begin to appear in the fossil record. Since the fossil record does not contain all types of organisms, we are left to speculate on what the earliest metazoans looked like. The first of these are the so-called Ediacaran organisms, named after the geological formation in which their fossils were first found.[39] Current hypotheses suggest they were immotile, like modern sponges but flatter and it remains unclear how they are related to later organisms. By the beginning of the Cambrian age (~545 x 10^6 years ago), a wide variety of organisms had appeared within the fossil record, many clearly related to modern organisms. Molecular level data suggest that their ancestors originated more than 30 million years earlier. These Cambrian organisms show a range of body types. Most significantly, many were armored. Since building armor involves expending energy to synthesize these components, the presence of armor suggests a need for armor, that is organisms gained something valuable from its presence. A plausible suggestion is that the appearance of armor was linked to the appearance of predators.

Viruses: Now, before we leave this chapter you might well ask, haven't we forgotten viruses? Well, no - viruses are often an important component of an ecosystem and an organism's susceptibility or resistance to viral infection is often an important evolutionary factor, but viruses are different from organisms in that they are non-metabolic. That means they do not carry out reactions and cannot replicate on their own, they can replicated only within a living cell. Basically they are not alive, so even though they are extremely important, we will discuss viruses only occasionally and in quite specific contexts.

Questions to answer & to ponder

- What factors would influence the probability that a particular organism, or type of organism, would be fossilized?
- What did Wöhler's synthesis of urea and the Miller/Urey experiment actually prove and what did they imply?
- Why can't we be sure about the stages that led to the origin of life?
- Can the origin of life be studied scientifically, and if so, how?
- What factors could drive the appearance of teeth, bones, shells, muscles, nervous systems, and eyes?
- What factors determine atmospheric O_2 levels?

[39] http://en.wikipedia.org/wiki/Ediacara_biota

Chapter 3: Evolutionary mechanisms and the diversity of life

In which we consider the rather exuberant diversity of organisms and introduce the primary evolutionary mechanisms responsible for it.

In medieval Europe there was a tradition of books known as bestiaries. These were illustrated catalogs of known and imagined organisms in which it was common for particular organisms to be associated with moral lessons. "Male lions were seen as worthy reflections of God the Father, for example, while the dragon was understood as a representative of Satan on earth."[40] One can see these books as an early version of a natural theology, that is, an attempt to gain an understanding of the supernatural through lessons from and studies of natural objects. In this case, the presumption was that each type of organism was created for a particular purpose, and that often this purpose was to provide people with a moral lesson. This way of thinking grew more and more problematic as more and more different types of organisms were recognized, many of which had no obvious significance to humans. Currently, scientists have identified approximately 1,500,000 different species of plants, animals, and microbes. The actual number of different types of organisms, referred to as species, may be as high as 10,000,000.[41] These numbers refer, of course, to the species that currently exist, but we know from the fossil record that many distinct species, which are now extinct, have existed in the past. So the obvious question is, why are there so many different types of organisms?[42] Do they represent multiple independent creation events, and if so, how many such events have occurred?

Catalogued and predicted species

doi:10.1371/journal.pbio.1001127.t002

Species	Earth	Ocean
	Catalogued	Catalogued
Eukaryotes		
Animalia	953,434	171,082
Chromista	13,033	4,859
Fungi	43,271	1,097
Plantae	215,644	8,600
Protozoa	8,118	8,118
Total	1,233,500	193,756
Prokaryotes		
Archaea	502	1
Bacteria	10,358	652
Total	10,860	653
Grand Total	**1,244,360**	**194,409**
Predicted	**8,750,000**	**2,210,000**

As the true diversity of organisms was discovered, a number of observations served to undermine the early concept that organisms were created to serve humanity. The first were the number of organisms that had very little obvious importance to the human condition. This was particularly obvious in the case of extinct organisms but extended further as a result of newly discovered organisms. At the same time students of nature, known generically as naturalists, discovered many different types of upsetting and cruel behaviors within the natural world. Consider the fungus *Ophiocordyceps unilateralis*, which infects the ant *Camponotus leonardi*. The fungus takes control of the ant's behavior, causing them to migrate to positions that favor fungal growth before killing the infected ant. Similarly, the nematode worm *Myrmeconema neotropicum* infects the ant *Cephalotes*

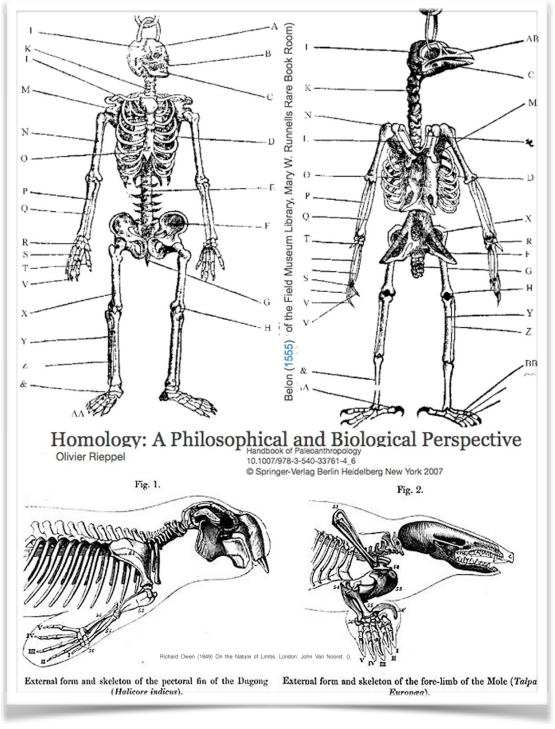
[40] http://www.getty.edu/art/gettyguide/artObjectDetails?artobj=304109

[41] How many species are there on Earth and in the ocean? http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001127

[42] As a technical point, which we will return to, we will refer to each distinct type of organism as a species.

atratus. This leads to dramatic changes in the morphology and behavior of the ant. The ant's abdomen turns red and is held up-raised, which makes the infected ant resemble a fruit and so increases the likelihood of it being eaten by birds. The birds transport the worms, which survive in their digestive systems until they are excreted and subsequently are eaten by ants to complete the worm's life cycle.[43] Perhaps the most famous example of this type of behavior are the wasps of the family *Ichneumonidae.* Female wasps deposit their fertilized eggs into the bodies of various types of caterpillars, where the eggs hatch out and produce larvae that feed on the caterpillar, keeping it alive while they eat it from the inside out. Charles Darwin remarked in a letter to Asa Gray, an American naturalist, *"There seems to me too much misery in the world. I cannot persuade myself that a beneficent & omnipotent God would have designedly created the Ichneumonidae with the express intention of their feeding within the living bodies of caterpillars, or that a cat should play with mice."* Rather than presume that a supernatural creator was responsible for such gratuitously (or at least apparently) cruel behaviors, Darwin and others sought alternative, morally neutral naturalistic processes that could generate biological diversity and explain biological behaviors.

As the diversity of organisms became increasingly apparent and difficult to ignore, another broad and inescapable conclusion began to emerge from anatomical studies of organisms, many different organisms displayed remarkable structural similarities. For example, as naturalists characterized various types of animals, they found that they either had an internal skeleton (the vertebrates) or did not (the invertebrates). Comparative studies among the vertebrates revealed that there were often striking similarities between quite different types of organisms. A classic work, published in 1555, compared the skeletons of a human and a bird.[44] While many bones have changed shape and relative sizes, what was most striking is how many bones are at least superficially similar between the two. This same type of "comparative anatomy" revealed many similarities between disparate organisms. For example, the skeleton of the dugong (a large aquatic mammal) appears quite similar to that of the european mole, which tunnels underground on land. In fact, there are general skeletal similarities between all vertebrates. The closer we look, the more similarities we find. These similarities run deeper than the anatomical, they extend to the cellular and the molecular as well. So the scientific question is, what explains such similarities? Why build an organism that walks, runs, and climbs, such



[43] The Life of a Dead Ant: The Expression of an Adaptive Extended Phenotype: http://www.jstor.org/stable/10.1086/603640

[44] Belon P (1555) L'Histoire de la Nature des Oyseaux. Paris, Guillaume Cavellat

as humans, with a skeleton similar to that of a organism that flies (birds), swims (dugongs), or tunnels (moles). Are these anatomical similarities just flukes or do they imply something deeper?

Organizing organisms (hierarchically)

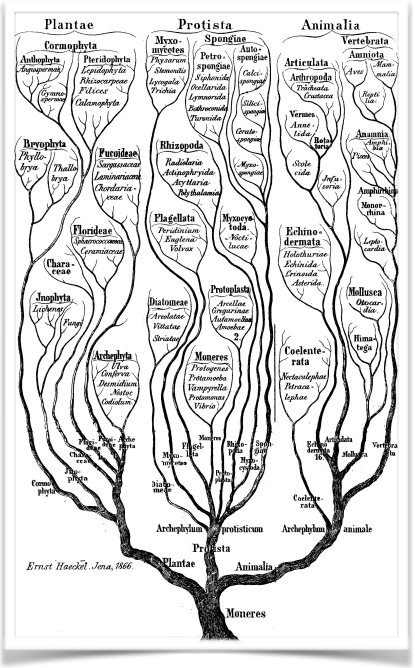
Carl Linnaeus (1707-1778) was the pioneer in taking the similarities between different types of organisms seriously. Based on similarities (and differences), he developed a system to classify organisms in a coherent and hierarchical manner. Each organism had a unique place in this scheme. What was, and occasionally still is, the controversial aspect of such a classification system is in deciding which traits should be considered significant and which are superficial or unimportant, at least for the purposes of classification. Linnaeus had no real theory to explain why organisms could be classified in such a hierarchical manner and could only go on observations. This might be a good place to reconsider the importance of hypotheses, models, and theories in biology. Linnaeus noticed the apparent similarities between organisms and used it to generate his classification scheme, but he had no explanatory model for why such similarities should exist (very much like Newton's law of gravitation did not explain why there was gravity). So what are the features of an explanatory model? Such a model has to go beyond just explaining, it also has to suggest observations or predict outcomes that have not yet been observed. It is these validity of these predictions that enable us to distinguish between different explanatory models. A model that makes no validated predictions is not particularly useful. A model that makes explicit predictions, even if they prove to be wrong, enables us to refine our model or force us to abandon the model and develop a new one. A model that, through its various hypotheses and their confirmation or refutation or revision, has been found to accurately explain a particular phenomena can become promoted to a theory. So this enables us to distinguish between a law and a theory. A law describes what we see but not why we see it. A theory provides the explanation for observable phenomena.[45]

Back to Linnaeus, whose classification system placed organisms of a particular type were placed together into a species. Of course, what originally counted as a discrete type of organism was based on Linnaeus's judgement as an observer and classifier; what particularly traits he felt defined the species and distinguished it from other, similar species. The choice of these key traits was subject to debate. Based on the percieved importance and presence of particular traits, organisms could be split into two or more types (species), or two types originally considered separate species could be reclassified into a single type.

As we will see, the individual organisms that make a up a species are not identical but share many traits. In organisms that reproduce sexually, there are often dramatic differences between males and females of the same species, a situation known as sexual dimorphism. In some cases, these differences can be so dramatic that without further evidence, it can be difficult to tell whether two animals are members of the same or different species. In this light the primary criteria for determining whether sexually reproducing organisms are members of the same or different species is whether they can and do successfully interbreed with one another. This criteria, that is reproductive compatibility, can be used to place species distinctions on a more empirical basis, but it cannot be used with asexual

[45] If we go back, Newton's law of gravity explained how objects behaved gravitationally, but it not why. In contrast, Einstein's theory of general relativity explained why there was gravity, and predicted behaviors that were not predicted by Newton's law.

species (such as most microbes). Within a species, there are sometimes regional differences that are distinct enough to be recognizable. Where this is the case, these groups are known as populations, races, or subspecies. While distinguishable, the organism in these groups retain the ability to interbreed and so are members of a single species.



After defining types of species, Linnaeus next grouped species that displayed similar traits into a larger group, known as a genus. While a species can be considered a natural, interbreeding population, a genus is a more artificial group. Which species are placed together within a particular genus depends on the common traits deemed important or significant by the person doing the classification. This can lead to conflicts which are generally resolved by the collection of more and more comparative data. In the Linnaean classification scheme, each organism has a unique name, which consists of its genus and species names. The accepted usage is to write out the name in italics with the genus name capitalized, for example, *Homo sapiens*. Following on this pattern, one or more genuses are placed into larger, more inclusive groups, and these groups, in turn, are themselves placed in larger groups. The end result of this process is the rather surprising observation that all organisms fall into a small number of "supergroups" or phyla. We will not worry about the traditional group names, because in most cases they really do not help in our understanding of basic biology. Perhaps most surprising of all, all organisms and all phyla fall into one and only one family - all of the organisms on earth can be placed into a single unified phylogenetic "tree" or perhaps better put, bush. That this should be the case is by no means obvious. This type of analysis could have produced multiple, distinct classification schemes, but it did not.

It is worth reiterating the fact that while a species can be seen as a natural group, the higher levels of classification are based on various hypotheses, specifically that certain traits are more important or informative than others. For example, having hair, four legs, and teeth is not enough to determine unambiguously whether an organism is in the genus *Canis*, which includes wolves and coyotes, or the genus *Vulpes*, which includes foxes. This is a choice based on various lines of evidence, but nothing as distinct as whether foxes normally mate with coyotes (they do not). Because genus and more inclusive group classifications are based on arguments about the significance of various shared traits. Where scientists place a species can change. New observations can lead to the reorganization of the classification scheme, a species or a genus can be moved from one place to another, or a larger group can be divided into two or more new groups. For example consider the types of organisms commonly known as bears. There are a number of different types of bear-like organisms, a fact that Linnaeus's classification scheme explicitly acknowledged even though it never attempted to explain why. Looking at all bear-like organisms we can recognize eight types.[46] We currently consider four of these, the brown bear (*Ursus arctos*), the Asiatic black bear (*Ursus thibetanus*), the American

[46] http://en.wikipedia.org/wiki/List_of_bears

bear (*Ursus americanus*), and the polar bear (*Ursus maritimus*) to be significantly more similar to one another, based on the presence of various traits, than they are to other types of bears. We therefore placed them in their own genus, *Ursus*. We have placed each of the other bears, the spectacled bear (*Tremarctos ornatus*), the sloth bear (*Melurus ursinus*), the sun bear (*Helarctos mayalanus*), and the giant panda (*Ailuropoda melanoleuca*) in their own separate genuses. Scientists considere these species more different from one another than are the members of the genus *Ursus*. That said, all of these bears clearly share a number of other traits, so we place them all in the larger group, the family Ursidae to reflect their undeniable similarities. Scientists originally considered the red panda (*Ailurus fulgens*) to be a bear, but it has now been moved into a distinct group, the Ailuridae. Both the Ursidae and the Ailuridda are part of a larger and more diverse group, the Carnivora, which includes cats, dogs, wolverines, and their relatives. A key for placing these species together is that they are all placental mammals. There are other bear-like organisms that are not bears or even members of the Carnivora group. Both the koala (*Phascolarctos cinereus*) and the extinct giant marsupial bears of the genus Proborhyaenid are marsupial mammals; their offspring are born relatively undeveloped and mature in a pouch on the mother. All marsupial mammals are more similar to one another in key ways than they are to *any* placental mammal. We consider placental and marsupial traits more significant, from a classification perspective, than the bear-like traits these organisms share. That said, both true (placental) bears and marsupial bears are placed in the larger group known as Mammalia, which includes monotreme (egg-laying), marsupial, and placental mammals. We group mammals together in part because they feed their young using a common substance, milk, secreted by the mammary glands of their mothers. We place mammals together with reptiles, birds, and fish into an even larger group known as the Chordates based on the presence of an internal skeleton and more specifically a backbone, and from there into larger and even more inclusive groups.

What is most significant for our purposes is *not* the particular place that an organism occupies within the classification system but rather the fact that we can place all organisms in a logical and self-consistent manner within such a system. As we will discover later on, the use of gene (DNA) sequencing methods has provided further support for this classification scheme, removing ambiguities,and supporting its underlying logic. As we gather more and more data, we find that Linnaeus was correct. These is an unambiguous hierarchical relationship between organisms.

Fossils and the Linnaean system

As mentioned previously, we continue to discover new fossils and new organisms.[47] In most cases, these fossils appear to represent organisms that lived many millions to hundreds of millions of years ago but which are now extinct. Clearly there are dramatic differences between the ability of different types of organisms to become fossilized. Perhaps the easiest to fossilize are those organisms with internal or external skeletons, yet it is estimated that between 85 to 97% of such organisms are not represented in the fossil record and various studies indicate that many other types of organisms have left no fossils whatsoever.[48] Some authors have estimated that the number of organisms at the genus

[47] Your inner fish: http://www.pbs.org/your-inner-fish/home/

[48] The incompleteness of the fossil record: http://www.donaldprothero.com/files/47440594.pdf

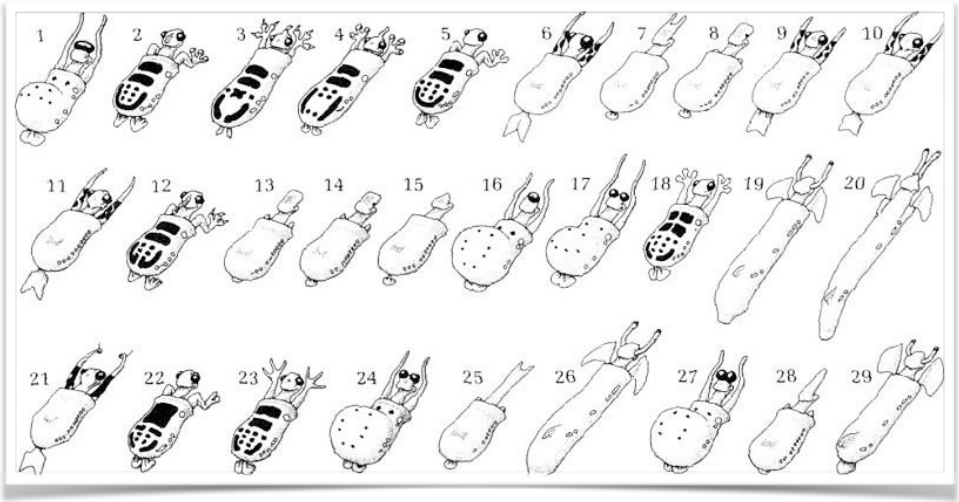
level that have been preserved as fossils may be less (often much less) than 5%.[49] For some categories of organisms, such as the wide range of microbes, essentially no informative fossils exist.

Once scientists recognized that fossils provide evidence for extinct organisms, the obvious question was do extinct organisms fit into the Linnaean classification scheme or do they form their own groups or their own separate trees? This can be a difficult question to answer, since many fossils are only fragments of the intact organism. The fragmentary nature of the fossil record can lead to ambiguities. Nevertheless, the conclusion that has emerged upon careful characterization is that we can place almost all fossilized organisms within the modern Linnaean classification scheme. There are possible exception, like the Ediacarian organisms that lived very long ago and appear structurally distinct from known living organisms. The presumption, however, is that if we had samples of these organisms for molecular analyses, we would find it that they too would fall nicely into the same classification scheme as all other organisms do.[50] For example, dinosaurs, along with modern birds, are clearly descended from a specific type of reptile, while living mammals are more closely related to a second, now extinct group, known as the "mammal-like reptiles."

In rare cases, particularly relevant to human evolution, one trait that can be recovered from bones is DNA sequence data. For example, it has been possible to extract and analyze DNA from the bones of Neanderthals and Denisovian-type humanoid organisms, that went extinct about 30,000 years ago, and to use that information to clarify their relationship to modern humans (*Homo sapiens*).[51] This type of data provides evidence for interbreeding and has led to the argument for the reclassification of Neanderthals and Denisovians as subspecies of *Homo sapiens.*

Questions to answer and ponder:

- Explain why you might expect that extinct species fit into the Linnaean classification scheme.
- What would make you decide that a particular trait was important or unimportant (secondary) from a classification perspective?
- Given the following imaginary animals →, place them in a plausible classification system and explain your reasoning.
- How could Neanderthals be a distinct species if evidence for in-breeding with H. sapiens exists?



[49] Absolute measures of the completeness of the fossil record: http://www.ncbi.nlm.nih.gov/pubmed/11536900

[50] On the eve of animal radiation: phylogeny, ecology and evolution of the Ediacara biota: http://users.unimi.it/paleomag/geo2/Xiao&Leflamme2008.pdf

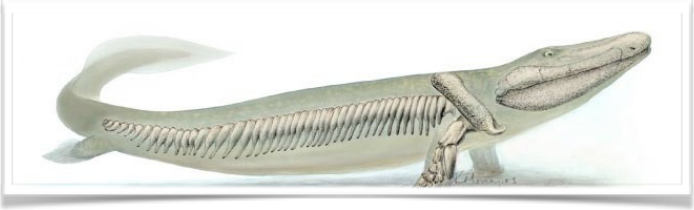
[51] Paleogenomics of archaic hominins:http://www.ncbi.nlm.nih.gov/pubmed/22192823

The theory of evolution and the organization of life

Perhaps surprisingly, Linnaeus never proposed a plausible (or even an implausible) naturalistic explanation for why organisms should be classifiable in a hierarchical way. Why is it that birds, whales, and humans share common features, such as the organization of their skeletons, that led Linnaeus to classify them together as vertebrates? Why are there extinct organisms, known from their fossils, that share these common features, even though they are otherwise quite different? We had to wait about 100 years for a plausible model that explained why the Linnaean classification scheme actually works and can be used it to make predictions about organisms that no longer exist. Charles Darwin (1809-1882) and Alfred Wallace (1823–1913) proposed such a model, described in great detail in Darwin's book *The Theory of Evolution by Natural Selection*, originally published in 1858.

The main unifying idea in biology is Darwin's theory of evolution through natural selection.
– John Maynard Smith

As we will see, evolutionary theory is based on a series of direct observations of the natural world and their logical implications. Evolutionary theory explains why similar organisms share similar traits and why we can place them easily into a hierarchical classification system. They are similar because they are related to one another – they share common ancestors. Moreover, we can infer that the more different two organisms are, the longer ago this common ancestor lived. We can even begin to make plausible and empirically-supportable deductions about what those common ancestors looked like. As an example, we can predict that the common ancestor of all terrestrial vertebrates will resemble a fish with leg-like limbs. Scientists have recently discovered fossils of such an organism, *Tiktaalik*.[52] This is just one more example of the fact that since its original introduction, and well before the mechanisms of heredity and any understanding of the molecular nature of organisms were resolved, evolutionary theory explained what was observed and made testable predictions about what would be found.

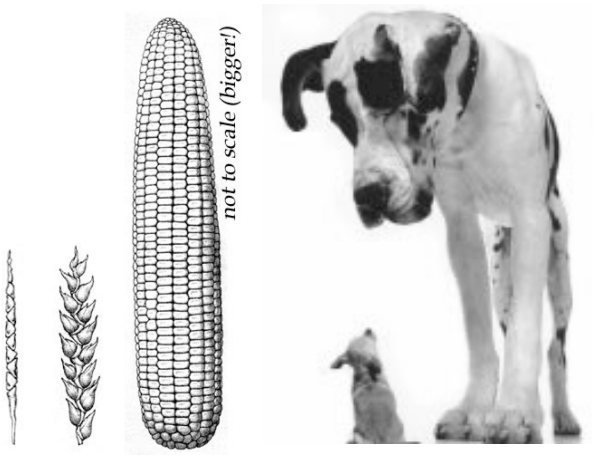


Tiktaalik roseae, an extinct fish-like organism that lived ~ 375 million years ago, is likely to be similar to the common ancestor of all terrestrial vertebrates.

So what are the facts and inferences upon which the Theory of Evolution is based? Two foundational observations are deeply interrelated and based on empirical observations associated with plant and animal breeding and the observed behaviors of natural populations. The first is the fact that whatever type of organism we examine, if we look carefully enough, making accurate measurements of visible and behavioral traits (this description of the organism is known as its **phenotype,** we find that individuals vary with respect to one another. More to the point, plant and animal breeders recognized that the offspring of a controlled mating between individuals often had phenotypes similar to those of their parents. Certain phenotypic traits can be inherited. Over many generations, domestic animal and plant breeders used what is now known as artificial selection to generate the range of domesticated plants and animals with highly exaggerated phenotypes that we now rely on (see picture on next page). For example, beginning about 10,000 years ago plant breeders in Mesoamerica developed modern

[52] Meet Tiktaalik roseae: An Extraordinary Fossil Fish: http://tiktaalik.uchicago.edu/meetTik.html

corn (maize) by the selective breeding of variants of the grass teosinte.[53] All of the various breeds of dogs, from the tiny to the rather gigantic, appear to be derived from a common ancestor that lived between 19,000 to 32,000 years ago (although as always, be skeptical; it could be that exactly where and when this common ancestor lived could be revised).[54] In all cases, the crafting of specific domesticated organisms followed the same pattern. Organisms with desirable traits (phenotypes) were selected for breeding with one another. Organisms that did not have these traits were discarded and not permitted to breed. This process, carried out over hundreds to thousands of generations, led to organisms that displayed distinct or exaggerated forms of the selected trait. What is crucial to understand is that this strategy could work only if different versions of the trait were present in the original selected population and at least a part of this phenotypic variation was due to genetic, that is inheritable, factors. What these inheritable factors were was completely unclear, but we can refer to it as the organism's **genotype** (even though plant and animal breeders would never have used that term).



This implies that different organisms have different genotypes, but where those differences come from was completely unclear to early plant and animal breeders. Were they imprinted on the organism in some way based on its experiences or induced by environmental factors? Was the genotype stable or could it be modified by experience? How were genotypic factors passed from generation to generation? And how, exactly, did a particular genotype produce or influence a specific phenotypic trait. As we will see, at least superficially, this last question remains poorly resolved for many phenotypes.

So what do we mean by genetic factors?

Here the answer is empirical. Traditional plant and animal breeders had come to recognize that offspring tended to display the same or similar traits as their parents. This observation led them to assume that there was some factor within the parents that was expressed within the offspring and could, in turn, be passed from the offspring to their own offspring. A classic example is the Hapsburg lip, which was passed through a European ruling family for generations.[55] In the case of artificial selection, an important point to keep in mind is that the various types of domesticated organisms that are produced are often dependent for their continued existence on their human creators. This relieves them from the

[53] Molecular Evidence and the Evolution of Maize: http://link.springer.com/article/10.1007/BF02860472

[54] From wild animals to domestic pets, an evolutionary view of domestication: http://www.pnas.org/content/106/Supplement_1/9971.full

[55] 'Imperial Stigmata!' The Habsburg Lip, A Grotesque 'Mark' Of Royalty Through The Centuries!: http://theesotericcuriosa.blogspot.com/2012/09/imperial-stigmata-habsburg-lip.html

constraints they would experience in the wild. Because of this dependence, artificial selection can produce quite exaggerated and, in the absence of human intervention, highly deleterious traits. Just look at domesticated chickens and turkeys which, while not completely flightless, can fly only very short distances and so are extremely vulnerable to predators. Neither modern corn (*Zea mays*) or chihuahuas, on of the smallest breeds of dog, also developed by Mesoamerican breeders, would be expected to survive for long in the wild, that is, without human assistance.[56]

Limits on populations

It is a given (that is, an empirically demonstrable fact) that all organisms are capable of producing many more than one copy of themselves. Consider, as an example, a breeding pair of elephants or a single asexually reproducing bacterium. Let us further assume that there are no limits to their reproduction. That is, that once born, the offspring will live a normal life-span and themselves reproduce. By the end of 500 years, a single pair of elephants could have produced ~15,000,000 living descendants.[57] Clearly if these 15,000,000 elephants then paired up to form 7,500,000 breeding pairs, within another 500 years (1000 years altogether) there would be $7.5 \times 10^6 \times 1.5 \times 10^7$ or 1.125×10^{14} elephants. Assuming that each adult elephant weighs ~6000 kilograms, which is the average between larger males and smaller females, the end result would be $\sim 6.75 \times 10^{18}$ kilograms of elephant. Allowed to continue unchecked, within a few thousand years a single pair of elephants could produce a mass of elephants larger than the mass of the Earth, an absurd conclusion. Clearly we must have left something out of our calculations! As another example, let us turn to a solitary bacterium, which needs no mate to reproduce. Let us assume that this is a photosynthetic bacterium that relies on sunlight and simple compounds, such as water, carbon dioxide, and some minerals, to grow. A bacterium is much smaller than an elephant but it can produce new bacteria at a much faster rate. Under optimal conditions, it could divide once every 20 minutes or so and would, within approximately a day, produce a mass of bacteria greater than that of Earth as a whole. Again, we are clearly making a number at least one mistake in our logic.

A single cell of the bacterium E. coli would, under ideal circumstances, divide every twenty minutes. That is not particularly disturbing until you think about it, but the fact is that bacteria multiply geometrically: one becomes two, two become four, four become eight, and so on. In this way it can be shown that in a single day, one cell of E. coli could produce a super-colony equal in size and weight to the entire planet Earth.
- Michael Crichton (1969) The Andromeda Strain

Elephants and bacteria are not the only types of organism on the Earth. In fact every known type of organism can produce many more offspring than are needed to replace themselves when they die. This trait is known as superfecundity. But unlimited growth does not and cannot happen for very long - other factors must constrain it. In fact, if you were to monitor the populations of most organisms, you would find that the numbers of a particular organism in a particular environment tend to fluctuate around a so-called steady state level. By steady state we mean that even though animals are continually being born and are dying, the number of organisms remains roughly constant.

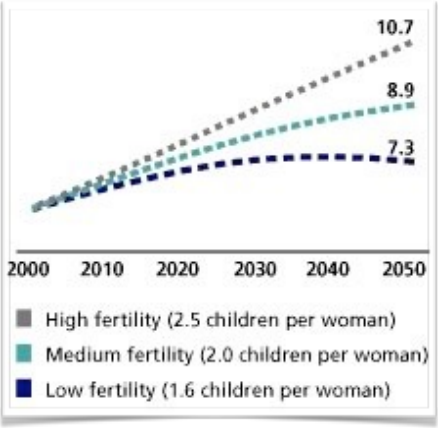
---

[56] How DNA sequence divides chihuahua and great dane: http://www.theguardian.com/science/2007/apr/06/uknews.sciencenews

[57] Darwin's elephants: http://www.idlex.freeserve.co.uk/idle/evolution/sex/elephant.html

So what balances the effects of superfecundity, what limits population growth? The obvious answer to this question is the fact that the resources needed for growth are limited and there are limited places for organisms to live. Thomas Malthus (1766-1834) was the first to clearly articulate the role of limited resources as a constraint on population. His was a purely logical argument. Competition between increasing numbers of organisms for a limited supply of resources would necessarily limit the number of organisms. Malthus painted a rather gloomy picture of organisms struggling with one another for access to these resources, with many living in an organismal version of poverty, starving to death because they could not out-compete others for the food or spaces they needed to thrive. One point that Malthus ignored, or more likely was ignorant of, is that organisms rarely behave in this way. It is common to find various types of behaviors that limit the direct struggle for resources. For example, in some organisms, an adult has to establish (and defend) a particular territory before it can successfully reproduce.[58] The end result of this type of behavior is to stabilize the population around a steady state level, which is a function of both environmental and behavioral constraints.

An organism's environment includes all factors that influence the organism and by which the organism influences other organisms and their environments. These include factors such as changes in climate, as well as changes in the presence or absence of other organisms. For example, if one organism depends in important ways upon another, the extinction of the first will necessarily influence the survival of the second.[59] Similarly, the introduction of a new type of organism or a new trait (think oxygenic photosynthesis) in an established environment can disrupt existing interactions and conditions. When the environment changes, the existing steady state population level may be unsustainable or many of the different types of organisms present may not be viable. If the climate gets drier or wetter, colder or hotter, if yearly temperatures reach greater extremes, or if new organisms (including new disease-causing pathogens) enter an area, the average population density may change or in some cases, if the environmental change is drastic enough, may even drop to zero, that is, certain populations could go extinct. Environmental conditions and changes will influence the sustainable steady state population level of an organism (something to think about in the context of global warming, whatever the cause).



An immediate example of this type of behavior involves the human population. Once constrained by disease, war, and periodic famine, human population increased dramatically with the introduction of better public health and sanitation measures, a more secure food supply, and reductions in infant mortality. Now, in many countries, populations appear to be heading to a new steady state, although

[58] Territorial Defense, Territory Size, and Population Regulation: https://iriss.stanford.edu/sites/all/files/shared/documents/Lopez-Sepulcre2005.pdf

[59] Why the Avocado Should Have Gone the Way of the Dodo http://www.smithsonianmag.com/arts-culture/why-the-avocado-should-have-gone-the-way-of-the-dodo-4976527/?no-ist and Neotropical Anachronisms: The Fruits the Gomphotheres Ate: http://www.sciencemag.org/content/215/4528/19.short

exactly what that final population total level will be is unclear.[60] Various models have been developed based on different levels of average fertility. In a number of countries, the birth rate has already fallen into the low fertility domain, although that is no guarantee that it will stay there![61] In this domain (ignoring immigration), a country's population actually decreases over time, since the number of children born is not equal to the number of people dying. This can generate its own social stresses. Decreases in birth rate per woman correlate with reductions in infant mortality (generally due to vaccination, improved nutrition, and hygiene) and increases in the educational level and the reproductive "self-determination" (that is, the emancipation) of women. Where women have the right to control their reproductive behavior, the birth rate tends to be lower. Clearly changes in the environment, and here we include the sociopolitical environment, can dramatically influence behavior and serve to limit reproduction and population levels.

The conceptual leap made by Darwin and Wallace

What Darwin and Wallace recognized were the implications and significance of these key facts: the hereditable nature of variation between organisms, the ability of organisms to reproduce many more offspring than are needed to replace themselves, and the constraints on population size due to limited environmental resources. Based on these facts, they drew a logical implication, namely that individuals would differ in their reproductive success – that is, different individuals would leave behind different number of descendants. Over time, we would expect that the phenotypic variations associated with greater reproductive success (and the genotypes associated with them) will increase in frequency within the population; they would replace those organisms with a less reproductively successful phenotype. Darwin termed this process natural selection, in analogy to the effects of artificial selection by plant and animal breeders. As we will see, natural selection is one of the major drivers of biological evolution.

Just to be clear, however, reproductive success is more, and more subtle, than survival of the fittest. First and foremost, from the perspective of future generations, surviving alone does not matter much if the organism fails to produce offspring. An organism's impact on future generations will depend not on how long it lives but on how many fertile offspring it generates. An organism that can produce many reproductively successful offspring at an early age will have more of an impact on subsequent generations than an organism that lives an extremely long time but has few offspring. Again, there is a subtle point here. It is not simply the number of offspring that matter but the relative number of reproductively successful offspring produced.

If we think about the factors that influence reproductive success, we can classify them into a number of distinct types. For example, organisms that reproduce sexually need access to mates, and must be able to deal successfully with the stresses associated with normal existence and reproduction. This includes the ability to obtain adequate nutrition and to avoid death from predators and pathogens. These are all parts of the organism's phenotype, which is what natural selection acts on. It is worth remembering, however, that not all traits are independent of one another. Often the mechanism (and

---

[60] Global population growth: https://www.ted.com/talks/hans_rosling_on_global_population_growth and The Joy of Stats: http://youtu.be/jbkSRLYSojo

[61] Hans Rosling: Religions and babies: http://www.youtube.com/watch?v=ezVk1ahRF78

genotype) involved in producing one trait also influences other traits – they are interdependent. There are also non-genetic sources of variation. For example, there are molecular level fluctuations that occur at the cellular level; these can lead genotypically identical cells to display different behaviors, that is, different phenotypes. Environmental factors can influence the growth, health, and behavior of organisms. These are generally termed physiological adaptations. An organism's genotype influences how it responds phenotypically to environmental factors, so the relationship between phenotype, genotype, and the organism's environment is complex.

Mutations and the origins of genotype-based variation

So now the question arises, what is the origin of genetic – that is inheritable-variation? How do genotypes change? As a simple (and not completely incorrect) analogy, we can think of an organism's genotype as a book. This book is also known as its **genome** (not to worry if this seems too simple, we will add needed complexities as we go along). An organism's genome is no ordinary book. For simplicity we can think of it as a single unbroken string of characters. In humans, this string is approximately 3.2 billion characters (or letters) long (~3,200,000,000). In case you are wondering, a character corresponds to a base pair, which we will consider in detail in Chapter 7. Within this string there are regions of what look like words and sentences, that is, regions that look like they have meaning. There are also long regions that appear to be meaningless. To continue our analogy, a few critical changes to the words in a sentence can change the meaning of a story, sometimes subtly, sometimes dramatically, and sometimes a change will lead to a story that makes no sense at all.

At this point we will define the meaningful regions (the words and sentences) to correspond to **genes** and the other intervening sequences as **intragenic** regions, that is, spaces between genes. We estimate that humans have approximately 25,000 genes (we will return to a molecular level discussion of genes and how they work in Chapters 7 through 9). As we continue to learn more about the molecular biology of organisms, our understanding of both genes and intragenic regions becomes increasingly sophisticated. The end result is that regions that appear meaningless can influence the meaning of the genome. Many regions of the genome are unique, they occur only once within the string of characters. Others are repeated, sometimes hundreds to thousands of times. When we compare the genotypes of individuals of the same type of organism, we find that they differ at a number of places. For example, we have found over 55,000,000 variations between human genomes and more are likely to be identified. When present within a population of organisms, these genotypic differences are known as **polymorphisms**, from the Latin meaning multiple forms. Polymorphisms are the basis for DNA-based forensic identification tests. One thing to note, however, is that only a small number of these variations are present within any one individual, and considering the size of the human genome, most people differ from one another less than 1 to 4 letters out of every 1000. That amounts to between 3 to 12 million letter differences between two unrelated individuals. Most of these differences are single characters, but there can be changes that involve moving regions from one place to another, or the deletion or duplication of a region. In sexually reproducing organisms, like humans, there are two copies of this book in each cell of the body, one derived from each of the organism's parents - organisms with two genomic "books" are known as **diploid**. When a sexual organism reproduces, it produces reproductive cells, known as sperm or eggs. Since each of these cells contains one copy of its own unique version of the genomic book, it is said to be **haploid**. This haploid genome is produced

through a complex process (known as meiosis) that leads to the significant shuffling between the organism's original parental genomes. The end result is that each new organism contains its own unique genomic book (or books). When the haploid sperm and haploid egg cells fuse a new and unique (diploid) organism is formed with its own unique pair of genomic books.

The origins of polymorphisms

So what produces the genomic variation between individuals found within current populations? Are these processes still continuing or have they ended? First, as we have alluded to (and will return to again and again), the sequence of letters in an organism's genome corresponds to the sequence of characters in DNA molecules. A DNA molecule in water (and over 70% of a typical cell is water) is thermodynamically unstable and can undergo various types of reactions that lead to changes in the sequences of characters within the molecule.[62] In addition, we are continually bombarded by radiation that can damage DNA (although not to worry, the radiation associated with cell phones, bluetooth, and wifi is too low in energy to damage DNA). Mutagenic radiation, that is, the types of radiation capable of damaging the genome, comes from various sources, including cosmic rays that originate from outside of the solar system, UV light from the sun, the decay of naturally occurring radioactive isotopes found in rocks and soil, including radon, and the ingestion of naturally occurring isotopes, such as potassium 40. DNA molecules can absorb such radiation, which can lead to chemical changes (mutations). Many but not all of these changes can be identified and repaired by cellular systems, which we will consider later in the book.

The second, and major source of change to the genome involves the process of DNA replication. DNA replication happens every time a cell divides and is remarkably accurate but it is not perfect. Copying creates mistakes. In humans, it appears that replication creates one error for every 100,000,000 (10^8) characters copied. A proof-reading error repair system corrects ~99% of these errors, leading to an overall error rate during replication of 1 in 10^{10} bases replicated. Since a single human cell contains about 6,400,000,000 (> 6 billion) bases of DNA sequence, that means that less than one new mutation is introduced per cell division cycle. Given the number of generations from fertilized egg to sexually active adult, that corresponds to 100-200 new mutations (changes) added to an individual's genome per generation.[63] These mutations can have a wide range of effects, complicated by the fact that essentially all of the various aspects of an organism's phenotype are determined by the action of hundreds to thousands of genes working in a complex network. And here we introduce our last new terms for a while; when a mutation leads to change in a gene, it creates a new version of that gene, which is known as an **allele** of the gene. When a mutation changes the DNA's sequence, whether or not it is part of a gene, it creates what is known as a **sequence polymorphism** (a different DNA sequence). Once an allele or polymorphism has been generated, it is stable - it can be inherited from a parent and passed on to an offspring. Through the various processes associated with reproduction (which we will consider in detail later on), each organism carries its own distinctive set of alleles and its own unique set of polymorphisms. Taken together these genotypic

---

[62] Instability and decay of the primary structure of DNA: http://www.nature.com/nature/journal/v362/n6422/pdf/362709a0.pdf and DNA has a 521-year half-life: http://www.nature.com/news/dna-has-a-521-year-half-life-1.11555

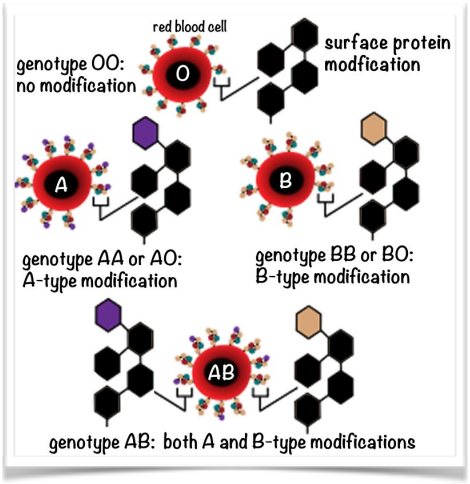
[63] Human mutation rate revealed: http://www.nature.com/news/2009/090827/full/news.2009.864.html

differences (different alleles and different polymorphisms) produce different phenotypes. The DNA tests used to determine paternity and forensic identity work because they identify the unique polymorphisms (and alleles) present within an individual's genome. We will return to and hopefully further clarify the significance of alleles and polymorphisms when we consider DNA in greater detail later on in this book.

Two points are worth noting about genomic changes or mutations. First, whether produced by mistakes in replication or chemical or photochemical reactions, it appears that these changes occur randomly within the genome. With a few notable and highly specific exceptions there are no known mechanisms by which the environment (or the organism) can influence where a mutation occurs. The second point is that a mutation may or may not influence an organism's phenotype. The effects of a mutation will depend on a number of factors, including exactly where the mutation is in the genome, its specific nature, the role of the mutated gene within the organism, the rest of the genome (the organism's genotype), and the environment in which the organism finds itself.

A short aside on the genotype-phenotype relationship

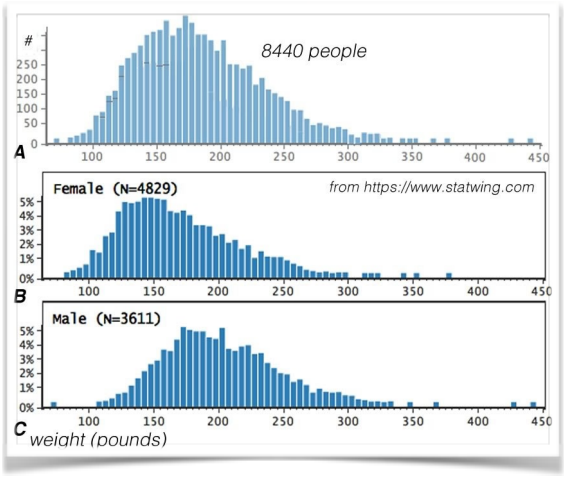
When we think about polymorphisms and alleles, it is tempting to assume simple relationships. In some ways, this is a residue from the way you may have been introduced to genetics in the past.[64] Perhaps you already know about Mendel and his peas. He identified distinct alleles of particular genes that were responsible for distinct phenotypes; yellow versus green peas, wrinkled versus smooth peas, tall versus short plants, etc. Other common examples might be the alleles associated with sickle cell anemia (and increased resistance to malarial infection) and the major blood types. Which alleles of the ABO gene you inherited determines whether you have O, A, B or AB blood type. Remember you are diploid, so you have two copies of each gene, including the ABO gene, in your genome, one inherited from your mom and one from your dad. There are a number of common alleles of the ABO gene present in the human population, the most common (by far) are the A, B, and O alleles. The two ABO alleles you inherited from your parents may be the same or different. If they are A and B, you have the AB blood type; if A and O or A and A, you have the A blood type, if B and O or B and B, you have the B blood type, or if you have O and O, you have the O blood type. These are examples of discrete traits; you are either A, B, AB, or O blood type – there are no intermediates. You cannot be 90% A and 10% B.[65] As we will see, this situation occurs when a particular gene determines the trait; in the case of the ABO gene, the nature of the gene product determines the modification of surface proteins on red blood cells. The O allele leads to no modification, the A allele leads to an A-type modification, while the B allele leads to a B-type modification. When A and B alleles are present, both types of modifications occur. However, most traits do not behave in such a simple way.



[64] We call this type of thinking didaskalogenic: http://en.wikipedia.org/wiki/Didaskalogenic

[65] Human blood types have deep evolutionary roots: https://www.sciencenews.org/article/human-blood-types-have-deep-evolutionary-roots

The vast majority of traits, however, are continuous rather than discrete. For example, people come in a continuous range of heights, rather than in discrete sizes. If we look at the values of the trait within a population, that is, if we can associate a discrete number to the trait (which one cannot always do), we find that each population can be characterized by a distribution. For example, let us consider the distributions of weights in a group of 8440 adults in the USA (see →). The top panel (A) presents a graph of the weights (along the horizontal or x-axis) versus the number of people with that weight (along the vertical or y-axis). We can define the "mean" or average of the population ($\bar{x}$) as the sum of the individual values of a trait (in this case each person's weight) divided by the number of individuals measured, as defined by the equation:



$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

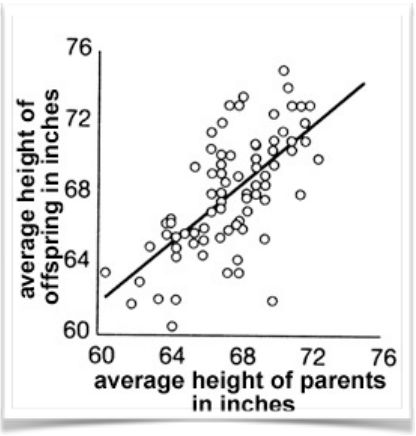
In this case, the mean weight of the population is 180 pounds. It is common to recognize another characteristic of the population, namely the median. The median is the point at which half of the individuals have a smaller value of the trait and half have a larger value. In this case, the median is 176. Because the mean does not equal the median, we say that the distribution is asymmetric, that is there are more people who are heavier than the mean value compared to those who are lighter. For the moment we will ignore this asymmetry, particularly since it is not dramatic. Another way to characterize the shape of the distribution is by what is known as its standard deviation (σ). There are different versions of the standard deviation that reflect the shape of the population distribution, but for our purposes we will take a simple one, the so-called uncorrected sample standard deviation.[66] To calculate this value, you subtract the mean value for the population ($\bar{x}$) from the value for each individual (x_i); since x_i can be larger or smaller than the mean, this difference can be a positive or a negative number. We then take the square of the difference which makes all values positive (hopefully this makes sense to you). We sum these squared differences together, divide that sum by the number of individuals in the population (N), and take the square root (which reverses the effects of our squaring x_i) to arrive at the standard deviation of the population. The smaller the standard deviation, the narrower the distribution - the more organisms in the population have a value similar to the mean. The larger is σ, the greater is the extent of the variation in the trait.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

So how do we determine whether a particularly complex trait like weight (or any other non-discrete, continuously varying trait) is genetically determined? We could imagine, for example, that an organism's weight is simply a matter of how easy it was for it to get food. The standard approach is to ask whether there is a correlation between the phenotypes of the parents and the phenotypes of the offspring. That such a correlation between parents and offspring exists for height is suggested by the graph on the next page. Such a correlation serves as evidence that height (or any other quantifiable trait) is at least to some extent genetically determined. What we cannot determine from such a

[66] http://en.wikipedia.org/wiki/Standard_deviation http://www.mathsisfun.com/data/standard-deviation.html

relationship, however, is how many genes are involved in the genetic determination of height or how their effects are influenced by the environment and environmental history which the offspring experience. For example, "human height has been increasing since at least the 19th century when comprehensive records first began. The mean height of Dutchmen, for example, has increased from 165cm in 1860 to a current 184cm. The spectacular rise in height probably reflects improvements in health care and diet", rather than changes in genes.[67] Geneticists currently estimate that allelic differences at more than 50 genes make significant contributions to the determination of height, with alleles at hundreds more having smaller effects that contribute to differences in height.[68] At the same time, specific alleles of certain genes can lead to extreme shortness or tallness. For example, mutations that inactivate or over-activate genes encoding factors required for growth can lead to dwarfism or giantism.



On a related didaskalogenic note, you may remember learning that alleles are often described as dominant or recessive. But the extent to which an allele is dominant or recessive is not necessarily absolute, it depends upon how well we define a particular trait and whether it can be influenced by other factors and other genes. These effects reveal themselves through the fact that people carrying the same alleles of a particular gene can display (or not display) the associated trait, which is known as its penetrance, and they can vary in the strength of the trait, which is known as its expressivity. Both the penetrance and expressivity of a trait can be influenced by the rest of the genome, that is, by which alleles of other genes are present. Environmental factors can also have significant effects on the phenotype associated with a particular allele or genotype.

Questions to answer & to ponder:

- Explain why superfecundity is required for evolution to occur.
- Why is the presence of inheritable variation important for any evolutionary model?
- How did plant and animal breeders inspire Darwin's thinking on evolution?
- From a practical point of view, what makes it possible for plant and animal breeding to produce distinctive types of organisms?
- What factors might lead to a new steady state level in the human population?
- How might the accumulation of mutations be used to determine the relationship between organisms?
- Why might the products of artificial selection not be competitive with "native" organisms?

Variation, selection, and isolation (speciation)

Darwin and Wallace's breakthrough conclusion was that genetic variation within a population would lead to altered reproductive success among the members of that population. Some genotypes, and the alleles of genes they contain, would become more common within subsequent generations because the individuals that contained them would reproduce more successfully. Other alleles and genotypes would become less common. The effects of specific alleles on an organism's reproductive

[67] "From Galton to GWAS: quantitative genetics of human height": http://www.ncbi.nlm.nih.gov/pubmed/21429269

[68] Genetics of human height: http://www.ncbi.nlm.nih.gov/pubmed/19818695

success would, of course, be influenced by the rest of the organism's genotype, its structure and behaviors (both selectable traits) and its environment. While some alleles can have a strong positive or negative impact on reproductive success, the effects of most alleles are subtle, assuming they produce any noticeable phenotypic effect at all. A strong positive effect will increase the frequency of the allele (and genotype) associated with it in future generations, while a strong negative effect can lead to the allele disappearing altogether from the population. At the same time, many alleles have more subtle, less strongly selectable effects. An allele that increases the probability of death before reproductive age is likely to be strongly selected against, whereas an allele that has only modest effects on the number of offspring an organism produces will be relatively weakly selected for.

Types of simple selection

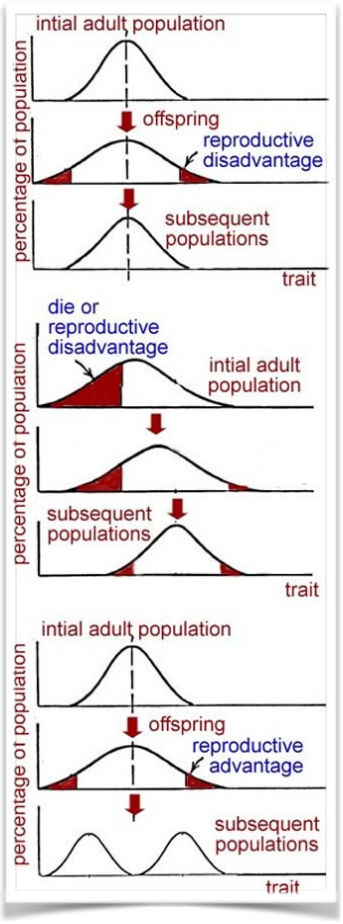
While it is something of an oversimplification (we will introduce the complexities associated with the random aspects of reproduction and the linked nature of genes shortly), we will begin with the three basic types of selection: conservative, directed, and disruptive. We start with a population composed of individuals displaying genetic variation in a particular trait. The ongoing processes of mutation continually introduces new genotypes, and their associated phenotypic effects. What is important to remember is that changes in the population and the environment can influence the predominant type of selection occurring over time, and that different types of selection may well (and most certainly are) occurring for different traits.

For each type of selection, we illustrate the effects as if they were acting along a single dimension, for example smaller to larger, or stronger to weaker, lighter to darker, slower to faster. In fact, most traits vary along a number of dimensions. For example, consider the trait of ear, paw, heart, or big toe shape. An appropriate type of graph would be a multi-dimensional surface, but that is harder to draw. Also, for simplicity, we start with populations whose distribution for a particular trait can be described by a simple and symmetrical curve, that is the mean and the median are equal. New variants, based on new mutations, generally fall more or less randomly within this distribution. Under these conditions, for selection NOT to occur we would have to make two seriously unrealistic assumptions: first that all organisms are equally successful at producing offspring, and second that each organism or pair of organisms produce only one or two (respectively) offspring. Whenever these are not the case, which is always, selective processes will occur, although the strength of selection may vary dramatically between traits.

Conservative selection: Sometimes a population of organisms appears static for extended periods of time, that is, the mean and standard deviation of a trait are not changing. Does that mean that selection has stopped? Obviously we can turn this question around, assume that there is a population with a certain stable mean and standard deviation of a trait. What would happen over time if selection disappeared?

Let us assume we are dealing with an established population living in a stable environment. This is a real world population, where organisms are capable of reproducing more, and some times, many more organisms than are needed to replace them when they die and that these organisms mate with one another randomly. Now we have to consider the factors that lead to the population distribution to being with: why is the mean value of the trait the value it is? What factors influence the standard

deviation? Assuming that natural selection is active, it must be that organisms that display a value of the trait far from the mean are (on average) at a reproductive disadvantage compare to those with the mean value of the trait. We do not know why this is the case (and don't really care at the moment). Now if selection (at least for this value of the trait) is acting, what happens? The organisms far from the mean are no longer at a reproductive disadvantage, so their numbers in the population will increase. The standard deviation will grow larger, until at the extreme, the distribution would be flat, characterized by a maximum and a minimum value. New mutations and existing alleles that alter the trait value will not be selected against, so they will increase in frequency. But in our real population, the mean and standard deviation associated with the trait remain constant. We can then predict selection against extreme values of the trait and can measure that selection "pressure" by following the reproductive success of individuals in the population with different values of the trait we have been considering. We would also predict that the more extreme the trait, that is, the further from the population mean, the greater its reproductive disadvantage would be, so that with each generation, the contribution of these outliers will be reduced. The distribution's mean will remain constant. The stronger the disadvantage the outliers face, the narrower the distribution will be – that is, the smaller the standard deviation. In the end, the size of the standard deviation will reflect both the strength of selection against outliers and the rate at which new variation enters the population through mutation. Similarly, we might predict that where a trait's distribution is broad, one might hypothesize that the impact of the trait on reproductive success is relatively weak.



Directed selection: Now imagine that the population's environment changes, and that it is no longer the case that the phenotype of the mean is the optimal phenotype, in terms of reproductive success. It could be that a smaller or a larger value is now more favorable. Under these conditions, we would expect that the mean of the distribution would shift toward the phenotypic value associated with maximum reproductive success over time. Once reached, and assuming the environment stays constant, conservative selection again becomes the predominant process. For directed selection to work, the environment must change at a rate and to an extent compatible with the changing mean phenotype of the population. Too big a change and the reproductive success of all members of the population could be dramatically reduced. The ability of the population to change will depend upon the variation already present within the population. While new mutations leading to new alleles are appearing, this is a relatively slow process. In some cases, the change in the environment may be so fast or so drastic and the associated impact on reproduction so severe that selection will fail to move the population and extinction will occur. One outcome to emerge from a changing environment leading to the directed selection is that as the selected population's mean moves, it may well alter the environment of other organisms.

Disruptive selection: A third possibility is that organisms find themselves in an environment in which traits at the extremes of the population distribution have a reproductive advantage over those nearer

the mean. If we think about the trait distribution as a multidimensional surface, it is possible that in a particular environment, there will be multiple and distinct strategies that lead to greater reproductive success compared to others. This leads to what is known as disruptive selection. The effect of disruptive selection in a sexually reproducing population will be opposed by the random mating between members of the population. But is random mating a good assumption? It could be that the different environments, which we will refer to as ecological niches, are physically distant from one another and organisms simply do not travel far to find a mate. The population will split into subpopulations in the process of adapting to the two different niches. Over time, two species could emerge, since whom one chooses to mate with and the productivity of that mating, are themselves selectable traits.

A short note on pedagogical weirdness

Many students are introduced into the field of population genetics and evolutionary mechanisms – that is, how phenotypes, genotypes, and allele frequencies change in the face of selective and environmental pressures – through what is known as the Hardy-Weinberg (H-W) equilibrium equation. Many H-W equation problems have been solved, but the question is why? From a historical perspective, the work of G.H. Hardy and Wilhelm Weinberg (published independently in1908) resolved the question of whether, in a *non-evolving population*, dominant alleles would replace recessive alleles over time. So what does that mean? Remember (and we will return to this later), in a diploid organism two copies of each gene are present. Each gene may be represented by different alleles. Where the two alleles are different, the one associated with the expressed (visible) phenotypic trait is said to be dominant to the other, which is termed recessive.[69] Geneticists previously believed that dominant alleles and traits were somehow "stronger" than recessive alleles or traits, but this is simply not the case and it is certainly not clear that this belief makes sense at the molecular level, as we will see. The relationship between allele and trait is complex. For example, an allele may be dominant for one phenotype and recessive for another (think about malarial resistance and sickle cell anemia, both due to the same allele in one or two copies.) What Hardy & Weinberg demonstrated was that in a non-*evolving* system, the original percentage of dominant and recessive alleles at various genetic loci (genes) stays constant. What is important to remember however is that this conclusion is based on five totally unrealistic assumptions, namely that: 1) the population is essentially infinite, so we did not have to consider processes like genetic drift (discussed below); 2) the population is isolated, no individuals left and none entered; 3) mutations do not occur; 4) mating between individuals is completely random (discussed further in Chapter 4); and 5) there are no differential reproductive effects, that is, no natural selection.[70] Typically H-W problems are used to drive students crazy and (more seriously) to identify situations where one of the assumptions upon which they are based is untrue (which are essentially all actual situations).

Questions to answer & ponder:

• Why does variation never completely disappear even in the face of conservative selection?

---

[69] In the context of the ABO gene for blood type, A and B alleles are dominant to O, which is recessive. Neither A nor B are dominant or recessive with respect to one another.

[70] Hardy-Weinberg Equilibrium: http://www.tiem.utk.edu/~gross/bioed/bealsmodules/hardy-weinberg.html

- What would lead conservative selection to be replaced by directed or disruptive selection?
- Explain the caveats associated with assuming that you know why a trait was selected.
- optional exercise: virtuallaboratory on adaptation:
  http://virtuallaboratory.colorado.edu/BioFun-Support/labs/Adaptation/Adaptation.html

Population size, founder effects and population bottlenecks

When we think about evolutionary processes from a Hardy-Weinberg perspective, we can ignore some extremely important situations that we would otherwise expect to impact populations. Things get more interesting when we take into consideration these non-exceptional processes. For example, what happens when a small number of organisms (derived from a much larger population) colonize a new environment? This is a situation, known as the founder effect, that is particularly relevant in island ecologies but also applies to pioneer populations migrating into new territories and then becoming isolated from their parent populations. Something similar happens when a large population is dramatically reduced, a situation known as a population bottleneck. Various types of environmental catastrophe, such as the appearance of a new pathogen, a new predator, or rapid climate change caused by volcanic activity, a cosmic collision, or a zombie apocalypse can cause population bottleneck. In both founder effect and population bottleneck situations, small populations become more susceptible to the effects of random fluctuations in survival and reproductive mechanisms, commonly referred to as genetic drift. In each case, given the dynamics of environmental change and population migrations, a population can come to develop unique traits through founder effects, population bottlenecks, and genetic drift. This can lead to the development of unexpected and advantageous traits that result in a selective advantage over the descendants of its parental population.

If we think of evolutionary changes as the movement of the population through a fitness landscape (the combination of the various factors that influence reproductive success), then isolation and evolutionary change of small populations can relieve, at least temporarily, the intensity of selective pressure and make possible the development and dispersal of new adaptations. For example, one effect of the major extinctions that have occurred during the evolution of life on Earth is that they provide a relaxed context for the evolution of new forms, a less densely-populated playing field, if you will. The expansion of the various types of mammals that followed the extinction of the dinosaurs is an example of one such opportunity, associated with changes in selection pressure.

Founder effects: What happens when a small subpopulation becomes isolated from its parent population? The original (large) population will contain a number of genotypes (and alleles), and if it is in a stable environment it will be governed primarily (as a first order approximation) by conservative selection. We can characterize this parental population in terms of the frequencies of the various alleles present within it. For the moment, we will ignore the effects of new mutations, which will continue to arise. Now assume that a small group of organisms from this parent population comes to colonize a new, geographically separate environment and that it is then isolated from the its parental population, so that no individuals travel between the parent and the colonizing population. The classic example of such a situation is the colonization of newly formed islands, but the same process applies more generally during various types of migrations. The small isolated group is unlikely to have the same distribution of alleles as the original parent population. Why is that? It is a question of the randomness of sampling of the population. For example, if rolled often enough (or an infinite number of times), a fair

six sided (cubical) die would be expected to produce the numbers 1, 2, 3, 4, 5, and 6 with equal probabilities. Each would appear 1/6th of the time. But imagine that the number of rolls is limited and relatively small. Would you expect to get each number appearing with equal probability? You can check your intuition using this applet [DiceExperiment]. See how many throws are required to arrive at an equal 1/6th probability distribution; the number is almost certainly much larger than you would guess. We can translate this onto populations in the following way: Imagine a population in which each individual carries one of six alleles and the percentage of each type is equal (1/6th). The selection of any one individual from this population is like a throw of the die, there is an equal 1/6th chance of selecting an individual with one of the six alleles. Since the parental population is large, the removal of one individual does not appreciably change the distribution of alleles remaining, so the selection of a second individual produces a result that is independent of the first just like rolls of die and equally likely to result in a 1/6th chance to produce any one of the six alleles. But producing a small subpopulation with 1/6th of each allele (or the same percentages of various alleles as are present in the parent population) is, like the die experiment above, very unlikely. The more genotypically complex the parent population, the more unlikely it is; imagine that the smaller colonizing population only has, for example, 3 members (three rolls of the die) – not all alleles present in the original population will be represented. Similarly, the smaller the subpopulation the more unlikely it is. So when a small group from a parent population invades or migrates into a new environment, it will very likely have a different genotypical profile from the parent population. This is a difference that is due not to natural selection but rather chance alone. Nevertheless, it will influence subsequent evolutionary events, first because the small subpopulation is likely to be significantly simpler genotypically than the original population and so likely to respond in different ways to new mutations and environmental pressures, and second, because the exact alleles present will influence the phenotypes associated with new combinations (genotypes) and new mutations.

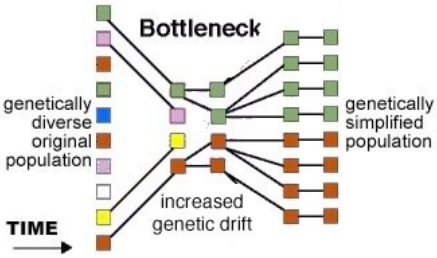
Because the human species appears to have emerged in Africa approximately 200,000 years ago, the people living in Africa represent the parent population of *Homo sapiens*. Genetic studies indicate that the African population displays a much greater genotypic complexity than do groups derived from the original African population, that is, everyone else. What remains controversial is the extent to which migrating populations of humans in-bred with what are known as archaic humanoids (such as Neanderthals and the Denisovians), which diverged from our lineage (*Homo sapiens*) approximately 1.2 million years ago.[71]

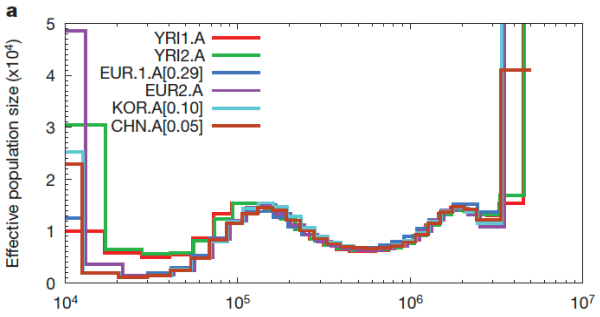
Population bottlenecks

A population bottleneck is similar in important ways to the founder effect. Population bottlenecks occur when some environmental change leads to the dramatic reduction of the size of a population. Catastrophic environmental changes, such as asteroid impacts, massive and prolonged volcanic eruptions (such as associated with continental drift), or the introduction of a particularly deadly pathogen, which kills a high percentage of the organisms that it infects can all create population bottleneck effects. Which organisms survive most types of bottlenecks will be random, that is unrelated to genotype (think of the immediate effects of an asteroid or the effects on a island-bound population

[71] Genetic Data and Fossil Evidence Tell Differing Tales of Human Origins: http://www.nytimes.com/2012/07/27/science/cousins-of-neanderthals-left-dna-in-africa-scientists-report.html?pagewanted=all

when the volcanic island they inhabit blows up or mostly blows up). There is compelling evidence that such drastic environmental events are responsible for population bottlenecks so severe that they led to mass extinctions. The most catastrophic of these extinction events was the Permian extinction that occurred ~251 million years ago, during which it appears that ~95% of marine organisms and ~75% of land species died off.[72] If most species were effected, we would not be surprised if the surviving populations experienced serious bottlenecks. The subsequent diversification of the surviving organisms, such as the dinosauria (which includes the extinct dinosaurs and modern birds) and the cynodontia, which includes the ancestors of modern mammals, including us, could be due in part to these bottleneck-associated effects, for example, through the removal of competing species or predators. A second catastrophic event occurred around 65 million years ago, which contributed to the extinction of the dinosaurs and led to the diversification of mammals, particularly the placental mammals.





In other cases, however, the effects of a bottleneck may not be random. Consider the effects of a severe drought or highly virulent bacterial or viral infection; the organisms that survive may have specific phenotypes (and associated genotypes) that increased their chances of survival. In such a case, the effect of the bottlenecking event would produce non-random changes in the distribution of genotypes (and alleles) in the post bottleneck population – these selective agents could continue to influence the population in various ways. For example, a trait associated with pathogen resistance may have other, even negative effects on phenotype, but these negative effects could be less important that the positive effect of surviving infection. In addition, the very occurrence of a rapid and extreme reduction in population size has its own effects. For example, it would be expected to increase the effects of genetic drift (see below).

We can identify extreme population reduction events such as founder effects and bottlenecks by looking at the variation in genotypes, particularly in genotypic changes not expected to influence phenotypes, mating preference, or reproductive success. These so-called neutral polymorphisms are expected to accumulate in the nonsense (intragenic) parts of the genome at a constant rate over time. The rate of the accumulation of such neutral polymorphisms is a type of population-based biological clock. Its rate can be estimated, at least roughly, by comparing the genotypes of individuals that are derived from populations in which the time of separation can be accurately estimated. For example, these types of studies indicate that the size of the human population dropped to a few thousands individuals between 20,000 to 40,000 years ago. This is a small number of people, likely to have been spread over a large area.[73] This bottleneck occurred around the time of the major migration of people

[72] The Permian extinction and the evolution of endothermy: http://www.nap.edu/openbook.php?record_id=11630&page=133

[73] Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans: http://ice2.uab.cat/argo/Argo_actualitzacio/argo_butlleti/ccee/geologia/arxius/1Ambrose%201998.pdf

out of Africa into Europe and Asia. Comparing genotypes, that is, neutral polymorphisms, between isolated populations also leads to estimates that aboriginal Australians reached Australia about 50,000 years ago, well before other human migrations[74] and that humans arrived in the Americas in multiple waves beginning around 15,000 to 16,000 years ago.[75] The arrival of humans into a new environment (another violation of the Hardy-Weinberg premises) has been linked to the extinction of a group of mammals known as the megafauna in those environments.[76] The presence of humans changed the environmental pressures on these organisms around the world.

Genetic drift

Genetic drift is an evolutionary phenomena that is difficult to comprehend in a strict Hardy-Weinberg world and explains the fact that most primates depend on the presence of vitamin C (ascorbic acid) in their diet. Primates are divided into two suborders, the Haplorhini (from the Greek meaning "dry noses") and the Strepsirrhini (from the Greek meaning "wet noses"). The Strepsirrhini contain the lemurs and lorices, while the Haplorhini include the tarsiers and the anthropoids (monkeys, apes, and humans). One characteristic trait of the Haplorhini is that they share a requirement for ascorbic acid (vitamin C) in their diet. In vertebrates, vitamin C plays an essential role in the synthesis of collagen, a protein involved in the structural integrity of a wide range of connective tissues. In humans, the absence of dietary vitamin C leads to the disease scurvy, which according to Wikipedia, "*often presents itself initially as symptoms of malaise and lethargy, followed by formation of spots on the skin, spongy gums, and bleeding from the mucous membranes. Spots are most abundant on the thighs and legs, and a person with the ailment looks pale, feels depressed, and is partially immobilized. As scurvy advances, there can be open, suppurating wounds, loss of teeth, jaundice, fever, neuropathy, and death.*"[77] The requirement for dietary vitamin C is due to a mutation in a gene, known as *gulo1*, which encodes the enzyme 1(-gulono-gamma-lactone oxidase (Gulo1)) required for the synthesis of vitamin C. One can show that the absence of a functional *gulo1* allele is the root cause of vitamin C dependence in Haplorrhini by putting a working copy of the *gulo1* gene, for example derived from the mouse, into human cells. The mouse-derived, *gulo1* allele, which encodes a functional form of the Gulo1 enzyme cures the human cells' need for exogenous vitamin C. But, no matter how advantageous a working *gulo1* allele would be (particularly for British sailors, who died in large numbers before the discovery of a preventative treatment for scurvy was discovered, a depressing story in its own right[78], no new *gulo1* allele appeared. Organisms do not always produce the alleles they need or that might be

[74] http://www.sciencemag.org/content/334/6052/94.short

[75] Reich et al., 2012. Reconstructing Native American population history. Nature; DOI: 10.1038/nature11258

[76] http://australianmuseum.net.au/Megafauna-extinction-theories-patterns-of-extinction and a very interesting video: http://youtu.be/8WZ5Q2JYbLY

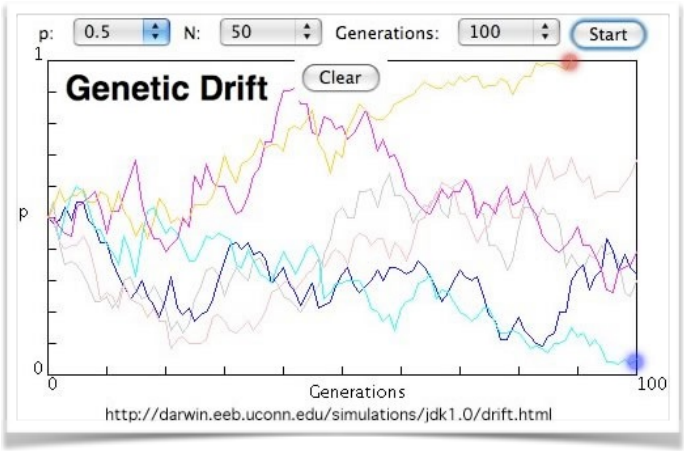
[77] One amazing fact is that it took various navies the deaths of thousands of sailors to understand the nutritional challenges of vitamin C. ADD REFERENCE

[78] http://mentalfloss.com/article/24149/how-scurvy-was-cured-then-cure-was-lost

beneficial, such alleles must be selected them from alleles already present in the population or that appear through mutation.

This mutant allele appears to have become fixed in the ancestral population that gave rise to the Haplorrhini ~40 million years ago. So the question is, how did we (that is our ancestors) come to loose a functional version of such an important gene? It seems obvious that when the non-functional allele became universal in that population, the inability to make vitamin C must not have been strongly selected against. We can imagine such an environment and associated behavior; namely, these organisms must have obtained sufficient vitamin C from their diet, so that the loss of the ability to synthesize vitamin C themselves had little negative effect on them.

So how were function alleles involved in vitamin C synthesis lost? In small populations, non-adaptive – that is, non-beneficial and even mildly deleterious – genotypic changes and their associated traits can increase in frequency through a process known as genetic drift. In such populations, selection continues to be active, but it has significant effects only for traits (and their associated alleles) when the trait strongly influences reproductive success. While genetic drift occurs in asexual populations, due to random effects on organismic survival, it is particularly prominent in sexually reproducing species. This is because cells known as gametes are produced during the process of sexual reproduction (Chapter 4). While the cell that generates these gametes contains two copies of each gene, and each gene can be one of a number of alleles within the population, any particular gamete contains only one allele of each gene. To generate a new organism, two gametes fuse to produce a diploid organism. This process combines a number of chance events: which two gametes fuse is generally a matter of chance, and which particular alleles each gamete contains is again a matter of chance. In a small population, over a reasonably small number of generations, one or the other alleles at a particular genetic locus will be lost, and given enough time, this allelic loss approaches a certainty. In this figure, six different experimental outcomes (each line) are analyzed over the course of 100 generations. In each case, the population size is set to 50, and at the start of the experiment half the individuals have one allele and half have the other. While we are watching only one genetic locus, this same type of behavior impacts every gene for which multiple alleles (polymorphisms) exist. In one of these six populations, one allele has been lost (red dot), in the other (blue dot), the other allele is close to being lost. When a particular allele becomes the only allele within a population, it is said to have been fixed. Assume that the two alleles convey no selective advantage, can you predict what will happen if we let the experiment run through 10,000 generations? If you are feeling mathematically inclined, you can even calculate the effect of mild to moderate positive or negative selective pressures on allele frequencies and the probability that a particular allele will be lost or fixed.



Since the rest of the organism's genotype often influences the phenotype associated with the presence of a particular allele, the presence or absence of various alleles within the population can influence the phenotypes observed. If an allele disappears because of genetic drift, future evolutionary

changes may be constrained (or perhaps better put, redirected). At each point, the future directions open to evolutionary mechanisms depend in large measure on the alleles currently present in the population. For example, what happens if drift leads to the fixation of a mildly deleterious allele, let us call this allele BBY. Now the presence of BBY will change the selective landscape: mutations and or alleles that ameliorate the negative effects of aBBY will increase reproductive success, selection pressures will select for those alleles. This can lead to evolution changing direction even if only subtly. With similar effects going on across the genome, one quickly begins to understand why evolution is something like a drunken walk across a selective landscape, with genetic drift and founder and bottleneck effects resulting in periodic staggers in random directions.[79]

This use of pre-existing variation, rather than the idea that an organism would invent variations in its genome as it needed them, was a key point in Darwin's view of evolutionary processes. The organism cannot create the alleles it might need nor are there any processes known that can produce specific alleles in order to produce specific phenotypes. Rather, the allelic variation generated by mutation, selection, and drift are all that evolutionary processes have with which to work. Only a rare mutation that recreates the lost allele can bring an allele back into the population once it has been lost. Founder and bottleneck effects, together with genetic drift combine to produce what are known as non-adaptive processes and make the history of a population a critical determinant of its future evolution.

Questions to answer & ponder:

- How does the extinction of one type of organism influence the evolution of others?
- How can a founder effect/bottleneck lead to a slightly deleterious mutation becoming common in a population?
- Why is the common need of a subclass of primates for vitamin C evidence for a common ancestor?
- Consider the various ways that the individuals that fail to pass through a bottleneck might differ from those that do. How many "reasons" can you identify?
- How does selection act to limit the effects of genetic drift? Under what conditions does genetic drift influence selection?
- Describe the relative effects of selection and drift following a bottleneck?
- How is it that drift can be quantified, but in any particular experiment, not predicted?
- Does passing through a bottleneck improve or hamper a population's chances for evolutionary success (that is, avoiding extinction)?

Gene linkage: one more complication

So far, we have not worried overly much about the organization of genes in an organism. It could be that each gene behaves like an isolate object, but in fact that is not the case. We bring it up here because the way genes are organized can, in fact, influence evolutionary processes. In his original genetic analyses, Gregor Mendel (1822 – 1884) spent a fair amount of time looking for "well behaved" genes and alleles, that is those that displayed simple recessive and dominant behaviors and that acted as if they were completely independent from one another. But it quickly became clear that these behaviors are not how most genes behave. In fact, they act as if they are linked together, because they are (as we will see, gene linkage arises from the organization of genes within the DNA molecules.) So what happens when a particular allele of a particular gene is highly selected for or against, based on its effects on reproductive success? That allele, together with whatever alleles are found in genes located

[79] Genetic drift: http://darwin.eeb.uconn.edu/simulations/jdk1.0/drift.html

near it, are also selected. We can think of this as a by stander (or sometimes termed a “piggy-back”) effect, where alleles are being selected not because of their inherent effects on reproductive success, but their location within the genome.

Linkage between genes is not a permanent situation. As we will see toward the end of the course, there are processes that can shuffle the alleles (versions of genes) on chromosomes, the end result of which is the further away two genes are from one another on a chromosome, the more likely alleles of those genes will appear to be unlinked. Over a certain distance, they will always appear unlinked. This means that effects of linkage will eventually be lost, but not necessarily before particular alleles are fixed. For example, extremely strong selection for a particular allele of gene A will lead to the fixation of alleles at neighboring genes; similarly, strong selection against a particular allele of gene A will lead to apparent selection against alleles in neighboring genes. This effect, together with other non-selective effects, such as genetic drift, can produce mildly non-advantageous traits. It is also possible that a trait that increases reproductive success, that is the number of surviving offspring, may have other not-so-beneficial, and sometime seriously detrimental effects - the key is to remember that evolutionary mechanisms do not result in what is best for an individual organism but what in the end enhances reproductive success. In this sense, they do not select for particular genes or versions of genes but rather for combinations of genes that optimize reproductive success. In this light, talking about selfish genes, as if a gene can exist outside of an organism, makes little sense. Evolution can be a rather dispassionate and even cruel process, if you personify it.

Of course, the situation gets more complex when evolutionary mechanisms generate organisms, like humans, who feel and can object to the outcomes of evolutionary processes. How such organisms come to be and the implications of their existence are deeply complex topics. In some cases, they may be the unintended side effects of selection for a particular trait; in other cases they arise from processes known as inclusive fitness and social evolution, which we will deal with in more detail in the next chapter.

A brief reflection on the complexity of phenotypic traits

We can classify traits into three general groups. Adaptive traits are those that, when present increase the organism’s reproductive success. These are the traits we normally think about when we think about evolutionary processes. Non-adaptive traits are those generated by stochastic processes, like drift and bottlenecks. These traits become established not because they improve reproductive success but simply because they happened to be fixed randomly within the population. Some of these non-adaptive traits can in fact be deleterious only in specific situations, for example when humans with a non-functioning *gulo-1* allele attempt to live on a diet from which vitamin C is absent. Of course, if an allele is extremely deleterious (particularly if it behaves in a dominant, genotypically and environmentally independent manner), it will disappear from the population due to selection. If it reappears, it is most likely to be due to a new (spontaneous) mutation that occurred within the affected individual or their parents. That said, when we consider an allele deleterious, we mean in terms of reproductive success. An allele can harm the individual organism carrying it yet persist in the population because it improves reproductive success. Similarly, an allele can be slightly positive in its effects, but again, its presence within the population is not directly due to these positive effects. Finally,

there are traits that could be seen as actively maladaptive, but which occur because they are linked, either genetically or mechanistically, to another positively-selected, adaptive trait. Many genes are involved in a number of distinct processes and their alleles can have multiple phenotypic effects. Such alleles are said to be pleiotropic, meaning they have many distinct effects on an organism's phenotype. Not all of the pleiotropic effects of an allele are necessarily of the same type; some traits can be beneficial, others deleterious. A trait that dramatically increases the survival of the young, and so their potential reproductive success, but leads to senility in older adults could well be positively selected for. In this scenario, the senility trait is maladaptive but is not eliminated by selection because it is mechanistically associated with the highly adaptive juvenile survival trait. It is also worth noting that a trait that is advantageous in one environment or situation can be disadvantageous in another. All of which is to say that when thinking about evolutionary mechanisms, do not assume that a particular trait exists independently of other traits or functions in the same way in all environments or even that its presence indicates that it is beneficial.

Questions to answer and ponder:

- Consider this quote from Charles Darwin, "*Natural selection will never produce in a being any structure more injurious than beneficial to that being, for natural selection acts solely by and for the good of each."* How would you modify it in light of our modern understanding of evolutionary mechanisms?
- Make a model of the factors that would influence a population isolated for 100 generations from its much larger parental population, assuming that it migrated back into its original habitat.

Speciation & extinction

As we have already noted, an important fact that any biological theory has to explain is why there are millions of different types of organisms currently present on Earth. The Theory of Evolution explains this observation through the process of speciation. The basic idea is that populations of organisms can split into distinct groups; over time evolutionary mechanisms acting on these populations will produce distinct types of organisms, that is, different species.[80] At the same time, we know from the fossil record and from modern experiences that types of organisms can disappear – they can become extinct. So the question is, what leads to the formation of a new species or the disappearance of an existing one?

So, naturalists observe, a flea has smaller fleas that on him prey; and these have smaller still to bite 'em; and so proceed ad infinitum.
- Jonathan Swift

To answer these questions, we have to consider how populations behave. A population of an organism will typically inhabit a particular geographical region. The size of these regions can range from extending over a continent or more, to a small region, such as a single isolated lake. Moreover, when we consider organisms that reproduce in a sexual manner, that is, that have to cooperate with one another to produce the next generation of organisms, we have to consider how far the organism (or its gametes) can travel. The range of some organisms is quite limited, whereas others can travel significant distances. Another factor we need to consider is how an organism makes its living, that is, where does

[80] The problem is, of course, more complex and subject with asexual species (such as bacteria), but here a more Linnaean analysis based on the comparison of traits is used. Among these traits are genomic sequence.

it get the food and space it needs to successfully reproduce?

The concept of an organism's **ecological niche**, which is the result of its past evolutionary history, that is, of the past selection pressures acting within a particular environment, combines all of these factors. In a stable environment, and a large enough population, reproductive success will reflect how organisms survive and exploit their ecological niche. Over time, conservative selection will tend to optimize the organism's adaptation to its niche. At the same time, it is possible that different types of organisms will compete for similar resources. This interspecies competition leads to a new form of selective pressure. If individuals of one population can exploit a different set of resources or the same resources differently, these organisms can minimize competition with other species and become more reproductively successful compared to individuals that directly compete with that species. This can lead to a number of outcomes. In one case, one species becomes much better than the other at occupying a particular niche, driving the other to extinction. Alternatively, one species may find a way to occupy a new or related niche, and within that particular niche, it can more effectively compete, so that the two species come to occupy distinct niches. Finally, one of the species may be unable to reproduce successfully in the presence of the other and become (at least) locally extinct. These scenarios are captured in what is known as the competitive exclusion principle or Gause's Law, which states that two species cannot (stably) occupy the same ecological niche - over time either one will leave (or rather be forced out) of the niche, or will evolve to fill a different, often subtly different niche. What is sometimes hard to appreciate is how specific a viable ecological niche can be. For example, consider the situation described by the evolutionary biologist Theodosius Dobzhansky (1900-1975):

> Some organisms are amazingly specialized. Perhaps the narrowest ecologic niche of all is that of a species of the fungus family Laboulbeniaceae, which grows exclusively on the rear portion of the elytra (the wing cover) of the beetle *Aphenops cronei*, which is found only in some limestone caves in southern France. Larvae of the fly *Psilopa petrolei* develop in seepages of crude oil in California oilfields; as far as is known they occur nowhere else.

While it is tempting to think of ecological niches in broad terms, the fact is that subtle environmental differences can favor specific traits and specific organisms. If an organism's range is large enough and each individual's range is limited, distinct traits can be prominent in different regions of the species' range. These different subpopulations (sometimes termed subspecies or races) reflect local adaptations. For example, it is thought that human populations migrating out of the equatorial regions of Africa were subject to selection based on exposure to sunlight in part through the role of sunlight in the synthesis of vitamin D.[81] In their original ecological niche, the ancestors of humans were thought to hunt in the open savannah (rather than within forests), and so developed adaptations to control their body temperature - human nakedness is thought to be one such adaptation (although there may be aspects of sexual selection involved as well, discussed in the next chapter). Yet, the absence of a thick coat of hair also allowed direct

[81]Genetics of skin color: http://humanorigins.si.edu/evidence/genetics/skin-color/modern-human-diversity-skin-color
image sources: http://hmg.oxfordjournals.org/content/18/R1/R9.full

exposure to the UV-light from the sun. While UV exposure is critical for the synthesis of vitamin D, too much exposure can lead to skin cancer. Dark skin pigmentation is thought to be a adaptive compromise. As human populations moved away from the equator, the dangers of UV exposure decreased while the need for vitamin D production remained. Under such condition, allelic variation that favored lighter skin pigmentation (but retaining the ability to tan, at least to some extent) appears to have been selected. Genetic analyses of different populations have begun to reveal exactly which mutations, and the alleles they produced, occurred in different human populations has they migrated out of Africa. Of course, with humans the situation has an added level of complexity. For example, the human trait of wearing clothing certainly impacts the pressure of "solar selection."

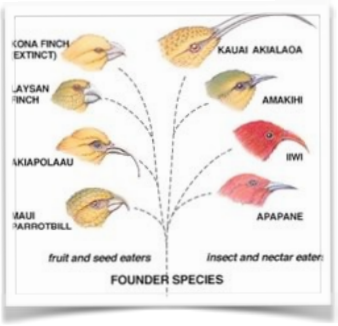
A number of variations can occur over the range of a species. Differences in climatic conditions, pathogens, predators, and prey can all lead to local adaptations, like those associated with human skin color. For example, many species are not continuously fertile and only mate at specific times of the day or year. When the range of a species is large, organisms in geographically and climatically distinct regions may mate at somewhat different times. As long as there is sufficient migration of organisms between regions and the organisms continue to be able to interbreed and to produce fertile offspring, the population remains one species.

Mechanisms of speciation

So now we consider the various mechanisms that can lead to a species giving rise to one or more new species. Remembering that species, at least species that reproduce sexually, are defined by the fact that they can and do interbreed to produce fertile offspring, you might already be able to propose a few plausible scenarios. An important point is that the process of speciation is continuous, there is no magic moment when one species changes into another, rather a new species emerges over time from a pre-existing species. Species are populations of organisms at a moment in time, they are connected to past species and can produce new species.

Perhaps the simplest way that a new species can form is if the original population is physically divided into isolated subpopulations. This is termed allopatric speciation. By isolated, we mean that individuals of the two subpopulations no longer mingle with one another, they are restricted to specific geographical areas. That also means that they no longer breed with one another. If we assume that the environments inhabited by the subpopulations are distinct, that they represent distinct sets of occupied and available ecological niches, distinct climate and geographical features, and distinct predators, prey, and pathogens, then these isolated subpopulations will be subject to different selection pressures, different phenotypes (and the genotypes associated with them) will have differential reproductive success. Assuming the physical separation between the populations is stable, and persists over a significantly long length of time, the populations will diverge. Both selective and non-selective processes will drive this divergence, and will influence by exactly what new mutations arise and give rise to alleles. The end result will be populations adapted to specific ecological niches, which may well be different from the niche of the parental population. For example, it is possible that while the parental population was more a generalist, occupying a broad niche, the subpopulations may be more specialized to a specific niche. Consider the situation with various finches (honeycreepers) found in the

Hawai'ian islands.[82] Derived from an ancestral population, these organisms have adapted to a number of highly specialized niches. These specializations give them a competitive edge in feeding off particular types of flowers [→]. As they specialize, however, they become more dependent upon the continued existence of their host flower or flower type. It is little like a fungus that can only grow on a particular place on a particular type of beetle, as well discussed earlier. We begin to understand why the drive to occupy a particular ecological niche also leads to vulnerability, if the niche disappears for some reason, the species adapted to it may not be able to cope, that is, be able to effectively and competitively exploit the remaining niches, and may become extinct. It is a sobering thought that current estimates are that greater that 98% of all species that have or now live on Earth are extinct, presumably due in large measure in changes in or the disappearance of their niche. You might speculate (and provide a logical argument to support your speculation) as to which of the honeycreepers illustrated above would be most likely to become extinct in response to environmental changes.[83] In a complementary way, the migration of organisms into a new environment can produce a range of effects as new competitions for existing ecological niches get resolved. If an organism influences its environment, the effects can be complex. As noted before, a profound and global example is provided by the appearance of photosynthetic organisms that released molecular oxygen (O_2) as a waste product early in the history of life on Earth. Because of its chemical reactivity the accumulation of molecular oxygen led to loss of some ecological niches and the creation of new ones. While dramatic, similar events occur on more modest levels all of the time, particularly in the microbial world. It turns out that extinction is a fact of life.



Gradual or sudden environmental changes, ranging from the activity of the sun, to the drift of continents and the impacts of meteors and comets, leads to the disappearance of existing ecological niches and appearance of new ones. For example, the collision of continents with one another leads to the formation of mountain ranges and regions of intense volcanic activity, both of which can influence climate. There have been periods when Earth appears to have been completely or almost completely frozen over. One such snowball Earth period has been suggested as playing an important role in the emergence of macroscopic multicellular life. These processes continue to be active today, with the Atlantic ocean growing wider and the Pacific ocean shrinking, the splitting of Africa along the Great Rift Valley, and the collision of India with Asia. As continents move and sea levels change, organisms that evolved on one continent may be able to migrate into another. All of these processes combine to lead to extinctions, which open ecological niches for new organisms, and so it goes.

At this point you should be able to understand that evolution never actually stops. Aside from various environmental factors, each species is part of the environment of other species. Changes in one species can have dramatic impacts on others as the selective landscape changes. An obvious example is the interrelationship between predators, pathogens, and prey. Which organisms survive to

---

[82] Hawaiian honeycreepers and their tangled evolutionary tree: http://www.theguardian.com/science/punctuated-equilibrium/2011/nov/02/hawaiian-honeycreepers-tangled-evolutionary-tree

[83] The Perils of Picky Eating: Dietary Breadth Is Related to Extinction Risk in Insectivorous Bats: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000672

reproduce will be determined in large part by their ability to avoid predators or recover from infection. Certain traits may make the prey more or less likely to avoid, elude, repulse, discourage, or escape a predator's attack. As the prey population evolves in response to a specific predator, these changes will impact the predator, which will also have to adapt. This situation is often call the Red Queen hypothesis, and it has been invoked as a major driver for the evolution of sexual reproduction, which we will consider in greater detail in the next chapter (follow the footnote to a video).[84]

As the Red Queen said to Alice ... "Here, you see, it takes all the running you can do to keep in the same place"
-Lewis Carroll, Though the Looking Glass

Isolating mechanisms

Think about a population that is on its way to becoming specialized to fill a particular ecological niche. What is the effect of cross breeding with a population that is, perhaps, on an adaptive path to another ecological niche? Most likely the offspring will be poorly adapted for either niche. This leads to a new selective pressure, selection against cross-breeding between individuals of the two populations. Even small changes in a particular trait or behavior can lead to significant changes in mating preferences and outcomes. Consider Darwin's finches or the Hawaiian honeycreepers mentioned previously. A major feature that distinguishes these various types of birds is the size and shapes of their beaks. These adaptations represent both the development of a behavior – that is the preference of birds to seek food from particular sources, for example, particular types of flowers or particular size seeds – and the traits needed to successfully harvest that food source, such as bill shape and size. Clearly the organism has to first display the behavior that makes selection of the physical trait beneficial. This is a type of loop, where behavioral and physical traits are closely linked. You can ask yourself, would a giraffe have a long neck if it did not like (want to) to eat the leaves of tall trees?

Back to finches and honeycreepers. Mate selection in birds is often mediated by song, generally males sing and females respond (or not). As beak size and shape change, so the song produced also changes.[85] This change is, at least originally, an unselected trait that accompanies the change in beak shape, but it can become useful if females recognize and respond to songs more like their own. This would lead to preferential mating between organisms with the same trait (beak shape). Over time, this preference could evolve into a stronger and stronger preference, until it becomes a reproductive barrier between organisms adapted to different ecological niches. Similarly, imagine that the flowers a particular subpopulation feeds on open and close at different times of the day. This could influence when an organism that feeds on a particular type of flower is sexually receptive. You can probably generate your own scenarios in which one behavioral trait has an influence on reproductive preferences. If a population is isolated from others, such effects may develop but are relatively irrelevant. They become important when two closely-related but phenotypically distinct populations come back into contact. Now matings between individuals in two different populations, sometimes termed hybridization, can lead to offspring poorly adapted to either niche. This creates a selective

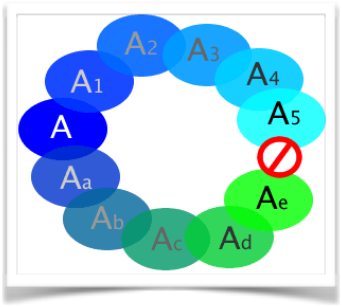
[84] The Red Queen: http://www.pbs.org/wgbh/evolution/library/01/5/l_015_03.html

[85] Beaks, Adaptation, and Vocal Evolution in Darwin's Finches: http://bioscience.oxfordjournals.org/content/54/6/501.short and Vocal mechanics in Darwin's finches: correlation of beak gape and song frequency: http://jeb.biologists.org/content/207/4/607.short

pressure to minimize hybridization. Again, this can arise spontaneously, such as the two populations mate at different times of the day or year or respond to different behavioral queues, such as mating songs. Traits that enhance reproductive success by reducing the chance of detrimental hybridization will be preferentially chosen. The end result is what is known as reproductive isolation.[86] Once reproductive isolation occurs, what was one species has become two. A number of different mechanisms ranging from the behavioral to the structural and the molecular are involved in generating reproductive isolation. Behaviors may not be "attractive," genitalia may not fit together, gametes might not survive, or embryos might not be viable - there are many possibilities.

Ring species

Ring species demonstrate a version of allopatric speciation. Imagine populations of the species A. Over the geographic range of A there exist a number of subpopulations. These subpopulations (A_1 to A_5) and (A_a to A_e) have limited regions of overlap with one another but where they overlap they interbreed successfully. But populations A_5 and A_e no longer interbreed successfully – are these populations separate species? In this case, there is no clear cut answer, but it is likely that in the link between the various populations will be broken and one or more species may form in the future. Consider the black bear, *Ursus americanus*. Originally distributed across North America, its distribution is now much more fragmented. Isolated bear populations are free to adapt to their own particular environments and migration between populations is limited. Clearly the environment in Florida is different from that in Mexico, Alaska, or Newfoundland. Different environments will favor different adaptations. If, over time, these populations were to come back into contact with one another, they might or might not be able to interbreed successfully - reproductive isolation may occur and one species might become many.



Sympatric speciation

While the logic and mechanisms of allopatric speciation are relatively easy to grasp (we hope), there is a second type of speciation, known as **sympatric speciation,** which was originally more controversial. It occurs when a single population of organisms splits into two reproductively isolated communities within the same physical area. How could this possibly occur, what would stop the distinct populations from in-breeding and reversing the effects of selection and nascent speciation? Recently a number of plausible mechanisms have been identified. One involves host selection.[87] In host selection, animals (such as insects) that feed off specific hosts may find themselves reproducing in distinct zones associated with their hosts. For example, organisms that prefer blueberries will mate in a different place, time of day, or time of year than those that prefer raspberries. There are blueberry and raspberry niches. Through a process of disruptive selection (see above), organisms that live primarily on a

[86] Beak size matters for finches' song: http://news.nationalgeographic.com/news/2004/08/0827_040827_darwins_finch.html

[87] Sympatric speciation by sexual selection: http://www.ncbi.nlm.nih.gov/pubmed/10591210?dopt=Abstract&holding=npg
Sympatric speciation in phytophagous insects: moving beyond controversy? http://www.ncbi.nlm.nih.gov/pubmed/11729091?dopt=Abstract&holding=npg

particular plant (or part of a plant) can be subject to different selective pressures, and reproductive isolation will enable the populations to more rapidly adapt. Mutations that reinforce an initial, perhaps weak, mating preference can lead to what known as reproductive isolation - as we will see this is a simple form of sexual selection.[88] One population has become two distinct, reproductively independent populations, one species as become two.

Questions to answer & ponder:

- Make a model of interactions of how non-adaptive factors could influence species formation.
- Describe the (Darwinian) cycle of selection associated with the development of the giraffe's neck.
- Provide a scenario that would explain why a small population associated with allopatric speciation would either speed evolutionary change or lead to extinction?
- Which comes first, the behavior or the ability to carry out the behavior?
- Make a model of the various effects of isolating mechanisms on allele frequencies between once isolated populations.
- How would you model the process by which an asexual organism would be assigned to a specific species?
- How would you go about determining whether an organism, identified through fossil evidence, was part of a new or a living species?
- How would you determine whether two species are part of the same genus?

Signs of evolution: homology and convergence

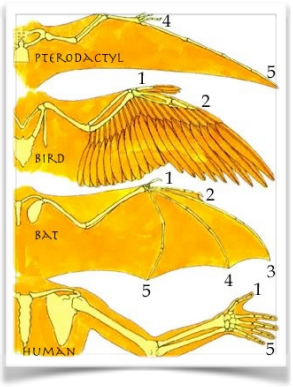
When we compare two different types of organisms we often find traits that are similar. On the basis of evolutionary theory, these traits can arise through either of two processes: the trait could have been present in the ancestral population that gave rise to the two species or the two species could have developed the traits independently. In this latter case, the trait was *not* present in the last common ancestor shared by the organism. Where a trait was present in the ancestral species it is said to be a **homologous** trait. If the trait was not present in the ancestral species but appeared independently within the two lineages, it is known as an **analogous** trait that arose through evolutionary **convergence**.

For example, consider the trait of vitamin C dependence, found in Haplorrhini primates discussed above. Based on a number of lines of evidence, we conclude that the ancestor of all Haplorrhini primates was vitamin C dependent and that vitamin C dependence in Haplorrhini primates is a homologous trait. On the other hand, Guinea pigs (*Cavia porcellus*), which are in the order Rodentia, are also vitamin C dependent, but other rodents are not. It is estimated that the common ancestor of primates and rodents lived more than 80 million years ago, that is, well before the common ancestor of the Halporrhini, and because other rodentia are vitamin C independent, that this common rodent/primate ancestor was itself vitamin C independent. We conclude that vitamin C dependence in Guinea pigs and Halporrhini are analogous traits.

As we look at traits, we have to look carefully, structurally, and more and more frequently in the 21st century, molecularly (genotypically) to determine whether they are homologous or analogous, that is the result of evolutionary convergence. Consider the flying vertebrates. The physics of flight (and many other behaviors that organisms perform) are constant. Organisms of similar size face the same aerodynamic and thermodynamic constraints. In general there are only a limited number of physically

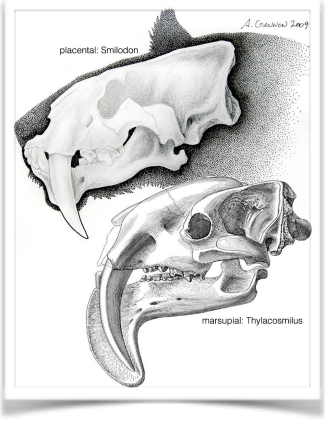
[88] The sexual selection: http://www.youtube.com/watch?v=JakdRczkmNo

workable solutions to deal with these constraints. Under these conditions different populations that are in a position to exploit the benefits of flight will, through the process of variation and selection, end up with structurally similar solutions. This process is known as convergent evolution. Convergent evolution occurs when only certain solutions to a particular problem are evolutionarily accessible.



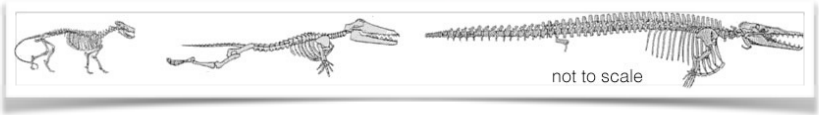
Consider the wing of a pterodactyl, which is an extinct flying reptile, a bird, and a bat, which is a flying mammal. These organisms are all tetrapod (four legged) vertebrates – their common ancestor had a structurally similar forelimb, so their forelimbs are clearly homologous. Therefore evolutionary processes (using the forelimb for flight) began from a similar starting point. But most tetrapod vertebrates do not fly, forelimbs have become adapted to different functions. An analysis of tetrapod vertebrate wings indicates that they took distinctly different approaches to generating wings. In the pterodactyl, the wing membrane is supported by the 5th finger of the forelimb, in the bird by the 2nd finger, and in the bat, by the 3rd, 4th and 5th fingers. The wings of pterodactyls, birds, and bats are clearly analogous structures, while their forelimbs are homologous.

As another example, the use of a dagger is an effective solution to the problem of killing another organism. Variations of this solution have been discovered or invented independently many times, with similar dagger-like teeth evolving independently (that is from ancestors without such teeth) in a wide range of evolutionarily distinct lineages. Consider, for example, the placental mammal Smilodon and the marsupial mammal Thyacosmilus [→]; both have similarly-shaped highly elongated canine teeth. Marsupial and placental mammals diverged from a common ancestor ~160 million years ago and this ancestor appeared to lack such teeth, as do most mammals. While teeth are a homologous feature of Smilodon and Thyacosmilus, elongated dagger-like teeth are analogous structures, the result of convergent evolution for this trait.



The loss of traits

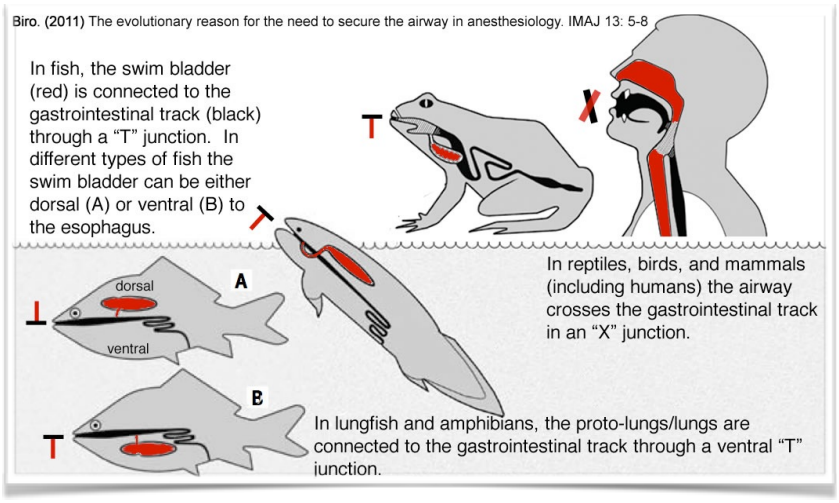
A major challenge when trying to determine the relationship between organisms based on anatomy has been to determine whether similar traits indicate common ancestry, that is whether the trait justifies putting two organisms into the same group, or whether it represents two independent solutions to a common problem, and so is irrelevant when it comes to placing an organism in a classification scheme. The loss of traits can confuse or complicate the positioning of an organism in a classification scheme. As organisms adapt to a specific environment and lifestyle, traits once useful can become irrelevant and may be lost (such as the ability to synthesize ascorbic acid). A classic example is the reduction of hind limbs during the evolution of whales. Another is the common loss of eyes often seen as populations adapt to environments in



which light is absent. The most dramatic case of loss involves organisms that become obligate parasites of other organisms. In many cases, these parasitic organisms become completely dependent on their hosts for many essential functions, and they can become quite simplified even though they are in fact highly evolved. For example, they lose many genes as they become dependent upon the host. The loss of traits can itself be an adaptation if it provides an advantage to organisms living in a particular environment. This fact can make it difficult to determine whether an organism is primitive (that is, retains ancestral features) or highly evolved.

Signs of evolutionary history

Evolution is an ongoing experiment in which random mutations are selected based on the effects of the resulting phenotypes on reproductive success. As we have discussed, various non-adaptive processes are also involved, which can impact evolutionary trajectories. The end result is that adaptations are based on past selective pressures and i) are rarely perfect and ii) may actually be outdated, if the environment the organisms live in has changed. One needs to keep this in mind when one considers the differences associated with living in a pre-technological world on the African savannah in small groups and living in New York City. In any case, evolution is not a designed process that reflects a predetermined goal but involves responses to current constraints and opportunities - it is a type of tinkering in which selective and non-selective processes interact with pre-existing organismic behaviors and structures and is constrained by cost and benefits associated with various traits and their effects on reproductive success.[89] What evolution can produce depends on the alleles present in the population and the current form of the organism. Not all desirable phenotypes (that is, leading to improved reproductive success) may be accessible from a particular genotype, and even if they are, the cost of attaining a particular adaptation, no matter how desirable to an individual, may not be repaid by the reproductive advantage it provides within a population. As an example, our ability to choke on food could be considered a serious design flaw, but it is the result of the evolutionary path that produced us (and other four-legged creatures), a path that led to the crossing of our upper airway (leading to the lungs) and our pharynx (leading to our gastrointestinal system). That is why food can lodge in the airway, causing choking or death. It is possible that the costs of a particular "imperfect" evolutionary design are offset by other advantages. For example, the small but significant possibility of death by choking may, in an evolutionary sense, be worth the ability to make more complex sounds (speech) involved in social communication [90].



[89] Evolutionary tinkering: http://virtuallaboratory.colorado.edu/Biofundamentals/lectureNotes/Readings/EvolutionTinkering.pdf

[90] How the Hyoid Bone Changed History: http://www.livescience.com/7468-hyoid-bone-changed-history.html

As a general rule, evolutionary processes generate structures and behaviors that are as good as they need to be for an organism to effectively exploit a specific set of environmental resources and to compete effectively with its neighbors, that is, to successfully occupy its niche. If being better than good enough does not enhance reproductive success, it cannot be selected for (at least via natural selection) and variations in that direction will be lost, particularly if they come at the expense of other important processes or abilities. In this context it is worth noting that we are always dealing with an organism throughout its life cycle. Different traits can have different values at different developmental stages. Being cute can have important survival benefits for a baby but be less useful in a corporate board room (although perhaps that is debatable). A trait that improves survival during early embryonic development or enhances reproductive success as a young adult can be selected for even, if it produces negative effects on older individuals. Moreover, since the probability of being dead (and so no longer reproductively active) increases with age, selection for traits that benefit the old will inevitably be weaker than selection for traits that benefit the young (although this trend can be modified in organisms in which the presence of the old can increase the survival and reproductive success of the young, for example through teaching and babysitting). Of course survival and fertility curves are also changing in response to changing environmental factors, which change selective pressures. In fact, lifespan itself is a selected trait, since it is the population not the individual that evolves.[91]

We see the evidence for various compromises involved in evolutionary processes all around us. It explains the limitations of our senses, as well as our tendency to get backaches, need hip-replacements, and our susceptibility to diseases and aging.[92] For example, the design of our eyes leaves a blind spot in the retina. Complex eyes have arisen a number of times during the history of life, apparently independently, and not all have a blind spot. We have adapted to this retinal blind spot through the use of saccadic movements because this is an evolutionarily easier fix to the problem than rebuilding the eye from scratch (which is essentially impossible). An "intelligently designed" human eye would presumably not have such an obvious design flaw, but because of the evolutionary path that led to the vertebrate eye, it may simply have been impossible to back up and fix this flaw. More to the point, since the vertebrate eye works very well, there is no reward in terms in reproductive success associated with fixing this flaw. This is a general rule: current organisms work, at least in the environment that shaped their evolution. Over time, organisms that diverge from the current optimal, however imperfect, solution will be at a selective disadvantage. The current vertebrate eye is maintained by conservative selection (as previously described).

Homologies provide evidence for a common ancestor

The more details two structures share, the more likely they are to be homologous. In the 21st century, molecular methods, particularly complete genome sequencing, have made it possible to treat gene sequences and genomic organization as traits that can be compared. Detailed analyses of many

[91] Methusaleh's Zoo: how nature provides us with clues for extending human health span: http://www.ncbi.nlm.nih.gov/pubmed/19962715 and Why Men Matter: Mating Patterns Drive Evolution of Human Lifespan: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000785

[92] http://www.pbs.org/wgbh/nova/evolution/what-evidence-suggests.html

different types of organisms reveals the presence of a common molecular signature that strongly suggests that all living organisms share a large numbers of homologies, which implies that they are closely related; that is, that they share a common ancestor. These universal homologies range from the basic structure of cells to the molecular machinery involved in energy capture and transduction, information storage and utilization. All organisms

- use double-stranded DNA as their genetic material;
- use the same molecular systems, transcription and translation, to access the information stored in DNA;
- use a common genetic code, with few variations, to specify the sequence of polypeptides (proteins);
- use ribosomes to translate the information stored in messenger RNAs into polypeptides; and
- share common enzymatic (metabolic) pathways.

Anti-evolution arguments

The theory of evolution has been controversial since its inception largely because it deals with issues of human origins and behavior, our place in the Universe, and life and its meaning. Its implications can be quite disconcerting, but many observations support the fact that organisms on Earth are the product of evolutionary processes and these processes are consistent with what we know about how matter and energy behave. As we characterize the genomes of diverse organisms, we see evidence for the interrelationships, observations that non-scientific (creationist) models would never have predicted and do not explain. That evolutionary mechanisms have generated the diversity of life and that all organisms found on Earth share a common ancestor is as well-established as the atomic structure of matter, the movement of Earth around the Sun, and the solar system around the Milky Way galaxy. The implications of evolutionary processes remain controversial, but not evolution itself.

Scientific knowledge is a body of knowledge of varying degrees of certainty-some most unsure, some nearly sure, but none absolutely certain ... Now we scientists are used to this, and we take it for granted that it is perfectly consistent to be unsure, that it is possible to live and not know.
- Richard Feynman.

...it is always advisable to perceive clearly our ignorance.–
Charles Darwin.

Questions to answer & to ponder:

- Justify the assumption that the mutations in Haplorrhini primates and guinea pigs were independent events?
- What typical mammalian traits have whales lost during their evolution?
- Model the factors that would influence the evolution whales back to a terrestrial lifestyle.
- Generate a model by which you could classify a trait as primitive or derived?
- How does the loss of a trait or convergent evolution complicate lineage analysis?
- If all organisms are descended from a common ancestor, what can we say about the diversity of pre-biogenic systems that existed before that ancestor?
- What conditions can lead to a complex organism becoming simpler?
- If the environment were constant, would extinction or evolution occur?
- In what ways can an organism direct its evolution?
- What are the benefits and drawbacks of a high degree of specialization for a species?
- How might the types of changes that lead to reproductive isolation be beneficial (overall) even if they were mildly deleterious?

- How do we know that a species is a species if we do not directly observe whether it can interbreed with other organisms?
- Consider Hawaiian honey creepers; which is most likely to become extinct and why?
- What testable predictions emerge from "intelligent design creationism"?
- Under what environmental conditions would a generalist be favored over a specialist?
- What benefit(s) could be linked to the loss of eyesight or other "advanced" traits?

4. Social evolution and sexual selection

In which we consider how organisms, even (some) unicellular organisms, have evolved to cooperate with one another, leading to the formation of multicellular organisms composed of distinct cell types. Similar evolutionary mechanisms have produced a range of cooperative (social) behaviors. One particularly important such behavior is sexual reproduction and we consider its effects on the morphology and behavior of organisms.

The naturalist Ernst Mayr made an important point when thinking about biology compared to physics and chemistry. The history of an electron, an atom, or a molecule is irrelevant to its physical and chemical properties. Each carbon isotope, for example, is identical to all others - one could be replaced by another and you could never, in theory or in practice, tell the difference. In contrast, each organism, how it is built, how it behaves, how it interacts with other organisms, and the future evolution of its descendants is the result of a continuous evolutionary process involving both selective and adaptive and non-selective and non-adaptive processes stretching back approximately 3.5 billion years. This history encompasses an unimaginable number of random events (mutations, accidents, environmental disasters, isolated and merging populations). Because of its molecular and cellular complexity and distinct history, each organism is unique and distinguishable from all others.

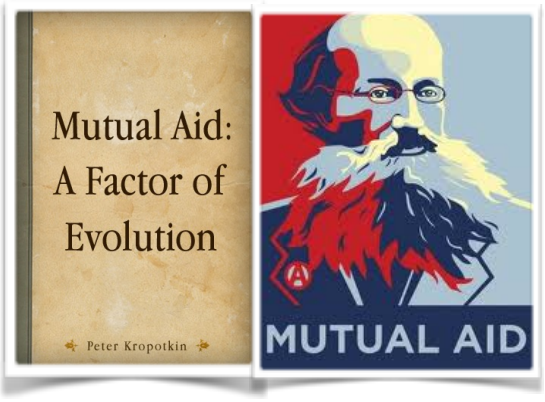
In biology, we normally talk about organisms, but this may be too simplistic. When does an organism begin? what are its boundaries? The answers can seem obvious, but then again, perhaps not. When a single-celled organism reproduces it goes through some form of cell division, and when division is complete, one of the two organisms present is considered a new organism and the other the old (preexisting) one, but generally it is not clear which is which. When an organism reproduces sexually, the new organism arises from the fusion of pre-existing cells and it itself produces cells that fuse to form the next generation. But if we trace the steps backward from any modern organism, where would we draw the lines between the different types (that is, species) of organisms? The answer is necessarily arbitrary, since cellular continuity is never interrupted. In a similar manner, we typically define the boundaries of an organism in physical terms, but organisms interact with one another, often in remarkably complex ways. A dramatic example of this are the eusocial organisms. While many of us are familiar with ants and bees, fewer (we suspect) are aware of the naked (*Heterocephalus glaber*) and the Damaraland (*Cryptomys damarensis*) mole rats. In these organisms, reproduction occurs at the group level; only selected individuals, termed queens because they tend to be large and female, produce offspring. Most members of the group are (often sterile) female workers, along with a few males to inseminate the queen.[93] So what, exactly, is the organism, the social group or the individuals

[93] An Introduction to Eusociality: http://www.nature.com/scitable/knowledge/library/an-introduction-to-eusociality-15788128

that make it up? From an evolutionary perspective, selection is occurring at a social level, rather than an organismic level. Similarly, consider yourself and other multicellular animals (and plants). Most of the cells in your body (known as the soma) do not directly contribute to the next generation, rather they cooperate to insure that a subset of cells, known as the germ line, have a chance to form a new organism. In a real sense, the somatic cells are sacrificing themselves so that the germ line cells can reproduce a new organism. They are the sterile workers to the germ line's queen.

We find examples of social behavior at the level of unicellular organisms as well. For example, think about a unicellular organism that divides but in which the offspring of that division stick together. As this process continues, we get what we might term a colony. Is it one or many organisms? If all of the cells within the group can produce new colonies, we could consider it a colony of organisms. So where does a colony of organisms turn into a colonial organism? The distinction is certainly not unambiguous, but we can adopt a set of guidelines or rules of thumb.[94] One criterion would be that a colony becomes an organism when it displays traits that are more than just sticking together or failure to separate, that is, when it acts more like an individual or a coordinated group. Conventionally this involves the differentiation of cells, one from the other, so that certain cells within the group become specialized to carry out specific roles, and reproducing the next generation is one such specialized role. Other cells may become specialized for feeding or defense. This differentiation of cells from one another has moved a colony of organisms to a multicellular organism. What is tricky about this process is that originally reproductively competent cells have given up their ability to reproduce, and are now acting, in essence, to defend or support the cells that do reproduce. This is a social event and is similar (analogous) to the behavior of naked mole rats. Given that natural selection acts on reproductive success, one might expect that the evolution of this type of cellular and organismic behavior would be strongly selected against or simply impossible to produce, yet multicellularity and social interactions have arisen independently dozens (or more likely millions) of times during the history of life on earth.[95] Is this a violation of evolutionary theory or do we have to get a little more sophisticated in our thinking?

Selecting social (cooperative) traits



The answer is that the origins and evolution of multicellularity do not violate evolutionary theory, but they do require us to approach evolutionary processes more broadly. The first new idea we need to integrate into our theoretical framework is that of inclusive fitness, which is sometimes referred to as kin selection. For the moment, let us think about traits that favor the formation of a multicellular organism - later we will consider traits that have a favorable effect on other, related organisms, whether or not they directly benefit the cell/organism that expresses that trait.

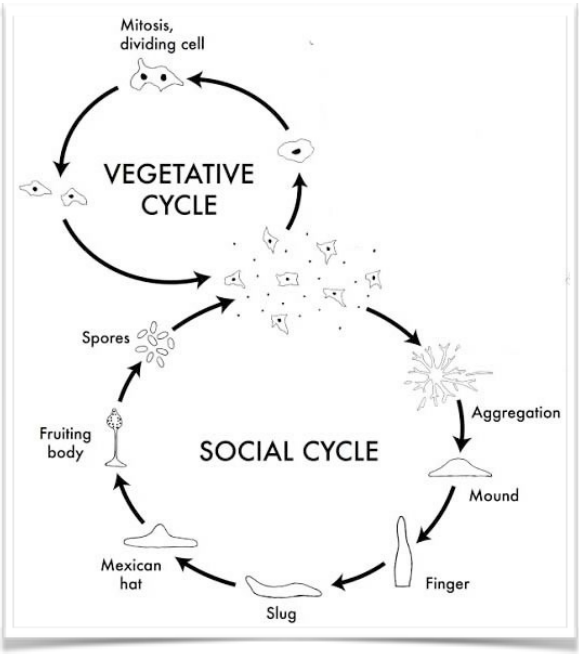
[94] A twelve-step program for evolving multicellularity and a division of labor: http://onlinelibrary.wiley.com/doi/10.1002/bies.20197/full

[95] The Origins of Multicellularity: https://bcrc.bio.umass.edu/courses/fall2010/biol/biolh100-03/sites/default/files/bonner_multicellularity_1998.pdf

Finally, we will consider social situations in which behaviors have become fixed to various extents, and are extended to strangers (humans can, but do not always, display such behaviors). The importance of mutual aid in evolutionary thinking, that is the roles of cooperation, empathy, and altruism in social populations, was a point emphasize by the early evolutionary biologist (and anarchist) (Prince) Peter Kropotkin (1842 – 1921).

All traits can be considered from an economic cost-benefit perspective. There is the cost (let us call that term "c") in terms of energetics needed to produce the trait and the risks associated with expressing the trait, and a benefit ("b") in terms of effects on reproductive success. To be evolutionarily preferred (or selected), the benefit b must be greater than the cost c ($b > c$). Previously we had tacitly assumed that both cost and benefit applied to a single organism, but in cooperative behaviors and traits, this is not the case. We can therefore extend our thinking as follows: assume that an organism displays a trait. That trait has a cost to produce and yet may have little or no direct benefit to the organism and may even harm it, *but* this same trait benefits neighboring cells. This is like (but not exactly the same as) the fireman who risks his life to save an unrelated child in a burning building. How is it possible for a biological system (the fireman), the product of evolutionary processes, to display this type of behavior?

Let us consider some examples of this type of behavior. A classic example is provided by social amoebae of the genus *Dictyostelium*.[96] These organisms have a complex life style that includes a stage in which unicellular amoeba-like organisms crawl around in the soil eating bacteria and dividing (watch http://youtu.be/bkVhLJLG7ug). These cells can divide asexually in what is known as a vegetative cycle (as if vegetables don't have sex, but we will come back to that!) If the environment turns hostile, the isolated amoeba begin to secrete a small molecule that influences their own and their neighbor's behaviors. They begin to migrate toward one another, forming aggregates of thousands of cells. Now something rather amazing happens: these aggregates begin to act as coordinated entities, they migrate around as "slugs" for a number of hours. Within the soil they respond to environmental signals, for example moving toward light, and then settle down and undergo a rather spectacular process of differentiation.[97] All through the cellular aggregation and slug migration stage, the original amoeboid cells remain as distinct cells. Upon differentiation approximately 20% of the cells in the slug differentiate to form stalk cells, which lift the rest of the cells above the soil. The stalk cells can no longer divide, in fact, they die; they



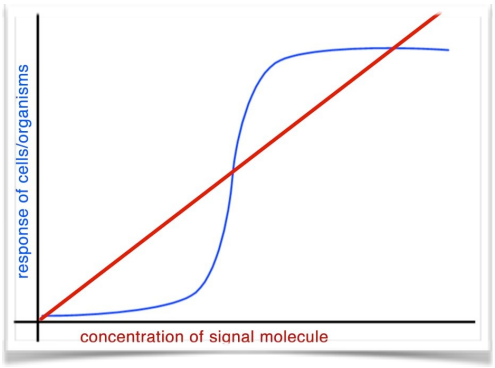
[96] Molecular phylogeny and evolution of morphology in the social amoebas: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2173941/#!po=38.8889 and A Simple Mechanism for Complex Social Behavior: http://youtu.be/vjRPla0BONA

[97] Behavior of cellular slime molds in the soil: http://www.mycologia.org/content/97/1/178.full

have sacrificed themselves for other cells, which go on to form spores. These spores are specialized cells that can survive harsh conditions and can be moved by the wind and other mechanisms into new environments. Once these spore cells land in a new environment, they convert back into unicellular amoeba that begin to feed and reproduce. The evidence indicates that within the slug, the "decision" on whether a cell will form a stalk or a spore cell is stochastic rather than innate. By stochastic we mean that the decision is controlled by underlying random processes, processes that we will consider in greater detail later on. What is important at this point is that this stochastic process is not based on genetic (genotypic) differences between the cells within a slug - two genotypically identical cells may both form spores, both stalk cells, or one might become a stalk and one a spore cell.

Quorum sensing

Another type of behavior at the unicellular level involves a behavior known as quorum sensing. This is a process by which such organisms can sense the density of other organisms in their immediate environment. Each individual secretes a molecule which they can also respond to, but their response is dependent upon the concentration of the secreted molecule and their response is non-linear. So what does that mean? As the concentration of signaling molecules increases, there is a discrete threshold concentration below which the cells/organisms do not change their behavior in response to the secreted compound and above which they do. When the cells/organisms are present at a low density, the concentration of the signaling molecule never reaches the threshold concentration. When the density (organisms per unit volume) increases sufficiently, the concentration of the signaling molecule rises above threshold and the cells/organisms change their behavior. Often this involves changes in the expression of specific genes (we will consider what that means exactly later on).[98]



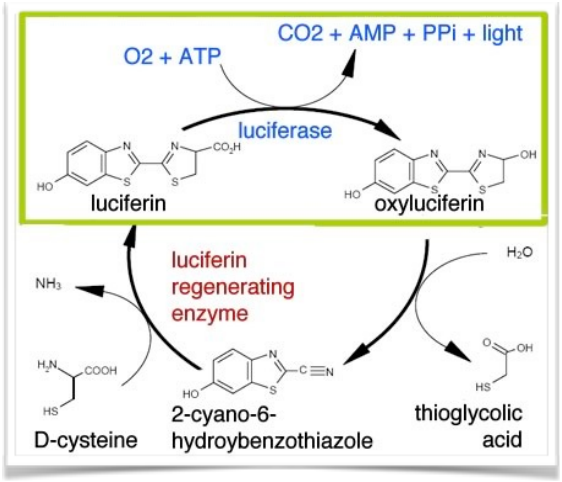
A classic example of a number of cooperative and quorum sensing behaviors is provided by the light emitting marine bacteria *Vibro fischeri. These* are marine bacteria that form a symbiotic arrangement with the squid *Euprymna scolopes*.[99] In these squid, the *V. fischeri* bacteria colonize a special organ known as a light organ. The squid uses light emitted from this organ to confuse and hide from its own predators as it hunts its prey. While their are many steps in the colonization process, and its regulation is complex, we will just consider a few to indicate how cooperative behaviors between the bacteria are critical. For the colonization of the squid's light organs the *V. fisherei* bacteria must bind to a specific region of the juvenile squid. As they divide, they sense the presence of their neighbors and begin to secrete molecules that form of gooey matrix - this leads to the formation of a specialized aggregate of cells (a type of biofilm) that is essential for the bacteria to colonize the squid's light organs. Within the biofilm, the bacteria acquire the ability to follow chemical signals produced by the squid's

[98] Quorum sensing in bacteria: http://www.ncbi.nlm.nih.gov/pubmed/11544353

[99] Gimme shelter: how Vibrio fischeri successfully navigates an animal's multiple environments: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3843225/pdf/fmicb-04-00356.pdf

light organ cells. The bacteria swim (through a process known as chemotaxis) toward these signals, thereby entering and colonizing the light organs.

The bacteria in the light organs can emit light through a reaction involving the luciferin molecule. This reaction is catalyzed (that is, sped up) by the protein luciferase, which is encoded by one of the bacteria's genes. We will discuss in some detail the thermodynamics of such reactions in the next section of the course. Given that bacteria are small, you can imagine that very little light would be emitted from a single bacteria. If there are only a small number of bacteria within the light organ, it would be ineffectual to carry out the light emitting reaction. The light emitting reaction occurs only when the number of bacteria within a light organ becomes sufficiently high. But how do the bacteria know that they are in the presence of sufficient numbers of neighbors? Here is where quorum sensing comes into play. A molecule secreted by the bacteria regulates the components of the light reaction. At high concentrations of bacteria, the concentration of the secreted molecule rises above a threshold, and the bacteria respond by turning on their light emitting system.



Mechanistically similar systems are involved in a range of processes including the generation of toxins (virulence factors and antibiotics directed against other types of organisms). These are produced only when the density of the bacteria rises above a threshold concentration. This insures that when a biologically costly toxin or other secreted molecule is made, it is effective – that is, it is produced at a level high enough to carry out its intended role. These high levels can only be attained through cooperative behaviors involving many individuals.

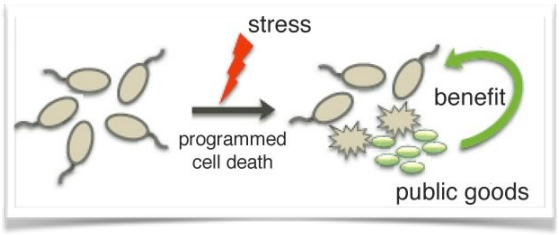
Active (altruistic) cell death

One type of behavior you might think would be impossible for evolutionary processes to produce would be the active, intentional or programmed death of a cell or an organism. Yet, such behaviors are found in a wide range of systems.[100] The death and release of leaves from deciduous trees in the autumn is an example of a programmed cell death process known generically as **apoptosis**. The process amounts to cellular suicide. It plays important roles in the formation of various structures within multicellular organisms, such as the fingers of your hands, which would develop as paddles without it, as well as playing a critical role in development of the immune and nervous systems, topics beyond the scope of this book. The process of programmed cell death is distinct from accidental cell death, such as occurs when a splinter impales a cell or you burn your skin. Such accidental death leads to what is known as necrosis, in which cellular contents are spilled out of the dying cell. It often provokes various organismic defense systems to migrate into the damaged area, primarily to fight off bacterial infections. The swelling and inflammation associated with injury is an indirect result of necrotic cell death. In contrast, apoptotic cell death occurs by a well defined pathway and requires energy to carry out. Cell

[100] See On the paradigm of altruistic suicide in the unicellular world: http://www.ncbi.nlm.nih.gov/pubmed/20722725

contents are retained during the process, and no inflammatory, immune system response is provoked. In general it appears to play specific and important roles within the context of the organism. Commitment to active cell death is generally very tightly controlled. A detailed discussion of the molecular mechanisms involved in apoptosis is beyond the scope of this course.

Here we will consider active/programmed cell death in the context of simpler systems, specifically those formed by unicellular organisms. In unicellular organisms, active cell death is a process triggered by environmental stresses together with quorum sensing. In this situation, a subset of the cells will "decide" to undergo active cell death by activating a pathway that will lead to the death of the cell. Now when one cell in a densely populated environment dies, its contents are released and can be used by the living cells that remain. These living cells gain a benefit, and we would predict that the increase in nutrients would increase their chances of their survival and successful reproduction. This strategy works because as the environment becomes hostile, not all cell die at the same time. As we will see later on, this type of individualistic behavior can occur even in a group of genetically identical cells through the action of stochastic processes.



So how do cells kill themselves (on purpose)? Many use a similar strategy. They contain a gene that directs the expression of a toxin molecule, which by itself will kill the cell. This gene is expressed in a continuous manner. Many distinct toxin molecules have been identified, so they appear to be analogous rather than homologous. Now you may well wonder how such a gene could exist, how does the cell survive in the presence of a gene that encodes a toxin. The answer is that the cell also has a gene that encodes an anti-toxin molecule, which typically binds to the toxin and renders it inactive. Within the cell, the toxin-anti-toxin complex forms exists but does no harm, since it is inactive (the toxin's activity is inhibited by the binding to the anti-toxin molecule.) The toxin and anti-toxin molecules differ however in one particularly important way. The toxin molecule is relatively stabile - once made it exists for a substantial period of time before it is degraded by other molecular systems within the cell. In contrast, the anti-toxin molecule is unstable. It is rapidly degraded. It can be maintained at a high enough level to inhibit the toxin only if new anti-toxin molecules are continually synthesized. In a sense the cell has become addicted to the toxin-anti-toxin module.

What happens if the cell is stressed, either by changes in its environment or perhaps infection by a virus? Often the synthesis of cellular components slows or stops. Now can you predict what happens? The level of the stable toxin molecule within the cell remains high and only slowly decreases, while the level of the unstable anti-toxin drops rapidly. It quickly drops below the threshold level required to keep the toxin inactive, so that the now active toxin initiates the process of active cell death.

In addition to the dying cell sharing its resources with its neighbors, active cell death can be used as a population-wide defense mechanism against viral infection. One of the key characteristics of viruses is that they must replicate within a living cell. Once a virus enters a cell, it typically disassembles itself and sets out to reprogram the cell's biosynthetic machines to generate new copies of the virus. During the period between viral disassembly and the appearance of newly synthesized viruses, the infectious virus disappears - it is said to be latent. If the cell were to kill itself before new

viruses were synthesized, it would also kill the infecting virus. By killing the virus (and itself) the infected cell acts to protect its neighbors from viral infection - this can be seen as the ultimate kind of altruistic, self-sacrificing behavior we have been considering.[101]

Inclusive fitness, group selection, and social evolution

Kin selection: The question that troubled Darwin and others was, how can evolutionary processes produce this type of social, self-sacrificing behavior? Consider, for example, the behavior of bees. Worker bees, who are sterile females, "sacrificed themselves to protect their hives" even though they do not themselves reproduce.[102] Another example, taken from the work of R.A. Fisher (1890–1962), involved the evolution of noxious taste as a defense against predators. Assuming that the organisms eaten by predators did not benefit from this trait, how could the trait of "distastefulness" arise in the first place? If evolution via natural selection is about an individuals differential reproductive success, how are such traits possible? W.D. Hamilton (1936–2000) provided the formal answer, expressed in the equation r x b > c (defined by Sewall Wright (1889–1988)), where "b" stands for the benefit of the trait to the organism and others, "c" stands for the cost of the trait to the individual and "r" indicates the extent to which two organisms within the population are related to one another.

Let us think some more about what this means. How might active cell death in bacterial cells be beneficial evolutionarily? In this case, reproduction is asexual and the cell's/organism's neighbors are likely to be closely related. They are likely to be clonally related, that is sets of cells or organisms derived from a common parent in an asexual manner. Aside from occasional mutations, the cells/organisms within a clone are genotypically identical. Their genotypic similarity arises from the molecular processes by which the genetic material (DNA) replicates and is delivered to the two daughter cells. We can characterize the degree of relationship or genotypic similarity through their r value, the coefficient of relationship. In two genetically identical organisms, r = 1. Two unrelated organisms, with minimum possible genotypic similarity would have an r very close to, but slightly larger than 0 (you should be able to explain why r is not equal to 0). Now let us return to our cost-benefit analysis of a trait's effect on reproductive success. As we introduced before, each trait has a cost = c to the organism that produces it, as well as a potential benefit = b in terms of reproductive success. Selection leads to a trait becoming prevalent or fixed within a population if b > c. But this equation ignores the effects of a trait on other related and neighboring organisms. In this case, we have to consider the benefits accrued by these organisms as well. Let us call the benefit to the cooperative/altruistic = b_i and the benefit to others/neighbors = b_o . To generate our social equation, known as Hamilton's rule (see above), we need to consider what is known as the inclusive fitness, namely the benefits provided to others as a function of their relationship to the cooperator. So b > c becomes $b_i + r \times b_o > c$. This leads to the conclusion that a trait can evolve if the cost to the cell or organism that displays it, in terms of metabolic, structural, or behavioral impact on its own reproductive ability, is offset by a sufficiently large increase in the reproductive success of individuals related to it. The tendency of an organism to sacrifice itself for another will increase (be selected for) provided that the reproductive success of closely enough related

[101] The evolution of eusociality: http://www.ncbi.nlm.nih.gov/pubmed/20740005.1

[102] Dugatkin, L.A. 2007. Inclusive Fitness Theory from Darwin to Hamilton. http://www.genetics.org/content/176/3/1375.full

organisms is increased sufficiently. We will see that we can apply this logic to a wide range of situations and it provides an evolutionary mechanism driving the appearance and preservation of various social behaviors.

That said, the situation can be rather more complex. Typically, to work, inclusive fitness requires a close relationship to the recipient of the beneficial act. So how can we assess this relationship? How does one individual know that it is making a sacrifice for its relatives and not just a bunch of (semi-) complete strangers? As social groups get increasingly large, this becomes a more and more difficult task. One approach is to genetically link the social trait (e.g. altruistic behavior) to a physically discernible trait, like smell or a detectable structure. This is sometimes called a "green beard" trait. Individuals that cooperate (that is, display social behavior) with other organisms do so only when the green beard trait is present. The presence of the green beard trait indicates that the organism is related to the cooperator. Assume a close linkage between the two traits (social and visible), one can expect social behavior from an apparent (distantly related) stranger. In some cases, a trait may evolve to such a degree that it becomes part of an interconnected set of behaviors. Once, for example, humans developed a brain sufficiently complex to do what it was originally selected for (assuming that it was brain complexity that was selected, something we might never know for sure), this complexity may have produced various unintended byproducts. Empathy, self-consciousness, and a tendency to neurosis may not be directly selected for but could be side effects of behavioral processes or tendencies that were. As a completely unsupported (but plausible) example, the development of good memory as an aid to hunting might leave us susceptible to nightmares. Assume, for the moment (since we are speculating here), that empathy and imagination are "unintended" products of selective processes. Once present, they themselves can alter future selection pressures and they might not be easy to evolve away from, particularly it they are mechanistically linked to a trait that is highly valued (that is, selected). The effects of various genetic mutations on personality and behavior strongly supports the idea that such traits have a basis in one's genotype. That said, this is a topic far beyond the scope of this book.

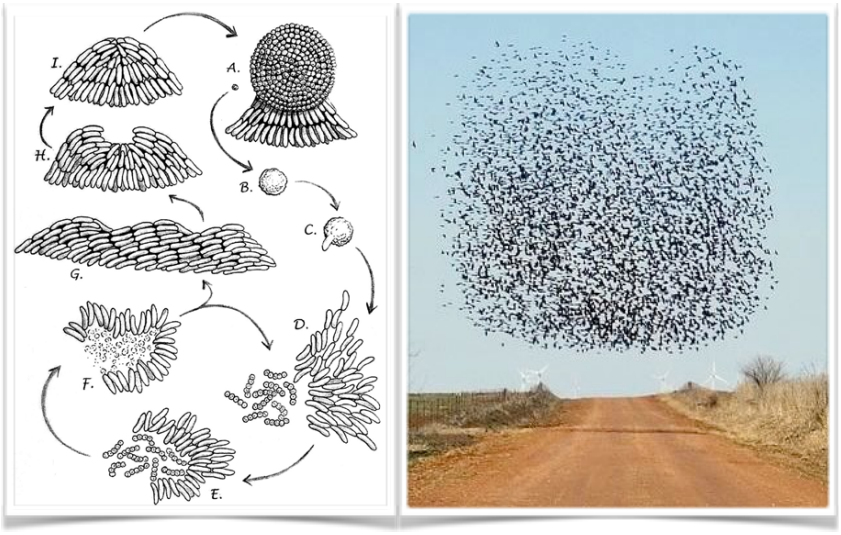
Group selection

A proposed alternative to inclusive fitness has been the concept of group selection. In this type of evolutionary scenario, small groups of organisms of the same species are effectively acting as single (perhaps colonial) organisms and it is the reproductive success of the group compared to other groups of the organism, that benefit from the presence of cooperative and altruistic traits. Again, the mathematical analysis is similar (and it is claimed that mathematically group and kin selection are equivalent).[103] The costs of a trait must be offset by the benefits, but now the key factor is membership in a particular group (and typically, members of a group tend to be related to one another). The life cycle of the bacterium *Myxococcus xanthus* provides an example of this type of behavior. When environmental conditions are harsh, the cells aggregate into dense, 100 µm diameter, "fruiting bodies" containing about 100,000 stress resistant spores each. When the environment improves, and prey becomes available, the spores are released en mass and return to active life. They move and feed in a

[103] Mathematics of kin- and group-selection: formally equivalent? http://www.ncbi.nlm.nih.gov/pubmed/19929970

cooperative manner through the release of digestive enzymes, which because they are acting in a quorum mode, can reach high levels.[104] A well coordinated group is expected to have a significant reproductive advantage compared to a more anarchic collection of cells.

While their functional roles are clearly different, analogous types of behavior are seen in flocks of birds, schools (or shoals) of fish, swarms of bees, and blooms of algae.[105] Each of these examples represent a cooperative strategy by which organisms can gain a reproductive advantage over those that do not display this behavior. While the original behavior is likely the result of kin selection, in the wild it is possible that different groups (communities) could be in competition with one another, and the group that produces the most offspring, that is, reproductively successful groups, will come to dominate.



Defense against social cheaters

Now an interesting question arises: within a social organization, such as a group of cooperating microbes or hunters,[106] we can expect that, through mutation (or through other behavioral mechanisms), cheaters will arise. What do we mean by a cheater? Imagine a bacterium within a swarm, a cell in an organism, or an animal in a social group that fails to obey the rules. In the case of slime mold aggregates, imagine that a cell can avoid becoming a non-reproductive stalk, but rather always differentiates to form a reproductively competent spore. What happens over time? One plausible scenario would be that this spore cell begins its own clone of migratory amoeba, but when conditions change so that aggregation and fruiting body formation occur, most of the cells avoid forming the stalk. We would predict that the resulting stalk, required to lift the spore forming region above the soil and necessary for spore dispersal, would be short or non-existent and so would reduce the efficiency of dispersion between different aggregates as a function of the number of individuals with a cheater phenotype present. If dispersion is important for reproductive success, there would be selection for those who maintain it and against cheaters.

Now the question is, once a social behavior has evolved, under what conditions can evolutionary mechanisms maintain it. One approach is to link the ability to join a social group with

---

[104] Evolution of sensory complexity recorded in a myxobacterial genome: http://www.ncbi.nlm.nih.gov/pubmed/17015832

[105] How Does Social Behavior Evolve? http://www.nature.com/scitable/knowledge/library/how-does-social-behavior-evolve-13260245

[106] An interesting read: The stag hunt and the evolution of social structure. http://bilder.buecher.de/zusatz/22/22362/22362426_lese_1.pdf

various internal and external mechanisms. This makes cooperators recognizable and works to maintain a cooperative or altruistic trait even in the face of with individual costs. There are a number of plausible mechanisms associated with specific social traits. This is, however, a topic that can be easily expanded into an entire course. We will focus on common strategies with occasional references to specific situations. To illustrate these mechanisms, we will use human tissues as an example. We can consider the multicellular organism as a social system. The cells that compose it have given up their ability to reproduce a new organism for the ability to enhance the reproductive success of the whole organism. In this context, cancer is a disease that arises from mutations that lead to a loss of social control. Cells whose survival and reproduction is normally strictly controlled lose that control; they become anti-social. They begin to divide in an uncontrolled manner, disrupt the normal organization of the tissue in which they find themselves, and can even breakaway, migrate, and colonize other areas of the body, a process known as metastasis. The controlled growth of the primary tumor and these metastatic colonies leads eventually to the death of the organism as a whole.

When we think about maintaining a social behavior, we can think of two general mechanisms: intrinsic and extrinsic policing. For example, assume that a trait associated with the social behavior is also linked to or required for cellular survival. In this case, a mutation that leads to the loss of the social trait may lead to cell death. Consider this in the context of cancer. Normal cells can be considered to be addicted to normality. When their normality is disrupted they undergo a type of active cell death, known as apoptosis. A cell carrying a mutation that would enable it to grow in an uncontrolled and inappropriate manner will likely kill itself before it can produce significant damage.[107] For a tumor to grow and progress, other mutations must somehow disrupt and inactivate the apoptotic process. The apoptotic process reflects an intrinsic-mode of social control. It is a little like the guilt experienced by (some) people when they break social rules or transgress social norms. The loss of social guilt or embarrassment is analogous to the inhibition of apoptosis in response to various cues associated with abnormal behavior.

In humans, and a number of other organisms, there is also an extrinsic social control system. This is analogous to the presence of external policeman (guilt and apoptosis are internal policemen). Mutations associated with the loss of social integration–that is, the transformation of a cell to a cancerous state–can lead to changes in the character of the cell. Specialized cells can recognize these changes by specialized cells of the organism's immune system; these cells recognize the mutant cell and kill it.[108] Of course, given that tumors occur and kill people, we can assume that there are mutations that enable tumor cells to avoid what is known as immune system surveillance. As we will see, one part of the cancerous phenotype is often a loss of normal mutation and genome repair systems; in effect, the mutant cell has increased the number of mutations, and consequently, the genetic variation in the cancer cell population. While many of these variants are lethal, the overall effect is to increase the rate of cancer cell evolution. This leads to an evolutionary race. If the cancer is killed by intrinsic and extrinsic control systems, no disease occurs. If, however, the cancer evolves fast enough to avoid death by these systems, the cancer will progress and spread. As we look at a range of

[107] Apoptosis in cancer: http://carcin.oxfordjournals.org/content/21/3/485.full

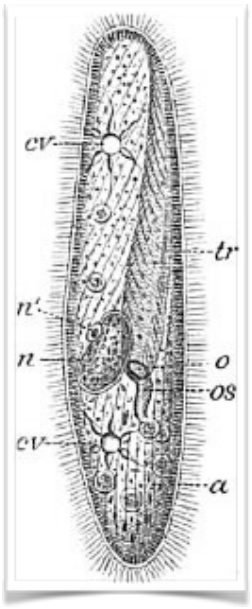
[108] Immune recognition of self in immunity against cancer: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC503781/

social systems, from cooperating bacteria to complex societies, we see examples of intrinsic and extrinsic control.

Driving the evolutionary appearance of multicellular organisms

Now that we have some idea about cooperative behaviors and how evolutionary mechanisms can select and maintain them, we can begin to consider their role in the evolution of multicellular organisms.[109] As we have mentioned there are a number of strategies that organisms take to exploit their environment. Most prokaryotes are unicellular, but some can grow to gigantic sizes. For example, the bacterium *Epulopiscium fishelsoni*, inhabits the gut of brown surgeonfish (*Acanthurus nigrofuscus*); it can grow to more than 600 µm in length. As we will see (from an experimental perspective), the cells of the unicellular eukaryotic algae of the genus *Acetabularia* can be more than 10 cm in length. Additionally, a number of multicellular prokaryotes exhibit quite complex behaviors. A particularly interesting one is a species of bacteria that form multicellular colonial organisms that sense and migrate in response to magnetic fields.[110] Within the eukaryotes, there are both unicellular and microscopic species (although most are significantly larger than the unicellular prokaryotes), as well as a range of macroscopic and multicellular species, includes those we are most likely to be familiar, namely animals, plants, and fungi.

What drove the appearance of multicellular organisms? Scientists have proposed a number of theoretical and empirically supported models. Researchers have suggested that predation is an important driver, either enabling the organisms to become better (or more specific) predators or to avoid predation. For example, Borass et al.,[111] reported that the unicellular algae *Chlorella vulgaris* (5-6 µm in diameter) is driven into a multicellular form when grown together with a unicellular predator *Ochromonas vallescia,* which typically engulfs its prey. They observed that over time, *Chlorella* were found in colonies that *Ochromonas* could not ingest.



At this point, however, what we have is more like a colony of organisms rather than a colonial organism or a true multicellular organism. The change from colony to organism appears to involve cellular specialization, so that different types of cells within the organism come to carry out different functions. The most dramatic specialization being that which gives rise to the next generation of organisms, the germ cells, and those that function solely within a particular organism, the somatic cells. At the other extreme, instead of producing distinct types of specialized cells, a number of unicellular eukaryotes, known as protists, have highly complex cells that display complex behaviors such as directed motility, predation, osmotic regulation, and digestion. But such specialization can be carried out much further in multicellular organisms, where there is a socially-based division of labor. The stinging cells of jellyfish

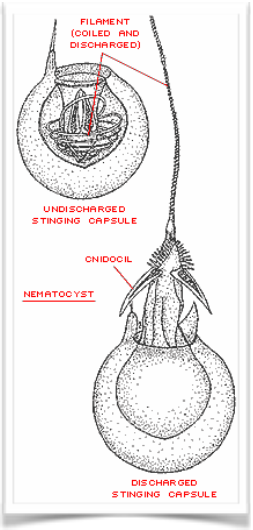
---

[109] The evolutionary-developmental origins of multicellularity: http://www.amjbot.org/content/101/1/6.long

[110] A novel species of ellipsoidal multicellular magnetotactic prokaryotes from Lake Yuehu in China. http://www.ncbi.nlm.nih.gov/pubmed/24725306

[111] Phagotrophy by a flagellate selects for colonial prey: A possible origin of multicellularity: http://link.springer.com/article/10.1023%2FA%3A1006527528063

provide a classic example where highly specialized cells deliver poison to any organism that touches them through a harpoon-like mechanism. The structural specialization of these cells makes processes such as cell division impossible and typically a stinging cell dies after it discharges. Such cells are produced by a process known as terminal differentiation, which we will consider later (but only in passing). While we are used to thinking about individual organisms, the same logic can apply to groups of distinct organisms. The presence of cooperation extends beyond a single species, into ecological interactions in which organisms work together to various degrees. From a related perspective, one can view cancer as a disease in which the cooperative behavior of cells breaks down.



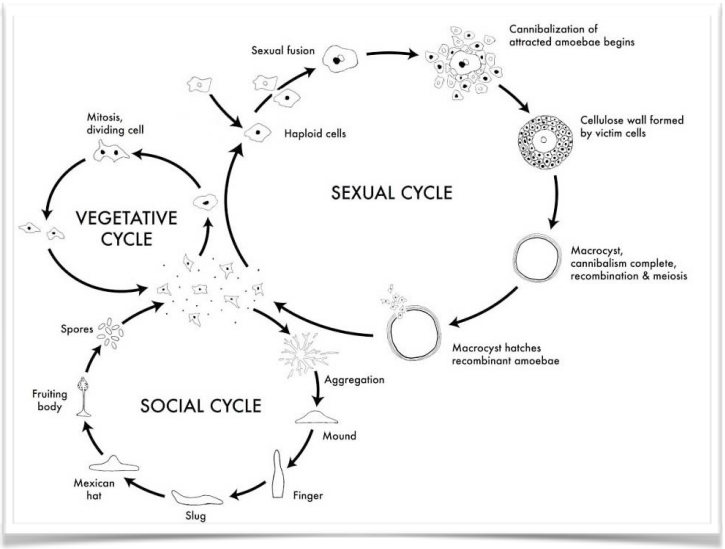
Based on the study of a range of organisms and their genetic information, we have begun to clarify the origins of multicellular organisms. Such studies indicate that multicellularity has arisen independently in a number of eukaryotic lineages. This strongly suggests that in a number of contexts, becoming multicellular is a successful way to establish an effective relationship with the environment.

Questions to answer & to ponder:

- Why does a quorum signal need to be secreted (released) from the organism?
- What components are necessary for quorum signaling?
- Why is r (the relationship between organisms) never 0 (although it can be quite small).
- What types of mechanisms can be used to address the effects of cheaters in a population?
- How would these mechanisms apply to social interactions?
- Make a model of the mechanisms that can lead to the evolution of social interactions within an organism and within a population.

Origins and implications of sexual reproduction

One type of social interaction that we have mentioned in passing is sex. In many of unicellular eukaryotes, there are a number of distinct mating types. Reproduction involves cooperative interactions between organisms of different mating types. Through mechanisms we will consider later, the outcome of sexual reproduction leads to increased diversity among offspring. So what are the common hallmarks of sexual reproduction? Let us return to the slime mold *Dictyostelium* as a exemplar. We have already considered its asexual life cycle, but *Dictyostelium* also has a sexual life cycle. Under specific conditions, two amoeboid cells of different mating types will fuse. The original cells will be haploid, meaning that they have a single copy of their genome. The resulting fused cell will have two copies of the genetic



material; it is diploid. This diploid cell will go through a series of events, eventually producing four haploid cells through the process known as meiosis. During meiosis, genetic material is shuffled, so that the genotypes of the haploid cells that emerge from the sexual process are different from those of the haploid cells that originally fused with one another.

Sexual dimorphism

What, biologically, defines whether an organism is female or male, and why does it matter? The question is largely irrelevant in unicellular organisms with multiple mating types. For example, the microbe *Tetrahymena* has seven different mating types, all of which appear morphologically identical. A individual *Tetrahymena* cell can mate with another individual of a different mating type but not of the same mating type as itself. Mating involves fusion and so the identity of the parents is lost. There is another cost of a sexual mode of reproduction in unicellular organisms, since they need to find a partner, something that is unnecessary in the asexual state.

In multicellular organisms, the parents do not themselves fuse with one another. Rather they produce cells, known as gametes, that do. Also, instead of two or more mating types, there are two sexes, male and female. This, of course, leads to the question, how do we define male and female? The answer is superficially simple but its implications are profound. Which sex is which is defined by the relative size of the fusing cells the organism's produce. The larger fusing cell is termed the egg and the organism that produces it is termed a female, while the smaller is termed a sperm and the organism that produces it is termed a male. At this point, we should note the limits of these definitions. There are organisms that can change their sex, which is known generically as sequential hermaphroditism. For example, in a number of fish it is common that all individuals originally develop into males, but based on environmental cues, the largest of these males changes sex to become female. Alternatively, one organism can produce both eggs and sperm; such an organism is known as a constitutive hermaphrodite.

The size difference between male and female gametes changes the stakes for the two sexes. Because of the larger size of the egg, the female invests more energy in its production (per egg) than a male invests in the production of a sperm cell. It is therefore relatively more important, from the perspective of reproductive success, that each egg produce a viable and fertile offspring. As the cost to the female of generating an egg increases, the more important the egg's reproductive success becomes. Because sperm are relatively cheap to produce, the selection pressure associated with their production is relatively less than that associated with producing an egg. The end result is that there emerges a conflict of interest between females and males. This conflict of interest increases as the disparity in the relative investment per gamete or offspring increases.

This is the beginning of an evolutionary economics cost-benefit analysis. First there is what is known as the two-fold cost of sex, which is associated with the fact that each asexual organism can produce offspring but that two sexually reproducing individuals must cooperate to produce offspring. Other, more specific factors influence an individual's reproductive costs. For example, the cost to a large female laying a small number of small eggs that develop independently is less than that of a small

female laying a large number of large eggs. Similarly, the cost to an organism that feeds and defends its young for some period of time after they are born (that is, leave the body of the female) is larger than the cost to an organism that lays eggs and leaves them to fend for themselves. Similarly, the investment of a female that raises its young on its own is different from that of the male that simply supply sperm and leaves. As you can imagine, there are many many different reproductive strategies (many more than we can consider), and they all have implications. For example, a contributing factor in social evolution is that where raising offspring is particularly biologically expensive, cooperation between the sexes or groups of organisms in child rearing can improve success and increase the return on the investment of the organisms involved. It is important to remember (and be able to apply in specific situations) that the reproductive investments, and so evolutionary interests, of the two sexes often diverge dramatically from one another.

Consider, for example, the situation in placental mammals, in which fertilization occurs within the female and relatively few new organisms are born from any one female. The female must commit resources to supporting the new organisms from the period from fertilization to birth. In addition, female mammals both protect their young and feed them with milk, using specialized glands (mammary glands). Depending on the species, the young are born at various stages of development, from the active and frisky (such as goats) to the relatively helpless (humans). During the period when the female feeds and protects its offspring, the female is more stressed and vulnerable than other times. Under specific conditions, cooperation with other females (as can occur within a pack) or with a specific male (typically the father) can greatly increase the rate of survival of both mother and offspring, as well as the reproductive success of the male. But consider this; how does a cooperating male know that the offspring he is helping to protect and nurture are his? Spending time protecting and gathering food for unrelated offspring is time when the male cannot produce new offspring and it will greatly reduce the male's reproductive success. Carrying this logic out to its conclusion can lead to behaviors such as males guarding of females from interactions with other males.

As we look at the natural world, we see a wide range of sexual behaviors, from males who sexually monopolize multiple females (polygyny) to polyandry, where the female has multiple male "partners." In some situations, no pair bond forms between male and female, whereas in others male and female pairs a re stable and (largely) exclusive. In some cases these pairs last for extremely long times; in others there is what has been called serial monogamy, where pairs form for a while, break up, and new pairs form (this seems relatively common among performing arts celebrities). Sometimes females will mate with multiple males, a behavior that is thought to confuse males (they cannot know which offspring are theirs) and so reduces infanticide by males.[112]

It is common that while caring for their young, females are generally reproductively inactive. Where a male monopolizes a female, the arrival of a new male who displaces the previous male can lead to behaviors such as infanticide. By killing the young, the female becomes reproductively active and able to produce offspring related to the new male. There are situations, for example in some spiders, in which the male will allow itself to be eaten during the course of sexual intercourse as a type of nuptial gift, which both blocks other males from mating with a female (who is busy eating) and

[112] Promiscuous females protect their offspring. http://www.ncbi.nlm.nih.gov/pubmed/16701243

increases the number of offspring that result from the mating. This is an effective reproductive strategy for the male if its odds of mating with a female are low: better (evolutionarily) to mate and die than never to have mated at all. An interesting variation on this behavior is described in a paper by Albo et al.[113] Male *Pisaura mirablis* spiders offer females nuptial gifts, in part perhaps to avoid being eaten during intercourse. Of course, where there is a strategy, there are counter strategies. In some cases, instead of an insect wrapped in silk, the males offers a worthless gift, an inedible object wrapped in silk. Females cannot initially tell that the gift is worthless but quickly terminate mating when they discover it is. This reduces the odds of a male's reproductive success. As deceptive male strategies become common, females are likely to display counter strategies. For example, a number of female organisms store sperm from a mating and can eject that sperm and replace it with that of another male (or multiple males) obtained from subsequent mating events.[114] There is even evidence that in some organisms, such as the wild fowl *Gallus gallus*, females can bias against fertilization from closely related males, a situation known as cryptic female choice, cryptic since it is not overtly visible in terms of who the female does or does not mate with.[115] And so it goes, each reproductive strategy can lead to counter measures.[116] For example, in species in which a male guards a set of females (its harem), groups of males can work together to distract the male, allowing members of their group to mate with the females. These are only a few of the mating and reproductive strategies that exist in the living world.[117] Molecular studies that can distinguish an offspring's parents suggest that cheating by both males and females is not unknown even among highly monogamous species. The extent of cheating will, of course, depend on the stakes. The more negative the effects on reproductive success, the more evolutionary processes will select against it.

In humans, a female can have at most one pregnancy a year, while a totally irresponsible male could, in theory at least, make a rather large number of females pregnant during this same period of time. Moreover, the biological cost of generating offspring is substantially greater for the female, compared to the male.[118] There is a low but real danger of the death of the mother during pregnancy, whereas males are not so vulnerable, at least in this context. So, if the female is going to have offspring, it would be in her evolutionary interest that those offspring be as robust as possible, meaning that they are likely to survive and reproduce. How can the female influence that outcome? One approach is to control fertility, that is the probability that a "reproductive encounter" results in pregnancy. This is accomplished physiologically, so that the odds of pregnancy increase when the female has enough resources to successfully carry the pregnancy to term. It should be noted that these are not

---

[113] Worthless donations: male deception and female counter play in a nuptial gift-giving spider: http://www.biomedcentral.com/1471-2148/11/329

[114] Evolution: Sperm Ejection Near and Far: http://www.sciencedirect.com/science/article/pii/S0960982204004452

[115] Cryptic female choice favors sperm from major histocompatibility complex-dissimilar males: http://rspb.royalsocietypublishing.org/content/280/1769/20131296.full

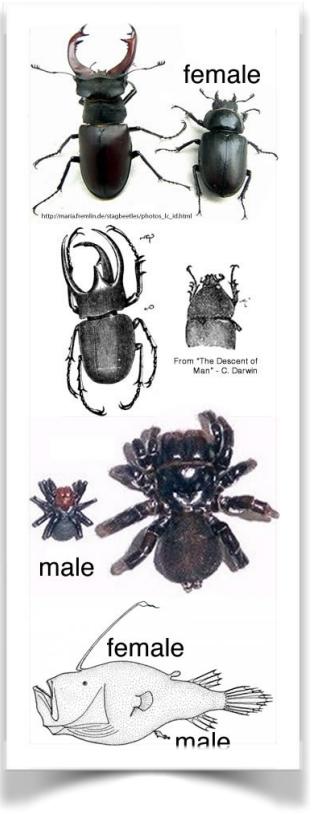
[116] Sperm Competition and the Evolution of Animal Mating Systems: http://www.sciencedirect.com/science/book/9780126525700

[117] The Evolution of Alternative Reproductive Strategies: Fitness Differential, Heritability, and Genetic Correlation Between the Sexes: http://jhered.oxfordjournals.org/content/92/2/198.full

[118] 'Parental investiment: http://www.anthro.utah.edu/PDFs/maynardsmith77parenting.pdf

conscious decisions on the part of the female but physiological responses to various cues. There are a number of examples within the biological world where females can control whether a particular mating is successful (that is, produces offspring). For example, female wild fowl are able to bias the success of a mating event in favor of dominant males by actively ejecting the sperm of subdominant males.[119] One might argue that the development of various forms of contraception are yet another facet of this type of behavior, but one in which females (and males) consciously control reproductive outcomes.

Sexual selection



As we have already noted, it is not uncommon to see morphological and behavioral differences between the two sexes. Sometimes the sexual dimorphism and associated behavioral differences between the sexes are profound; they can even obscure the fact that the two sexes are actually members of the same species. In some cases, specific traits associated with one sex can appear to be maladaptive, that is, they might be expected to reduce rather than enhance an organism's reproductive potential.[120] The male peacock's tail, the gigantic antlers of male moose, or the bright body colors displayed by some male birds are classic examples. Darwin recognized the seriousness of this problem for evolutionary theory and addressed it in his book *The Descent of Man and Selection in Relation to Sex* (1871). Where the investment of the two sexes in successful reproduction is not the same, as is often the case, the two sexes may have different and potentially antagonistic reproductive strategies. Organisms of different sexes may be "looking" for different traits in their mates. In general, the larger parental investment in the production and rearing of offspring, the less random is mating and the more prominent are the effects of sexual selection.[121] It is difficult not to place these behaviors in the context of conscious behaviors, (looking, wanting, etc.), in fact these are generally the result of evolved behaviors and do not imply self-conscious decision-making. This may even be the case among organisms, like humans, who are self-conscious. What is happening is an interaction between costs, benefits, and specific behaviors.

Consider an example in which the female does not require help in raising offspring but in which the cost to the female is high. Selection would be expected to favor a behavior in which females mate preferentially with the most robust males available. Females will select their mates based on male phenotype on the (quite reasonable) assumption that the most robust appearing male will be the most likely to produce the most robust offspring. In the context of this behavior, the reproductive success of a male would be enhanced if they could advertise their genetic robustness, generally through visible and

---

[119] Female feral fowl eject sperm of subdominant males: http://www.ncbi.nlm.nih.gov/pubmed/10866198

[120] "Flaunting It' - Sexual Selection and the Art of Courtship: http://youtu.be/g3B8hS80k6A

unambiguous features.[122] To be a true sign of the male's robustness, this advertisement needs to be difficult to fake and so reflects the true state of the male. For example consider scenarios involving territoriality. Individuals, typically males, establish and defend territories. Since there are a limited number of such territories and females only mate with males that have a territory, only the most robust–as defined in terms of the ability to establish and defend a territory–are reproductively successful. An alternative scenario involves males monopolizing female's sexually. Because access to females is central to their reproductive success, males will interact with one another to establish a dominance hierarchy, typically in the form of one or more alpha males. Again, the most robust males are likely to emerge as alpha males, which in turn serves the reproductive interests of the females. This type of dominance behavior is difficult or impossible to fake. But, cooperation between non-alpha males can be used to thwart the alpha male's monopolization of females.

Now consider how strategies change if the odds of successful reproduction are significantly improved if the male helps the female raise their joint offspring. In this situation, there is a significant reproductive advantage if females can accurately identify those males that display this type of reproductive loyalty.[123] Under these conditions, that is the shared rearing of offspring with a committed male, females will be competing with other females for access to these males. Moreover, it is in the male's interest to cooperate with fertile females, and often females (but not human females) advertise their state of fertility (that is the probability that mating with them will produce offspring) through external signals. There are of course, alternative strategies. For example, groups of females (sisters, mothers, daughters, aunts, and grandmothers) can cooperate with one another, thereby reducing the importance of male cooperation. At the same time, there may be what could be termed selection conflicts. What happens if the most robust male is not the most committed male? A female could maximize their reproductive success by mating with a robust male and bonding with a committed male, who helps rear another male's offspring. Of course this is not in the committed male's reproductive interest. Now selection might favor male's that cooperate with one another to ward off robust but promiscuous and transient males. Since these loyal males already bond and cooperate with females, it may well be a simple matter for them to bond and cooperate with each other. In a semi-counter intuitive manner, the ability to bond with males could be selected for based on its effect on reproductive success with females. On the other hand, a male that commits himself to a cooperative (loyal and exclusive) arrangement with a female necessarily limits his interactions with other females. This implies that he will attempt to insure that the offspring he is raising are genetically related to him.

The situation quickly gets complex and many competing strategies are possible. Different species make different choices depending upon their evolutionary history and environmental constraints. As we noted above, secondary sexual characteristics, that is, traits that vary dramatically between the two sexes, serve to advertise various traits, including heath, loyalty, robustness, and fertility. The size and symmetry of a beetle's or an elk's antlers or a grasshopper's song communicate

---

[122] In Male Rhinoceros Beetle, Horn Size Signals Healthy Mate: http://www.aaas.org/news/releases/2012/0726sp_plumage.shtml

[123] http://www.madsci.org/posts/archives/2012-05/1336600952.Ev.r.html

rather clearly their state of health.[124] The tail of the male peacock is a common example, a male either has large, colorful and symmetrical tail, all signs of a health or it does not – there is little room for ambiguity. These predictions have been confirmed experimentally in a number of systems; the robustness of offspring does correlate with the robustness of the male, a win for evolutionary logic.[125]

It is critical that both females and males can correctly read, that is respond to, various traits. For example, males can often read the traits of other males in order to determine whether they are likely to win a fight with another male, a fight that could end up crippling both males. A more complex question is how does a female determine whether a male is committed, and vice versa? As with advertisements of overall robustness, we might expect the female to look for behaviors that are difficult to fake. So how does one unambiguously signal one's propensity to loyalty and willingness to cooperate? As noted above, one could use the size and value of nuptial gifts. The more valuable (that is, the more expensive and difficult the gift is to attain), the more loyal the female can expect the gift giver to be. On the other hand, once valuable gift-giving is established, one can expect the evolution of traits in which the cost of the gift given is reduced and by which the receiver needs to be skeptical about the actual nature of the gift.

One of the most robust and reliable findings in the scientific literature on interpersonal attraction is the overwhelming role played by physical attractiveness in defining the ideal romantic partner (Hatfield & Sprecher, 1986; Jackson, 1992). Both men and women express marked preference for an attractive partner in a noncommitted short-term (casual, one night stand) relationship (Buss & Schmitt, 1993).

For committed long-term relationships, females appear to be willing to relax their demand for a partner's attractiveness, especially for males with high social status or good financial prospects (for a review see Buss, 1999).

Males also look for various personality qualities (kindness, understanding, good parental skills) in their search for long-term mating partners, but unlike females, they assign disproportionately greater importance to attractiveness compared to other personal qualities (Buss, 1999).

This points out a general pattern. When it comes to sexual (and social) interactions, organisms have evolved to know the rules involved. If the signs an organism must make to another are expensive, there will be selective pressure to cheat. Cheating can be suppressed by making the sign difficult or impossible to fake, or by generating counter-strategies that can be used to identify fakes. These biological realities produce many behaviors, some of which are disconcerting. These include sexual cannibalism and male infanticide, both mentioned above. What we have not considered as yet is the conflict between parents and offspring. Where the female makes a major and potentially debilitating investment in its offspring, it may respond to signs of reproductive distress that might threaten the survival of the female by spontaneously aborting the offspring. Of course, this is not in the interest of the offspring and mechanisms exist to maintain pregnancy, even if it risks the

The paramount importance of attractiveness in males' mate choices has been recently demonstrated by using the distinction between necessities (i.e., essential needs, such as food and shelter) and luxuries (i.e., objects that are sought after essential needs have been satisfied, such as a yacht or expensive car) made by economists.

Using this method, Li et al., (2002) reported that males treat female attractiveness as a necessity in romantic relationships; given a limited "mating budget," males allocate the largest proportion of their budget to physical attractiveness rather than to other attributes such as an exciting personality, liveliness, and sense of humor.
- from Mating strategies for young women by Devendra Singh (2004).

---

[124] Attractiveness of grasshopper songs correlates with their robustness against noise:http://beheco.oxfordjournals.org/content/early/2011/05/08/beheco.arr064.full

[125] Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1688278/pdf/2UET8WJ5TD06WFUT_264_297.pdf

life of the mother. There are many variations of reproductive behavior to be found in the biological world and a full discussion is beyond the scope of this course, but it is a fascinating subject with clear and complex implications for human behavior. Part of the complexity arises from the fact that the human brain (and the mind it generates) can respond in a wide range of individualistic behaviors, not all of which seem particularly rational. It may well be many of these are emergent behaviors. That means that they were not directly selected for but emerged in the course of evolution and once present, play important roles in subsequent organismic behavior (and presumably evolution).

Curbing runaway selection

Sexual selection can lead to what has been termed runaway selection. For example, the more prominent the peacock male's tail the more likely he will find a mate even though larger and larger tails may also have negative effects. There will be both positive and negative selection for tail size, which will be influenced by the overall probability that a particular male males successfully. Selection does not ever really run away, but settles down when the positive (in terms of sexual success) and negative (in turns of various costs) of a trait come to equal each other. Sufficient numbers of male peacocks emerge as reproductively successful even if many males are handicapped by their tails and fall prey to predators. For another example, consider the evolution of the extremely large antlers associated with male dominance and mate accessibility such as occurred in *Megaloceros giganteous.* These antlers could also act to inhibit the animal's ability to move through heavily wooded areas. In a stable environment, the costs and benefits associated with the development of sexual advertising would be expected to balance out; selection would produce an optimal solution. But if the environment changes, pre-existing behavior and phenotypes could act to limit an organism's ability to adapt or to adapt fast enough to avoid extinction. In the end, as with all adaptations, there is a balance between the positive effects of a trait, which lead to increased reproductive success, and their negative effects, which can influence survival. The optimal form of a trait may not be stable over time, particularly if the environment is changing.

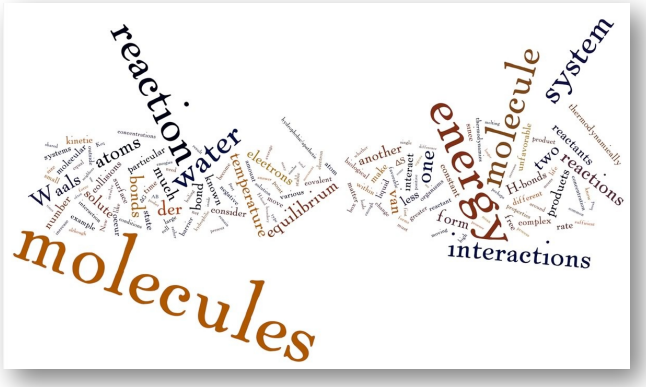
Summary: Social and ecological interactions apply to all organisms, from bacteria to humans. They serve as a counter-balance to the common caricature of evolution as a ruthless and never ceasing competition between organisms. This hyper-competitive view, often known as the struggle for existence or Social Darwinism, was not in fact supported by Darwin or by scientifically-established evolutionary mechanisms, but rather by a number of pundits who used it to justify various political positions, particularly arguing against social programs that helped the poor at the "expense" of the wealthy. Assuming that certain organisms were inherently less fit, and that they could be identified, this view of the world gave rise to Eugenics, the view that inferior people should be killed, removed, or sterilized, less they overwhelm a particular culture. Eugenics was a particularly influential idea in the United States in the early part of the 20th century and inspired the Nazis in Germany. What is particularly odd about this evolutionary perspective is that it is actually anti-evolutionary, since if the unfit are actually unfit, they could not possibly take over a population.

Questions to answer & ponder

- What does it mean to cheat, in terms of sexual selection - is the "cheating" organism actually being consciously deceptive?
- Why do the different sexes (of the same species) often display different secondary sexual traits?
- If the two sexes appear phenotypically identical, what might you conclude about their reproductive behaviors?
- What types of "cheating" behaviors do females use with males? What about males with females?
- What are the costs involved when a male tries to monopolize multiple females? What are the advantages?
- What limits runaway selection?
- Why would you expect female infanticide to be extremely rare? When might it make evolutionary sense?
- Is Devendra Singh right about mating budgets?
- Is the schooling or herd behavior seen in various types of animals (such as fish and cows) a homologous or an analogous trait?

5. Molecular interactions, thermodynamics, and reaction coupling

In which we drastically change gears and move from evolutionary mechanisms to the physicochemical properties of organisms. We consider how molecules interact and react with one another, how these interactions and reactions determine the properties of substances and systems.



A very little thermodynamics

While the diversity of organisms and the unique properties of each individual organism are the products of evolutionary processes, initiated billions of years ago, it is equally important to recognize that all biological systems and processes, from growth and cell division to thoughts and feelings, obey the rules of chemistry and physics - in particular the laws of thermodynamics. What makes biological systems unique is that, unlike simpler physicochemical systems which, under the right conditions move toward thermodynamic equilibrium, organisms must maintain a non-equilibrium state in order to remain alive. While a chemical reaction system is easy to assemble *de novo*, every biological system has been running continuously for billions of years. So, before we continue we have be clear about what we mean and imply when we say that a system is at equilibrium versus being in a obligate non-equilibrium state.

To understand the meaning of thermodynamic equilibrium we have to learn the see the world differently, and learn new meanings for a number of words. First we have to make clear the distinction between the macroscopic world that we directly perceive and the sub-microscopic, molecular world that we can understand based on scientific observations and conclusions - it is this molecular world that is particularly important in the context of biological systems. The two, the macroscopic and the molecular, behave very differently. To illustrate this point, we will use a simpler model that displays the basic behaviors that we want to consider but is not as complex as a biological system. In our case, let us consider a small, well insulated air-filled room in which there is a table with a bar of gold (we use gold since it is chemically rather inert, that is, un-reactive. Iron bars could rust, which would complicate things). In our model at the beginning the room is a cosy 70 ºF (~21 ºC) and the bar of gold is at 200ºC. What will happen? Our first task is the define the system – that is the part of the universe are we interested in? We could define the system as the gold bar or the room with the gold bar in it (notice, we are not really concerned about how the system came to be the way it is.) We could, if we wanted to, demonstrate quite convincingly that its history will not influence its future behavior – an important difference between biological and physical systems. For now we will use the insulated room as the system, but it doesn't really matter as long as we clearly define what we consider the system to be.

Common sense tells us that there will be an energy transfer between the gold bar and the rest of the room and that the temperature of the gold bar will decrease over time. Why do you think that is? Why doesn't the hot bar get hotter and the room get cooler? We will come back to this question shortly. What may not be quite as obvious is that the temperature of the room will increase slightly as well.

Eventually the block of gold and the room will reach the same temperature and the system will be said to be at equilibrium.

Remember we defined the system as isolated from the rest of the universe, but what does that mean? Basically no matter or energy passes into or out of the room – such a system is said to be a closed system. Because it is a closed system, once the system reaches its final temperature, NºC, no further change will occur. At that point the system is said to be at equilibrium. At the macroscopic level nothing more happens. This does not mean, however, that nothing is going on. If we could look at the molecular level we would see that molecules of air are moving, colliding with one another and the bars and the table constantly. (You could predict what would happen if there was no air in the room.) The molecules within the bars and the table are vibrating. The speed of these molecular movements is a function of temperature, the higher (or lower) the temperature the faster (or slower) these motions would be. As we will consider further on, all of the molecules in the system have kinetic energy, which is the energy of motion. Through their interactions, the kinetic energy of any one particular molecule will be constantly changing. At the molecular level the system is dynamic, even though at the macroscopic level it is static. We will come back to this insight repeatedly in our considerations of biological systems.

What is important about a system at equilibrium is that it is static. Even at the molecular level, while there is still movement, there is no net change. The energy of two colliding molecules is the same after a collision as before, even though the energy may be distributed differently between the colliding molecules. The system as a whole cannot really do anything. In physical terms, it cannot do work - no macroscopic changes are possible. This is a weird idea, since (at the molecular level) things are still moving. So, if we return to living systems, which are clearly able to do lots of things, including moving macroscopically, growing, thinking, and such, it is clear that they cannot be at equilibrium.

We can ask, what is necessary to keep a system from reaching equilibrium? The most obvious answer (we believe) is that unlike our toy system of a closed room, the system must be open, that is, energy and matter must be able to enter and leave it. An open system is no longer isolated from the rest of the universe, it is part of it. For example, we could imagine a system in which energy, in the form of radiation, can enter and leave our room. We could maintain a difference in the temperature between the two bars by illuminating one bar and removing heat from the room as a whole. A temperature difference between two bars could then (in theory) produce what is known as a heat engine, which can do work. As long as we continue to heat one block and remove heat from the rest of the system, we can continue to do work - macroscopically observable changes can happen.

Cryptobiosis: At this point we have characterized organisms as dynamic, open, non-equilibrium systems. An apparent exception to the dynamic aspect of life are organisms that display a rather special phenotypic adaptation, known generically as cryptobiosis. Organisms, such as the tardigrad (or water bear), can be freeze-dried and persist in a “suspended animation” state for decades. What is critical, however, is to note that when in this cryptobiotic state the organism is not at equilibrium, in much the same way that a piece of wood in air is not at equilibrium, but capable of reacting. An organism in a cryptobiotic state is certainly not dead; it can be reanimated when returned

to normal conditions.[126] Cryptobiosis is an genetically-based adaptation that takes energy to produce and energy is used to emerge from the stasis state. While the behavior of tardigrads is extreme, many organisms display a range of adaptive behaviors that enable them to survive hostile environmental conditions.

Reactions: favorable, unfavorable, and their dynamics

As we will see, biological systems are extremely complex and both their overall structural elements and many of their molecular components (including DNA) are the products of thermodynamically unfavorable processes and reactions. How do these reactions take place in living systems? The answer comes from the coupling of thermodynamically favorable reactions to a thermodynamically unfavorable reactions. This is a type of work, although not in the standard macroscopic physics model of work (w) = force x distance. In the case of (chemical) reaction coupling, the work involved drives thermodynamically unfavorable reactions, typically the synthesis of large and complex molecules and macromolecules (that is, very large molecules). Here we will consider the thermodynamics of these processes.

Thinking about energy: Thermodynamics is at its core about energy and changes in energy. This leads to the non-trivial question, what is energy? Energy comes in many forms. There is energy associated with the movement and vibrations of objects with mass. At the atomic and molecular level there is energy associated with the (quantum) state of electrons. There is energy associated with fields that depends upon an object's nature (for example its mass or electrical charge) and its position within the field. There is the energy associated with electromagnetic radiation, the most familiar form is visible light, but electromagnetic radiation extends from microwaves to X-rays. Finally, there is the energy that is present in the very nature of matter, such energy is described by the equation

e (energy) = m (mass) x c^2 (c = speed of light).

From a simple perspective, we can call on our day to day experiences. Energy can be used to make something move. Imagine a system of a box sitting on a rough floor. You shove the box so that it moves and then stop pushing – the box travels a short distance and then stops. The first law of thermodynamics is that the total energy in a system is constant. So the question is where has the energy gone? One answer might be that the energy was destroyed. This is wrong. Careful observations lead us to deduce that the energy still exists, but it has been transformed. One obvious change is the transformation of energy from a mechanical force to some other form, so what are those other forms? It is unlikely that the mass of the box has increased, so we have to look at more subtle forms – the most likely is heat. The friction generated by moving the box represents an increase in the movements of molecules of the box and the floor over which the box moved. Through collisions and vibrations, this energy will, over time, be distributed throughout the

[126] On dormancy strategies in tardigrades: http://www.ncbi.nlm.nih.gov/pubmed/21402076 and Towards decrypting cryptobiosis--analyzing anhydrobiosis in the tardigrade Milnesium tardigradum using transcriptome sequencing.: http://www.ncbi.nlm.nih.gov/pubmed/24651535

system. This thermal motion can be seen in what is known as Brownian motion. In 1905 Albert Einstein explained Brownian motion in terms of the existence, size, and movements of molecules.[127]

In the system we have been considering, the concentrated energy used to move the box has been spread out throughout the system. While one could use the push to move something (to work), the diffuse thermoenergy cannot. While the total amount of energy is conserved, its ability to do things has been decrease (almost abolished). This involves the concept of entropy, which we will turn to next.

Thinking entropically (and thermodynamically)

We certainly are in no a position to teach you (rigorously) the basics of chemistry and chemical reactions, but we can provide a short refresher that focusses on the key points we will be using over and over again.[128] The first law of thermodynamics is that while forms of energy may change, that is, can be interconverted between distinct forms, the total amount of energy within a closed system remains constant. Again, we need to explicitly recognize the distinction between a particular system and the universe as a whole. The universe as a whole is itself (apparently) a closed system. If we take any isolated part of the system we must define a system boundary, the boundary and what is inside it is part of the system, while the rest of the universe outside of the boundary layer is not. While we will consider the nature of the boundary in greater molecular detail in the next chapter, we can anticipate that one of the boundary's key features are its selectivity in letting energy and or matter to pass into and out of the system and what constraints it applies to those movements.

Assuming that you have been introduced to chemistry, you might recognize the Gibb's free energy equation: $\Delta G = \Delta H - T\Delta S$, where T is the temperature of the system.[129] From our particularly biological perspective, we can think of ΔH as the amount of heat released into (or absorbed from) the environment in the course of a reaction, and ΔS as the change in a system factor known as entropy. To place this equation in a context, let us think about a simple reaction:

oil mixed with water $\rightleftarrows$ oil + water (separate) ΔG is negative

While a typical reaction involves changes in the types and amounts of the molecules present, we can extend that view to all types of reactions, including those that involve changes in temperature of distinct parts of a system (the bar model) and the separation of different types of molecules in a liquid (the oil-water example). No matter what the type of reaction, every reaction is characterized by its equilibrium constant, K_{eq}, which is a function of both the reaction itself and the conditions under which the reaction is carried out. These conditions include parameters such as the initial state of the system, the concentrations of the reactants, and system temperature and pressure. In biological systems we generally ignore pressure, although pressure will be important for organisms that live on the sea floor (and perhaps mountain tops).

---

[127] Albert Einstein: The Size and Existence of Atoms http://youtu.be/nrUBPO6zZ40

[128] Of course, we recommend a chemistry course sequence based on Cooper & Klymkowsky, 2014. Chemistry, Life, the Universe and Everything: pdf available on request (see http://besocratic.colorado.edu/CLUE-Chemistry/index.html)

[129] in the real world, the value of ΔG depends upon the concentrations of solute and solvent, but we will ignore that complexity for the moment.

The equilibrium constant for a reaction is defined as the rate of the forward reaction k_f (reactants to products) divided by the rate of the reverse reaction k_r (products to reactants). At equilibrium (where nothing macroscopic is happening), k_f times the concentrations of the reactants equals k_r times the concentration of the products. For a thermodynamically favorable reaction, that is one that favors the products, k_f will be greater than k_r and K_{eq} will be greater, often much greater than one. The larger K_{eq} is, the more product and the less reactant there will be when the system is at equilibrium. If the equilibrium constant is less than 1, then at equilibrium, the concentration of reactants will be greater than the concentration of products.

$$K_{eq} = \frac{k_f}{k_r} \qquad k_f\,[\text{reactants}] = k_r\,[\text{products}]$$

While the concentration of reactants and products of a reaction at equilibrium remains constant it is a mistake to think that the is system is static. If we were to peer into the system at the molecular level we would find that, at equilibrium, reactants are combining to form products and products are rearranging to form reactants at similar rates.[130] That means that the net flux, the rate of product formation minus the rate of reactant formation, will be zero. If, at equilibrium, a reaction has gone almost to completion and $K_{eq} >> 1$, there will be very little of the reactants left and lots of the products. The product of the forward rate constant times the small reactant concentrations will equal the product of the backward rate constant times the high product concentrations. Given that most reactions involve physical collisions between molecules, the changes in the frequency of productive collisions between reactants or products increases as their concentrations increase. Even improbable events can occur, albeit infrequently, if the rate of precursor events are high enough.

Reaction rates

Knowing whether a reaction is thermodynamically favorable and its equilibrium constant does not tell us much (or really anything) about whether the reaction actually occurs to any significant extent under the conditions we are concerned with. To know the reaction's rate we need to know the reaction kinetics for the specific system we are dealing with. Reaction kinetics tells us the rate at which the reaction actually occurs under a particular set of conditions. For example, consider a wooden log, which is composed mainly of the carbohydrate polymer cellulose ($(CH_2O)_n$. In the presence of molecular oxygen (O_2) the reaction,

$$nO_2 + \text{wooden log } ((CH_2O)_n) \leftrightarrows nCO_2 + nH_2O + \text{heat}$$

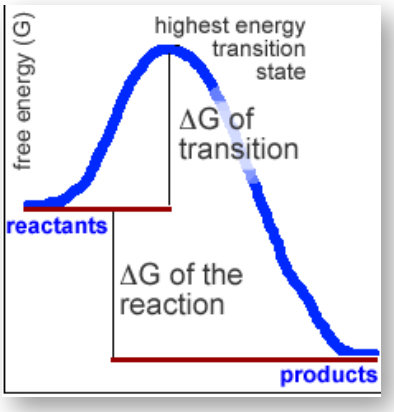
is extremely favorable thermodynamically, that is, it has a negative ΔG and a large equilibrium constant, yet the log is stable - it does not burst into flames spontaneously. The question is, of course, why ever not, why is the world so annoyingly complex?

The answer lies in the details of the reaction, how exactly are the reactants converted into the products? At this point, for simplicity or perhaps better put accessibility, let us consider another non-chemical but rather widespread, biologically, type of reaction. In this reaction system there is a barrier between two compartments, specifically the barrier membrane that separates the inside from the

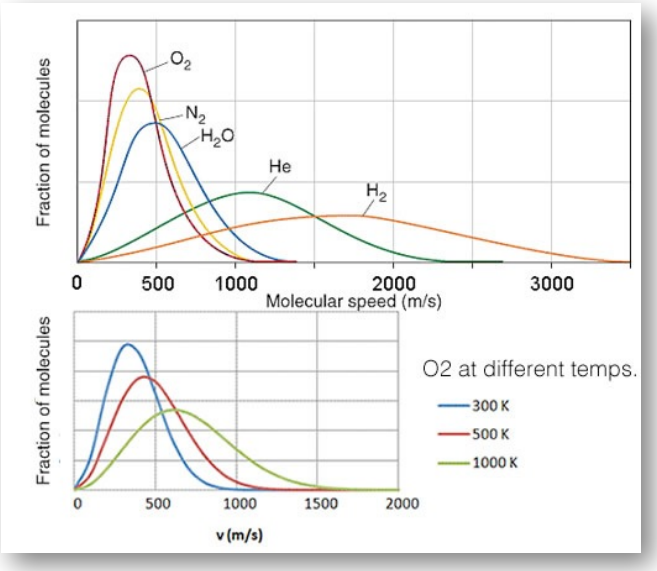
[130] This, of course, assumes that we have a closed system, that is, that neither the products or the reactants can leave the system, and that the volume of the system also remains constant. If the reactants can "leave the scene" of the reaction, then of course the back reaction, Products ⇆ Reactants, will be much less likely to occur.

outside of a cell. At this point, we do not need to consider the exact details of the barrier's structure (although we will in next chapter). In our particular example, outside the cell the concentration of molecule A is high, while inside the cell its concentration is low. We can write out this reaction equation as $A_{outside} \leftrightarrows A_{inside}$ (perhaps you make a prediction of the ΔG of this reaction and what it depends upon.)

The reaction consists of moving A molecules across the barrier between the inside and the outside of the cell. In our example the concentration of A outside the cell (written $[A_{outside}]$, with the square brackets indicating concentration) is much greater than $[A_{inside}]$. At any moment in time, the number of collisions between $A_{outside}$ and the barrier will be much greater than the number of collisions between A_{inside} and the barrier. Assuming that the probability of crossing the barrier is a function of the collision frequency, there will be net movement of $A_{outside}$ to A_{inside}. The real question is how large this net flux will be. This will depend on the amount of energy a molecule needs to cross the barrier. We can represent this energy as the highest peak in a reaction graph (here we assume a simple process with a single peak, in the real world it can involve a number of sub-reactions and look more like a roller-coaster than a simple hill)[→]. In such a graph, we begin with the free energy of the reactants (along the Y-axis), and plot the changing free energies of the various intermediates (along the X-axis), leading to the free energy of the products. The difference between the intermediate with the highest free energy and the free energy of the reactants ($\Delta G_{transition}$) corresponds (roughly in our simplified view of the subject) to the rate limiting step in the reaction and reflects the reaction's activation energy.



For a reaction to move from reactants ($A_{outside}$) to products (A_{inside}) the reactants must capture enough energy from their environment to traverse the barrier between outside and inside. In biological systems there are two major sources for this energy. The reactants can absorb electromagnetic energy, that is, light, or energy can be transferred to it from other molecules through collisions. In liquid water, molecules are moving; at room temperature they move on average at about 640 m/s. That is not to say that all molecules are moving with the same speed. If we were to look at the population of molecules, we would find a distribution of speeds known as a Boltzmann distribution. As they collide with one another, they exchange kinetic energy, and one molecule can emerge from the collision with much more energy than it entered with. Since reactions occur at temperatures well above absolute zero, there is plenty of energy available in the form of the kinetic energy of molecules, and occasionally a molecule with extremely high energy will emerge. If such an energetic A molecule gains sufficient energy and collides with the boundary layer, it could cross the boundary layer, that is, move from outside to inside. If not, it will probably loose that energy to other molecules very quickly through collisions. It is this dynamic exchange of kinetic energy that drives



the movement of molecules (as well as the breaking of bonds associated with chemical reactions.)

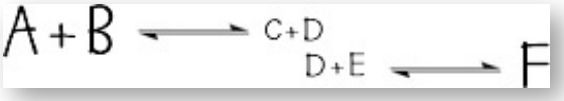
The difference between the free energies of the reactants and products ($\Delta G_{reaction}$) determines the equilibrium constant for a particular reaction system. In the case of our barrier system, since the A molecules are the same whether inside or outside the cell, the difference in the free energies of the reactants and products reflects (primarily) the difference in their concentrations. Higher concentration correlates with higher free energy (remember, we are interested in the ΔG of the $A_{outside} \leftrightarrows A_{inside}$ reaction.) Clearly the more molecules of A are present, the higher the ΔG of A. One point is worth emphasizing, it is possible for a reaction to have a large $\Delta G_{reaction}$ and either a large or small $\Delta G_{transition}$. So assuming that there is enough energy in the system, and $\Delta G_{transition}$ is small enough for the reaction to proceed at a noticeable rate, you should be able to predict what happens to the system as it moves toward equilibrium. If the $\Delta G_{transition}$ is high enough, however the $A_{outside} \leftrightarrows A_{inside}$ reaction will not occur to any significant extent.

Coupling reactions

There are large numbers of different types of reactions that occur within cells. As a rule of thumb, a reaction that produces smaller molecules from larger ones will be thermodynamically favored, while reactions that produce larger molecules from smaller ones will be unfavorable. Similarly a reaction that leads to a molecule moving from a region of higher concentration do a region of lower concentration will be favored. So how exactly can we build the big molecules, such as DNAs and proteins, that life depends upon.

As we noted before reactions can be placed into two groups, those that are thermodynamically favored (negative ΔG, equilibrium constant is greater, typically much greater, than 1) and those that are unfavorable (positive ΔG, equilibrium constant less, often much less than 1). Thermodynamically favored reactions are typically associated with the release of energy from, and the breakdown of, various forms of food (known generically as catabolism), while reactions that build up biomolecules (known generically as anabolism) are typically thermodynamically unfavorable. An organism's metabolism is the sum total of all of these various reactions.

To get unfavorable reactions to occur they are coupled to thermodynamically favorable reactions. This requires that the two reactions share a common intermediate. In this example [→] the two reactions share the component "D". Let us assume that the upper reaction is unfavorable while the lower reaction is favorable. What happens? Let us assume that both reactions are occurring at measurable rates, perhaps through the mediation of appropriate catalysts, which act to lower the activation energy of a reaction, and that E is present within the system. At the start of our analysis, the concentrations of A and B are high. We can then use Le Chatelier's principle to make our predictions.[131]



Let us illustrate how Le Chatelier's principle works. Assume for the moment that the reaction

[131] http://en.wikipedia.org/wiki/Le_Chatelier's_principle

A + B ⇆ C + D has reached equilibrium. Now consider what happens to the reaction if, for example, we removed (somehow, do not worry about how) all of the C from the system. Alternatively, consider what happens if we add more B. The answer is, that the reaction would move to the right (even though that reaction is thermodynamically unfavorable), in order to re-establish the equilibrium condition. If all C were removed, the C + D to A + B reaction could not occur, so the A + B reaction would continue in an unbalanced manner until the level of C + D increased and C + D to A + B reaction became balanced with the A + B to C + D reaction. In the second case, the addition of B would lead to the increased production of C + D, until their concentration reached a point where the C + D to A + B reaction became balanced with the A + B to C + D reaction. This type of behavior arises directly from the fact that at equilibrium reaction systems are not static at the molecular level, but dynamic – things are still occurring, they are just balance so that no net change occurs. When you add or take something away from the system, it becomes unbalanced, that is, it is no longer at equilibrium. Because the reactions are occurring at a measurable rate, the system will, over time, return to equilibrium.

So back to our reaction system. As the unfavorable A+B reaction occurs and approaches equilibrium it will produce a small amount of C+D. However, the D+E reaction is favorable; it will produce F while at the same time removing D from the system. As D is removed, it influences the A+B reaction (because it makes the C+D "back reaction" less probable even though the A+B "forward reaction" continues.) The result is that more C and D will be produced. Assuming that sufficient amounts of E are present, more D will be removed. The end result is that, even though it is energetically unfavorable, more and more C and D will be produced, while D will be used up to make F. It is the presence of the common component D and its utilization as a reactant in the D + E reaction that drives the synthesis of C from A and B, something that would normally not be expected to occur to any great extent. Imagine then, what happens if C is also a reactant in some other favorable reaction(s)? In this way reactions systems are linked together, and the biological system proceeds to use energy and matter from the outside world to produce the complex molecules needed for its maintenance, growth, and reproduction.[132]

Questions to answer & to ponder:

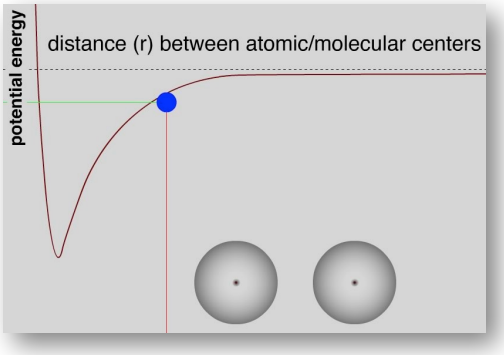
- What are the common components of a non-equilibrium system and how does a dried out tardigrad fulfill those requirements?
- You use friction to ignite a fire, where is the energy of the fire derived from?
- A reaction is at equilibrium and we increase the amount of reactant, what happens in terms of the amount of reactant and product?
- A reaction is at equilibrium and we increase the amount of product, what happens in terms of the amount of reactant and product?
- What does the addition of a catalyst do to a system already at equilibrium
- What does the addition of a catalyst do to a system far from equilibrium?
- Where does the energy come from to reach the activation state/reaction intermediate?
- Why does a catalyst not change the equilibrium state of a system?
- Why are catalysts required for life?

[132] http://haha.nu/science/the-amazing-human-body/

Molecules and molecular interactions

We have already briefly (rather absurdly briefly) defined what energy is and begun to consider how it can be transformed from form to form. Now we need to consider what we mean by matter, which implies an understanding of the molecules that compose matter. As you hopefully know by now, all matter is composed of atoms. The structure of atoms is the subject of quantum physics; typically atoms interact with one another through a number of different types of interactions. The first are van der Waals interactions that are mediated by London Dispersion Forces. These arise from the fact that the positively charged components of atoms (protons) are localized in the extremely small central nucleus, while equal numbers of negatively charged components (electrons) are located around this nucleus. The charges on the protons and electrons are equal in magnitude, so that the atom is electrically neutral. That would be that, except that the electrons are moving around the nucleus. If you are of a curious disposition you might wonder why the negatively charged electrons are not simply attracted to and become localized to the positively charged nucleus; the answer is because of quantum principles, which we will not consider here. In any case, the movements of electrons means that over time, a observer outside of the atom will experience a fluctuating electrical field. This same phenomena applies to molecules, which are collections of atoms bonded together through covalent bonds, which we will consider further on.

London Dispersion Forces and Van der Waals interactions



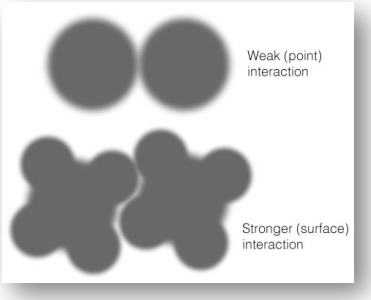
As two molecules approach one another, their fluctuating electric fields interact and attract one another. This force varies as ~$1/R^6$ where R is the distance between the molecules. The implication of this force equation is that this London dispersion force (LDF), named after its discoverer Fritz Wolfgang London (1900 – 1954), acts only over very short distances. The magnitude of this attractive force reaches its maximum when the two molecules are separated by what is known as the sum of their van der Waals radii. If they move closer, the weak attractive LDF is quickly overwhelmed by a rapidly increasing, and extremely strong repulsive force that arises from the electrostatic interactions between the positively charged nuclei and the negatively charged electrons of the two molecules.[133]

Each atom and molecule has its own characteristic van der Waals radius, although since most molecules are not spherical, it is perhaps better to refer to a molecule's van der Waals surface. This surface is the closest distance that two molecules can normally approach one another before repulsion kicks in. It is common to see molecules displayed in terms of the van der Waals surfaces. Because each molecule generates LDF when approached by any other molecule, van der Waals interactions are universal, all molecules interact with one another in this manner. The one exception occurs when pairs of small similarly charged "ionic" molecules, that is molecules with a permanent net positive or negative charge, approach each other. The strength of their electrostatic repulsion will be greater than the LDF.

---

[133] this can be explored further at http://besocratic.colorado.edu/CLUE-Chemistry/LondonDispersionForce%20copy/1.2-interactions-0.html

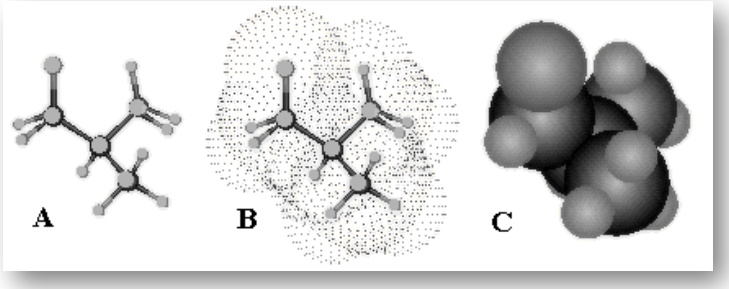
The extent to which molecules interact, via LDFs, is determined in large part by their shapes. Compare the interaction between two monoatomic Noble atoms, such as helium or neon, and two molecules with more complex shapes (figure →). The two monoatomic particles interact via LDFs at a single point, so the strength of the interaction is minimal. On the other hand, the two more complex molecules interact over extended surfaces, so the LDF between them is much stronger.



Covalent bonds

In van der Waals interactions, the atoms and molecules involved retain their hold on their electrons, they remain distinct and discrete. There are cases, however, where atoms come to "share" each other's electrons. This sharing involves pairs of electrons, one from each atom. When electron pairs are shared, the atoms stop being distinct in that their shared electrons are no longer restricted to one or the other. In fact, since one electron cannot even in theory be distinguished from any other electron, they become a part of the molecule's electron system.[134] This sharing of electrons leads to what is known as a covalent bond. Covalent bonds are 20 to 50 times stronger than van der Waals interactions. What does that mean, exactly? It means that it takes much more energy, from either collisions with surrounding molecules or the absorption of energy (light), to break them apart again. While the bonded form is always more stable than the unbounded form; the extent of that change in stability can be considered the bond energy. The size of any particular bond energy depends upon the specific molecule in which it occurs. Whether the bonded form remains intact will depend upon the ratio of bond energy to energy that is delivered to the molecule upon collisions with other molecules (or through the absorption of light).

When atoms form a covalent bond, their individual van der Waals surfaces merge to produce a new molecular van der Waals surface. There are a number of ways to draw molecules, but the space-filling or van der Waals surface view is the most realistic (at least for our purposes). While realistic it can also be confusing, since it obscures the underlying molecular structure, that is, how the atoms in the molecule are linked together. This can be seen in this set of representations of the simple molecule 2-methylpropane.[135] As molecules become larger, it can become impossible (or at least quite difficult) to recognize them based on a van der Waals surface representation.



Because they form a new stable entity, it is not surprising (perhaps) that the properties of a molecule are quite distinct from, although certainly influenced by, the properties of the atoms from which they are composed. A molecule's properties are based on its shape, which is dictated by how the various atoms that compose the molecule are connected to one another. These geometries are

[134] Unlike organisms, each of which is unique in practice and theory, all electrons are identical in theory.

[135] Explicit Concepts of Molecular Topology: http://www.chem.msu.ru/eng/misc/babaev/match/top/top02.htm

imposed by each atom's quantum mechanical properties and the interactions between atoms within the molecule. Some atoms, common to biological systems, such as hydrogen (H), can form only one covalent bond with another atom. Others can make two (oxygen (O) and sulfur (S)), three (nitrogen (N)), four (carbon (C)), or five (phosphorus (P)). In addition to smaller molecules, biological systems contain a number of distinct types of extremely large molecules, known as macromolecules. Such macromolecules are not rigid; they can fold back on themselves to form what are known as intramolecular interactions. There can also be intermolecular interactions (mediated primarily by van der Waals interactions) with other molecules (small and large). These interactions can vary dramatically in terms of stability, that is, how long they persist before the interacting molecules come apart, or dissociate.

Molecules are dynamic. Collisions with other molecules can lead to parts of a molecule rotating around a single bond.[136] The presence of a double bond restricts these kinds of movements, rotation around a double bond requires what amounts to breaking and then reforming one of the bonds. In addition, and if you have mastered some chemistry you already know this, it is often incorrect to consider bonds as distinct entities, isolated from their surroundings. Adjacent bonds can interact forming what are known as resonance structures that behave as mixtures of single and double bonds. Again this restricts free rotation around the bond axis and acts as a constraint on molecular geometry. The peptide bond common to protein is an example of such a resonance structure, as are the various "bases" found in nucleic acids (we will return to this type of bond later in our discussion of proteins and nucleic acids). These complexities combine to make predicting a particular molecule's three dimensional structure increasingly difficult as its size increases. Molecules undergo complex and dynamic interactions with one another, that is they can associate, a process that involves van der Waals interactions, and dissociate in response to various factors, including thermal motion.

Bond stability and thermal motion (a non-biological moment)

Molecules do not exist out of context. In the real world they are not sitting alone in a vacuum. We always need to consider the system in which the molecules occur. For example, most biological molecular interactions occur in aqueous solution. That means, biological molecules are surrounded by other molecules, mostly water molecules. As you may already know from physics, there is a lowest possible temperature, known as absolute zero (0 K or –273.15 ºC or –459.67 °F). At this, completely biologically irrelevant temperature, molecular movements are minimal, but not apparently absent all together. When we think about a system, we inevitably think about its temperature. Temperature is a concept that makes sense only at the system level. Individual molecules do not have a temperature. The temperature of the system is a measure of the average kinetic energy

$$E_k = 1/2 \text{ (average mass x (average velocity)}^2$$

of the molecules within it. It does not matter whether the system is composed of only a single type of molecule or many different types of molecules, at a particular temperature the average kinetic energy of the molecules has one value. This is not to say that all molecules have the same kinetic energy, they certainly do not.

[136] This could be basis of a square dance like in class activity!

In a gas we can largely overlook the attractive interactions between molecules (their intermolecular interactions) because the average kinetic energies of the molecules of the system are sufficient to disrupt those interactions - that is, after all, why they are a gas. As we cool the system, we remove energy from it, and the average kinetic energy of the molecules in the system decreases. If the average kinetic energy gets low enough, the molecules will form a liquid. In a liquid, the movement of molecules is not enough to disrupt the interactions between them. This is a bit of a simplification, however. Better to think of it more realistically. Consider a closed box partially filled with a substance in a liquid state. What is going on? Assuming there are no changes in temperature over time, the system will be a equilibrium. What we will find, if we think about it, is that there is a reaction going on, that reaction is Molecule (gas) ⇆ Molecule (liquid), and that at the particular temperature, the liquid phase is favored, although there will be molecules in the gaseous phase within the system. The point is that at equilibrium, the number of molecules moving from liquid to gas will be equal to the number of molecules moving from the gas to the liquid phase. If we increase or decrease the temperature of the system, we will alter the equilibrium state, that is, the relative amounts of molecules in the gaseous versus the liquid states.

In a liquid, while molecules remain associated with one another, they can also move with respect to one another relatively easily. That is why liquids can be poured, and why they assume the shape of the (solid) containers into which they are poured. This is in contrast to the solid container. In a solid the molecules are tightly associated with one another and so do not move with respect to one another. Solids do not flow. The cell, or more specifically, the cytoplasm, acts primarily as a liquid and biological processes take place in the liquid phase. This has a number of implications. First molecules, even very large macromolecules, can move with respect to one another. Driven by thermal motions, molecules will move around in a Brownian manner, a behavior known as a random walk.

This thermal motion will also influence whether molecules associate with one another. We can think about this process in the context of an ensemble of molecules, let us call them A and B that interact to form a complex, AB. Assume that this complex is held together by van der Waals interactions. In an aqueous solution, the complex is colliding with water molecules with various energies, as described by the Boltzmann distribution. There is a probability that in any unit of time, one or more of these collisions will deliver sufficient energy to disrupt the interaction between A and B, and the AB complex will disassociate into separate A and B molecules. Assume we start with all AB molecules, the time for 50% of these molecules to dissociate into A and B would be considered the half life of the complex. Now here is the tricky part, much like the situation with radioactive decay, but subtly different. While we can confidently conclude that 50% of the AB complexes will have disassembled into A and B at the half-life time, we can not predict which of these AB complexes will have disassembled and which will have remained intact. Why? Because we cannot predict which collisions will lead to disassociation (by providing sufficient energy to overcome the van der Waals interactions between A and B) and which will not.[137] This type of process is known as a stochastic process, since it is driven by random events. Genetic drift is another form of a stochastic process, since in a particular drifting

[137] It should be noted that, in theory at least, we might be able to make this prediction if we mapped the movement of every water molecule. This is different from radioactive decay, where it is not even theoretically possible to predict the behavior of an individual radioactive atom.

population it is not possible to predict which alleles will be lost and which fixed, or when exactly fixation will occur. This is a hallmark of stochastic processes, they are best understood in terms of probabilities.

Stochastic processes are particularly important within biological systems because, generally, cells are small and often contain only a small number of molecules of a particular type. If, for example, the expression of a gene depends upon a protein binding (reversibly) to specific sites on a DNA molecule, and if there are relatively small numbers of the protein and (usually) only one or two copies of the gene (that is, the DNA molecule), we will find that whether or not the protein is bound to the DNA behaves as a stochastic process. If there are enough cells, then the group average will be predictable, but the behavior of any one cell will not be predictable. In an individual cell, sometimes the protein will be bound and the gene will be expressed, and sometimes not, all because of thermal motion and the small numbers of interacting components involved. This noisy (or stochastic) property of cells often plays an important role in the control of cell and organismic behavior. It can even transform a genetically identical population of organisms into subpopulations that display two or more distinct behaviors, a property with important implications, that we will return to.

Bond polarity, inter- and intramolecular interactions

So far, we have been considering covalent bonds in which the sharing of electrons between atoms is more or less equal, but that is not always the case. Because of their atomic structures, which arise from quantum mechanical principles (not to be discussed here), different atoms have different affinities for their own electrons. When an electron is removed or added to an atom (or molecule) that atom/molecule becomes an ion. Atoms of different elements differ in the amount of energy it takes to remove an electron from them; this is, in fact, the basis of the photoelectric effect explained by Albert Einstein, in another of his 1905 papers.[138] Each type of atom (element) has a characteristic electronegativity, the measure of how tightly the atom holds onto its electrons. If the electronegativities of the two atoms in a bond are equal or similar, then the electrons are shared more or less equally between them and the bond is said to be non-polar. There are no stable regions of net negative or positive charge on the surface of the resulting molecule. If the electronegativities of the two bonded atoms are unequal, however, then the electrons will not be shared equally. On average, there will be more electrons around the more electronegative atom and less around the less electronegative atom. This leads to stable partially negatively and positively-charged regions to the bond; this charge separation produces an electrical field, known as a dipole. A bond between atoms of differing electronegativities is said to be polar.

In biological systems, atoms of O and N will sequester electrons when bonded to atoms of H and C, the O and N become partly negative compared to their H and C bonding partners. Because of the quantum mechanical organization of atoms, these partially negative regions are organized in a non-uniform manner, which we will return to. In contrast, there is no significant polarization of charge in bonds between C and H atoms, and such bonds are non-polar. The presence of polar bonds leads to the possibility of electrostatic interactions between molecules. Such interactions are stronger than van der Waals interactions but much weaker than covalent bonds, but like covalent bonds, they have a

[138]Albert Einstein: Why Light is Quantum: http://youtu.be/LWIi7NO1tbk

directionality to them. For such an electrostatic interaction to form, the three atoms involved have to be arranged more or less along a straight line. There is no similar geometric constraint on van der Waals intermolecular interactions.

Since the intermolecular forces arising from polarized bonds often involve an H interacting with an O or an N, these have become known generically (at least in biology) and perhaps unfortunately as hydrogen or H-bonds. Why unfortunate? Because H atoms can take part in covalent bonds, but H-bonds are not covalent bonds, they are very much weaker. It takes much less energy to break an H-bond between molecules or between parts of (generally macro-) molecules that it does to break a covalent bond involving a H atom.

The implications of bond polarity

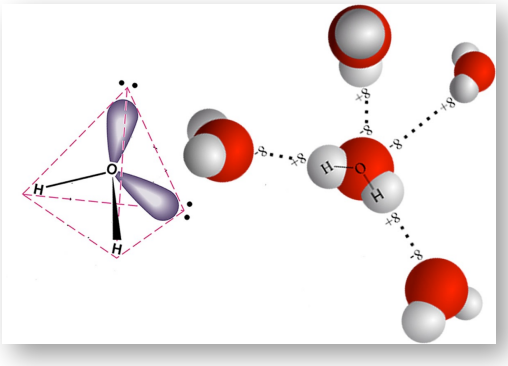
Two important physical properties of molecules (although this applies primarily to small molecules and not macromolecules) are their melting and boiling points. Here we are are considering a pure sample of the molecule. Let us start at a temperature at which the sample is liquid. The molecules are moving with respect to one another, there are interactions between the molecules, but they are transient - the molecules are constantly switching neighbors. As we increase the temperature of the system, the energetics of collisions are now such that all interactions between neighboring molecules are broken, and the molecules fly away from one another. If they happen to collide with one another, they do not adhere, the bond that might form is not strong enough to resist the kinetic energy of the molecules. They are said to be a gas, and the transition from liquid to gas is said to be the boiling point. Similarly, starting with a liquid, when we reduce the temperature, the interactions between molecules become longer lasting until such a time as the energy transferred through collisions is no longer sufficient to disrupt these interactions. As more and more molecules interact, neighbors become permanent - the liquid has been transformed into a solid. While liquids flow and assume the shape of their containers, because neighboring molecules are free to move with respect to one another, solids maintain their shape, and neighboring molecules stay put. The temperature at which a liquid changes to a solid is known as the melting point. These temperatures mark what are known as phase transitions.

Compounds	CH_4	NH_3	OH_2	FH	Ne
molecular weight	16.04	17.02	18.02	20.01	20.18
bond electronegativity	0.45	0.94	1.34	1.88	N/A
# of electrons	10	10	10	10	10
# of bonds	4	3	2	1	0
melting point	-182°C	-77.7°C	**0°C**	-83°C	-248.6°C
boiling point	-161.5°C	-33.4°C	**100°C**	19.5°C	-246.1°C

At the macroscopic level, we see the rather dramatic effects of bond polarity on melting and boiling points by comparing molecules of similar size with and without polar bonds. For example, CH_4 (methane) and Ne (neon) have no polar bonds and cannot form H-bond-type electrostatic interactions, whereas NH_3 (ammonia), H_2O (water), and FH (hydrogen fluoride) have three, two and one polar bonds, respectively, and can take part in one or more H-bond-type electrostatic interactions. All five

compounds have the same number of electrons, ten. When we look at their melting and boiling temperatures, we see rather immediately how bond polarity influences these properties.

In particular water stands out as dramatically different from the rest, with a > 70ºC higher melting and boiling point than its neighbors. So why is water weird? Well, in addition to the presence of polar covalent bonds, we have to consider the molecule's geometry. Each water molecule can take part in four hydrogen bonding interactions with neighboring molecules - it has two partially positive Hs and two partially negative sites on the O. These sites of potential H-bond-type electrostatic interactions are arranged in a nearly tetragonal geometry. Because of this arrangement, each water molecule can interact through H-bond-type electrostatic interactions with four other water molecules. To remove a molecule from its neighbors, four H-bond-type electrostatic interactions must be broken, which is relatively easy since they are each rather weak. In the liquid state, molecules jostle one another and change their H-bond-type electrostatic interaction partners constantly. Yet, each water molecule remains linked to multiple neighbors via H-bond-type electrostatic interactions.



This molecular hand-holding leads to water's high melting and boiling points as well as its high surface tension. We can measure the strength of surface tension in various ways. The most obvious is the weight that the surface can support. Water's surface tension has to be dealt with by those organisms that interact with a liquid-gas interface. Some, like the water strider, use it to cruise along the surface of ponds. As the strider walks on the surface of the water, the molecules of its feet do not form H-bond-type electrostatic interactions with water molecules, they are said to be **hydrophobic**, although that is clearly a bad name - they are not afraid of water, rather they are simply apathetic to it. They interact with other molecules, including water molecules, through van der Waals interactions only. Molecules that can make H-bonds with water are termed **hydrophilic**. As molecule's increase in size they can have regions that are hydrophilic and regions that are hydrophobic (or hydroapathetic). Molecules that have distinct hydrophobic and hydrophilic regions are termed amphipathic and we will consider them in more detail in the next chapter.

Interacting with water

We can get an idea of the hydrophilic, hydrophobic/hydroapathetic, and amphipathic nature of molecules through their behaviors when we try to dissolve them in water. Molecules like sugars (carbohydrates), alcohols, and most amino acids are primarily hydrophilic. They dissolve readily in water. Molecules like fats are highly hydrophobic/hydroapathetic, and they do not. So why the difference? To answer this question we have to be clear what we mean when we say that a molecule is soluble in water. We will consider this from two perspectives. The first is what the solution looks like at the molecular level, the second is how the solution behaves over time. To begin, we need to understand what water alone looks like. Because of its ability to make and donate multiple H-bond-type

electrostatic interactions in a tetrahedral arrangement, water molecules form a dynamic three-dimensional intermolecular interaction network. In liquid water the H-bond-type electrostatic interactions between the molecules break and form rapidly.

To insert a molecule A, known as a solute, into this network you have to break some of the H-bond-type electrostatic interactions between the water molecules (known as the solvent). If the A molecules can make H-bond-type electrostatic interactions with water molecules, that is, it is hydrophilic, then there is little net effect on the free energy of the system. Such a molecule is soluble in water. So what determines how soluble the solute is. As a first order estimate, each solute molecule will need to have at least one layer of water molecules around it, otherwise it will be forced to interact with other solute molecules. If the number of these interacting solute molecules is large enough, the solute will no longer be in solution. In some cases, aggregates of solute molecule can, because they are small enough, remain suspended in the solution. This is a situation known as a colloid. While a solution consists of individual solute molecules surrounded by solvent molecules, a colloid consists of aggregates of solute molecules in a solvent. So we might predict that all other things being equal (a unrealistic assumption), the larger the solute molecule the lower its solubility. You might be able to generate a similar rule for the size of particles in a colloid.

Now we can turn to a conceptually trickier situation, the behavior of a hydrophobic/apathetic solute molecule in water. Such a molecule cannot make H-bond-type electrostatic interactions with water, so when it is inserted into water the total number of H-bond-type electrostatic interactions in the system decreases - the energy of the system increases (remember, bond forming lowers potential energy). However, it turns out that much of this “enthalpy” change, conventionally indicated as ΔH, is compensated for by van der Waals interactions (that is, non-H-bond-type electrostatic interactions) between the molecules. Generally, the net enthalpic effect is minimal. Something else must be going on to explain the insolubility of such molecules.

Turning to entropy: Typically, in a liquid water molecules will be found in a state that maximizes the number of H-bond-type electrostatic interactions present. And because these interactions have a directionality, their presence constrains the possible orientations of the molecules with respect to one another. This constraint is captured when water freezes, and is the basis for ice crystal formation and why the density of water increases before freezing, so that ice floats in liquid water.[139] In the absence of the hydrophobic/hydroapathetic solute molecule, there are many many equivalent ways that liquid water molecules can interact to produce these geometrically specified orientations. But the presence of a solute molecule that cannot form H-bond-type electrostatic interactions restricts this number, a much smaller number of configurations results in the maximizing of H-bond formation between water molecules. The end result is that the water molecules become arranged in a limited number of ways around each solute molecule. These water molecules are in a more ordered, that is, a more improbable state, than they would be in the absence of solute. The end result is that there will be a decrease in entropy (indicated as ΔS), the measure of the probability of a state. ΔS will be negative compared to arrangement of water molecules in the absence of the solute.

[139] http://youtu.be/UukRgqzk-KE

How does this influence whether dissolving a molecule into water is thermodynamically favorable or unfavorable. It turns out that the interaction energy (ΔH) of placing most solutes into the solvent is near 0, so that it is the ΔS that makes the difference. Keeping in mind that ΔG = ΔH - TΔS, if ΔS is negative, then -TΔS will be positive. The ΔG of a thermodynamically favorable reaction is, by definition, negative. This implies that the reaction:

$$\text{water + solute} \leftrightarrows \text{solution (water + solute)}$$

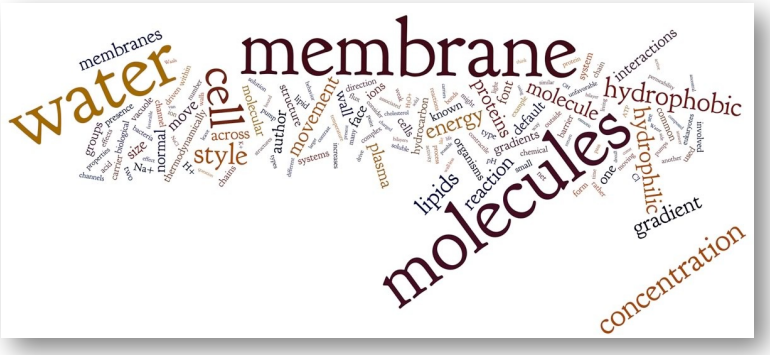
is thermodynamically unfavorable; the reaction will move to the left. That is, if we start with a solution, it will separate so that the solute is removed from the water. How does this happen? The solute molecules aggregate with one another. This reduces their effects on water, and so the ΔS for aggregation is positive. If the solute is oil, and we mix it into water, the oil will separate from the water, driven by the increase in entropy associated with minimizing solute-water interactions. This same basic process plays a critical influence on macromolecular structures.

Questions to answer & to ponder:

- Given what you know about water, why is ice less dense than liquid water?
- Make of model relating the solubility of a molecule with a hydrophilic surface to the volume of the molecule?
- Use your model to predict the effect on solubility if your molecule with a hydrophilic surface had a hydrophobic/apathetic interior.
- Under what conditions might entropic effects influence the interactions between two solute molecules?
- Based on your understanding of various types of intermolecular and intramolecular interactions, propose a model for why the effect of temperature on covalent bond stability is not generally significant in biological systems?
- How does temperature influence intermolecular interactions?

6. Membrane boundaries and capturing energy

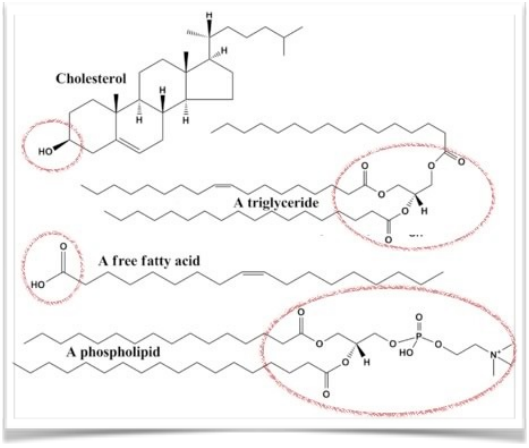
In which we consider how the aqueous nature of biological systems drives the formation of lipid-based barrier membranes and how such membranes are used to capture and store energy from the environment and chemical reactions. We consider how coupled reactions are used to drive macromolecular synthesis and growth.



Defining the cell's boundary

A necessary step in the origin of life was the generation of a discrete barrier, a boundary layer that serves to separate the living non-equilibrium reaction system from the rest of the universe. This boundary layer, the structural ancestor of the plasma membrane of modern cells, serves to maintain the integrity of the living system and mediates the movement of materials and energy into and out of the cell. Based on our current observations, the plasma membrane of all modern cells appears to be a homologous structure derived from a precursor present in the last common ancestor of life. So what is the structure of this barrier (plasma) membrane? How is it built and how does it work?

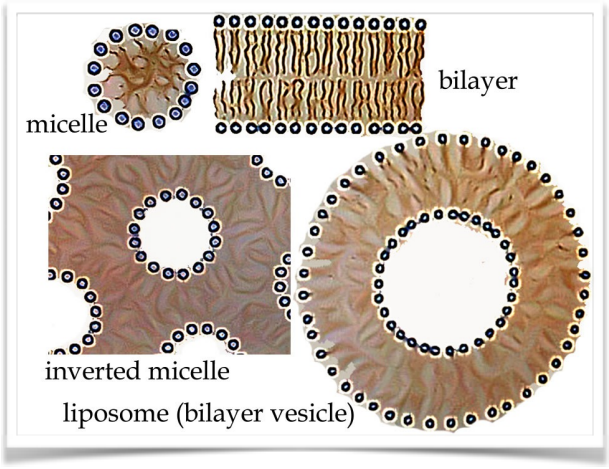
As we have already seen, when a new cell is formed, its plasma membrane is derived from the plasma membrane of the parental cell. As the cell grows, new molecules must be added into the membrane to enable it to increase its surface area. Biological membranes are composed of two general classes of molecules, proteins (which we will discuss in much greater detail in the next section of the course) and lipids. It is worth noting explicitly here that, unlike a number of other types of molecules we will be considering, such as proteins, nucleic acids, and carbohydrates, lipids are not a structurally coherent group, that is they do not have one particular basic structure. Such apparently diverse molecules as cholesterol and phospholipids, are both considered lipids, and while there is a relatively small set of common lipid types, there are many different lipids found in biological systems and the characterization of their structure and function(s) has led to a new area of specialization known as lipidomics.[140]



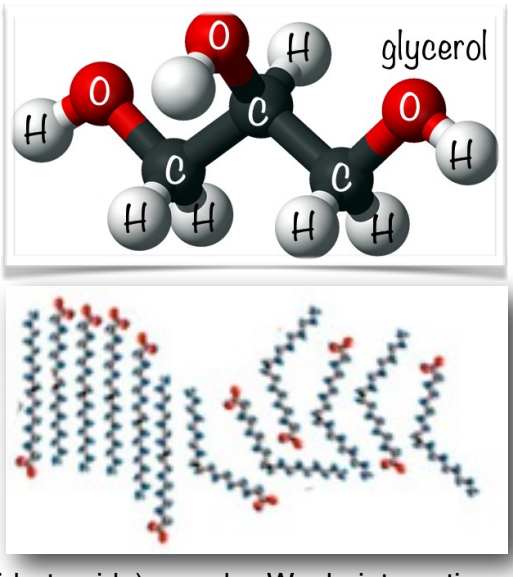
All lipids have two distinct domains: a hydrophilic (circled in red in this figure →) domain characterized by polar regions and hydrophobic/hydroapathetic domains that are usually just made up of C and H and are non-polar. Lipids are **amphipathic**. In aqueous solution, entropic effects will drive the hydrophobic/hydroapathetic parts of the lipid out of solution. But in contrast to totally non-polar molecules, like oils, the hydrophobic/hydroapathetic part of the lipid is connected to a hydrophilic domain that is soluble in water. Lipid molecules deal with this dichotomy by associating with

[140] On the future of "omics": lipidomics: http://www.ncbi.nlm.nih.gov/pubmed/21318352 and Lipidomics: new tools and applications http://www.ncbi.nlm.nih.gov/pubmed/21145456

other lipid molecules in multimolecular structures in which the interactions between the hydrophilic parts of the lipid molecule and water molecules are maximized and the interactions between the lipid's hydrophobic/hydroapathetic parts and water are minimized. Many different multi-molecular structures can be generated that fulfill these constraints. The structures that form depend upon the details of the system, including the shapes of the lipid molecules and the relative amounts of water and lipid present, but the reason these structures self-assemble is because their formation leads to an increase in the overall entropy of the system, a somewhat counterintuitive idea. For example, in a micelle the hydrophilic region is in contact with the water, while the hydrophobic regions are inside, away from direct contact with water. This leads to a more complete removal of the hydrophobic domain of the lipid from contact with water than can be arrived at by a purely hydrophobic oil molecule, so unlike oil, lipids can form stable structures in solution. The diameter and shape of the micelle is determined by the size of its hydrophobic domain. As this domain gets longer, the center of the micelle becomes more crowded. Another type of organization that avoids "lipid tail crowding" is known as a bilayer vesicle. Here there are two layers of lipid molecules, pointing in opposite directions. The inner layer surrounds a water filled region (the lumen of the vesicle), while the outer layer interacts with the external environment. In contrast to the situation within a micelle, the geometry of a vesicle means that there is significantly less crowding as a function of lipid tail length. Crowding is further reduced as a vesicle increases in size to become a cellular membrane. Micelles and vesicles can form a colloid-like system with water, that is they exist as distinct structures that can remain suspended in a stable state. We can think of the third type of structure, the planar membrane, as simply an expansion of the vesicle to a larger and more irregular size. Now the inner layer faces the inner region of the cell (which is mostly water) and the opposite region faces the outside world. For the cell to grow, new lipids have to be inserted into both inner and outer layers of the membrane; how exactly this occurs typically involves interactions with proteins. For example, there are proteins that can move a lipid from the inner to the outer domain of a membrane (they flip the lipid between layers, and are known as flipases), but the molecular details are beyond our scope here. While there are a number of distinct mechanisms that are used to insert molecules into membranes they always involve a pre-existing membrane – this is another aspect of the continuity of life. Totally new cellular membranes do not form, membranes are built on pre-existing membranes. For example, a vesicle (that is a spherical lipid bilayer) could fuse into or emerge from a planar membrane. These processes are typically driven by thermodynamically favorable reactions involving protein-based molecular machines. When the membrane involved is the plasma (boundary) membrane, these processes are known as exocytosis and endocytosis, respectively. These terms refer explicitly to the fate of the material within the vesicle. Exocytosis releases that material from the vesicle interior into the outside world, whereas endocytosis captures material from outside of the cell and brings it into the cell. Within a cell, vesicles can fuse and emerge from one another.

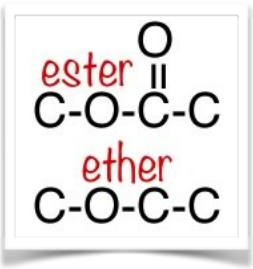


As noted above, there are hundreds of different types of lipids, generated by a variety of biosynthetic pathways catalyzed by proteins encoded in the genetic material. We will not worry too much about all of these different types of lipids, but we will consider two, the glycerol-based lipids and cholesterol, because considerations of their structures illustrates general ideas related to membrane behavior. In bacteria and eukaryotes, glycerol-based lipids are typically formed from the highly hydrophilic molecule glycerol combined with two or three fatty acid molecules. Fatty acids contain a long chain hydrocarbon with a polar (carboxylic acid) head group. The nature of these fatty acids influences the behavior of the membrane formed. Often these fatty acids have what are known as saturated hydrocarbon tails. A saturated hydrocarbon contains only single bonds between the carbon atoms of the tail domain. While these chains can bend and flex, they tend to adopt a more or less straight configuration. In this straight configuration, they pack closely, which maximizes the lateral (side to side) van der Waals interactions between them. Because of the extended surface contact between the chains, lipids with saturated hydrocarbon chains are typically solid around room temperature. On the other hand, there are cases where the hydrocarbon tails are "unsaturated", that is they contain double bonds (–C=C–) in them. These are typically more fluid and flexible. This is because unsaturated hydrocarbon chains have permanent kinks in them (because of the rigid nature and geometry of the C=C bonds), so they cannot pack as regularly as saturated hydrocarbon chains. The less regular packing means that there is less interaction area between the molecules, which lowers the strength of the van der Waals interactions between them. This in turn, lowers the temperature at which they change from a solid (no movement of the lipids relative to each other within the plane of the membrane) to a liquid (much freer movements). Recall that the strength of interactions between molecules determines how much energy is needed to overcome a particular type of interaction. Because these van der Waals intermolecular interactions are relatively weak, changes in environmental temperature influence the physical state of the membrane. The liquid like state is often referred to as the fluid state. The importance of membrane state is that it can influence the behavior and activity of membrane proteins. If the membrane is in a solid state, such proteins will be immobile, while in the liquid state they move by diffusion, that is, by thermally driven movement. Alternatively, since lipids are closely associated with proteins in the membrane, the physical state of the membrane can influence the activity of a protein embedded within it (a topic to which we will return).

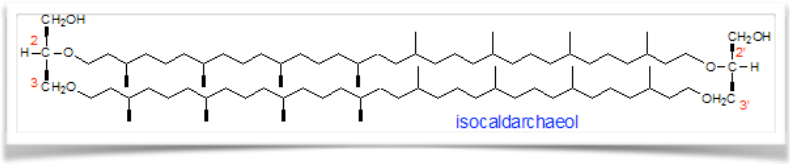


Cells can manipulate the solid-to-liquid transition temperature of their membrane by altering the membrane's lipid composition. For example, by altering the ratio of saturated to unsaturated chains present. This level of control involves altering the activities of the enzymes involved in saturation/desaturation reactions. That these enzymes can be regulated implies a feedback mechanism, by which either temperature or membrane fluidity acts to regulate metabolic processes. This type of feed back mechanism is part of what is known as the homeostatic and adaptive system of the cell (and the organism) and is another topic we will return to toward the end of the course.

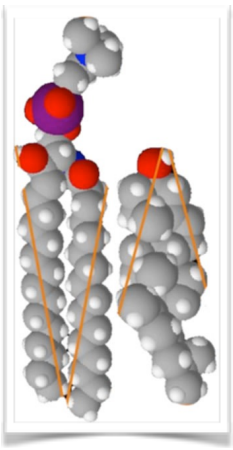
There are a number of differences between the lipids used in bacterial and eukaryotic organisms and archaea.[141] For example, instead of hydrocarbon chains, archaeal lipids are constructed of isoprene (CH_2=C(CH_3)CH=CH_2) polymers linked to the glycerol group through an ether (rather than an ester) linkage. The bumpy and irregular shape of the isoprene groups (compared to the relatively smooth saturated hydrocarbon chains) means that archaeal membranes will tend to melt (go from solid to liquid) at lower temperatures.[142] At the same time the ether linkage is more stable (requires more energy to break) than the ester linkage. It remains unclear why it is that while all organisms use glycerol-based lipids, the bacteria and the eukaryotes use hydrocarbon chain lipids, while the archaea use isoprene-based lipids. One speculation is that archaeal were originally adapted to live at higher temperatures, where the greater stability of the ether linkage would provide a critical advantage.



At the highest temperatures, thermal motion might be expected to disrupt the integrity of the membrane, allowing small charged molecules (ions) through.[143] Given the importance of membrane integrity, we will (perhaps) not be surprised to find "double-headed" lipids in organisms that live at high temperatures (thermophiles and hyperthermophiles).These lipid molecules have a two distinct hydrophilic glycerol moieties, one located at each end of the molecule; this enables them to span the membrane. The presumption is that such lipids act to stabilize the membrane against the disruptive effects of high temperatures - important since some archaea live (happily, apparently) at temperatures up to 110 ºC[144]. Similar double-headed lipids are also found in bacteria that live in high temperature environments.



That said, the solid-fluid nature of biological membranes, as a function of temperature, is complicated by the presence of cholesterol and structurally similar lipids. For example, in eukaryotes the plasma membrane can contain as much as 50% (by number of lipid molecules present) cholesterol. Cholesterol has a short bulky hydrophobic domain that does not pack well with other lipids (**FIG**: a hydrocarbon chain lipid (left) and cholesterol (right)). When present, it dramatically influences the solid-liquid behavior of the membrane. The diverse roles of lipids is a complex subject that goes beyond our scope here.[145]



[141]A re-evaluation of the archaeal membrane lipid biosynthetic pathway: http://www.nature.com/nrmicro/journal/v12/n6/full/nrmicro3260.html

[142]The origin and evolution of Archaea: a state of the art: http://rstb.royalsocietypublishing.org/content/361/1470/1007.full

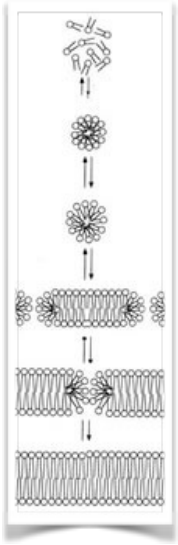
[143] Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea. : http://www.ncbi.nlm.nih.gov/pubmed/8825096

[144] You might want to consider how this is possible and under want physical conditions you might find these "thermophilic" archaea.

[145] At this point, such a search recovers 636 papers (and there are many more than concern lipid function but do not contain lipidomics in the title.

The origin of biological membranes

The modern cell membrane is composed of a number of different types of lipids. Those lipids with one or more hydrophobic “tails” have tails that typically range from 16 to 20 carbons in length. The earliest membranes, however, were likely to have been composed of similar, but simpler molecules with shorter hydrophobic chains. Based on the properties of lipids, we can map out a plausible sequence for the appearance of membranes. Lipids with very short hydrophobic chains, 2 to 4 carbons in length, can dissolve in water (can you explain why?) As the lengths of the hydrophobic chains increases, the molecules begin to self-assemble into micelles. By the time the hydrophobic chains reach ~10 carbons in length, it becomes increasingly more difficult to fit the hydrocarbon chains into the interior of the micelle without making larger and larger spaces between the hydrophilic heads. Water molecules can begin to move through these spaces and interact with the hydrocarbon tails. At this point, the hydrocarbon-chain lipid molecules begin to associate into semi-stable bilayers. One interesting feature of these bilayers is that the length of the hydrocarbon chain is no longer limiting in the same way that it was limiting in a micelle. One problem, though, are the edges of the bilayer, where the hydrocarbon region of the lipid would come in contact with water, a thermodynamically unfavorable situation. This problem is avoided by linking edges of the bilayer to one another, forming a balloon-like structure. Such bilayers can capture regions of solvent, that is water and any solutes dissolved within it.



Bilayer stability increases further as hydrophobic chain length increases. At the same time, membrane permeability decreases. It is a reasonable assumption that the earliest biological systems used shorter chain lipids to build their "proto-membranes" and that these membranes were relatively leaky.[146] The appearance of more complex lipids, capable of forming more impermeable membranes must therefore have depended upon the appearance of mechanisms that enabled hydrophilic molecules to pass through membranes. The process of interdependence of change is known as co-evolution. Co-evolutionary processes were apparently common enough to make the establishment of living systems possible. We will consider the ways through a membrane in detail below.

Questions to answer & to ponder:

- Draw diagrams to show how increasing the length of a lipid's hydrocarbon chains affects the structures that it can form.
- How are the effects at the hydrophobic edges of a lipid bilayer minimized?
- What types of molecules might be able to go through the plasma membrane on their own?
- In the light of the cell theory, what can we say about the history of cytoplasm and the plasma membrane?
- Why do fatty acid and isoprene lipids form similar bilayer structures?
- Speculate on why it is common to see phosphate and other highly hydrophilic groups attached to the glycerol groups of lipids?
- Are the membranes of bacteria and archaea homologous or analogous? What type of data would help you decide?
- Why is the movement of materials through the membrane essential for life?
- Why do membrane lipids solidify at low temperature? How are van der Waals interactions involved? Are H-bond type electrostatic interactions involved?

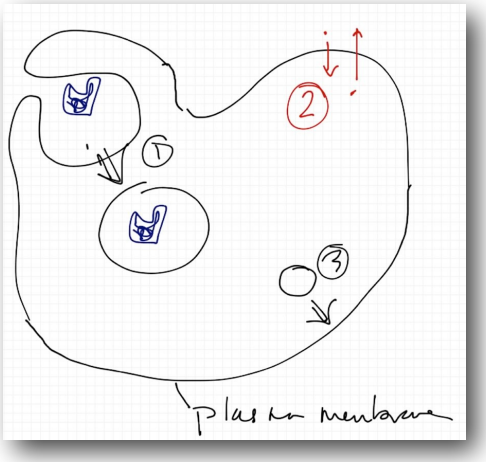
[146] http://astrobiology.arc.nasa.gov/workshops/1996/astrobiology/speakers/deamer/deamer_abstract.html

- Predict (and justify) the effect of changing the position of a double bond in a hydrocarbon chain on the temperature of membrane solidification.
- Would a membrane be more permeable to small molecules at high or low temperatures and why?

Transport across membranes

As we have said before (and will say again), the living cell is a continuous non-equilibrium system. To maintain its living state both energy and matter have to move into and out of the cell, which leads us to consider both the intracellular and extracellular environments and the membrane that separates them. The differences between the inside and the outside of the plasma membrane are profound. Outside, even for cells within a multicellular organism, the environment is generally mostly water, with relatively few complex molecules. Inside, the membrane-defined space, is a highly concentrated (> 60 mg/ml) solution of proteins, nucleic acids, smaller molecules, and thousands of interconnected chemical reactions, known collectively as cytoplasm. Cytoplasm (and the membrane around it) is inherited by the cell when it was formed, and represents an uninterrupted continuous system that first arose billions of years ago.

A lipid bilayer membrane poses an interesting barrier to the movement of molecules. First for larger molecules, particles or other organisms, it acts as a physical barrier. Typically when larger molecules, particles (viruses), and other organisms enter a cell, they are actually engulfed by the membrane, in a range of processes from pinocytosis (cell drinking) to endocytosis (cell entry) and phagocytosis (cell eating)(process 1). A superficially similar process, running in "reverse", known as endocytosis (process 3), is involved in moving molecules to the cell surface and releasing them into the extracellular space. Both endocytosis and exocytosis involve membrane vesicles emerging from or fusing into the plasma membrane. These processes leave the topology of the cell unaltered, in the sense that a molecule within a vesicle is still "outside" of the cell, or at least outside of the cytoplasm. These movements are driven by various molecular machines that we will consider only briefly; they are typically considered in greater detail in courses on cell biology. We are left with the question of how molecules can enter or leave the cytoplasm, this involves passing directly through a membrane (process 2).



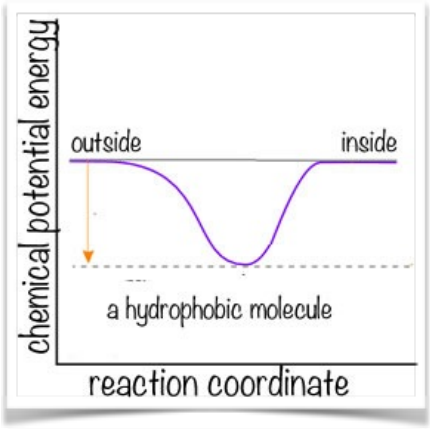
Transport to and across the membrane

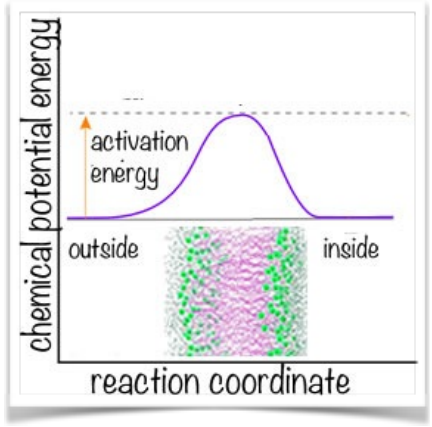
So the question is, how does the membrane "decide" which molecules to allow into and out of the cell. If we think about it, there are three possible general mechanisms (let us know if you can think of more). Molecules can move on their own through the membrane, they can move passively across the membrane using some type of specific "carrier" or "channel", or they could be moved actively using some kind of "pump". In particular, which types of carriers, channels, and pumps are present will determine what types of molecules move through the membrane. As you might deduce pumps require a source of energy to drive them. As we will see, in the vast majority of cases, these carriers, channels,

and pumps are protein-based molecular machines, the structure of which we will consider in detail later on. We can think of this molecular movement reaction generically as:

$$Molecule_{outside} \rightleftharpoons Molecule_{inside\ membrane} \rightleftharpoons Molecule_{inside}.$$

As with standard chemical reactions, movement through a membrane involves an activation energy, which amounts to the energy needed to pass through the membrane. So, you might well ask, why does the membrane, particularly the hydrophobic center of the membrane, pose a barrier to the movement of hydrophilic molecules. Here the answer involves the difference in the free energy of the moving molecule within an aqueous solution, including the hydrophilic surface region of the membrane, where H-bond type electrostatic interactions are common between molecules, and the hydrophobic region of the membrane, where only van der Waals interactions are present. The situation is exacerbated for charged molecules, since water molecules are typically organized in a dynamic shell around an ion. Instead of reactants and products we can plot the position of the molecule relative to the membrane. We are considering molecules of one particular substance moving through the membrane and so the identity of the molecule does not change. If the concentrations of the molecules are the same on both sides of the membrane, then their Gibbs free energies are also equal, the system will be in equilibrium with respect to this reaction. In this case, as in the case of chemical reactions, there will be no net flux of the molecule across the membrane, but molecules will be moving back and forth at an equal rate. The rate at which they move back and forth will depend on the size of the activation energy associated with moving across the membrane.



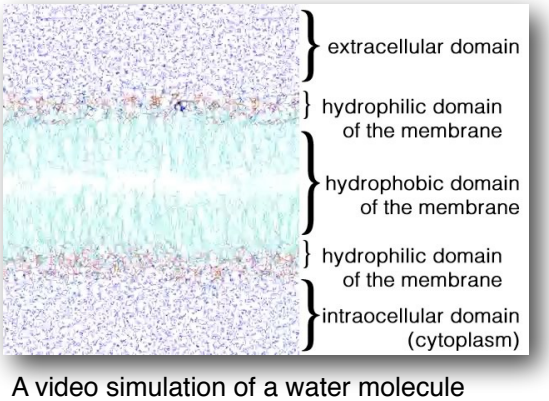


http://youtu.be/JxtnneWWzHo

If a molecule is hydrophobic (non-polar) it will be more soluble in a hydrophobic environment in the center of the membrane than it is in an aqueous environment. In contrast the situation will be distinctly different for hydrophilic molecules. By this point, we hope you will recognize that in a simple lipid-only membrane (a biologically unrealistic case), the shape of this graph, and specifically the height of the activation energy peak will vary depending upon the characteristics of the molecule we are considering moving as well as the membrane itself. If the molecule is large and highly hydrophilic, for example, if it is charged, the activation energy associated with crossing the membrane will be higher than if the molecule is small and uncharged. Just for fun, you might consider what the reaction diagram for a single lipid molecule might look like; where might it be located, and what energy barriers are associated with its movement (flipping) across a membrane.

Let us begin with water itself, which is small and uncharged. When a water molecule begins to leave water and enter the hydrophobic (central) region of the membrane, there are no H-bonds to take the place of those that are lost, no strong handshakes, and often the molecule is “pulled back” into the water. Nevertheless, there are so many molecules of water outside (and inside) the cell, and water molecules are so small, that once they enter the membrane, they can pass through it. The activation

energy for the $Water_{outside} \rightleftharpoons Water_{inside}$ reaction is low enough that water can pass through a membrane (in both directions) at a reasonable rate.

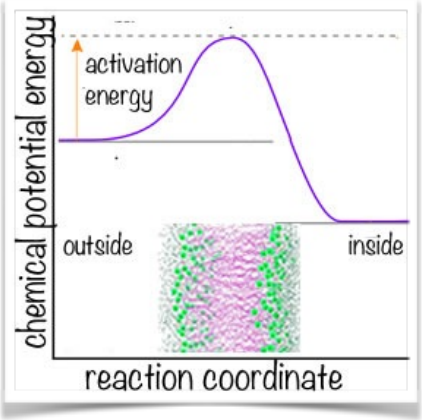


A video simulation of a water molecule moving through a membrane: http://youtu.be/ePGqRaQiBfc

Small non-polar molecules, like O_2 and CO_2, can (very much like water) pass through a biological membrane relatively easily. There is more than enough energy available through collisions with other molecules (thermal motion) to provide them with the energy needed to overcome the activation energy and pass through the membrane. However now we begin to see changes in free energies of the molecules on the inside and outside of the cell. For example, in organisms that depend upon O_2 (obligate aerobes), the O_2 outside of the cell comes from the air (it is generated by plants that release O_2 as a waste product.) Once O_2 enters the cell, it takes part in the reactions of respiration (we will get back to both processes further on.) The result is that the concentration of O_2 outside the cell will be greater than the concentration of O_2 inside the cell. That means that the free energy of O_2 outside will be greater than the free energy of O_2 inside. The reaction $O_{2\ outside} \rightleftharpoons O_{2\ inside}$ is now thermodynamically favorable and there will be a net flux of O_2 into the cell. We can consider how a similar situation applies to water. The intracellular domain of a cell is a concentrated solution of proteins and other molecules. Typically, the concentration of water outside of the cell is greater than the concentration of water inside the cell. Our first order presumption is that the reaction:

$$H_2O_{\ outside} \rightleftharpoons H_2O_{\ inside}$$

is favorable, so water will flow into a cell. So the obvious question is, what happens over time? We will return to how cell's (and organisms) resolve this important problem shortly.



Channels and carriers

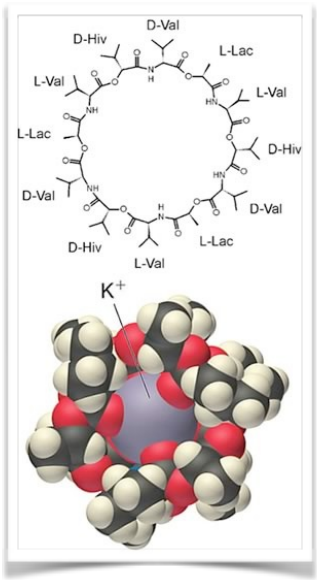
Beginning around the turn of the last century, a number of scientists began working to define the nature of cell's boundary layer. In the 1930's it was noted that small, water soluble molecules entered cells faster than predicted based on the assumption that the membrane acts like a simple hydrophobic barrier - an assumption known as Overton's Law. Collander et al., postulated that membranes were more than simple hydrophobic barriers, specifically that they contained features that enabled them to act as highly selective molecular sieves. Most of these are proteins (never fear, we are getting closer to a more thorough discussion of proteins) that can act as channels, carriers, and pores. If we think about crossing the membrane as a reaction, then the activation energy of this reaction for highly hydrophilic and larger molecules will be quite high, we will need a catalyst to reduce it. There are two generic types of membrane permeability catalysts available: carriers and channels.

Carrier proteins are membrane proteins that can shuttle back and forth across the membrane. They can bind to specific hydrophilic molecules when they are located in the hydrophilic region of the membrane, hold on to the bound molecule as they traverse the hydrophobic region of the membrane, and then release their "cargo" when they again reach the hydrophilic region of the membrane. These movements of carrier and cargo across the membrane are driven by thermal motions, so no other energy source is necessary. We can write this class of reactions as:

$$Molecule_{outside} + carrier_{empty} \rightleftharpoons carrier\text{–}\ Molecule_{outside} \rightleftharpoons carrier\text{–}\ Molecule_{inside} \rightleftharpoons Molecule_{inside} + carrier_{empty}.$$

There are many different types of carrier molecules and each type of carrier has a preferred cargo molecule. Related molecules may be bound and transported, but with much less specificity (and so at a much lower rate). So exactly which molecules a particular cell will allow to enter will be determined in part by which carrier protein genes it expresses. Mutations in a gene encoding a carrier can change (or abolish) the range of molecules that that carrier can transport across a membrane.

Non-protein carriers: An example of a carrier is a class of antibiotics that carry ions across membranes. These molecules are known generically as ionophores. They kill cells by disrupting the normal ion balance across the membrane and within the cytoplasm, which in turn is thought to disrupt normal metabolic activity.[147] One of these is valinomycin (→), a molecule made by *Streptomyces* type bacteria. The valinomycin molecule has a hydrophobic periphery and a hydrophilic core. It binds K^+ ions approximately 10^5 times more effectively than it binds Na^+. It shuttles (with the bound ion) back and forth across the membrane. In the presence of a K^+ gradient, that is a higher concentration of K^+ on one side of the membrane compared to the other, the presence of valinomycin will produce a net flux of K^+ across the membrane. Again, to be clear, in the absence of a gradient, K^+ ions will still move across the membrane (in the presence of the carrier), but there will be no net change in the concentration of K^+ ion inside the cell. For the experimentally inclined, you might consider how you could prove that movements are occurring even in the absence of a gradient. In a similar manner, there are analogous carrier systems that move hydrophobic molecules through water.



Channel molecules sit within a membrane. They contain a channel that spans the membrane's hydrophobic region. Hydrophilic molecules of particular sizes and shapes can pass through this "aqueous" channel and their movement involves a much lower activation energy than would be associated with moving through the lipid part of the membrane. Channels are generally very selective in terms of which particles pass through them. For example, there are channels in which 10,000 potassium ions will pass through for every one sodium ion.

The channels in these proteins can be regulated; they can exist in two or more distinct structural states. For example, in one state the channel can be open and allow particles to pass through or it can be closed, that is the channel can be turned on and off. The transition between open and closed states

[147] That said, there is little data in the literature on exactly which cellular processes are disrupted by which ionophore; in mammalian cells (as we will see) these molecules are by disrupting ion gradients in mitochondria and chloroplasts, apparently.

can be regulated through a number of processes, including the reversible binding of small molecules and various other molecular changes (which we will consider when we talk about proteins) or changes in electrochemical gradients across the membrane.

Another method of channel control depends on the fact that channel proteins are embedded within a membrane and are contain of charged groups. As we will see, cells can (and generally do) generate ion gradients, that is a separation of charged species, across their membranes. For example if the concentration of K^+ ions is higher on one side of the membrane, there will be an ion gradient where the natural tendency is for the ions to move to the region of lower K^+ concentration[148]. The ion gradient in turn can produce electrical fields across the plasma membrane. As these fields change, they can produce (induce) changes in channel structure, which can switch the channel from open to closed and vice versa. Organisms typically have many genes that encode specific channel proteins which are involved in a range of processes from muscle contraction to thinking. As in the case of carriers, channels do not determine the direction of molecular motion. The net flux of molecular movement is determined by the gradients of molecules across the membrane, with the thermodynamic driver being entropic factors. That said, the actual movement of the molecules through the channel is driven by thermal motion.

Questions to answer & to ponder:

- What does it mean to move up a concentration gradient?
- Are there molecules that can move up their concentration gradients spontaneously?
- Where does the energy involved in moving molecules come from? Is there a "force" driving the movement of molecules "down" their concentration gradient?
- If there is no net flux of A, even if there is a concentration gradient between two points, what can we conclude?
- What happens to the movement of molecules through channels and transporters if we reverse the concentration gradients across the membrane?
- Is energy needed to maintain gradients across a membrane (what is your thermodynamic logic)?
- Why do we need to add energy to maintain gradients?
- Which (and why) would you think would transport molecules across a membrane faster, a carrier, a channel, or a pump?

Generating gradients: using coupled reactions and pumps

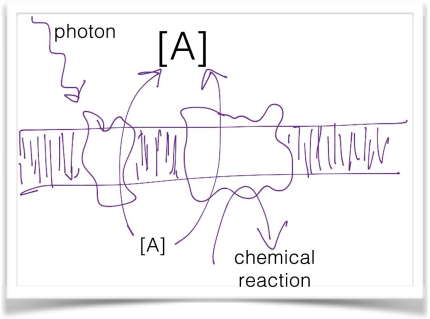
Both carriers and channels can allow the directional movement (net flux) of molecules across a membrane, but only when a concentration gradient is present. If a membrane contains active channels and carriers (as all membranes do), without the input of energy eventually concentration gradients across the membrane will disappear (disperse). The $[molecule]_{outside}$ will become equal to $[molecule]_{inside}$. Yet, when we look at cells we find lots of concentration gradients, which raises the question, what produces and then maintains these gradients.

The common sense answer is that there must be molecules (proteins) that can move specific molecules through a membrane against their concentration gradient. We will call this type of molecule a pump and write the reaction it is involved in as:

$$[Molecule]_{low\ concentration} + pump \longleftrightarrow [Molecule]_{high\ concentration} + pump$$

[148] In fact this tendency for species to move from high to low concentration until the two concentrations are equal can be explained by the Second Law of Thermodynamics. Check with your chemistry instructor for more details

As you might already suspect this is a thermodynamically unfavorable reaction. Like a familiar macroscopic pump, it will require the input of energy. We will have to "plug in" our molecular pump into a source of energy. What energy sources are available to biological systems? Basically we have two choices: the system can use electromagnetic energy, that is light, or it can use chemical energy. In a light-driven pump, there is a system that captures (absorbs) light which is then coupled to the pumping system. Where the pump is driven by a chemical reaction, the thermodynamically favorable reaction is often catalyzed by the pump itself and coupled to the movement of a molecule against its concentration gradient. An interesting topological point is that for a light or chemical reaction driven pump to work to generate a concentration gradient, all of the pump molecules within a membrane must be oriented in the same direction. If the pumps were oriented randomly there probably would be no overall flux (the molecules would move in both directions) and no gradient would develop.



Chemical-reaction driven pumps are also oriented within membranes in the same direction. A number of chemical reactions can be used to drive such pumps and these pumps can drive various reactions (remember reactions can move in both directions). The most common ones are the movement of energetic electrons through a membrane-bound, protein-based "electron transport" system, leading to the creation of an H^+ electrochemical gradient. The movement of H^+ down its concentration gradient through the pump then drives the synthesis of ATP:

$$H^+ \text{ from } [H^+]_{high} \rightleftharpoons H^+ \text{ to } [H^+]_{low}$$

which is coupled to

$$\text{adenosine diphosphate (ADP) + phosphate} \rightleftharpoons \text{adenosine triphosphate (ATP) + } H_20$$

or through the hydrolysis of adenosine triphosphate, a highly thermodynamically favorable reaction:

$$H^+ \text{ from } [H^+]_{low} \rightleftharpoons H^+ \text{ to} [H^+]_{high}$$

is coupled to

$$\text{adenosine triphosphate (ATP) + } H_20 \rightleftharpoons \text{adenosine diphosphate (ADP) + phosphate.}$$

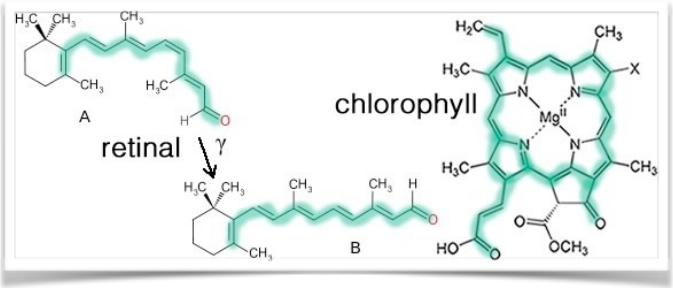
By coupling a ATP hydrolysis reaction to the pump, the pump can move molecules from a region of low concentration to one of high concentration, a thermodynamically unfavorable reaction.

Simple Phototrophs

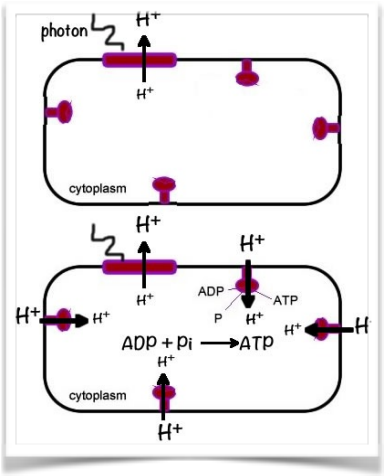
Phototrophs are organisms that capture light particles (photons) and transform their electromagnetic energy into energy stored in unstable molecules, such as ATP and carbohydrates. Light can be considered as both a wave and a particle (that is quantum physics for you) and the wavelength of a photon determines its color and the amount of energy it contains. Again, because of quantum mechanical factors, a particular molecule can only absorb photons of specific wavelengths (energies) - in fact, we can identify molecules based on the photons they absorb, this is the basis of spectroscopy. Our atmosphere allows mainly visible light from the sun to reach the earth's surface, but most biological molecules do not absorb visible light very effectively or at all. To capture this energy, organisms have evolved the ability to synthesize special molecules, known as pigments to capture, and

therefore allow organisms to use visible light. The color we see for a typical pigment is the color of the light that is not absorbed but rather is reflected. For example chlorophyl appears green because light in the red and blue regions of the spectrum is absorbed and green light is reflected. The question we need to answer is how does the organism use the electromagnetic energy that is absorbed?

One of the simplest examples of a phototrophic system, that is, a system that directly captures the energy of light and transforms it into the energy stored in the chemical system, is provided by the archaea *Halobacterium halobium*.[149] *Halobacteria* are extreme halophiles or salt-loving organisms. They live in waters that contain up to 5M NaCl. *H. halobium* uses the membrane protein, bacteriorhodopsin to capture light. Bacteriorhodopsin consists of two components, a polypeptide, known generically as an opsin, and a non-polypeptide prosthetic group, the pigment retinal, a molecule derived from vitamin A.[150] Together the two, opsin + retinal, form the functional bacteriorhodopsin protein.



Retinal absorbs visible light. This is because its electrons are located in extended molecular orbitals that have energy gaps between them that are of the same order as the energy of visible light. This extended molecular orbital (highlighted in the figure) is associated with a region of the molecule drawn as containing alternating single and double bonds between carbons, which we call a conjugated pi orbital system. Similar conjugated pi systems are responsible for the absorption of light by other pigments, like chlorophyll and heme. When a photon of light is absorbed by the retinal group, it undergoes a reaction that leads to a change in the pigment molecule's shape and composition, which in turn leads to a change in the structure of the polypeptide to which the retinal group is attached. This is called a photoisomerization reaction.



The bacteriorhodopsin protein is embedded within the plasma membrane, where it associates with other bacteriorhodopsin proteins to form patches of proteins. These patches of membrane protein give the organisms their purple color and are known as purple membrane. When one of these proteins absorbs light, the change in the associated retinal group produces a light-induced change in protein structure that results in the movement of a H^+ ion from the inside of the cell to the outside of the cell. The protein (and its associate pigment) then return to its original low energy state, that is, its state before it absorbed the photon of light. Because all of the bacteriorhodopsin molecules are oriented in the same way in the membrane, as light is absorbed all of the H^+ ions move in the

[149] http://youtu.be/4OkN1QC4hyY

[150] As we will return to later, proteins are functional entities, composed of polypeptides and prosthetic group. The prosthetic group is essential for normal protein function. The protein without the prosthetic group is known as the apoprotein.

same direction, leading to the formation of a H^+ concentration gradient across the plasma membrane with $[H^+]_{outside} > [H^+]_{inside}$. This H^+ gradient is based on two sources. First there is the gradient of H^+ ions. As light is absorbed the concentration of H^+ outside the cell increases and the concentration of H^+ inside the cell decreases. The question is – where is this H^+ coming from? As you (perhaps) learned in chemistry water undergoes the reaction (although this reaction is quite unfavorable):

$$H_2O \rightleftharpoons H^+ + OH^-.$$

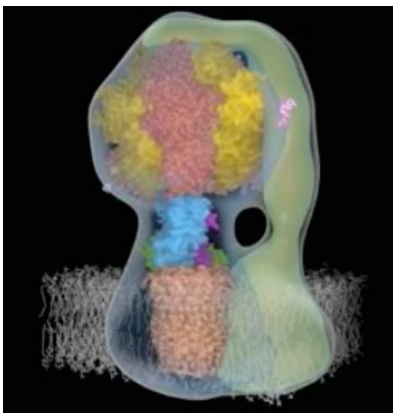
So H^+ is already present in water and it is these H^+ s that move.

In addition to the chemical gradient in H^+ ions that forms as H^+ ions are pumped out of the cell by the bacteriorhodopsin + light + water reaction, an electrical field is also established. There are excess + charges outside of the cell (from H^+ being moved there) and excess – charges inside the cell (from –OH being left behind). As you know from your physics, positive and negative charges attract, but the membrane stops them from reuniting. The result is the accumulation of positive charges on the outer surface of the membrane and negative charges on the inner surface. This charge separation produces an electric field. Now, a H^+ outside of the cell will experience two forces. If there is a way across the membrane, the $[H^+]$ gradient will lead to its movement back into the cell. Similarly the electrical field will also drive the positively charged H^+ back into the cell. The formation of the $[H^+]$ gradient basically generates a battery, a source of energy, into which we can plug in our pump.

So how does the pump tap into this battery? The answer is a second membrane protein, an enzyme known as the H^+-driven ATP synthase. H^+ ions move through the ATP synthase molecule, which is a thermodynamically favorable reaction. The ATP synthase couples this favorable movement to an unfavorable chemical reaction, a condensation reaction:

$$H^+_{outside} + ADP + \text{inorganic phosphate} \xrightleftharpoons{\text{ATP synthase}} ATP + H_2O + H^+_{inside}$$

This reaction will continue as long as light is absorbed and bacteriorhodopsin acts to generate a H^+ gradient. It will also continue for a time after the light goes off (that is, night time) because it takes time for H^+ ions to move through the ATP synthase and for the H^+ gradient to dissipate, but after a short while (in the dark), net ATP synthesis will slow and stop. The point of this process is that, in the light, the cell generates (and stores for later use in various coupled reactions) ATP. ATP acts as a type of chemical battery, in contrast to the electrochemical battery of the H^+ gradient.



link: http://youtu.be/J8lhPt6V-yM

An interesting feature of the ATP synthase is that as H^+ ions move through it (driven by the electrochemical power of the H^+ gradient), it rotates. It is worth noting that there is no thermodynamic reason that the ATP synthase cannot run in the opposite direction. In fact, it can catalyze (and couple) the hydrolysis of ATP to the pumping of H^+ out of the cell:

$$ATP + H_2O + H^+_{inside} \xrightleftharpoons{\text{ATP synthase}} H^+_{outside} + ADP + \text{inorganic phosphate}$$

Because it catalyzes the hydrolysis of ATP, the enzyme can be called an ATP hydrolase. Again, when it catalyzes the hydrolysis of ATP, it rotates, although in the opposite direction compared to when it catalyzes the synthesis of ATP. Now its energy driven rotation (by either the electrochemical H^+ battery or ATP hydrolysis) raises in interesting possibility. This enzyme (or rather a variant) could be used to drive the swimming movement of cells (imagine connecting it to some kind of propeller.)

beSocratic exercise: Draw a membrane, place bacteriorhodopsin molecules in it, mark their orientation and the direction of the H^+ gradient that arises in the light. Draw the ATP synthase, indicate how movement of H^+ leads to ATP synthesis. Indicate how ATP hydrolysis or tapping into the H^+ gradient could lead to cell movement. Can you imagine and describe other mechanisms that could move cells?

Chemo-osmosis (an overview)

One of the most surprising discoveries in biology was the wide spread, almost universal use of H^+ gradients to generate ATP. It was originally known as the chemiosmotic hypothesis by the eccentric British scientist, Peter Mitchell (1920 – 1992).[151] Before the significance of H^+ membrane gradients was known, Mitchell proposed that energy captured through the absorption of light (by phototrophs) or the breakdown of molecules into more stable molecules (by various types of chemotrophs) relied on the same basic (homologous) mechanism, namely the generation of H^+ gradients across membranes (the plasma membrane in prokaryotes or the internal membranes of mitochondria or chloroplasts (intracellular organelles, derived from bacteria)(see below) in eukaryotes.

What makes us think that these processes have a similar evolutionary root, that they are homologous? It is that in both light and chemical based processes, captured energy is transferred through the movement of electrons through a membrane-embedded "electron transport chain." This chain involves a series of reactions, specifically reduction-oxidation or redox reactions (see below) during which electrons move from a high energy to a lower energy state. Some of this energy difference is used to move H^+ ions across the membrane and so generate a H^+ concentration gradient. The thermodynamically favorable movement of H^+ down this concentration gradient is then used to drive ATP synthesis (a thermodynamically unfavorable process.) ATP synthesis itself involves the rotating ATP synthase. The movement of H^+ ions down the H^+ gradient through the ATP synthase drives the reaction:

$$H^+_{outside} + ADP + Pi \rightleftharpoons ATP + H_2O + H^+_{inside},$$

where "inside" and "outside" refer to compartments defined by the membrane containing the electron transport chain and the ATP synthase. Again, this reaction can run backwards. When this occurs, the ATP synthase acts as an ATPase that can pump H^+ (or other molecules) against their concentration gradient. In fact the action of such pumping ATPases establishes many biologically important molecular gradients across membranes. In such a reaction:

$ATP + H_2O$ + molecule in low concentration region $\rightleftharpoons$ ADP + Pi + molecule in low concentration region.

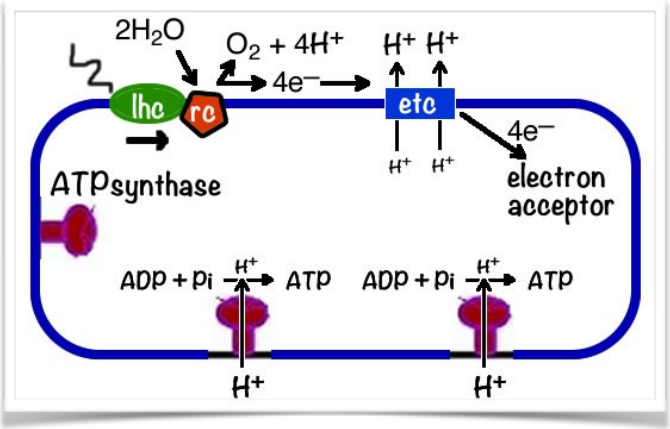
In an sense, the most important difference between phototrophs and chemotrophs is how high energy electrons enter the electron transport chain.

[151] http://en.wikipedia.org/wiki/Peter_D._Mitchell

Oxygenic photosynthesis

Compared to the salt loving archaea *Halobium*, with its purple, bacteriorhodopin-rich membranes, photosynthetic cyanobacteria (which are bacteria), green algae, and higher plants (both eukaryotes) use more complex systems to capture and utilize light. In all of these organisms, their photosynthetic systems appear to be homologous, that is derived from a common ancestor, a topic we will return to later in this chapter. For simplicity's sake, we will describe the photosynthetic system of cyanobacterium; the system in eukaryotic algae and plants, while more complex, follows the same basic logic. At this point, we consider only one aspect of this photosynthetic system, known as the oxygenic or non-cyclic system (look to more advanced classes for more details.) The major pigment in this system, chlorophyll, is based on a complex molecule, a porphyrin (see above) and it is primarily these pigments that give plants their green color. As in the case of retinal, they absorb visible light due to the presence of a conjugated structure (drawn as a series of single and double) carbon-carbon bonds. Chlorophyll is synthesized by a conserved biosynthetic pathway that is also used to synthesize heme, which is found in the hemoglobin of animals and in the cytochromes within the electron transport chain present in both plants and animals (which we will come to shortly), vitamin B12, and other biologically important prosthetic (that is non-polypeptide) groups associated with proteins and required for their normal function.[152]

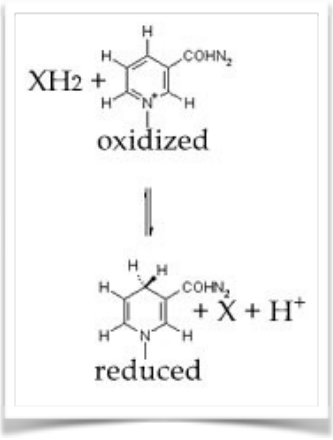
Chlorophyll molecules are organized into two distinct protein complexes that are embedded in membranes. These are known as the light harvesting and reaction center complexes. Light harvesting complexes (lhc) act as antennas to increase the amount of light the organism can capture. When a photon is absorbed, an electron is excited to a higher molecular orbital. An excited electron can be passed between components of the lhc and eventually to the reaction center ("rc") complex. Light harvesting complexes are important because photosynthetic organisms can compete with one another for light, so their presence can enable a photosynthetic organism to flourish at lower light levels.



In the oxygenic, that is molecular oxygen (O_2) generating (non-cyclic) photosynthesis reaction system, high energy (excited) electrons are passed from the reaction center to a complex of membrane proteins known as the electron transport chain ("etc"). As an excited electron moves through the etc its energy is used to move H^+s from inside to outside of the cell. This is the same geometry of H^+ movement that we saw previously in the case of the purple membrane system. The end result is the formation of a H^+ based electrochemical gradient. As with purple bacteria, the energy stored in this H^+ gradient is used to drive the synthesis of ATP within the cell's cytoplasm.

[152] Mosaic Origin of the Heme Biosynthesis Pathway in Photosynthetic Eukaryotes: http://mbe.oxfordjournals.org/content/22/12/2343.full.pdf

Now you might wonder, what happens to the originally excited electrons, and the energy that they carry. In what is known as the cyclic form of photosynthesis, low energy electrons from the electron transport chain are returned to the reaction center, where they return the pigments to their original (before absorbing light) state. In contrast, in the non-cyclic process that we have been considering, electrons from the electron transport chain are delivered to an electron acceptor. Generally this involves the absorption of a second photon, a mechanistic detail that need not trouble us here. This is a general type of chemical reaction known as an oxidation-reduction (redox) reaction. Where electrons are within a molecule's electron orbital system determines the amount of energy present in the molecule. In this light, it makes sense that adding an electron to a molecule will (generally) increase the molecule's overall energy, and make it less stable. When an electron is added to a molecule, that molecule is said to have been "reduced", and yes, it does seem weird that adding an electron "reduces" a molecule. If an electron is removed, the molecule's energy is changed and the molecule is said to have been "oxidized".[153] Since electrons, like energy, are neither created nor destroyed in biological systems (remember, no nuclear reactions), the reduction of one molecule is always coupled to the oxidation of another. For this reason, reactions of this type are referred to as "redox" reactions. During such a reaction, the electron acceptor is said to be "reduced". Reduced molecules are generally unstable, so the reverse, thermodynamically favorable reaction, in which electrons are removed (known as oxidation) can be used to drive various types of thermodynamically unfavorable metabolic reactions.



Given the conservation of matter in biological systems, if electrons are leaving the photosynthetic system (in the non-cyclic process) they must be replaced. So where do they come from? Here we see what appears to be a major evolutionary breakthrough. During the photosynthetic process, the reaction center couples light absorption with the oxidation (removal of electrons) from water molecules:

$$\text{light} + 2H_2O \rightleftharpoons 4H^+ + 4e^- + O_2.$$

The four electrons, derived from two molecules of water, pass to the reaction center, while the $4H^+$s contribute to the proton gradient across the membrane.[154] O_2 is a waste product of this reaction. Over millions of years, the photosynthetic release of O_2 changed the Earth's atmosphere from containing essentially 0% molecular oxygen to the current ~21% level. Because O_2 is highly reactive, this transformation is thought to have been a major driver of subsequent evolutionary change. However, there remain even today organisms that cannot use O_2 and cannot survive in its presence. They are known as obligate anaerobes (to distinguish them from organisms that normally grow in the absence of O_2 but which can survive in its presence, which are known as facultative anaerobes. In the past the level of atmospheric O_2 has changed dramatically based on how much O_2 was released into the atmosphere by oxygenic photosynthesis and how much was removed by various reactions, such as the decomposition of plant materials. When large amounts of plant materials are buried before they could decay, such as occurred with the formation of coal beds, during the Carboniferous period (from ~360 to

[153] you can review redox here: http://www.biologie.uni-hamburg.de/b-online/e18/18b.htm or in CLUE: http://besocratic.colorado.edu/CLUE-Chemistry/chapters/chapter7txt.html

[154] Photosystem II and photosynthetic oxidation of water: an overview: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1693055/

299 million years ago), the level of atmospheric O_2 increased dramatically, to an estimated level of ~35%. It is speculated that such high levels of molecular oxygen made it possible for organisms without lungs (like insects) to grow to gigantic sizes.[155]

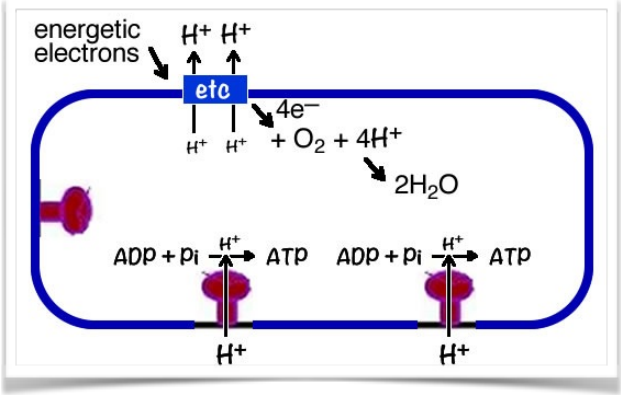
Chemotrophs

Those organisms that are not phototrophic capture energy from other sources, specifically by transforming thermodynamically unstable molecules into more stable species. These organisms are known generically as chemotrophs. They can be divided into various groups, depending upon the types of food molecules they use. They include organotrophs, which use carbon-containing molecules (you are an organotroph) and lithotrophs (or rock eaters), which use various inorganic molecules. In the case of organisms that can "eat" H_2, the electrons that result are delivered along with accompanying H^+ ions to CO_2, to form methane (CH_4) following the reaction:

$$CO_2 + 4H_2 \rightleftharpoons CH_4 + 2H_2O;$$

Because of this they are referred to as methanogens (methane-producers).[156] In the modern world methanogens (typically archaea) are found in environments with low O_2 such as your gut. In many cases, such reactions can occur only in the absence of O_2. In fact, O_2 is so reactive, that it can be thought of as a poison, particularly for organisms that cannot actively "detoxify" it. When we think about the origins and subsequent evolution of life, we have to consider how organisms that arose in the absence of molecular O_2 adapted to its introduction into their environment. It is commonly assumed that modern strict obligate anaerobes might still have features common to the earliest organisms.

The amount of energy that an organism can capture is determined by the energy of the electrons that the electron acceptor(s) they use can accept. If only high amounts of energy can be captured, then inevitably smaller amounts of energy have to be left behind. On the other hand, the lower the amount of energy that an electron acceptor can accept, the more energy can be captured from the original "food" molecules used and the less energy must be left behind. Molecular oxygen is unique in its ability to accept low energy electrons. For example, consider an organotroph that eats carbohydrates $[C_6H_{10}O_5]_n$, a class of molecules that includes various sugars, starches, and wood. In the absence of O_2, that is under anaerobic conditions, the end product of the breakdown of a carbohydrate leaves about 94% of the theoretical amount of energy present in the original carbohydrate molecule in molecules that cannot be broken down further by most organisms. However, when O_2 is present, the carbohydrate can be broken down completely into CO_2 and H_2O, a process known as glycolysis, from the Greek words meaning sweet (glyco) and splitting (lysis). In these organisms the energy stored in energetic electrons is used to generate a membrane-



[155] When Giants Had Wings and 6 Legs: http://www.nytimes.com/2004/02/03/science/when-giants-had-wings-and-6-legs.html

[156] http://en.wikipedia.org/wiki/Lithotroph

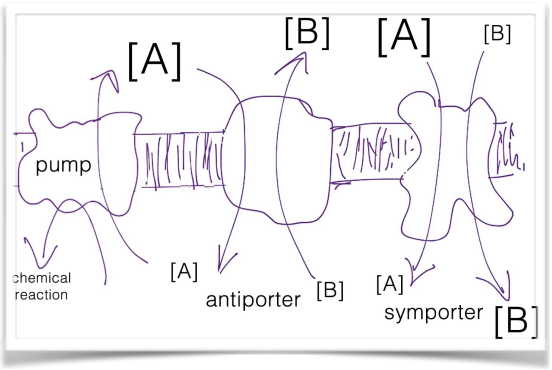
associated H+ based electrochemical gradient which in turn drives ATP synthesis, through the membrane-based ATP synthase. In an environment that contains molecular oxygen, organisms that use O_2 as an electron acceptor have a distinct advantage; instead of secreting energy rich molecules, like ethanol, they release the energy poor (stable) molecules CO_2 and H_2O.

No matter how cells (and organisms) capture energy, to maintain themselves and to grow, they must make a wide array of various complex molecules. Understanding how these molecules are synthesized lies within the purview of biochemistry. That said, in each case, thermodynamically unstable molecules (like lipids, proteins, and nucleic acids) are built through series of coupled reactions that rely on energy capture from light or the break down of food molecules.

Questions to answer & to ponder

- In a phototroph, why does the H+ gradient across the membrane dissipate when the light goes off? What happens to the rate of ATP production?
- What would limit the "size" of the H+ gradient that bacteriorhodopsin could produce?
- What would happen if bacteriorhodopsin molecules were oriented randomly within the membrane?
- What is photoisomerization? Is this a reversible or an irreversible reaction?
- How (do you suppose) does an electron move through an electron transport chain? Make a graph that describes its energy as it moves through the chain.
- In non-cyclic photosynthesis, where do electrons end up?
- What would happen to a cell's ability to make ATP if it where exposed to an H+ carrier or channel?
- Why are oxidation and reduction always coupled?
- Why are carbohydrates good for storing energy?
- If "photosynthesis is glycolysis run backward", why does glycolysis not emit light?
- Which do you think would have an evolutionary advantage, an organism growing aerobically or anaerobically? How do environmental conditions influence your answer?

Using the energy stored in membrane gradients



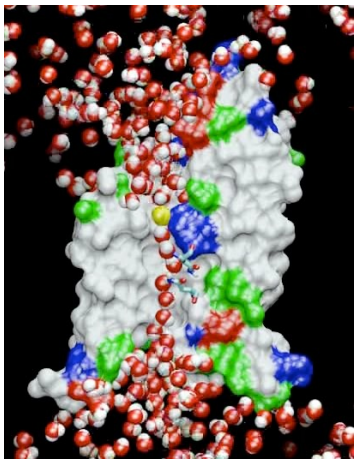
The energy captured by organisms (and their cells), is used to drive a number of processes in addition to synthesis reactions. For example, we have already seen that ATP synthases can act as pumps (ATP-driven transporters), coupling the favorable ATP hydrolysis reaction to the movement of molecules against their concentration gradients.The resulting gradient is a form of stored (potential) energy. This energy can be used to move other molecules, that is molecules that are not moved directly by a ATP-driven transporter. This involves what is known as coupled transport.[157] It uses membrane-bound proteins that allow a molecule to move down its concentration gradient. In contrast to simple carriers and channels, however, this thermodynamically favorable movement is physically coupled to the movement of a second molecule across the membrane and *against* its concentration gradient. When the two transported molecules move in the same direction, the transporter is known as a **symporter**,

[157] Structural features of the uniporter/symporter/antiporter superfamily: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2143070/

when they move in opposite directions, it is known as an **antiporter**. Which direction(s) the molecules move will be determined by the relative sizes of the concentration gradients of the two types of molecules moved. There is no inherent directionality associated with the transporter itself - the net movement of molecules reflects the relative concentration gradients of the molecules that the transporter can productively bind. What is important here is that energy stored in the concentration gradient of one molecule can be used to drive the movement of a second type of molecule against its concentration gradient. In mammalian systems, it is common to have Na^+, K^+, and Ca^{2+} gradients across the plasma membrane, and these are used to transport molecules into and out of cells. Of course, the presence of these gradients implies that there are ion-specific pumps that couple an energetically favorable reaction, typically ATP hydrolysis, to ion movement. Without these pumps (and the chemical reactions that drive them), the membrane battery would run down quite fast. Many of the immediate effects of death are due to the loss of membrane gradients and much of the energy needs of cells (and organisms) involves running such pumps.

Osmosis and living with and without a cell wall

Cells are packed full of molecules. These molecules take up space, space no longer occupied by water. The concentration of water outside of the cell $[H_2O]_{out}$ will necessarily be higher than the concentration of water inside the cell $[H_2O]_{in}$. This concentration gradient in solvent leads to the net movement of water into the cells[158]. Such a movement of solvent is known generically as osmosis. Much of this movement occurs through the membrane, which is somewhat permeable to water (see above). A surprising finding, which won Peter Agre a share of the 2003 Noble prize in chemistry. was that the membrane also contains water channels, known as aquaporins.[159] [This links to a molecular simulation of a water molecule (yellow) moving through an aquaporin →] It turns out that the rate of osmotic movement of water is dramatically reduced in the absence of aquaporins - they are important for cellular function. In addition to water, aquaporins can also facilitate the movement of other small uncharged molecules across a membrane.



http://www.ks.uiuc.edu/Research/aquaporins/waterpermeation.mpg

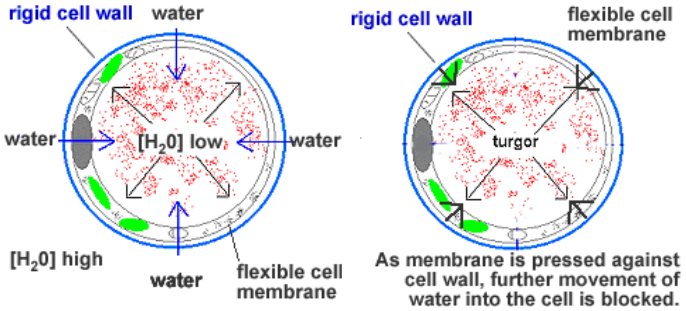
The difference or gradient in the concentrations of water (together with the presence of aquaporins) leads to a system that is capable of doing work, it can lift a fraction of the solution against the force of gravity. How is this possible? If we think of a particular molecule in solution, it will be moved around through collisions with its neighbors. These collisions drive the movement of particles randomly. But if there is a higher concentration of molecules on one side of a membrane compared to the other, then the random movement of molecules will lead to a net flux of molecules from the area of high concentration to that of low concentration, even though each molecule on its own moves randomly, that

[158] One important note here is that if you learn about osmosis in chemistry classes you will almost certainly be taught that water moves from a region of low SOLUTE concentration to a region of high SOLUTE concentration. These two definitions mean the same this but it is easy to get confused.

[159] Water Homeostasis: Evolutionary Medicine: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540612/

is, without a preferred direction [this video [160] is good at illustrating this behavior]. At equilibrium, the force generated by the net flux of water moving down its concentration gradient is balanced by forces acting in the other direction.

The water concentration gradient across the plasma membrane of most organisms leads to an influx of water into the cell. As water enters, the plasma membrane expands (you might want to think about how that occurs, in terms of membrane structure). If the influx of water continued unopposed, the membrane would eventually burst like an over-inflated balloon, killing the cell. One strategy to avoid this lethal outcome, adopted by a range of organisms, is to build a semi-rigid "cell wall" exterior to the plasma membrane. The synthesis of this cell wall is based on the controlled assembly of macromolecules secreted by the cell through the process of exocytosis (see above). As water passes through the plasma membrane and into the cell (driven by osmosis), the plasma membrane is pressed up against the cell wall. The force exerted by the rigid cell wall on the membrane balances the force of water entering the cell. When the two forces are equal, the net influx of water into the cell stops. Conversely, if the $[H_2O]_{outside}$ decreases, this pressure is reduced, the membrane moves away from the cell wall and (because they are only semi-rigid) the walls flex. It is this behavior that causes plants to wilt when they do not get enough water. These are passive behaviors, based on the structure of the cell wall. They are essentially built into the wall as it is first assembled. Once the cell wall has been built, a cell with a cell wall does not need to expend energy to resist osmotic effects. Plants, fungi, bacteria and archaea all have cell walls. A number of antibiotics work by disrupting the assembly of bacterial cell walls. This leaves the bacteria osmotically sensitive, water enters these cells until they burst and die.



Questions to answer & to ponder:

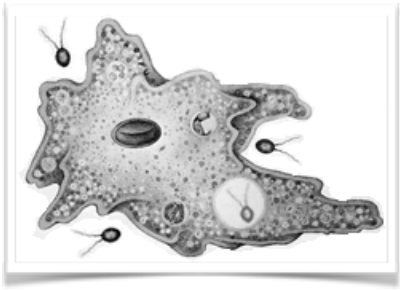
- Using the U-tube applet (in beSocratic), how would you get water to move from the right side to the left side of the membrane? How could such a system be used to purify water? (not currently active)
- Where does the energy involved in moving molecules come from?
- Plants and animals are both eukaryotes; how would you decide whether the common ancestor of the eukaryotes had a cell wall
- Why does an aquaporin channel not allow a Na+ ion to pass through it?
- If there is no net flux of A, even if there is a concentration gradient between two points, what can we conclude?

An evolutionary scenario for the origin of eukaryotic cells

When we think about how life arose, and what the first organisms looked like, we are moving into an area where data is fragmentary and speculation is often rampant. These are also, dare we remind you, events that took place billions of years ago. But these obstacles do not mean we cannot draw interesting conclusions – there is relevant data present in each organisms' genetic data (its

[160] http://youtu.be/ePGqRaQiBfc

genotype) and the structure of its cells and their ecological interactions that can be used as a basis for our speculations.



Animal cells do not have a rigid cell wall. This allows them to be active predators, moving rapidly and engulfing their prey whole or in macroscopic bits through phagocytosis (see above). They use complex "cytoskeletal" and "cytomuscular" systems to drive these thermodynamically unfavorable behaviors (again, largely beyond our scope here). Organisms with a rigid cell wall can't do that. Given that bacteria and archaea have cell walls, it is possible that cell walls were present in the common ancestral organism. But this leads us to think more analytically about the nature of the earliest organisms and the path to the common ancestor. A cell wall is a complex structure that would have had to be built through evolutionary processes before it would be useful. If we assume that the original organisms arose in an osmotically friendly (that is, non-challenging environment), then a cell wall could have been generated in steps, and once adequate it could enable the organisms that possessed it to invade new, more osmotically challenging (dilute) environments - like most environments today.

For example, one plausible scenario is that the ancestors of the bacteria and archaea developed cell walls originally as a form of protection against predation. So who were the predators. Where they the progenitors of the eukaryotes? If so, we would come to assume that the organisms in the eukaryotic lineage never had a cell wall, rather than that they once shared a cell wall with bacteria and archaea. In this scenario, the development of eukaryotic cell walls by fungi and plants represents an example of convergent evolution and these structures are analogous (rather than homologous) to the cell walls of prokaryotes.

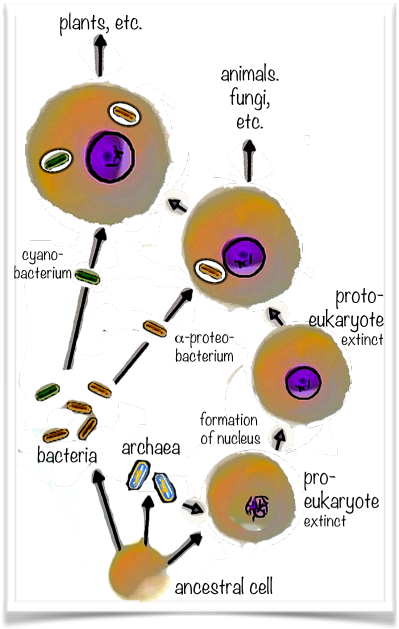
But now a new complexity arises, there are plenty of eukaryotic organisms, including microbes like the amoeba, that live in osmotically challenging environments. How do they deal with the movement of water into their cells? They actively pump the water that flows in back out again using an organelle known as the contractile vacuole. Water accumulates within the contractile vacuole, a membrane-bounded structure within the cell, which inflates. To expel the water, the vacuole connects with the plasma membrane and is squeezed by cytomuscular systems within the cytoplasm. This squirts the water out of the cell. The process of vacuole contraction is an active one, it involves work and requires energy. One might speculate that this cytomuscular system was originally involved in predation, that is, enabling the cell to surround and engulf other organisms (phagocytosis). The resulting vacuole became specialized to aid in killing and digesting the engulfed prey. When digestion is complete, it can fuse with the plasma membrane to discharge the waste, using either a passive or an active "contractile system". It turns out that the molecular systems involved in driving active membrane movement are related to the systems involved in dividing the eukaryotic cell into two; distinctly different systems are used in the division of prokaryotes.[161] So a question is which came first, different cell division mechanisms, which led to differences in the membrane behavior of cells, one leading to a predatory active membrane and the other that led to a passive membrane, perhaps favoring the formation of a cell wall?

---

[161] The cell cycle of archael: http://www.ncbi.nlm.nih.gov/pubmed/23893102 and Bacterial cell division: http://www.ncbi.nlm.nih.gov/pubmed/17098054

Making a complete eukaryote

Up to this point we have only touched on a few of the ways that prokaryotes (bacteria and archaea) differ from eukaryote. The majors ones are the fact that eukaryotes have their genetic material isolated from the cytoplasm by a complex double-layered membrane/pore system known as the nuclear envelope (which we will discuss in some detail later on) and the location of chemo-osmotic and photosynthetic systems between the two types of organisms. In prokaryotes, these systems (light absorbing systems, electron transport chains and ATP synthases) are found either within the plasma membrane or within internal membranes clearly derived from the plasma membrane. In contrast, in eukaryotes (plants, animals, fungi, protozoa, and other forms) these structural components are not located on the plasma membrane, but rather within discrete intracellular structures. In the case of the system associated with aerobic respiration, these systems are located in the inner membranes of a double-membrane bound cytoplasmic organelles known as **mitochondria**. Photosynthetic eukaryotes (algae and plants) have a second type of cytoplasmic organelle (in addition to mitochondria), known as **chloroplasts**. Like mitochondria, chloroplasts are also characterized by the presence of a double membrane and an electron transport chain associated with the inner membrane and membranes apparently derived from it. These are just the type of structures one might expect to see if a bacterial cell were engulfed by the ancestral pro-eukaryotic cell (→), with the host cell's membrane surrounding the engulfed cells plasma membrane. A closer molecular analysis reveals that the mitochondrial and chloroplast electron transport systems as well as the ATP synthase proteins more closely resemble those found in one type of bacteria, rather than archaea. In fact, detailed analysis of the genes and proteins involved suggest that the electron transport/ATP synthesis systems of eukaryotic mitochondria are homologous to those of α-proteobacteria while the light harvesting/reaction center complexes, electron transport chains and ATP synthesis proteins of photosynthetic eukaryotes (algae and plants) appear to be homologous to those of a second type of bacteria, the photosynthetic cyanobacteria.[162] In contrast, many of the nuclear systems appear more similar to systems found in archaea. How do we make sense of these observations?



Clearly when a eukaryotic cell divides it must also have replicated its mitochondria and chloroplasts, otherwise they would eventually be lost. In 1883, Andreas Schimper (1856-1901) noticed that chloroplasts divided independently of their host cells. Building on Schimper's observation, Konstantin Merezhkovsky (1855-1921) proposed that chloroplasts were originally independent organisms and that plant cells were chimeras, really two independent organisms living together. In a similar vein, in 1925 Ivan Wallin (1883-1969) proposed that the mitochondria of eukaryotic cells were derived from bacteria. This "endosymbiotic hypothesis" for the origins of eukaryotic mitochondria and chloroplasts fell out of favor, in large part because the molecular methods needed to unambiguously

[162] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC138944/

resolve there implications were not available. A breakthrough came with the work of Lynn Margulis (1938-2011) and was further bolstered when it was found that both the mitochondrial and chloroplast protein synthesis machineries were sensitive to drugs that inhibited bacterial but not eukaryotic protein synthesis and by the discovery that mitochondria and chloroplasts contained DNA molecules that were organized like, and contained genes similar to genes found in bacteria (we will consider DNA and its organization soon).

All eukaryotes appear to have mitochondria. Suggestions that some eukaryotes, such as the human anaerobic parasites *Giardia intestinalis*, *Trichomonas vaginalis* and *Entamoeba histolytica* [163] do not failed to recognize cytoplasmic organelles known as mitosomes as degenerate mitochondria. Based on these and other data it is now likely that all eukaryotes are derived from an ancestor that engulfed an aerobic α-proteobacteria-like bacterium. Instead of being killed and digested, these (or even one) of these bacteria survived within the eukaryotic cell, replicated, and were distributed into the progeny cell when the parent cell divided. This process resulted in the engulfed bacterium becoming an endosymbiont, which over time became mitochondria. At the same time the engulfing cell became dependent upon the presence of the endosymbiont to initially detoxify molecular oxygen, and then to utilize molecular oxygen to break down molecules, and so maximize the energy that could be derived from their metabolism. All eukaryotes (including us) are descended from a mitochondria-containing eukaryote. This event is thought to have occurred around 2 billion years ago. The next step in eukaryotic evolution involved a second endosymbiotic event in which a cyanobacteria-like bacterium formed an endosymbiotic relationship with a mitochondria-containing eukaryote. This lineage gave rise to the glaucophytes, the red and the green algae. The green algae, in turn, gave rise to the plants.

As we look through modern organisms there are a number of examples of similar events, that is, one organism becoming inextricably linked to another through endosymbiotic processes. There are also examples of close couplings between organisms that are more akin to parasitism rather then mutually beneficial symbiosis.[164] For example, a number of insects have intracellular bacterial parasites, and some pathogens and parasites live inside human cells.[165] In some cases, even these parasites can have parasites. Consider the mealybug *Planococcus citri*; this organism contains cells known as bacteriocytes. Within these cells are *Tremblaya princeps* type β-proteobacteria. Surprisingly, within these bacterial cells, which lie within the eukaryotic mealybug cells, live *Moranella endobia*-type γ-proteobacteria.[166] In another example, after the initial endosymbiotic event that formed the proto-algal cell, the ancestor of red and green algae and the plants, there have been endocytic events in which a eukaryotic cell has engulfed and formed an endosymbiosis with a eukaryotic green algal cell, to form a "secondary" endosymbiont. Similarly, secondary endosymbionts have been engulfed by yet another

---

[163] The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica: http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2958.1999.01414.x/full

[164] Mechanisms of cellular invasion by intracellular parasites: http://www.ncbi.nlm.nih.gov/pubmed/24221133

[165] Intracellular protozoan parasites of humans: the role of molecular chaperones in development and pathogenesis. http://www.ncbi.nlm.nih.gov/pubmed/20955165

[166] Mealybugs nested endosymbiosis: going into the 'matryoshka' system in Planococcus citri in depth. http://www.ncbi.nlm.nih.gov/pubmed/23548081

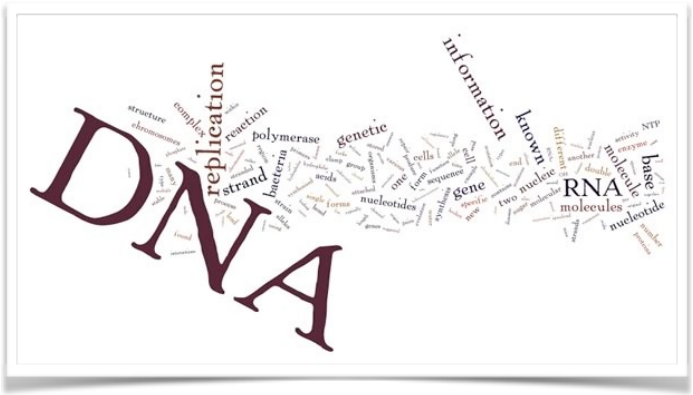
eukaryote, to form a tertiary endosymbiont.[167] The conclusion is that there are combinations of cells that can survive better, in a particular ecological niche, than either could alone. In these phenomena we see the power of evolutionary processes to populate extremely obscure ecological niches in rather surprising ways.

Questions to answer & to ponder:

- Are the mitochondria of plants and animals homologous or analogous?
- Did the earliest eukaryote have a cell wall? why or why not? Where did this organism live?
- What advantage would the host cell get from the early proto-mitochondrial or proto-chloroplastic symbionts?
- Was there an advantage for the engulfed bacteria? If so, what could it be?
- Define the difference between a symbiotic and a parasitic relationship?
- Why does the number of membranes around an eukaryotic organelle matter? Where do these membranes come from?
- What evidence would lead you to suggest that there were multiple symbiotic events that gave rise to the mitochondria of different eukaryotes?
- Why might a plant cell not notice the loss of its mitochondria?

[167] Photosynthetic eukaryotes unite: endosymbiosis connects the dots: http://dblab.rutgers.edu/home/downloads/Files/Bhattacharya%20et%20al%20BioEssays%202004.pdf

7. The molecular nature of heredity



In which we discover how the physical basis of inheritance, DNA, was discovered, and learn about the factors that influence its structure, how it encodes genetic information, how that information is replicated and read, how mutations occur and are often repaired, and how such an extravagantly long molecule is organized in such small cells.

One of the most amazing facts associated with Darwin and Wallace's original evolutionary hypothesis was their complete lack of a coherent understanding of genetic mechanisms. While it was very clear, based on the experiences of plant and animal breeders, that organisms varied and that part of that variation was inherited, the mechanism by which genetic information was stored and transmitted was not clear and at the time could not have been known. Nevertheless there were a number of hypotheses, some of which relied on supernatural or metaphysical mechanisms.[168] For example, some thought that evolutionary variation was generated by a type of inner drive or logic within the organism. This had the comforting implication that evolutionary processes reflected some kind of over-arching design, that things were going somewhere. Well before the modern theory of evolution was proposed in 1859, Jean-Baptiste Lamarck's (1744 – 1829) proposed that inheritance somehow reflected the desires and behaviors of the parent. This would have predicted a type of "directed" evolution. In contrast, Darwin's model, based on completely random variations seemed more arbitrary and unsettling. It implied a lack of an over-arching purpose to life in general, and human existence in particular.

Another surprising realization is that modern genetics had its origins beginning with the work of Gregor Mendel (1822 – 1884). He published his work on sexually reproducing peas in 1865, shortly after the introduction of the modern theory of evolution. Since Darwin published revised editions of "On the Origin of Species" through 1872, one might ask why did he not incorporate a Mendelian view of heredity? The simplest explanation would be that Darwin was unaware of Mendel's work - in fact, Mendel's work was essentially ignored until the early years of the 20th century. One might ask why was the significance of Mendel's observations not immediately recognized? It turns out that Mendel's conclusions were actually quite specialized and could be attributed to the design details of his experiments and his choice of organism. Mendel carefully selected discrete traits (phenotypes) displayed by the garden pea *Pisum sativum*: smooth versus wrinkled seeds, yellow versus green seeds, grey versus white seed coat, tall versus short plants, etc. In the plants he used, he found no intermediate phenotypes of these traits. In addition, these traits were independent, the presence of one trait did not influence any of the other traits he was considering. Each was controlled (as we now know) by a single genetic locus (position or gene). However, the vast majority of traits do not behave in this way. Most genes play a role in a number of different traits and a particular trait is generally controlled (and influenced) by many genes. Allelic versions of genes interact in complex and non-additive ways. For example, the extent to which a trait is visible, even assuming the underlying genetic factor is

[168] http://en.wikipedia.org/wiki/The_eclipse_of_Darwinism

present, can vary dramatically depending upon the rest of an organism's genotype. Finally, in an attempt to established the general validity of his conclusions, after working with peas, which reproduce sexually, Mendel examined the behavior of a number of other plants, including hawkweed. Unfortunately, hawkweed uses a specialized, asexual reproductive strategy, known as apomixis, in which Mendel's laws are not followed.[169] This did not help reassure Mendel or others that his genetic laws were universal.

Subsequent work, however, led to the recognition of the general validity of Mendel's basic conclusions (there are organisms that display exceptions, but we will ignore these for now.) Mendel deduced that there are stable hereditary "factors" - which became known as genes - and that these are present as discrete objects within organisms. Each gene can exist in a number of different forms, known as alleles. In many cases specific alleles (a specific version of a gene) are associated with specific forms of a trait, or the presence or absence of a trait. For example, whether you are lactose tolerant as an adult is influenced by which allele of the MCM6 gene you carry. The allele that promotes lactose tolerance acts to maintain the expression of the gene that encodes the enzyme lactase, which is necessary to digest lactose.[170] When a cell divides, its genes must be reproduced so that each daughter cell receives a full set of genes (a genome). The exact set of alleles it inherits determines its genotype (note, words like genomes and genotypes, are modern terms, that reflect underlying Mendelian ideas). Later it was recognized that sets of genes were linked together in some way, but that this linkage was not permanent - that is, processes existed that could shuffle linked genes (or rather the alleles of genes).

In sexually reproducing organisms, like the peas that Mendel originally worked with, two copies of each gene were present in each somatic (body) cell. Such cells are said to be diploid. During sexual reproduction, cells are produced that contain only a single copy of each gene, they are referred to as haploid (although monoploid would be a better term). Two such haploid cells (typically known as egg and sperm in animals and ovule and pollen in plants), derived from different parents, fuse to form a new diploid organism. An important feature of this model is that the alleles inherited from the two parents are shuffled through various mechanisms (and to various extents) when the new organism is formed, so that offspring are genetically distinct from their parents. This makes sense from a conceptual standpoint, it creates increasing levels of genetic and phenotypic variation. It leaves unanswered the question, what is the molecular mechanism by which these inherited traits are transmitted from generation to generation? How is it that offspring are in some sense very similar to their parents (that is, they are the same species), but yet are also different and distinguishable? The answer lies in the way this information is encoded, stored, and transmitted at the molecular level - and to understand that we have to move to the atomic molecular scale.

Discovering how nucleic acids store genetic information

To follow the historical pathway that led to our understanding of how heredity works, we have to start back at the cell. As it became more firmly established that all organisms were composed of cells,

---

[169] Apomixis in hawkweed: Mendel's experimental nemesis: http://www.ncbi.nlm.nih.gov/pubmed/21335438

[170] http://www.hhmi.org/biointeractive/making-fittest-got-lactase-co-evolution-genes-and-culture

and all cells were derived from pre-existing cells, it became more and more likely that inheritance had to be a cellular phenomena. As part of their studies, cytologists (students of the cell) began to catalog the common components of cells. One such component was the nucleus. At this point it is worth remembering that most cells do not contain pigments. Under a microscope, they appear clear, after all they are ~70% water. To be able to discern structural details cytologists had to stabilize the cell and to visualize its various components. As you might suspect, stabilizing the cell means killing it. To be observable, the cell had to be killed (known technically as "fixed") in such a way as to insure that its structure was preserved as close to the living state as possible. Originally, this process involved the use of chemicals, such as formaldehyde, that could cross-link various molecules together, which stopped them from moving with respect to one another. Alternatively, the cell could be treated with organic solvents such as alcohols; this leads to the precipitation of the water soluble components. As long as the methods used to visualize the fixed tissue were of low magnification and resolution, the results were generally acceptable. In more modern studies, using various optical methods[171] and electron microscopes, such crude fixation methods are unacceptable, and have been replaced by various alternatives, including rapid freezing. Even so it was hard to resolve the different subcomponents of the cell. To do this the fixed cells were treated with various dyes. Some dyes bind preferentially to molecules located within particular parts of the cell. The most dramatic of these cellular sub-sections was the nucleus, which could be readily identified because it was stained very differently from the surrounding cytoplasm. One standard stain involves a mixture of hematoxylin (actually oxidized hematoxylin and aluminum ion) and eosin, which leaves the cytoplasm pink and the nucleus dark blue.[172] The nucleus was first described by Robert Brown (1773-1858)(the person after which Brownian motion was named). The presence of a nucleus was characteristic of eukaryotic (true nucleus) organisms.[173] Prokaryotic cells (before a nucleus) are typically much smaller and originally it was impossible to determine whether they had a nucleus or not (they do not).

The careful examination of fixed and living cells revealed that the nucleus underwent a dramatic reorganization as a cell divided, losing its typical roughly spherical shape; it was replaced by discrete stained strands, known as chromosomes (or colored bodies). In 1887 Edouard van Beneden reported that the number of chromosomes was constant for each species and that different species had different numbers of chromosomes. Within a particular species the chromosomes have distinctive sizes and shapes. For example, in the fruit fly *Drosophila melanogaster* there are four chromosomes each with a distinctive length and shape. That means that chromosomes could be followed as cellular transformations occurred. In 1902, Walter Sutton published his observation that chromosomes obey Mendel's rules of inheritance, that is that during the formation of the cells that fuse during sexual reproduction (gametes: sperm and

species	chromosome #
Ophioglossum reticulatum (a fern)	1260 (630 pairs)
Canis familiaris (dog)	78 (39 pairs)
Cavia cobaya (guinea pig)	60 (30 pairs)
Solanum tuberosom (potato)	48 (24 pairs)
Homo sapiens (humans)	46 (23 pairs)
Macaca mulatta (monkey)	42 (21 pairs)
Mus musculus (mouse)	40 (20 pairs)
Felis domesticus (house cat)	38 (19 pairs)
Saccharamyces cervisae (yeast)	32 (16 pairs)
Drosophila melanogaster (fruit fly)	8 (4 pairs)
Myrmecia pilosula (ant)	2 (1 pair)

[171] **Optical microscopy beyond the diffraction limit:** http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2645564/

[172] The long history of hematoxylin: http://www.ncbi.nlm.nih.gov/pubmed/16195172

[173] There are some eukaryotic cells, like human red blood cells, that do not have a nucleus, they are unable to divide.

egg), each cell received one and only one copy of each chromosome. This strongly suggested that Mendel's genetic factors were associated with chromosomes.[174] Of course by this time, it was recognized that there were many more Mendelian factors than chromosomes, which means that many factors must be present on each chromosome. These observations provided a physical explanation for the fact that many traits did not behave independently but acted as if they were linked together. The behavior of the nucleus, and the chromosomes that appeared to exist within it, mimicked the type of behavior that a genetic material would be expected to display.

These cellular anatomy studies were followed by studies on the composition of the nucleus. As with many scientific studies, progress is often made when one has the right "model system" to work with. It turns out that some of the best systems for the isolation and analysis of the components of the nucleus were sperm and pus (isolated from discarded bandages from infected wounds (yuck)). It was therefore assumed, quite reasonably, that components enriched in this material would likely be enriched in nuclear components. Using sperm and pus as a starting material Friedrich Miescher (1844 – 1895) was the first to isolate a phosphorus-rich compound, called nuclein.[175] At the time of its original isolation there was no evidence linking nuclein to genetic inheritance. Later nuclein was resolved into an acidic component, deoxyribonucleic acid (DNA), and a basic component, primarily proteins known as histones. Because they have different properties (acidic DNA, basic histones), chemical "stains" that bind or react with specific types of molecules and absorb visible light, could be used to visualize the location of these molecules within cells using a light microscope. The nucleus stained for both highly acidic and basic components - which suggested that both nucleic acids and histones were localized to the nucleus, although what they were doing there was unclear.

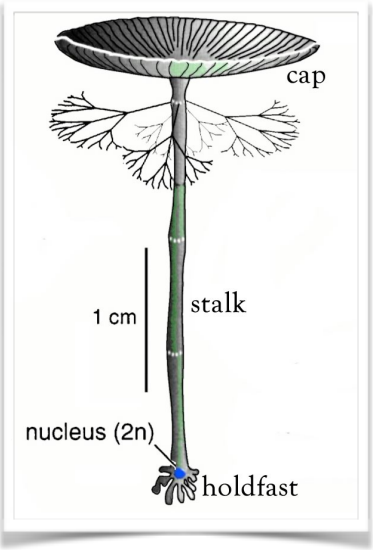
Locating hereditary material within the cell

Further evidence suggesting that hereditary information was probably localized in the nucleus emerged from transplantation experiments carried out by Joachim Hammerling in the 1930's using the giant unicellular green alga *Acetabularia,* known as the mermaid's wineglass. Hammerling's experiments (video: http://youtu.be/tl5KkUnH6y0) illustrate two important themes in the biological sciences. The idiosyncrasies of specific organisms can be exploited to carry out useful studies that are simply impossible to perform elsewhere. At the same time, the underlying evolutionary homology of organisms makes it possible to draw broadly relevant conclusions from such studies. In this case, Hammerling exploited three unique features of *Acetabularia*. The first is the fact that each individual is a single cell, with a single nucleus. It is therefore possible to isolate nuclear and anucleate (not containing a nucleus) regions of the organism. Second, these cells are very large (1 to 10 cm in height), which makes it possible to carry out various microsurgical operations on them. You can remove and transplant regions of one organism (cell) to another. Finally, different species of *Acetabularia* have distinctively different "caps" that regrow faithfully following amputation. In his experiments, he removed the head and stalk regions from one individual, leaving a region that was much smaller but, importantly, it contained the nucleus. He then transplanted large regions of anuclear stalk derived from an organism

[174] http://www.nature.com/scitable/topicpage/developing-the-chromosome-theory-164

[175] Friedrich Miescher and the discovery of DNA: http://www.sciencedirect.com/science/article/pii/S0012160604008231

of another species, with a distinctively different cap morphology, onto the nucleus-containing holdfast region. When the cap regrew it had the morphology characteristic of the species that provided the nucleus - no matter that this region was much smaller than the transplanted (anucleate) stalk region. The conclusion was that the information needed to determine the cap morphology in *Acetabularia* was located within the region of the cell that contained the nucleus, rather than dispersed throughout the cytoplasm. Its just a short step from these experimental results to the conjecture that all genetic information is located within the nucleus.

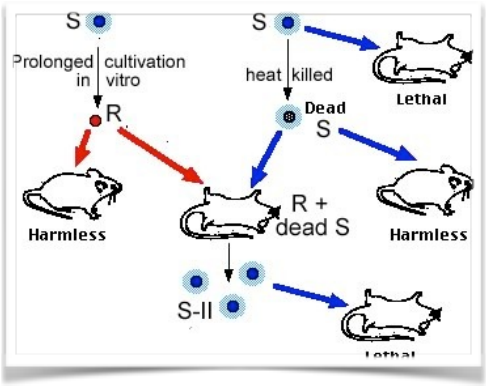


.

Identifying DNA as the genetic material

The exact location, and the molecular level mechanism of the storage and transmission of the genetic information were still to be determined. Two kinds of experiment led to the realization that genetic information was stored in a chemically stable form. In one set of studies, H.J. Muller (1890 – 1967) was able to show that exposing fruit flies to X-rays (a highly energetic form of light) generated mutations that could be inherited from generation to generation. This suggested that genetic information was stored in a chemical form that could be altered through interactions with radiation, and that once altered it was again stable. The second experimental evidence supporting the idea that genetic information was encoded in a stable chemical form came from a series of experiments initiated in the 1920s by Fred Griffith (1879–1941). He was studying two strains of the bacterium *Streptococcus pneumoniae.* This type of bacteria causes bacterial pneumonia and, when introduced, killed mice.

He grew these bacteria in the laboratory. This is known as culturing the bacteria; often we say the bacteria grown in culture have been grown *in vitro* or in glass as opposed to *in vivo* or within a living animal. Following common methods, he grew bacteria on plates covered with solidified agar (a jello-like substance derived from sea water alga) containing various nutrients. Typically, a liquid culture of bacteria is diluted and spread on these plates. Individual bacteria bind to the plate independently of, and separated from, one another. Bacteria are asexual and so each individual bacterium can grow up into a colony, a clone of the original bacteria that landed on the plate. The disease-causing strain of bacteria grew up into smooth or S-type colonies, due to the fact that the bacteria secrete a slimy mucus-like substance. He found that mice injected with S strain the mice quickly sickened and died. However, if he killed the bacteria with heat before injection, the mice did not get sick, indicating that it was the living bacteria that produced (or evoked) the disease symptoms, not some chemical toxin.

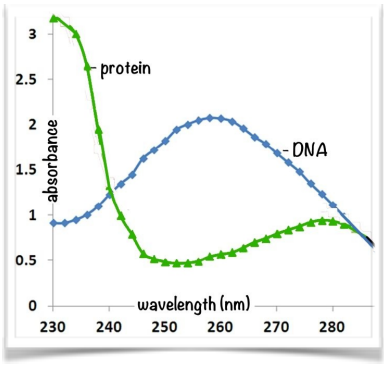


During extended cultivation *in vitro*, however, cultures of S strain bacteria sometimes gave rise to rough (R) colonies. These were not smooth and shiny, but rather rough in appearance. This was a genetic change because once isolated, R-type strains continued to produce R-type colonies, a process that could be repeated many, many times. More

importantly, mice injected with R strain bacteria did not get sick. BUT, weirdly enough, mice co-injected with the living R (which did not cause the disease) and dead S (which did not cause the disease) bacteria did get sick and died! Griffith was able to isolate and culture bacteria from these dying mice, he found that when grown *in vitro* they produced smooth colonies - he termed such strains S-II smooth strains. His hypothesis was that a stable chemical (that is, non-living) component derived from the dead S bacteria had "transformed" the avirulent (benign) R strain to produce a new virulent S-II strain.[176] Unfortunately Fred Griffith died in 1941 during the bombing of London, which put an end to his studies.

In 1944, Griffith's studies were continued and extended by Oswald Avery, Colin McLeod and Maclyn McCarty. They set out to use Griffith's assay to isolate what they termed the “transforming principle” responsible for turning R into S strains. Their approach was to make cell extracts. They ground up cells and isolated various components, such as proteins, nucleic acids, carbohydrates, and lipids. They then digested these extracts with various enzymes and asked whether the transforming principle was still intact.

Treating cellular extracts with proteases (which degrade proteins), lipases (which degrade lipids), or RNAases (which degrade RNAs) had no effect on transformation. In contrast, treatment of the extracts with DNAases, which degrade DNA, destroyed the activity. Further support for the idea that the “transforming substance” was DNA was suggested by the fact that it had the physical properties of DNA, for example it absorbed light like DNA rather than protein. Subsequent studies confirmed this conclusion. Furthermore DNA isolated from R strain bacteria did not produce S-strain bacteria, whereas DNA from S strain bacteria could transform S strains into R strains. They concluded that DNA derived from S cells contains the information required for the conversion -- it is, or rather contains, a gene required for the S strain phenotype. This information had been lost by mutation during the formation of R strains. The phenomena exploited by Griffiths and Avery et al., known as transformation, is an example of horizontal gene transfer, which we will discuss in greater detail later on. It is the movement of genetic information from one organism to another (as opposed to vertical gene transfer, which is the process by which the progeny of an organism inherit their DNA, their genetic material, from their parent(s). In fact variants of horizontal gene transfer occur commonly within the microbial world and allow genetic information to move between species. For example horizontal gene transfer is responsible for the rapid expansion of populations of antibiotic resistant bacteria. Viruses use a highly specialized (and optimized) form of horizontal gene transfer.[177] The question is, why is this even possible? While we might readily accept that genetic information must be transferred from parent to offspring (we can see the evidence for this process with our eyes), the idea that genetic information can be transferred between different organisms that are not (apparently) related is quite a bit more difficult to swallow. As we will see, horizontal transfer is possible primarily because all organisms share the same system for reading and replicating genetic information. The hereditary machinery is homologous.



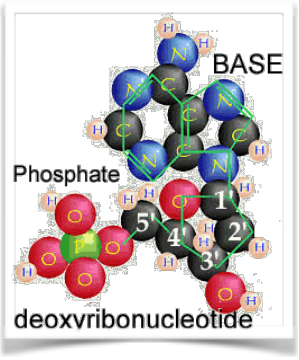
[176] http://en.wikipedia.org/wiki/Griffith's_experiment

[177] Virus-like particles speed bacterial evolution: http://www.nature.com/news/2010/100930/full/news.2010.507.html

Questions to answer & to ponder

- Is there a correlation between the number of chromosomes and the complexity of an organism? What might the complexity of an organism be related to?
- What is meant by complexity of an organism?
- What caused the change from S to R strains in culture?
- In Griffith's study, he found that dead smooth *S. pneumoniae* could transform living rough strains of *S. pneumoniae* when co-injected into a mouse. Would another species of dead bacteria give the same result? Explain your reasoning.
- How would Hammerling's observations have been different if hereditary information was localized in the cytoplasm?
- How might horizontal gene transfer confuse molecular phylogenies (family trees)?
- Where did the original genes come from?

Unraveling Nucleic Acid Structure



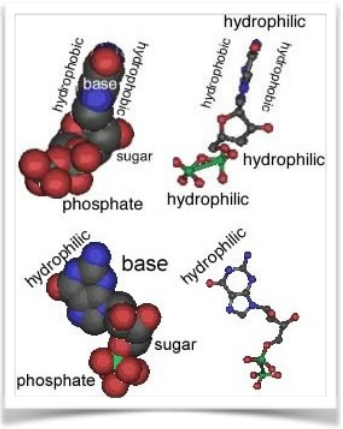
Knowing that the genetic material was DNA was a tremendous break through, but it left a mystery - how was genetic information stored and replicated. Nucleic acids were thought to be aperiodic polymers, that is molecules built from a defined set of subunits (also known as monomers), but without a simple overall repeating pattern. The basic monomeric units of nucleic acids are known as nucleotides. A nucleotide consists of three distinct types of molecules joined together, a 5-carbon sugar (ribose or deoxyribose), a nitrogen-rich "base" that is either a purine (guanine (G) or adenine (A)) or a pyrimidine (cytosine (C), or thymine (T)) in DNA or uracil (U) instead of T in RNA, and a phosphate group. The carbon atoms of the sugar are numbered 1' to 5'. The nitrogenous base is attached to the 1' carbon and the phosphate is attached to the 5' carbon. The other important group attached to the sugar is a hydroxyl group attached to the 3' carbon. RNA differs from DNA in that there is hydroxyl group attached to the 2' carbon of the ribose in RNA, but this hydroxyl is absent in DNA, which is why it is "deoxy" ribonucleic acid! We take particular note of the 5' phosphate and 3' hydroxyl groups because they are directly involved in the polymerization of nucleotides to form nucleic acids.

Discovering the structure of DNA

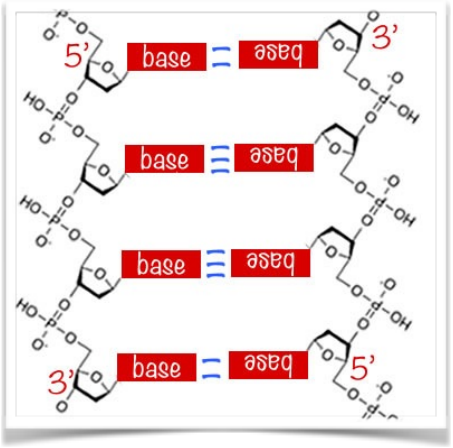
A critical clue to understanding the structure of nucleic acids came from the work of Erwin Chargaff (1905 – 2002). When analyzing DNA from various sources, he found that the relative amounts of G, C, T and A varied between organisms but were the same (or very similar) for organisms of the same type or species. On the other hand, the ratios of A to T and G to C were always equal to 1, no matter where the DNA came from. Knowing these rules, James Watson and Francis Crick (1916 –2004) built a model of DNA that fit what was known about the structure of nucleotides and structural data from Rosalind Franklin (1920 – 1958).[178] Franklin got these data by pulling DNA into oriented strands, fibers of many molecules aligned parallel to one another. By passing X-rays through these fibers she was able to obtain a diffraction pattern. This pattern is based on the structure of DNA molecules, and

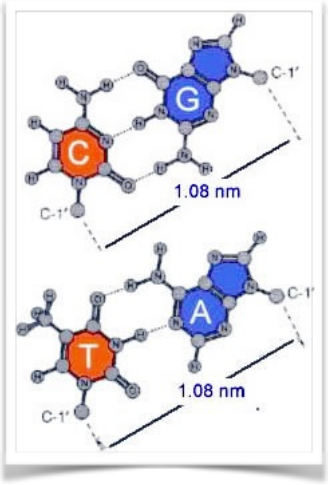
[178] An interesting depiction of this process is provided by the movie "Life Story" http://en.wikipedia.org/wiki/Life_Story_(TV_film)

defines key parameters that constrain any model of the molecule's structure. But making a model of the molecule that would produce the observed X-ray data was not simple.



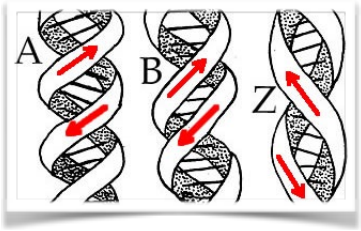
To understand this process, let us consider the chemical nature of a nucleotide and nucleotide polymer like DNA. First the nucleotide bases (bases A, G, C and T) have a number of similar properties. Each nucleotide has three hydrophilic regions: the negatively charged phosphate group, a sugar which has a lot of O–H groups, and the hydrophilic edge of the base (where the N–H and N groups lie). While the phosphate and sugar are three-dimensional moieties, the bases are flat, the atoms in the rings are all in one plane. The upper and lower surfaces of the rings are hydrophobic (non-polar) while the edges have groups that can interact via hydrogen bonds. This means that the amphipathic factors that favor the assembly of lipids into bilayer membranes are also at play in nucleic acid structure. To reduce their interactions with water, in their model Watson and Crick had the bases stacked on top of one another, hydrophobic surface next to hydrophobic surface. This left each base's hydrophilic edge, with -C=O and -N-H groups that can act as H-bond acceptors and donors, to be dealt with. How were these hydrophilic groups to be arranged? Their great insight, which led to a direct explanation of why Chargaff's rules were universal, was to recognize that pairs of nucleotide bases, in two DNA strands could be arranged in an anti-parallel and complementary orientation. So what does that mean? Each DNA polymer strand has a directionality to it, it runs from the 5' phosphate group at one end to the 3' hydroxyl group at the other, each nucleotide monomer is connected to the next through a phosphodiester linkage. When the two strands were arranged in opposite orientations, that is, anti-parallel to one another: one from 5' → 3' and the other 3' ← 5', the bases attached to the sugar-phosphate backbone could interact with one another in highly specific ways. An A would form two hydrogen bonding interactions with a T on the opposite (anti-parallel) strand, while a G would form three hydrogen bonding interactions with a C. A key feature of this arrangement was that the lengths of the A::T and G:::C base pairs are almost identical. The hydrophobic surfaces of the bases were stacked on top of each other, while the hydrophilic sugar and phosphate groups were in contact with the surrounding water. The possible repulsion between negatively charged phosphate groups was neutralized (or shielded) by the presence of positively charged sodium ions present in the solution from which the X-ray measurements were made.





In their final model, Watson and Crick depicted what is now known as B-form DNA. Under different salt conditions, DNA can form two other double helical forms, known as the A and Z forms. A and B forms of DNA are "right-handed" helices, the Z-form of DNA is a left-handed helix. In cells, DNA is

usually in the B form, although it can assume other forms locally (and as we will see, it can open up - the two strands can separate from one another) under some conditions.



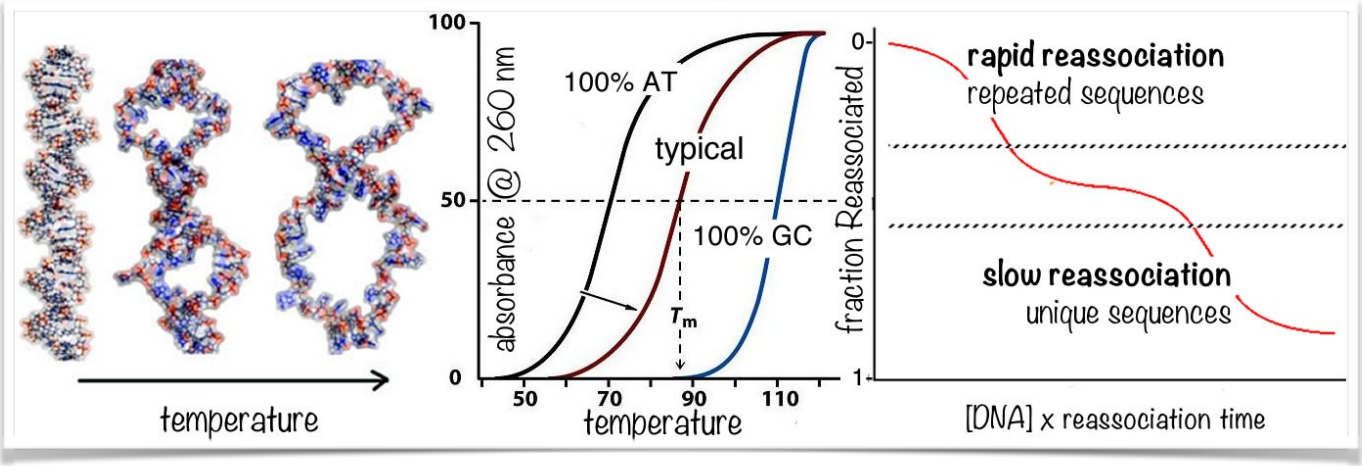
As soon as the structure of DNA was proposed its explanatory power was obvious. Because the A:T and G:C base pairs are of the same length, the sequence of bases along the length of a DNA molecule (written in the 5' to 3' direction) has little effect on the overall three-dimensional structure of the molecule. That implies that essentially any possible sequence could be found, at least theoretically, in a DNA molecule. If information were encoded in the sequence of nucleotides along a DNA molecule, any information could be placed there and that information would be as stable as the DNA molecule itself. This is similar to the storage of information in various modern computer memory devices, that is, any type of information can be stored, because storage does not involve any dramatic change in the basic structure of the storage material. The structure of a flash drive is not altered by whether in contains photos of your friends or a song or a video or a textbook. At the same time, the double-stranded nature of the structure and complementary nature of base pairing (A to T and G to C) immediately suggested a simple model for DNA (and information) replication - that is, pull the two strands of the molecule apart and build new (anti-parallel) strands using the original two strands as templates. The two strands of the parental molecule are held together only by hydrogen bonding interactions, so no chemical reaction is needed to separate them, no covalent bond needs to be broken. In fact, at physiological temperatures DNA molecules are often opening up over short stretches and then closing, a process known as DNA breathing.[179] This makes the replication of the information stored in the molecule conceptually straightforward (even though the actual biochemical process is complex.) The existing strands determine the sequence of nucleotides on the newly synthesized strands. The newly synthesized strand can, in turn, direct the synthesis of a second strand, identical to the original strand. Finally, the double stranded nature of the DNA molecule means that the information is stored in a redundant fashion. If one strand is damaged, that is its DNA sequence is lost or altered, the second undamaged strand can be used to repair that damage. A number of mutations in DNA are repaired using this type of mechanism (see below).

DNA, sequences, and information

We can now assume that somehow the sequence of nucleotides in the DNA molecule encodes information but the question remains what kind(s) of information is stored in DNA? Early students of DNA could not read DNA sequences, as we can now, so they relied on various measurements to better understand the behavior of the molecule. For example, the way a double stranded DNA molecule interacts with light is different from the way that of a single stranded DNA molecule does. Since the two strands of double stranded DNA molecules (often written dsDNA) are attached only by hydrogen bonding interactions, increasing the temperature of the system can lead to their separation into two single stranded molecules (ssDNA)(left panel figure below). ssDNA absorbs light at 260nm (in the ultraviolet) more strongly than does dsDNA, so the absorbance of a DNA solution can be used to determine the relative amounts of single and double stranded DNA in a sample at a particular temperature. What we find is that the temperature at which 50% of dsDNA molecules have separated

[179] Dynamic approach to DNA breathing: http://www.ncbi.nlm.nih.gov/pubmed/23345902

into ssDNA varies between organisms. This is not particularly surprising given Chargaff's observation that the ratio of AT to GC varied between various organisms and the fact that GC base pairs, mediated by three H-bonds, are predicted to be more stable than AT base pairs, which are held together by only two H-bonds. In fact, one can estimate the AT:GC ratio based on melting curves (middle panel).

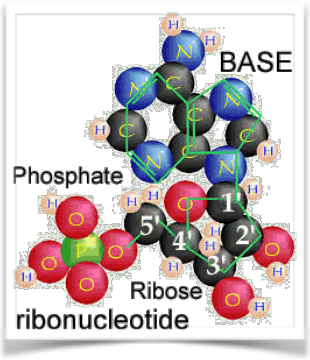


It quickly became clear that things were more complex than previously expected. Here a technical point needs to be introduced. Because of the extreme length of the DNA molecules found in biological systems, it is almost impossible to isolate them intact. In the course of their purification, the molecules will be sheared into shorter pieces, typically thousands of base pairs in length compared to the millions to hundreds of millions of base pairs in intact molecules. In another type of experiment, one could look at how fast ssDNA (the result of a melting experiment) would reform dsDNA. The speed of these "reannealing reactions" is dependent on DNA concentration. When such experiments were carried out, it was found that there was a fast annealing population of DNA fragments and various slower annealing populations (right panel above). How to explain this result, was it a function of AT:GC ratio? Subsequent analysis revealed that it was due to the fact that within the DNA isolated from organisms, particularly eukaryotes, there were many (hundreds to thousands) of regions (fragments) that contained similar nucleotide sequences. Because the single strands of these fragments can associated with one another, these sequences occurred in much higher effective concentrations compared to regions of the DNA with unique sequences. This type of analysis revealed that much of the genome of eukaryotes was composed of various families of repeated sequences and that unique sequences amounted to less than 5% of the total DNA. While a complete discussion of these repeated sequence elements is beyond our scope here, we can make a few points. As we will see, there are repair mechanisms that can move regions of a DNA molecule from one position to another within the genome. The end result is that the genome (the DNA molecules) of a cell/organism are dynamic, a fact with profound evolutionary implications.

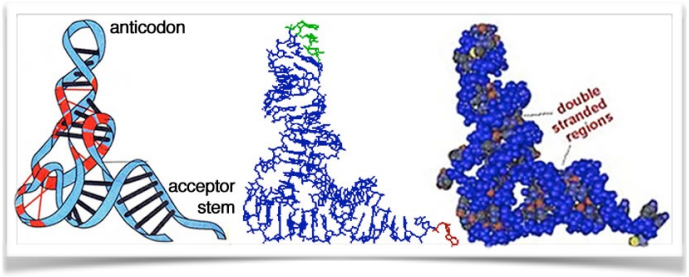
Questions to answer & to ponder

- Which do you think is stronger (and why), an AT or a GC base pair?
- Why does the ratio of A to G differ between organisms?
- Why is the ratio of A to T the same in all organisms?
- What does it mean that the two strands of a DNA molecule are anti-parallel?
- Normally DNA exists inside of cells at physiological salt concentration (~140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$ and some minor ions). Predict what will happen (what is thermodynamically favorable) if you place DNA into distilled water (that is, no dissolved salts.)

Discovering RNA: structure and some functions



DNA is not the only nucleic acid found in cells. A second class of nucleic acid is known as ribonucleic acid (RNA.) RNA differs from DNA in that RNA contains i) the sugar ribose (with a hydroxyl group on the 2' C) rather than deoxyribose; ii) it contains the pyrimidine uracil instead of the pyrimidine thymine found in DNA; and iii) RNA is typically single rather than double stranded. Nevertheless, RNA molecules can associate with an ssDNA molecule with the complementary nucleotide sequence. Instead of the A-T pairing in DNA we find A pairing with U instead. This change does not make any difference when the RNA strand interacts with DNA since the number of hydrogen bonding interactions are the same. When RNA was isolated from cells, one population was found to reassociate with unique sequences within the DNA. As we will see later, this class of RNA, includes molecules, known as messenger or mRNAs, that carry information from DNA to the molecular machinery that mediates the synthesis of proteins. In addition to mRNAs there are other types of RNAs in cells. These include structural, catalytic, and regulatory RNAs. As you might have already suspected, the same hydrophobic/hydrophilic/H-bond considerations that were relevant to DNA structure apply to RNA, but because RNA is generally single stranded, the structures found in RNA are somewhat different. A single-stranded RNA molecule can fold back on itself to create double stranded regions. Just as in DNA, these folded strands are anti-parallel to one another. This results in double-stranded "stems" that end in single-stranded "loops". Regions within a stem that do not base pair will bulge out. The end result is that RNA molecules can adopt complex three-dimensional structures in solution. Such RNAs often form complexes with other molecules, particularly proteins, to carry out specific functions. For example, the ribosome, the macromolecular machine involved in the synthesis of proteins, is a complex of structural and catalytic RNAs (known as ribosomal or rRNAs) and proteins. Transfer RNAs (tRNAs) are integral components of the protein synthesis system. RNAs, in combination with proteins, also play a number of regulatory functions including recognizing and regulating the behaviors of mRNAs, subjects typically considered in greater detail in courses in molecular biology.



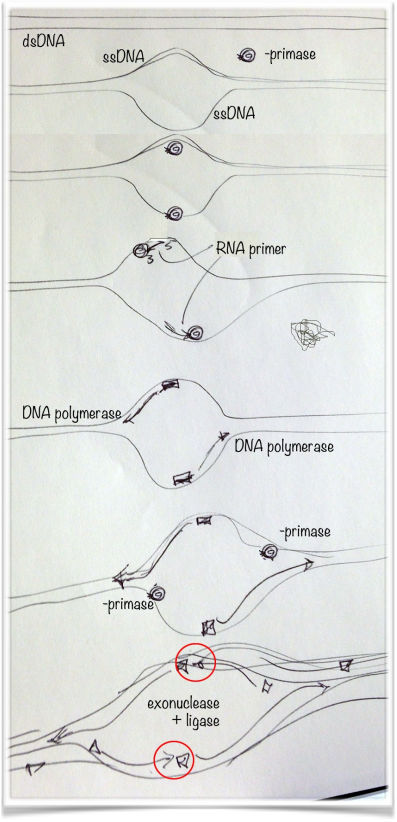
The ability of RNA to both encode information in its base sequence and to mediate catalysis through its three dimensional structure has led to the "RNA world" hypothesis. It proposes that early in the evolution of life various proto-organisms relied on RNAs, or more likely simpler RNA-like molecules, rather than DNA and proteins, to store genetic information and to catalyze reactions. Some modern day viruses use single or double stranded RNAs as their genetic material. According to the RNA world hypothesis, it was only later in the history of life that organisms developed the more specialized DNA-based systems for genetic information storage and proteins for catalysis and other structural functions. While this idea is compelling, there is no reason to believe that simple polypeptides and other molecules were not also present and playing a critical role in the early stages of life's origins. At the

same time, there are many unsolved issues associated with a simplistic RNA world view, the most important being the complexity of RNA itself, its abiogenic (that is, without life) synthesis, and the survival of nucleotide triphosphates in solution. Nevertheless, it is clear that catalytic and regulatory RNAs play a key role in modern cells and their throughout their evolution. The catalytic activity of the ubiquitous ribosome, which is involved in protein synthesis, is based on a ribozyme, a RNA-based catalyst.

DNA replication

Once it was proposed, the double-helical structure of DNA immediately suggested a simple mechanism for the accurate duplication of the genetic information stored in DNA. Each strand contains all of the information necessary to specify the sequence of its complementary strand. The process begins when a dsDNA molecule opens to produce two single-stranded regions. Where DNA is free, that is, not associated with other molecules (proteins), this can occur easily. Normally, the single strands simply rebind to one another. To replicate DNA the open region has to be stabilized and the catalytic machinery organized. We will consider how this is done only in general terms, in practice this is a complex and highly regulated process involving a number of components.

The first two problems we have to address may seem arbitrary, but they turn out to be common features of DNA synthesis. The enzymes that catalyze the synthesis of a new DNA strand (DNA polymerases) cannot start synthesis on their own. In contrast, the catalysts that synthesize RNA do not require a pre-existing strand, they can start the synthesis of new RNA strand de novo, although they do require an existing nucleic acid strand to determine the order in which nucleotides are added. Both DNA and RNA synthesis require a pre-existing 3' end of a nucleic acid molecule. The polymerases involved in both RNA and DNA synthesis can add nucleotides only to the 3' OH group of an existing nucleic acid strand. Later on we will consider how nucleic acid synthesis, which includes DNA replication and RNA synthesis are regulated, but for now let us assume that some process has determined where replication starts. We begin our discussion with DNA replication.



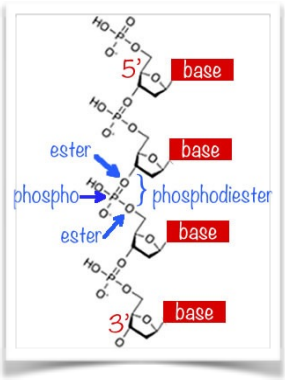
The first step is to locally open up the dsDNA molecule. An enzyme that synthesizes a short RNA molecule, known as the primer (the enzyme is known as primase), must collide with and engage the DNA. Because the two strands of the DNA molecule point in opposite directions, one primase complex must associate with each strand. These synthesize a short RNA molecule. Once these are in place, the appropriate nucleotide, determined by its match with the nucleotide present at that position of the existing DNA strand, needs to be added to the 3' end of the RNA primer.

Nucleotides exist in various phosphorylated forms within the cell, including nucleotide monophosphate (NMP), nucleotide diphosphate (NDP), and nucleotide triphosphate (NTP). To make the nucleic acid

polymerization reaction thermodynamically favorable, the reaction uses the NTP form of the nucleotide monomers, in the reaction:

(**5'P**)NTP(3'OH) + (5'P)NTP(**3'OH**) + H_20 ⟷ (**5'P**)NTP-NMP(**3'OH**) + diphosphate.

This NTP hydrolysis driven polymerization reaction leads to the loss of the added nucleotide's 5' phosphate while a phosphodiester bond [-C-O-P-O-C] is formed, and a new 3' OH end, which can react with another NTP is generated. In theory, this process can continue until the newly synthesized strand reaches the end of the DNA molecule. For the process to continue, however, the double stranded region of the original DNA will have to open up, exposing more single stranded DNA. Keep in mind that this process is moving in both directions along the DNA molecule. Because the polymerization reaction only proceeds by 3' addition, as new single stranded regions are opened new primers must be created (by primase) and then extended (by DNA polymerase). If you try drawing what this looks like, you will realize that i) this process is asymmetric in relation to the start site of replication; ii) the process generates RNA-DNA hybrid molecules, and RNA regions are not found in "mature" DNA molecules; and iii) that eventually an extending DNA polymerase will run into the RNA primer part of an "upstream" molecule. For a dynamic look check out this video[180] which is nice, but very "flat" to reduce the complexity of the process. These issues are resolved by the fact that the DNA polymerase complex contains more than one catalytic activity. When it reaches the upstream nucleic acid chain it uses an RNA exonuclease activity to remove the RNA nucleotides. It then replaces them with DNA nucleotides using the existing DNA strand as the primer. Once the RNA portion is removed, a DNA ligase activity acts to join the two DNA molecules. These reactions, driven by nucleotide hydrolysis, end up producing a continuous DNA strand.



Evolutionary considerations: At this point you might well ask yourself, why (for heavens sake) is the process so complicated. Why not use a DNA polymerase that does not need an RNA primer, or any primer for that matter, since RNA polymerase does not need a primer? Why not have polymerases that add nucleotide equally well to either end of a polymer? That such a mechanism is possible is suggested by the presence of enzymes in eukaryotic cells that can carry out the 5' capping reaction associated with mRNA synthesis, briefly considered later on, but such activities are not used in DNA replication. The real answer is that we are not sure of the reasons. These could be evolutionary relics, a process established within the last common ancestor and extremely difficult or impossible to change through evolutionary mechanisms. Alternatively, there could be strong selective advantages associated with the system that preclude such changes. What is clear is that this is how the system is set up in all known organisms, so for practical purposes, we have to remember the particular details involved.

Replication machines

We have presented DNA replication (the same, apparently homologous process is used in all known organisms) in as conceptually simple terms as we can, but it is important to keep in mind that

[180] http://www.biostudio.com/d_%20DNA%20Replication%20Coordination%20Leading%20Lagging%20Strand%20Synthesis.htm

the actual machinery involved is complex. In part the complexity arises because the process is topologically constrained and needs to be highly accurate. In the bacterium *Escherichia coli* over 100 genes are involved in DNA replication and repair. To insure that replication is controlled and complete, replication begins at specific sequences along the DNA strand, known as origins of replication or origins for short. Origin DNA sequences are recognized by specific DNA binding proteins. The binding of these proteins initiates the assembly of an origin recognition complex, an ORC.

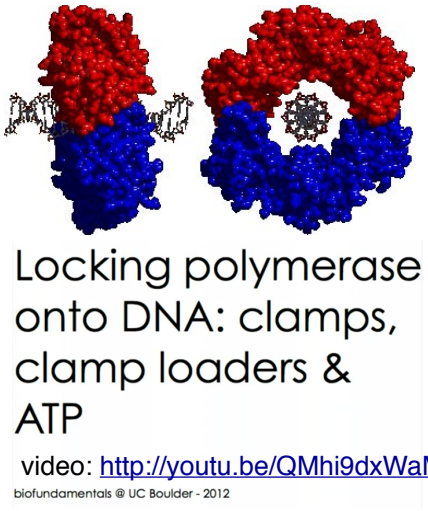
In the laboratory, increasing temperature is used to separate dsDNA into single strands that can be replicated. In the cell, various proteins act on the DNA to locally denature (unwind) and block the single strands from reannealing. This leads to the formation of a replication bubble. A multiprotein complex then assembles at each end of the replication bubble, these structures are known as replication forks. Using a single replication origin and two replication forks moving in opposite directions, a rapidly growing *E. coli* can replicate its ~4,700,000 base pairs of DNA (which are present in a single circular DNA molecule) in ~40 minutes. Each replication fork moves along the DNA adding ~1000 base pairs of DNA per second to the newly formed DNA polymer.

Synthesis (replication) is a highly accurate process; the polymerase makes about one error for every 10,000 bases it adds. But that level of error would almost certainly be highly deleterious, and in fact most of these errors are quickly recognized. To understand how, remember that correct AT and GC base pairs have the same molecular dimensions, that means that incorrect AG, CT, AC, and GT base pairs are either too long or too short. By responding to base pair length, molecular machines can recognize a base pairing mistake as a structural defect in the DNA molecule. When a mismatched base pair is formed, the DNA polymerase reverses and removes it using an “DNA exonuclease” activity. It then resynthesizes it, (hopefully) correctly. This process is known as proof-reading; the proof-reading activity of the DNA polymerase complex reduces the total DNA synthesis error rate to ~1 error per 1,000,000,000 (10^9) base pairs synthesized.

At this point let us consider nomenclature, which can seem arcane and impossible to understand, but which in fact obeys reasonably clear rules. An exonuclease is an enzyme that can bind to the free end of a nucleic acid polymer and remove nucleotides through a hydrolysis reaction of the phosphodiester bond. A 5' exonuclease cuts the nucleotide off the 5' end of the molecule, a 3' exonuclease, off the 3' end. A circular nucleic acid molecule is immune to the effects of an exonuclease. To break the bond between two nucleotides in the interior of a nucleic acid molecule (or in a circular molecule, which has no ends), one needs an endonuclease activity.

As you think about the processes involved, you come to realize that once DNA synthesis begins, it is important that it continues uninterrupted. But the interactions between nucleic acid chains are based on weak H-bonding interactions, and the enzymes involved in the process can be expected to dissociate from the DNA because of the effects of thermal motion, imagine the whole system jiggling and vibrating - held together by relatively weak interactions. We can characterize how well a DNA polymerase remains productively associated with a DNA molecule in terms of the number of nucleotides it adds to a new molecule before it falls off; this is known as its “processivity”. So if you think of the DNA replication complex as a molecular machine, you can design ways to insure that the replication complex has high processivity, basically by keeping it bound to the DNA. One set of such machines is the polymerase sliding clamp and clamp loader (see video below). The DNA polymerase complex is held onto the DNA by a doughnut shaped protein, known as a sliding clamp. This protein

encircles the DNA double helix and is strongly bound to the DNA polymerase. So the question is, how does a protein come to encircle a DNA molecule? The answer is that the clamp protein is added to DNA by another protein molecular machine known as the clamp loader.[181] Once closed around the DNA the clamp can move freely along the length of the DNA molecule, but it cannot leave the DNA. The clamp's sliding movement along DNA is diffusive – that is, driven by thermal motion. Its movement is given a direction because the clamp is attached to the DNA polymerase complex which is adding monomers to the growing nucleic acid polymer. This moves the replication complex (inhibited from diffusing away from the DNA by the clamp) along the DNA in the direction of synthesis. Processivity is increased since, in order to leave the DNA the polymerase has to disengage from the clamp or the clamp as to be removed by the clamp loader acting in reverse.



video: http://youtu.be/QMhi9dxWaM8

Further replication complexities

There are important differences between DNA replication in prokaryotes and eukaryotes. The DNA molecules found in eukaryotic nuclei are double-stranded, linear molecules, with free ends, a fact that leads to problems replicating the ends of the molecule, known as its telomeres (see below). In contrast the DNA molecules found in bacteria and archaea are circular; there are no free ends.[182] This creates a topological complexity. After replication, the two circles are linked together. Long linear DNA molecules can also become knotted together within the cell. In addition, the replication of DNA unwinds the DNA, and this unwinding leads to supercoiling of the DNA molecule. Left unresolved, supercoiling and knotting would inhibit DNA synthesis and the separation of replicated strands. These topological issues are resolved by enzymes known as topoisomerases. There are two types. Type I topoisomerases bind to the DNA, catalyze the breaking of a single bond in one sugar-phosphate-sugar backbone, and allow the release of overwinding through rotation around the bonds in the intact chain. When the tension is released, and the molecule has returned to its "relaxed" form, the enzyme catalyzes the reformation of the broken bond. Both bond breaking and reformation are coupled to ATP hydrolysis.

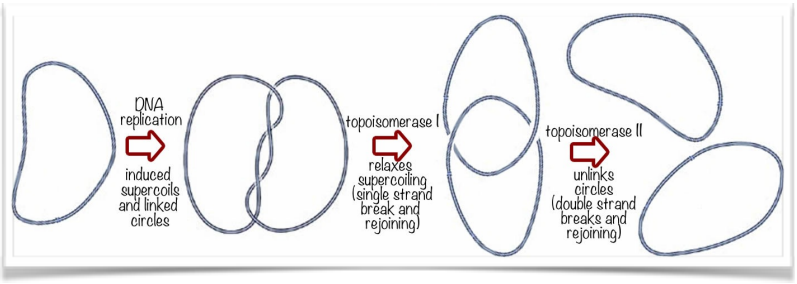
see http://youtu.be/EYGrElVyHnU

Type II topoisomerases are involved in "unknotting" DNA molecules. These enzymes bind to the DNA, catalyze the hydrolysis of both backbone chains, but hold on to the now free ends. This allows

[181] see http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3331839/?tool=pubmed and http://youtu.be/QMhi9dxWaM8

[182] The mitochondria and chloroplasts of eukaryotic cells also contain circular DNA molecules, another homology with their ancestral bacterial parents. ,

another strand to “pass through” the broken strand. The enzyme also catalyzes the reverse reaction, reforming the bonds originally broken.



Eukaryotic cells can contain more than 1000 times the DNA found in a typical bacterial cell. Instead of circles, they contain multiple linear molecules that form the structural basis of their chromosomes. Their linearity creates problems when it comes to replicating their ends. This is solved by a catalytic system composed of proteins and RNA known as telomerase which we will not discuss further here.[183] The eukaryotic DNA replication enzyme complex is slower (about 1/20th as fast) as prokaryotic systems. While a bacterial cell can replicate its circular ~3 x 10^6 base pair chromosome in about 1500 seconds using a single origin of replication, the replication of the billions of base pairs of eukaryotic DNAs involves the use of multiple origins of replication, scattered along the length of each chromosome. Another required function is a specific molecular machine that acts when replication forks "crash" into one another. In the case of circular DNA molecules, with their single origins of replication, the replication forks resolve in a specific region known as the terminator. At this point type II topoisomerase allows the two circular DNA molecules to disengage from one another, and move to opposite ends of the cell. The cell division machinery forms between the two DNA molecules. The system in eukaryotes is much more complex, with multiple linear chromosomes and involves a more complex molecular machine, which we will return to, although only superficially, later.

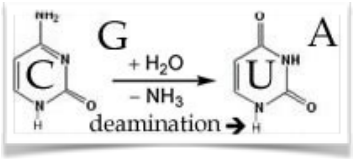
Questions to answer & to ponder:

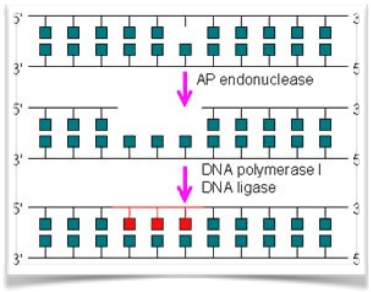
- On average, during DNA/RNA synthesis, what is the ratio of productive to unproductive interactions between nucleotides and the polymerase?
- Where would variation come from if DNA were totally stable and DNA replication was error-free?
- Draw a diagram to explain how the DNA polymerase recognizes a mismatched base pair.
- Why do you need to denature (melt) the DNA double-helix to copy it?
- What would happen if H-bonds were “real” covalent bonds?
- How does the DNA polymerase complex know where to start replicating DNA?
- Make a cartoon of a prokaryotic chromosome, indicate where replication starts and stops. Now make a cartoon of eukaryotic chromosomes.
- List all of the unrealistic components in the replication video
- Is an RNA primer needed to make an mRNA?
- Why is only a single RNA primer needed to synthesize the leading strands, but multiple primers are needed to synthesize the lagging strands?
- During the replication of a single circular DNA molecule, how many leading and lagging strands are there? What is the situation in a linear DNA molecule?
- Assume that there is a mutation that alters the proof-reading function of the DNA polymerase complex - what will happen to the cell?
- Explain how the absence of the clamp would influence DNA replication?
- How do you think the clamp is removed?

[183] http://en.wikipedia.org/wiki/Telomerase

Mutations, deletions, duplications & repair

While DNA is used as the genetic material, it is worth remembering that it is a thermodynamically unstable molecule. Eventually it will decompose into simpler (more stable) components. For example, at a temperature of ~13ºC, half of the phosphodiester bonds in a DNA sample would break after ~500 years. But there is more. For example, cytosine can react with water, which is present at a concentration of ~54 M inside a cell. This leads to a deamination reaction that transforms cytosine into uracil. If left unrepaired, the original CG base pair would be replaced by an AU base pair. But, uracil is not normally found in DNA, so its presence can be easily recognized by an enzyme that severs the bond between the uracil moiety and the deoxyribose group.[184] The absence of a base, due either to spontaneous loss or enzymatic removal, acts as a signal for another enzyme system (the Base Excision Repair complex) that removes a section of the DNA strand with the missing base.[185] DNA polymerase binds to the open DNA and uses the undamaged strand as a template to fill in the gap. Finally, another enzyme (a DNA ligase) joins the newly synthesized segment to the pre-existing strand. In the human genome there are over 130 genes devoted to repairing damaged DNA.[186] [video with lots of misspelled words:http://youtu.be/g4khROaOO6c].





Another type of hydrolysis reaction involves the removal of a base from the DNA. These are known as depurination - the loss of an cytosine or thymine group and depyrimidination - the loss of an adenine or guanine group. The reaction rate is increased at acidic pH, which is probably one reason that the cytoplasm is not acidic. How frequent are such events? A human body contains ~10^{14} cells. Each cell contains about ~10^{9} base pairs of DNA. Each cell (whether it is dividing or not) undergoes ~10,000 base loss events per day or ~10^{18} events per day per person. That's a lot! The basic instability of DNA (and the lack of repair after an organism dies) means that DNA from dinosaurs (the last of which went extinct about 65,000,000 years ago) has disappeared from the earth. This makes it impossible to actually clone (or resurrect) a true dinosaur.[187] In addition, mistakes are also made during DNA synthesis and DNA can be damaged by environmental factors, such as radiation, ingested chemicals, and reactive compounds made by the cell itself. Many of the most potent known mutagens are natural products, often produced by organisms to defend themselves against being eaten or infected by parasites, predators, or pathogens.

Genes and alleles

Up to now we have been considering genes as abstract entities and mentioning, only in passing, what they actually are. We think about genes encoding traits, but this is perhaps the most incorrect possible

[184] uracil-DNA-N-glycosidase

[185] absent purine/absent pyrimidine endonuclease http://omim.org/entry/300773

[186] Human DNA Repair Genes: http://www.sciencemag.org/content/291/5507/1284.full

[187] DNA has a 521-year half-life: http://www.nature.com/news/dna-has-a-521-year-half-life-1.11555

view of what they are and what they do. A gene is a region of DNA. That region can encode a gene product. The gene also includes the sequences required for its proper expression or activity. While we have not consider it in any significant detail, it is worth noting that genes can be quite complex. There can be multiple regulatory regions controlling the same coding sequence and particularly in eukaryotes a single gene can produce multiple, functionally distinct gene products.[188] How differences in gene sequence influence the role of a gene is often not simple. One critical point to keep in mind is that a gene has meaning only in the context of an organism. Change the organism and the same, or rather, more accurately put, homologous genes (that is gene that share a common ancestor, a point we will return to) can have different roles.

Once we understand that a gene corresponds to a specific sequence of DNA, we understand that alleles of a gene correspond to different sequences. Two alleles of the same gene can differ from one another by as little as a single nucleotide position. The most common version of an allele is often referred to as the wild type allele, but that is really just because it is the most common. There can be multiple "normal" alleles of a particular gene within any one population. Genes can overlap with one another, particularly in terms of their regulatory regions, and defining all of the regulatory regions of a gene can be difficult. A gene's regulatory regions may span many kilobases of DNA and be located upstream, downstream, or within the coding region. In addition, because DNA is double stranded, one gene can be located on one strand and another, completely different gene can be located on the anti-parallel strand. We will return to the basic mechanisms of gene regulation later one, but as you probably have discerned, gene regulation is complex and typically the subject of its own course.

Alleles: Different alleles of the same gene can produce quite similar gene products or their products can be different. The functional characterization of an allele is typically carried out with respect to how its presence influences a specific trait(s). Again, remember that most traits are influenced by multiple genes, and a single gene can influence multiple traits and processes. An allele can produce a gene product with completely normal function or absolutely no remaining functional activity, referred to as a null or amorphic allele. It can have less function than the "wild type" allele (hypomorphic), more function than the wild type (hypermorphic), or a new function (neomorphic). Given that many gene products function as part of multimeric complexes and that many organisms (like us) are diploid, there is one more possibility, the product of one allele can antagonize the activity of the other - this is known as an antimorphic allele. These different types of alleles were defined genetically by Herbert Muller, who won the Nobel prize for showing that X-rays could induce mutations, that is, new alleles.

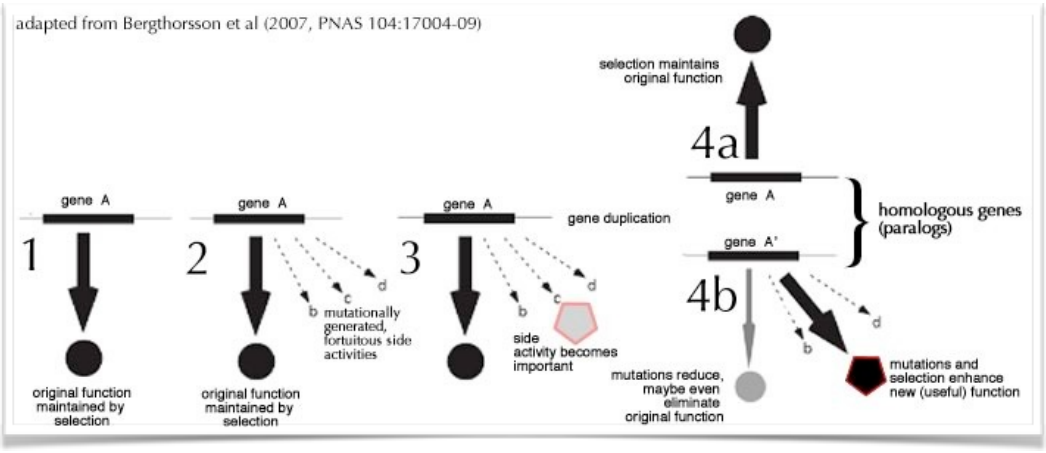
Mutations and evolution

That said, there are often multiple common alleles in the population, and they all may be equally normal in terms of the phenotypes they produce. If there is no significant selective advantage between them, their relative frequencies within a population will drift. Often the history of populations is tracked by the alleles present within it, since this can reflect events such as bottlenecks associated with migrations. At the same time, they may produce different phenotypes in the presence of specific alleles at other genetic loci. Since most traits are the results of hundreds or thousands of genes functioning together,

[188] Expansion of the eukaryotic proteome by alternative splicing: http://www.nature.com/nature/journal/v463/n7280/full/nature08909.html

and different combinations of alleles can produce different effects, the universe of variation is large. This can make identifying the genetic basis of a disease difficult, particularly when variation at a specific locus can have only a minor contribution to the disease phenotype. On top of that, environmental and developmental differences can outweigh genetic influence on phenotype.

Mutations are the ultimate source of genetic variation – without them evolution would be impossible. Mutations can lead to a number of effects; they can create new activities. At the same time these changes may reduce the original activity of an important gene. Left unresolved such molecular level conflicts would greatly limit the flexibility of evolutionary mechanisms. For example, it is common to think of a gene (or rather the particular gene product it encodes) as having one and only one function or activity, but in fact, when examined closely many catalytic gene products (typically proteins) can catalyze "off-target" reactions or carry out, even if rather inefficiently, other activities - they interact with other molecules within the cell and the organism. Assume for the moment that a gene encodes a gene product with an essential function as well as potentially useful (from a reproductive success perspective) activities. Mutations that enhance these "ancillary functions" will survive (that is be passed on to subsequent generations) only to the extent that they do not negatively influence the gene's primary and essential function. The evolution of ancillary functions may be severely constrained or blocked altogether.



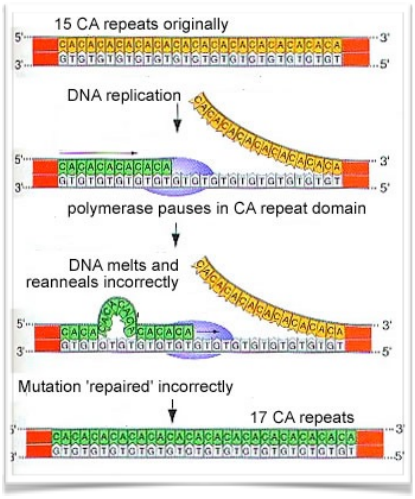
This problem is circumvented to a significant extent by the fact that the genome, that is, DNA molecules, is not static. There are processes through which regions of DNA (and the genes that they contain) can be deleted, duplicated, and moved from place to place within the genome. Such genomic rearrangements occur continuously. Such events even occur during embryonic development. This means that while most of the cells in your body have very similar genomes (perhaps containing some single base pair changes that arose during DNA replication), some have genomes with different arrangements of DNA. Not all cells in your body have exactly the same genome.[189]

In the case above, imagine that the essential gene is duplicated. Now one copy can continue to carry out its essential function, while the second is free to change. While most mutations will inactivate the duplicated gene, some might increase and refine its favorable ancillary function. A new trait can emerge freed from the need to continue to perform an essential function. We see evidence of this type of process around the biological world. When a gene is duplicated, the two copies are known as paralogs. Such paralogs can evolve independently.

---

[189] Copy Number Variation in Human Health, Disease, and Evolution: http://www.annualreviews.org/doi/abs/10.1146/annurev.genom.9.081307.164217 and LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? http://www.academia.edu/328644/LINE-1_retrotransposons_mediators_of_somatic_variation_in_neuronal_genomes

Triplet repeat diseases and genetic anticipation



While they are essential for evolution, defects in DNA synthesis and genomic rearrangements more frequently lead to a genetic (that is inherited) disease than any benefit to an individual. You can explore the known genetic diseases by using the web based On-line Mendelian Inheritance in Man (OMIM) database.[190] To specifically illustrate diseases associated with DNA replication, we will consider a class of genetic diseases known as the trinucleotide repeat disorders. There are a number of such "triplet repeat" diseases, including several forms of mental retardation, Huntington's disease, inherited ataxias, and muscular dystrophys. These diseases are caused by slippage of DNA polymerase and the subsequent duplication of sequences. When these "slippable" repeats occur in a region of DNA encoding a protein, it can lead to regions of a repeated amino acid. For example, expansion of a domain of CAGs in the gene encoding the polypeptide Huntingtin causes the neurological disorder Huntingdon's chorea.

Fragile X: This DNA replication defect is the leading form of autism of known cause. Sadly, there are many forms of autism in which the cause is not known. Only ~6% of all autistic individuals have fragile X. Fragile X can also lead to anxiety disorders, attention deficit hyperactivity disorder, psychosis, and obsessive-compulsive disorder. Because the mutation involves the FMR-1 gene, which is located on the X chromosome, the disease is sex-linked and effects mainly males (who are XY, compared to XX females).[191] In the unaffected population, the FMR-1 gene contains between 6 to 50 copies of a CGG repeat. Individuals with 6 to 50 repeats are phenotypically normal. Those with 50 to 200 repeats carry what is known as a premutation; these individuals rarely display symptoms but can transmit the disease to their children. Those with more than 200 repeats typically display symptoms and often have what appears to be a broken X chromosome – from which the disease derives its name. The pathogenic sequence in Fragile X is downstream of the FMR1 gene's coding region. When this region expands, it inhibits the gene's activity.

Defects in DNA repair can lead to severe diseases and often a susceptibility to cancer. A OMIM search for DNA repair returns 654 entries! For example, defects in mismatch repair lead to a susceptibility to colon cancer, while defects in translation-coupled DNA repair are associated with Cockayne syndrome. People with Cockayne's syndrome are sensitive to light, short and appear to age prematurely.[192]

Summary: Our introduction to genes has necessarily been quite foundational. There are lots of variations and associated complexities that occur within the biological world. The key ideas are that genes represent biologically meaningful DNA sequences. To be meaningful, the sequence must play a

---

[190] http://www.ncbi.nlm.nih.gov/omim/

[191] You will probably want to learn how to use the On-line Mendelian Inheritance in Man (OMIM) to explore various disease and their genetic components. OMIM is a part of PubMed: http://www.ncbi.nlm.nih.gov/pubmed

[192] Cockayne syndrome: http://omim.org/entry/278760

role within the organism, typically by encoding a gene product (which we will consider next) and the information needed to insure its correct “expression”, that is, where and when the information in the gene is accessed. A practical problem is that most studies of genes are carried out using organisms grown in the lab or in otherwise artificial or unnatural conditions. It might be possible for an organism to exist with an amorphic mutation in a gene in the lab, but organisms that carry that allele may well be at a significant reproductive disadvantage in the real world (what ever that is). Moreover, a particular set of alleles, a particular genotype, might have a reproductive advantage in one environment (one ecological/behavioral niche) but not another. Measuring these effects can be quite difficult. All of which should serve as a warning to consider skeptically pronouncements that a gene, or more accurately a specific allele of a gene, is responsible for a certain trait, particularly if the trait is complex, ill-defined, and likely to be significantly influenced by genomic context (the rest of the genotype) and environmental factors.

Questions to answer & to ponder:

- What happens in cells with defects in DNA repair systems when they attempt to divide?
- I thought RNA primers were used to make DNA! So why is there no uracil in a DNA molecule?
- A base is lost, how is this loss recognized by repair systems?
- How could a DNA duplication lead to the production of a totally new gene (rather than just two copies of a preexisting gene)?
- How does a mutation generate a new allele? And what exactly is the difference between a gene and an allele?
- What would be a reasonable way to determine that you had defined an entire gene?
- Given that DNA is unstable, why hasn't evolution used a different type of molecule to store genetic information?
- Is it possible to build a system (through evolutionary mechanisms) in which mutations do not occur?
- Would such an "error-free" memory system be evolutionarily successful?

8. Peptide bonds, polypeptides and proteins

In which we consider the nature of proteins, how they are synthesized, how they are folded and assembled, how they get to where they need to go, how they function, how their activities are regulated, and how mutations can influence their behavior.

We have mentioned proteins many times, since there are few biological processes that do not rely on them. Proteins act as structural elements, signals, regulators, and catalysts in a wide arrange of molecular machines. Up to this point, however, we have not said much about what they are, how they are made, and how they do what they do. The first scientific characterization of what are now known as proteins was published in 1838 by the Dutch chemist, Gerardus Johannes Mulder (1802–1880).[193] After an analysis of a number of different substances, he proposed that they all represented versions of a common chemical core, with the molecular formula $C_{400}H_{620}N_{100}O_{120}P_1S_1$, and that the differences between them were primarily in the numbers of phosphate (P) and sulfur (S) atoms they contained. The name "protein", from the Greek word πρώτα ("prota"), meaning "primary", was suggested by the Swede, Jons Jakob Berzelius (1779–1848) based on the presumed importance of these compounds in biological systems.[194] As you can see, Mulder's molecular formula is not very informative, it tells us little or nothing about protein structure, but suggested that all proteins are fundamentally similar, which is confusing since they carry out so many different roles. Subsequent studies revealed that protein could be dissolved in either water or dilute salt solutions but aggregated and became insoluble when the solution was heated; as we will see this aggregation reflects a change in the structure of the protein. Mulder was able to break down proteins through an acid hydrolysis reaction into amino acids, named because they contained amino ($-NH_2$) and carboxylic acid (-COOH) groups. Twenty different amino acids could be identified in hydrolyzed samples of proteins. Since their original characterization as a general class of compounds, we now understand that while they share a common basic structure, proteins are remarkably diverse. They are involved in roles from the mechanical strengthening of skin to the regulation of genes, to the transport of oxygen, to the capture of energy, to the catalysis and regulation of essentially all of the chemical reactions that occur within cells and organisms.

Polypeptide and protein structure basics

While all proteins have a similar bulk composition, this obscures rather than illuminates their dramatic structural and functional differences. With the introduction of various chemical methods, it was discovered that different proteins were composed of distinct and specific sets

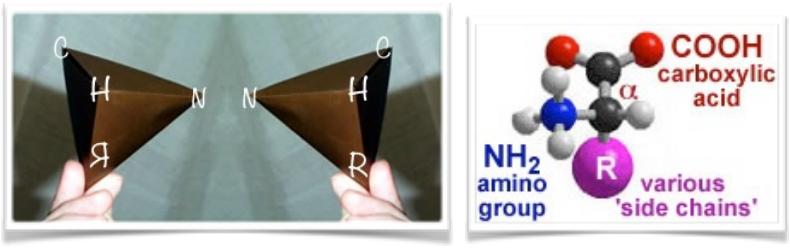
[193] From 'protein' to the beginnings of clinical proteomics: http://www.ncbi.nlm.nih.gov/pubmed/21136729

[194] While historically true, the original claim that proteins get their name from "the ancient Greek sea-god Proteus who, like your typical sea-god, could change shape. The name acknowledges the many different properties and functions of proteins." seems more poetically satisfying to us.

of subunits, and that each subunit is an unbranched polymer of amino acids with a specific sequence. Because the amino acids in these polymers are linked by what are known as peptide bonds, the polymers are known generically as polypeptides. At this point, it is important to reiterate that proteins are functional objects, In addition to polypeptides many proteins also contain other molecular components, known as co-factors or prosthetic groups (we will call them co-factors for simplicity's sake.) These co-factors can range from metal ions to various small molecules.

Amino acid polymers

As you might remember from chemistry, carbon atoms (C) form four bonds, and where these are all single bonds, the basic structure of the atoms bound to a C is tetrahedral. We can think of an amino acid as a (highly) modified form of methane (CH_4), with the C referred to as the alpha carbon (C_α). Instead of four hydrogens attached to the central C, there is one H, an amino group ($-NH_2$), a carboxylic acid group (-COOH), and a final, variable (R) group attached to the central C_α atom. The four groups attached to the α-carbon are arranged at the vertices of a tetrahedron. If all four groups attached to the α-carbon are different from one another, as they are in all amino acids except glycine, the resulting amino acid can exist in two possible stereoisomers, which are known as enantiomers. Enantiomers are mirror images of one another and are termed the L- and D- forms. Only L-type amino acids are found in proteins, even though there is no obvious reason that proteins could not have also been made using both types of amino acids or using only D-amino acids.[195] It appears that the universal use of L-type amino acids in the polypeptides found in biological systems is yet another example of the evolutionary relatedness of organisms, it appears to be a homologous trait. Even though there are hundreds of different amino acids known, only 22 amino acids (these include the 20 common amino acids and two others, selenocysteine and pyrrolysine) are found in proteins.

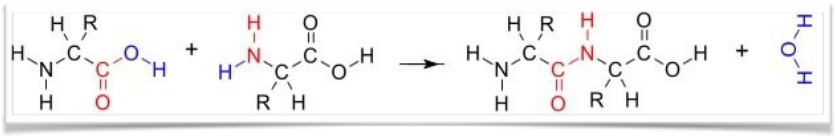


Amino acids differ from one another by their R-groups, which are often referred to as "side-chains". Some of these R-groups are large, some are small, some are hydrophobic, some are hydrophilic, some of the hydrophilic R-groups contain weak acidic or basic groups. The extent to which these weak acidic or basic groups are positively or negatively charged will change in response to environmental pH. Changes in charge will (as we will see) influence the structure of the polypeptide/protein in which they find themselves. The different R-groups provide proteins with a broad range of chemical properties, which are further extended by the presence of co-factors.

As we noted for nucleic acids, a polymer is a chain of subunits, amino acid monomers linked together by peptide bonds. Under the conditions that exist inside the cell, this is a thermodynamically unfavorable dehydration reaction, and so must be coupled to a thermodynamically favorable reaction. A

[195] It is not that D-amino acids do not occur in nature, or in organisms, they do. They are found in biomolecules, such as the antibiotic gramicidin, which is composed of alternating L-and D-type amino acids - however gramicidin is synthesized by a different process than that used to synthesize proteins.

molecule formed from two amino acids, joined together by a peptide bond, is known as a dipeptide. As in the case of each amino acid, the dipeptide has an N-terminal (amino) end and a C-terminal (carboxylic acid) end. To generate a polypeptide, new amino acids are added (exclusively) to the C-terminal end of the polymer. A peptide bond forms between the amino group of the added amino acid and the carboxylic acid group of the polymer. This reaction generates a new C-terminal carboxylic acid group. It is important to note that while some amino acids have a carboxylic acid group as part of their R-groups, new amino acids are not added there. Because of this fact, polypeptides are unbranched, linear polymers. This process of amino acid addition can continue, theoretically without limit. Biological polypeptides range from very short (5-10) to many hundreds (thousands) of amino acids in length. For example, the protein Titin (found in muscle cells) can be more than 30,000 amino acids in length. Because there is no theoretical constraint on which amino acids occur at a particular position within a polypeptide, there is a enormous universe of possible polypeptides that could exist. In the case of a 100 amino acid long polypeptide, there are 20^{100} possible different polypeptides that could be formed.



Specifying a polypeptide's sequence

Perhaps at this point you are asking yourself, if there are so many different possible polypeptides, and there is no inherent bias favoring the addition of one amino acid over another, what determines the sequence of a polypeptide, clearly it is not random. Here we connect to the information stored in DNA. We begin with a description of the process in bacteria and then extend it to archaea and eukaryotes. We introduce them in this order because, while basically similar, the system is simpler in bacteria (although you might find it complex enough for your taste.) Even so, we will leave most of the complexities for subsequent courses. One thing that we will do that is not common is that we will consider the network dynamics of these systems. We will even ask you to do a little analytics, with the goal of enabling you to make plausible predictions about the behavior of these systems, particularly in response to various perturbations. Another important point to keep in mind, one we have made previously, is that the system is continuous. The machinery required for protein synthesis is inherited by the cell, so each new polypeptide is synthesized in an environment full of pre-existing proteins and ongoing metabolic processes.

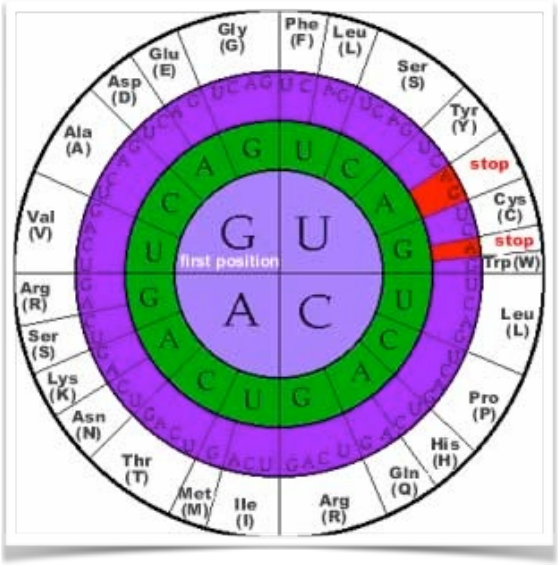
Making a polypeptide in a bacterial cell

A bacterial cell synthesizes thousands of different polypeptides. The sequences of these polypeptides are encoded within the DNA of the organism. The genome of most bacteria is a double-stranded circular DNA molecule that is millions of base pairs in length. Each polypeptide is encoded by a specific region of this DNA molecule. So, our questions are how are specific regions in the DNA recognized and how is nucleic acid-encoded information translated into polypeptide sequence.

To address the first question, thinking back to the structure of DNA, it was immediately obvious that the one-dimensional sequence of a polypeptide could be encoded in the one-dimensional

sequence of the polynucleotide chains in a DNA molecule.[196] The real question was how to translate the language of nucleic acids, which consists of sequences of four different nucleotide bases, into the language of polypeptides, which consists of sequences of the 20 different amino acids. As pointed out by the physicist George Gamow (1904-1968), when he was a professor at UC Boulder, the minimum set of nucleotides needed to encode all 20 amino acids is three; a sequence of one nucleotide (4^1) could encode at most four different animo acids, a two nucleotide length sequence could encode (4^2) or 16 different amino acids (not enough), while a three nucleotide sequence (4^3) could encode 64 different amino acids (more than enough).[197] Although the actual coding scheme that Gamow proposed was wrong, his thinking about coding capacity influenced those who experimentally determined the actual rules of the "genetic code".

The genetic code is not the information itself, but the algorithm by which nucleotide sequences are "read" to determine polypeptide sequences. A polypeptide is encoded by the sequence of nucleotides. This nucleotide sequence is read in groups of three nucleotides, known as a codon. Codons are read in a non-overlapping manner, with no spaces (that is, non-coding nucleotides) between them. Since there are 64 possible codons but only 20 (or 22 - see above) different amino acids used in organisms, the code is redundant, that is, certain amino acids are encoded by more than one codon. In addition, there are three codons, UAA, UAG and UGA that encode "stops"; they do not encode any amino acid but are used to mark the end of a polypeptide. The region of the nucleic acid that encodes a polypeptide begins with what is known as the "start" codon and continues until a stop codon is reached. This sequence is known as an open reading frame or an ORF.



There are a number of hypotheses on the origin of the genetic code. One is the frozen accident model in which the code used in modern cells is the result of an accident, a bottleneck event. Early in the evolution of life on Earth, there may have been multiple types of organisms, using different codes, but the code used reflects the fact that only one of these organisms gave rise to all modern organisms. Alternatively, the code could reflect specific interactions between RNAs and amino acids that played a role in the initial establishment of the code. What is clear is that the code is not necessarily fixed, there are examples in which certain codons are "repurposed" in various organisms. What these variations in the genetic code illustrate is that evolutionary mechanisms can change the genetic code.[198] Since the genetic code does not appear to be predetermined, the

---

[196] Nature of the genetic code finally revealed!: http://www.nature.com/nrmicro/journal/v9/n12/full/nrmicro2707.html

[197] The Big Bang and the genetic code: Gamow, a prankster and physicist, thought of them first: http://www.nature.com/nature/journal/v404/n6777/full/404437a0.html:

[198] The genetic code is nearly optimal for allowing additional information within protein-coding sequences: http://genome.cshlp.org/content/17/4/405 and Stops making sense: translational trade-offs and stop codon reassignment: http://www.ncbi.nlm.nih.gov/pubmed/21801361

general conservation of the genetic code among organisms is seen as strong evidence that all organisms (even the ones with minor variations in their genetic codes) are derived from a single common ancestor. It appears that the genetic code is a homologous trait between organisms.

An important feature of the genetic system is that the information stored in DNA is not used directly to direct polypeptide synthesis. Rather it has to be copied through the formation of an RNA molecule, known as a messenger RNA or mRNA. In contrast to the process involved in the transformation of the information stored in a nucleic acid sequence into a polypeptide sequence, both DNA and RNA use the same nucleotide language. Because of this fact, the process of DNA-directed RNA synthesis is known as ***transcription.*** The process of RNA-directed polypeptide synthesis is known as ***translation***, because the language of nucleic acids is different from the language of polypeptides.

Protein synthesis: transcription (DNA to RNA)

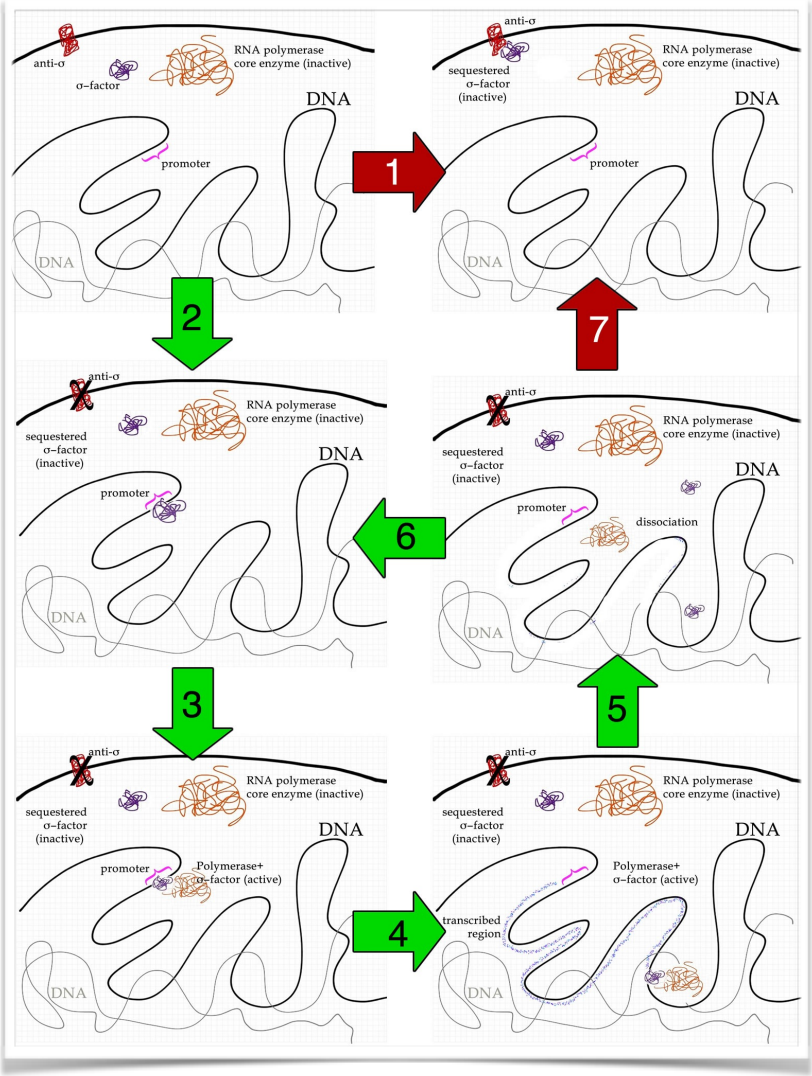
Having introduced the genetic code and RNA, however, briefly, we now return to the process by which a polypeptide is specified by a DNA sequence. Our first task is to understand how it is that we can find the specific region of the DNA molecule that encodes a specific polypeptide, since we are looking for a short region of DNA within millions or in eukaryotes, typically billions of base pairs of sequence). So while the double stranded nature of DNA makes the information stored in it redundant (a fact that makes DNA replication straightforward), the specific nucleotide sequence that will be decoded using the genetic code is present in only one of the two strands. From the point of view of polypeptide sequence the other strand is nonsense.

As we have noted, a gene is the region(s) of a larger DNA molecule. Part of the gene's sequence, its regulatory region, is used (as part of a larger system involving the products of other genes) to specify when, where, and how much the gene is "expressed". Another part of the gene's sequence is used to direct the synthesis of an RNA molecule (the transcribed or coding region). Once a gene's regulatory region is engaged, the synthesis of an RNA molecule is the next step in the expression of the gene. As a general simplification, we will say that a gene is expressed when the RNA that it encodes is synthesized. We can postpone further complexities to later (and subsequent classes). It is important to recognize that an organism as "simple" as a bacterium can contain thousands of genes, and that different sets of genes are used in different environments to produce specific behaviors. In some cases, these behaviors may be mutually antagonistic. For example, a bacterium facing a rapidly drying out environment might turn on genes that allow it to stop growing and dividing, and prepare it to survive in such a hostile environment. That means some genes (involved in active growth and replication) need to be turned off, while others, involved in survival, need to be turned on. Our goal is not to have you accurately predict the particular behavior of an organism, but rather to be able to make plausible predictions about how gene expression will change in response to various perturbations. This requires us to go into some detail about mechanisms, but rather superficially, in order to illustrate a few of the regulatory processes that are active in cells.

So you need to think, what are the molecular components that can recognize a gene's regulatory sequences? The answer is proteins. The class of proteins that do this are known generically

as transcription factors. Their shared property is that they bind with high affinity to specific sequences of nucleotides within DNA molecules. For historical reasons, in bacteria these transcription factor proteins are known as sigma (σ) factors. The next question is how is an RNA made based on a DNA sequence? The answer is DNA-dependent RNA polymerase, which we will refer to as RNA polymerase. In bacteria, groups of genes share regulatory sequences recognized by specific σ factors. As we will see this makes it possible to regulate groups of specific genes in a coordinated manner. Now let us turn to how, exactly (although at low resolution), this is done, first in bacteria and then in eukaryotic cells.

At this point, we need to explicitly recognize common aspects of biological systems. They are highly regulated, adaptive and homeostatic - that is, they can adjust their behavior to changes in their environment (both internal and external) to maintain the living state. These types of behaviors are based on various forms of feedback regulation. In the case of the bacterial gene expression system, there are genes that encode specific σ factors. Which of these genes are expressed determines which σ factor proteins are present and which genes are actively expressed. Of course, the gene encoding a specific σ factor is itself regulated. At the same time, there are other genes that encode what are known as anti-σ factors. One class of anti-σ factors are membrane-associated proteins. For a σ factor to activate a gene, it must be able to bind to the DNA, which it cannot do if it is bound to the anti-σ factor. So a gene may not be expressed (we say that it is "off") because the appropriate σ factor is not expressed or because even though that σ factor is expressed, the relevant anti-σ factor is also expressed, and its presence acts to block the action of the σ factor (arrow 1). We can, however, turn on our target gene if we inactivate the anti-σ factor. Inactivation can involve a number of mechanisms, including the destruction or modification of the anti-σ factor so that it no longer interacts with the σ factor. Once the σ factor is released, it can diffuse through out the cell and bind to its target DNA sequences (arrow 2). Now an inactive RNA polymerase can bind to the DNA-σ factor complex (arrow 3). This activates the RNA polymerase, which initiates DNA-dependent RNA synthesis (arrow 4). Once RNA polymerase has been activated, it will



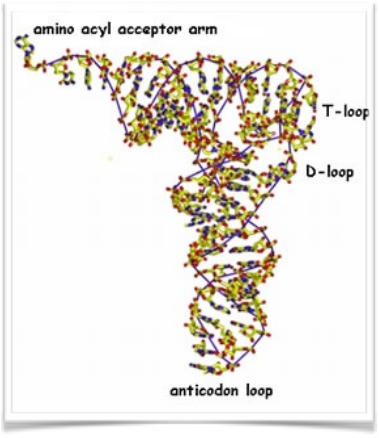
move away from the σ factor. The DNA bound σ factor could bind another polymerase (arrow 6) or the σ factor could release from the DNA and then diffuse around and rebind to other sites in the DNA or to the anti-σ factor if that protein is present (arrow 7).

As a reminder, RNA synthesis is a thermodynamically unfavorable reaction, so for it to occur it must be coupled to a thermodynamically favorable reaction, in particular nucleotide triphosphate hydrolysis (see previous chapter). The RNA polymerase moves along the DNA (or the DNA moves through the RNA polymerase, your choice), to generate an RNA molecule (the transcript). Other signals lead to the termination of transcription and the release of the RNA polymerase (arrow 5). Once released, the RNA polymerase returns to its inactive state. Another gene can be transcribed if the RNA polymerase interacts with a σ factor bound to its promoter (arrow 6). Since multiple types σ factor proteins are present within the cell and RNA polymerase can interact with all of them, which genes are expressed within a cell will depend upon the relative concentrations of σ factors and anti-σ factor proteins present and active, and the binding affinities of particular σ factors for specific DNA sequences (compared to their general low-affinity binding to DNA in general).

Protein synthesis: translation (RNA to polypeptide)

Translation involves a complex cellular organelle, the ribosome, which together with a number of accessory factors reads the code in a mRNA molecule and produces the appropriate polypeptide.[199] The ribosome is the site of polypeptide synthesis. It holds the various components (the mRNA, tRNAs, and accessory factors) in appropriate juxtaposition to one another to catalyze polypeptide synthesis. But perhaps we are getting ahead of ourselves. For one, what exactly is a tRNA?

While we have focussed on mRNA up to now, the process of transcription is also used to generate other types of RNAs; these play structural, catalytic, and regulatory roles within the cell. Of these non-mRNAs, two are particularly important in the context of polypeptide synthesis. The first are molecules known as transfer RNAs (tRNAs). These small single stranded RNA molecules fold back on themselves to generate a compact L-shaped structure (→). In the bacterium *E. coli*, there are 87 tRNA encoding genes (there are over 400 such tRNA encoding genes in human). For each amino acid and each codon there are one or more tRNAs. The only exception being the stop codons. A tRNA specific for the amino acid phenylalanine would be written $tRNA^{Phe}$. Two parts of the tRNA molecule are particularly important and functionally linked: the part that recognizes the codon on the mRNA and the amino acid acceptor stem, which is where an amino acid is attached to the tRNA. Each specific type of tRNA can recognize a particular codon in an mRNA through base pairing interactions with what is known as the anti-codon. The rest of the tRNA molecule mediates interactions with protein catalysts (enzymes) known as amino acyl tRNA synthetases. There is a distinct amino acyl tRNA synthetase for each amino acid, so that there is a phenylalanine-tRNA synthetase and a proline-tRNA synthetase, etc. An amino acyl tRNA synthetase binds the appropriate tRNA and amino acid and, through a reaction coupled to a thermodynamically favorable nucleotide triphosphate hydrolysis

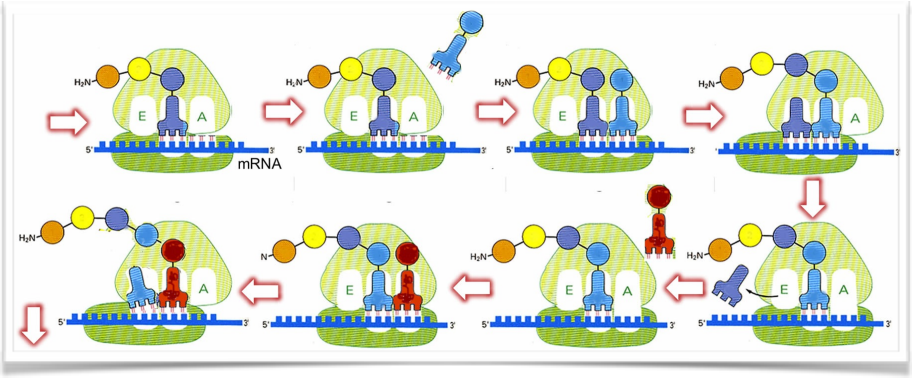


[199] Can't stop yourself? go here for a more detailed description of translation.http://www.nature.com/nsmb/journal/v19/n6/full/nsmb.2313.html?WT.ec_id=NSMB-201206

reaction, catalyzes the formation of a covalent bond between the amino acid acceptor stem of the tRNA and the amino acid, to form what is known as a charged or amino acyl-tRNA. The loop containing the anti-codon is located at the other end of the tRNA molecule. As we will see, in the course of polypeptide synthesis, the amino acid group attached to the tRNA's acceptor stem will be transferred from the tRNA to the growing polypeptide.

Ribosomes: Ribosomes are composed of roughly equal amounts (by mass) of ribosomal (rRNAs) and ribosomal polypeptides. An active ribosome is composed of a small and a large ribosomal subunit. In the bacterium *E. coli*, the small subunit is composed of 21 different polypeptides and a 1542 nucleotide long rRNA molecule, while the large subunit is composed of 33 different polypeptides and two rRNAs, one 121 nucleotides long and the other 2904 nucleotides long.[200] It goes without saying (so why are we saying it?) that each ribosomal polypeptide and RNA is itself a gene product. The complete ribosome has a molecular weight of ~3 x 10^6 daltons. One of the rRNAs is an evolutionarily conserved catalyst, known as a ribozyme (in contrast to protein based catalysts, which are known as enzymes). This catalytic rRNA lies at the heart of the ribosome - it catalyzes the transfer of an amino acid bound to a tRNA to the carboxylic acid end of the growing polypeptide chain.

The growing polypeptide chain is bound to a tRNA, known as the peptidyl tRNA. When a new aa-tRNA enters the ribosome's active site (site A), the growing polypeptide is added to it, so that it becomes the peptidyl tRNA (with a newly added amino acid, the amino acid originally associated with incoming aa-tRNA). This attached polypeptide group is now one amino acid longer. Again, the use of an RNA based catalysts is a conserved feature of polypeptide synthesis in all known organisms, and appears to represent an evolutionarily homologous trait.



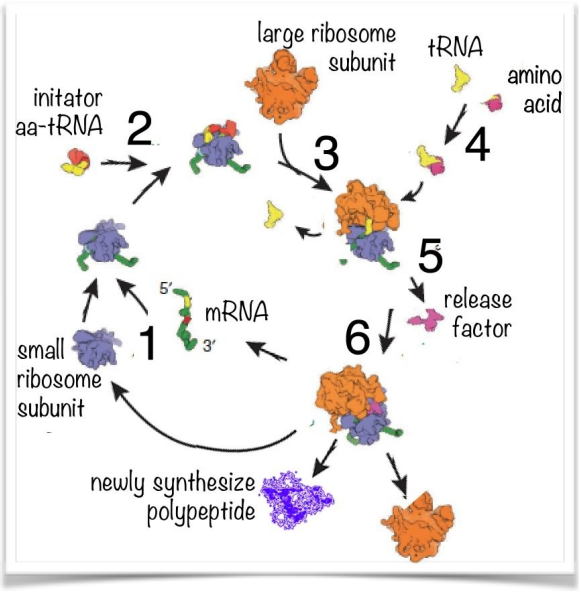
The cytoplasm of cells is packed with ribosomes. In a rapidly growing bacterial cell, approximately 25% of the total cell mass is ribosomes. Although structurally similar, there are characteristic differences between the ribosomes of bacteria, archaea, and eukaryotes. This is important from a practical perspective. For example, a number of antibiotics selectively inhibit polypeptide synthesis by bacterial, but not eukaryotic ribosomes. Both chloroplasts and mitochondria have ribosomes of the bacterial type. This is yet another piece of evidence that chloroplasts and mitochondria are descended from bacterial endosymbionts and a reason that translational blocking anti-bacterial antibiotics are mostly benign, since most of the ribosomes inside a eukaryotic cell are not effected by them.

[200] In the human, the small ribosomal subunit is composed of 33 polypeptides and a 1870 nucleotide rRNA, while the large ribosomal subunit contains 47 polypeptides, and three rRNAs of 121, 156, and 5034 nucleotides in length.

The translation (polypeptide synthesis) cycle

In bacteria, there is no barrier between the cell's DNA and the cytoplasm, which contains the ribosomal subunits and all of the other components involved in polypeptide synthesis. Newly synthesized RNAs are released directly into the cytoplasm, where they can begin to interact with ribosomes. In fact, because the DNA is located in the cytoplasm in bacteria, the process of protein synthesis (translation) can begin before mRNA synthesis (transcription) is complete.

We will walk through the process of protein synthesis, but at each step we will leave out the various accessory factors involved in regulating the process and coupling it to the thermodynamically favorable reactions that make it possible. These can be important if you want to re-engineer or manipulate the translation system, but are unnecessary conceptual obstacles that obscure a basic understanding. Here we will remind you of two recurring themes. The first is to recognize that all of the components needed to synthesize a new polypeptide (except the mRNA) are already present in the cell; another example of biological continuity. The second is that all of the interactions we will be describing are based on stochastic, thermally driven movements. For example, when considering the addition of an amino acid to a tRNA, random motions have to bring the correct amino acid and the correct tRNA to their binding sites on the appropriate amino acyl tRNA synthetase, and then bring the correct amino acid charged tRNA to the ribosome. Generally, many unproductive collisions will occur before a productive (correct) one, since there are more than 20 different amino acid/tRNA molecules bouncing around in the cytoplasm.

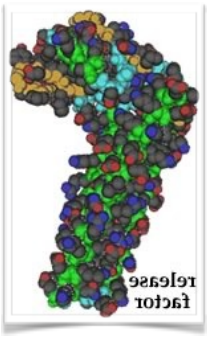


The first step in polypeptide synthesis is the synthesis of the specific mRNA that encodes the polypeptide. (1) The mRNA contains a sequence[201] that mediates its binding to the small ribosomal subunit. This sequence is located near the 5' end of the mRNA. (2) the mRNA-small ribosome subunit complex now interacts with and binds to a complex containing an initiator (start) amino acid:tRNA. In both bacteria and eukaryotes the start codon is generally an AUG codon and inserts the amino acid methionine (although other, non-AUG start codons are possible).[202] This interaction defines the beginning of the polypeptide and the reading frame within the mRNA. (3) The met-tRNA:mRNA:small ribosome subunit complex can now form a functional complex with a large ribosomal subunit to form the functional mRNA:ribosome complex. (4) Catalyzed by amino acid tRNA synthetases, charged amino acyl tRNAs will be present and can interact with the mRNA:ribosome complex to generate a polypeptide. Based on the mRNA sequence and the reading frame defined by the start codon, amino acids will be added

[201] Known as the Shine-Delgarno sequence for its discovers

[202] Hidden coding potential of eukaryotic genomes: nonAUG started ORFs: http://www.ncbi.nlm.nih.gov/pubmed/22804099

sequentially. With each new amino acid added, the ribosome moves along the mRNA. An important point, that we will return to when we consider the folding of polypeptides into their final structures, is that the newly synthesized polypeptide threads through a molecular tunnel in the ribosome. Only after the N-terminal end of the polypeptide begins to emerge from this tunnel can it begin to fold. (5) The process of polypeptide polymerization continues until the ribosome reaches a stop codon, that is a UGA, UAA or UAG.[203] Since there is no tRNA for this codon, the ribosome pauses, waiting for a charged tRNA which will never arrive. Instead, a polypeptide known as release factor, which has a shape something like a tRNA, binds to the polypeptide:mRNA:ribosome complex instead.
(6) This leads to the release of the polypeptide, the disassembly of the ribosome into small and large subunits, and the release of the mRNA.



When associated with the ribosome, the mRNA is protected against interaction with proteins (ribonucleases) that could degrade it, that is, break it down into nucleotides. Upon its release the mRNA may interact with a new small ribosome subunit, and begin the process of polypeptide synthesis again or it may interact with a ribonuclease and be degraded. Where it is important to limit the synthesis of particular polypeptides, the relative probabilities of these two events (new translation or RNA degradation) will be skewed in favor of degradation. Typically this is mediated by specific nucleotide sequences in the mRNA. The relationship between mRNA synthesis and degradation will determine the half-life of a population of mRNA molecules, the steady state concentration of the mRNA in the cell, and indirectly, the level of polypeptide present.

Bursting synthesis and alarm generation

At this point, let us consider a number of interesting behaviors associated with translation. First, the onset of translation begins with the small ribosomal subunit interacting with the 5' end of the mRNA. Multiple ribosomes can interact with a single mRNA, each moving down the mRNA molecule, synthesizing a polypeptide. Turns out, the initial interaction between an mRNA and the first ribosomal subunit makes it more likely that other ribosomal subunits can add, once the first ribosome begins moving away from the ribosomal binding site on the mRNA. This has the result that the synthesis of polypeptides from an RNA often involves a burst of multiple events. Since the number of mRNA molecules encoding a particularly polypeptide can be quite small (less than 10 per cell in some cases), this can lead to noisy protein synthesis. Bursts of new polypeptide synthesis can then be followed by periods when no new polypeptides are made.

The translation system is dynamic and a major consumer of energy within the cell.[204] When a cell, particularly a bacterial cell, is starving, it does not have the energy to generated amino acid charged tRNAs. The result is that uncharged tRNAs accumulate. Since uncharged tRNAs fit into the

[203] In addition to the common 19 amino and 1 imino (proline) acids, the code can be used to insert two other amino acids selenocysteine and pyrrolysine. In the case of selenocysteine, the amino acid is encoded by a stop codon, UGA, that is in a particular context within the mRNA. Pyrrolysine is also encoded by a stop codon. In this case, a gene that encodes a special tRNA that recognizes the normal stop codon UAG is expressed. see Selenocysteine: http://www.ncbi.nlm.nih.gov/pubmed/8811175

[204] Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources: http://www.ncbi.nlm.nih.gov/pubmed/24766808

amino-acyl-tRNA binding sites on the ribosome, their presence increases the probability of unproductive tRNA interactions with the mRNA-ribosome complex. When this occurs, the stalled ribosome generates a signal (see [205]) that can lead to adaptive changes in the cell that enable it to survive for long periods in a “dormant” state.[206]

Another response that can occur is a more social one. Some cells in the population can “sacrifice” themselves for their (generally closely related) neighbors (remember kin selection and inclusive fitness.) This mechanism is based on the fact that proteins, like nucleic acids, differ in the rates that they are degraded within the cell. Just as ribonucleases can degrade mRNAs, proteases degrade proteins and polypeptides. How stable a protein/polypeptide is depends upon its structure, which we will be turning to soon.

A common system within bacterial cells is known as an addiction module. It consists of two genes, encoding two distinct polypeptides. One forms a toxin molecule which when active can kill the cell. The second is an anti-toxin (a common regulatory scheme, think back to σ factors and anti-σ factors.) The key feature of the toxin-anti-toxin system is that the toxin molecule is stable, it is has a long half life. The half-life of a molecule is the time it takes for 50% of the molecules in a population to be degraded (or otherwise disappear from the system.) In contrast, the anti-toxin molecule’s half-life is short. The result is that if protein synthesis slows or stops, the level of the toxin will remain high, while the level of the anti-toxin will drop rapidly, which leads to loss of inhibition of the toxin, and cell death. Death leads to the release of the cell’s nutrients which can be used by its neighbors. A similar process can occur if a virus infects a cell, if the cell kills itself before the virus replicates, it destroys the virus and protects its neighbors (who are likely to be its relatives.

Questions to answer & to ponder:

- What are the “natural” limits to the structure of an R-group in a polypeptide?
- How would a condensation reaction be effected by the removal of water from a system?
- Why do we think that the use of a common set of amino acids is a homologous trait?
- What factors can you imagine influenced the set of amino acids used in organisms?
- Why so many tRNA genes?
- Why does the ribosome tunnel inhibit the folding of the newly synthesized polypeptide?
- What types of molecules does DNA directly encode? How about indirectly?
- How might a DNA molecule encode the structure of a lipid?
- How, in the most basic terms, do different tRNAs differ from one another?
- What is the minimal number of different tRNA-amino acid synthetases in a cell?
- What could happen if a ribosome started translating an mRNA at the "wrong" place?
- Why don’t release factors cause the premature termination of translation at non-stop codons?
- What does it mean to say the genetic code is an algorithm?
- What is meant when people call the genetic code a "frozen accident"?
- What is (seriously) unrealistic about this tutorial [http://youtu.be/TfYf_rPWUdY]?
- Design a process (and explain the steps) by which you might reengineer an organism to use a new (non-biological) type of amino acid in its proteins.

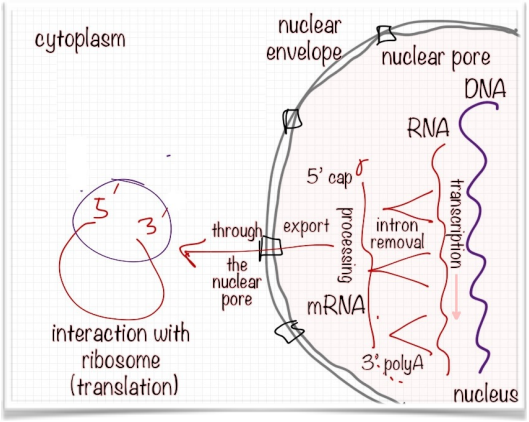
---

[205] http://virtuallaboratory.colorado.edu/BioFun-Support/labs/Adaptation/section_03.html

[206] Characterization of the Starvation-Survival Response of Staphylococcus aureus*:* http://www.ncbi.nlm.nih.gov/pmc/articles/PMC107086/

Getting more complex: gene regulation in eukaryotes

At this point, we will not take vary much time to go into how gene expression in particular, and polypeptide synthesis in general differ between prokaryotes and eukaryotes except to point out a few of the major differences, some of which we will return to, but most will be relevant only in more specialized courses. The first and most obvious difference is the presence of a nucleus, a distinct domain within the eukaryotic cell that separates the cell's genetic material, its DNA, from the cytoplasm. The nucleus is a distinct compartment, with a distinct environment. This distinction is maintained by active processes that serve to both restrict the movement of molecules into and out of the nucleus (from the cytoplasm), and to re-establish the nuclear environment in situations in which it breaks down. As we will see later on this can occurs during cell division (mitosis). The barrier between nuclear interior and cytoplasm is known as the nuclear envelope (no such barrier exists in prokaryotic cells, the DNA is in direct contact with the cytoplasm.) The nuclear envelope consists of two lipid bilayer membranes that are punctuated by macromolecular complexes (protein machines) known as nuclear pores. While molecules of molecular weight less than ~40,000 atomic units (known as daltons) can generally pass through the nuclear pore, larger molecules must be transported actively, that is, in a process that is coupled to a thermodynamically favorable reaction, in this case the hydrolysis of guanosine triphosphate (GTP) instead of adenine triphosphate (ATP). The movement of larger molecules into and out of the nucleus through nuclear pores is regulated by what are known as nuclear localization and nuclear export sequences, present in polypeptides. These are recognized by proteins associated with the pore complex, and lead to movement of the polypeptide into or out of the nucleus.

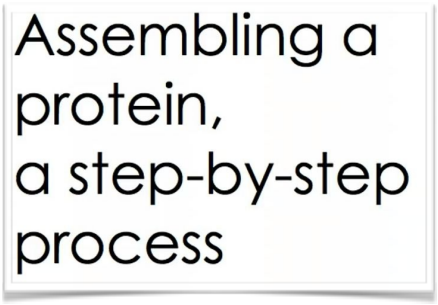


Aside from those within mitochondria and chloroplasts, the DNA molecules of eukaryotic cells are located within the nucleus. One difference between eukaryotic and bacterial genes is that the transcribed region of eukaryotic genes often contains what are known as intervening sequences or introns. After the RNA is synthesized, these non-coding introns are removed enzymatically, resulting in a shorter mRNA. As a point of interest, which sequences are removed can be regulated, this can result in mRNAs that encode somewhat (and often dramatically) different polypeptides. In addition to removing introns, the mRNA is further modified at both its 5' and 3' ends. Only after RNA processing as occurred is the mature mRNA exported out of the nucleus, through a nuclear pore into the cytoplasm, where it can interact with ribosomes. One further difference from bacteria is that the mRNA recognition of the small ribosomal subunit involves the formation of a complex in which the 5' and 3' ends of the mRNA are brought together into a circle. The important point here is that unlike the situation in bacteria, where mRNA is synthesized into the cytoplasm and so can immediately interact with ribosomes and begin translation (even before the synthesis of the RNA is finished), the coupling of transcription and translation does not occur in eukaryotes because of the nuclear envelope. Transcription occurs within the nucleus and the mRNA must be transported to the cytoplasm (where the ribosomes are located) before it can be translated. This makes processes like RNA splicing, and the generation of multiple,

functionally distinct RNAs from a single gene possible. This leads to significantly greater complexity from only a relatively small increase in the number of genes.

Turning polypeptides into proteins

Protein structure is commonly presented in a hierarchical manner. While this is an over-simplification, it is a good place to start. When we think about how a polypeptide folds, we have to think about the environment it will inhabit, how it interacts with itself, and were it is part of a *multi*-polypeptide protein, how its interactions with other subunits are established. As we think about polypeptide structure, it is typical to see it referred to in terms of primary, secondary, tertiary, and quaternary structure. The primary structure of a polypeptide is the sequence of amino acids in a polypeptide, chain, written from its N- or amino terminus to its C- or carboxyl terminus. As we will see below, the secondary structure of a polypeptide consists of local folding motifs: the α-helix, the β-sheet, and connecting domains. The tertiary structure of a polypeptide is the overall three dimensional shape a polypeptide takes in space (as well as how its R-chains are oriented). Quaternary structure refers to how the various polypeptides and co-factors that combine to make up a functional protein are arranged with respect to one another. In a protein that consists of a single polypeptide and no co-factors, its tertiary and quaternary structures are the same. As a final complexity, a particular polypeptide can be part of a number of different proteins. This is one reason that a gene can play a role in a number of different processes and be involved in a number of different phenotypes.



see http://youtu.be/Rq7DwrX0Uoc

Polypeptide synthesis (translation), like most all processes that occur within the cell, is a stochastic process, meaning that it is based on random collisions between molecules. In the specific case of translation, the association of the mRNA with ribosomal components occurs stochastically; similarly, the addition of a new amino acid depends on the collision of the appropriate amino acid-charged tRNA with the RNA-ribosome complex. Since there are many different amino-acid charged tRNAs in the cytoplasm, the ribosomal complex must be able to productively bind only the tRNA that the mRNA specifies, that is the tRNA with the right anticodon. This enables its attached amino acid to interact productively to add the amino acid to the growing polypeptide chain. In most illustrations of polypeptide synthesis, you rarely see this fact illustrated. From 12 to 21 amino acids are added per second in bacterial cells (and about half that rate in mammalian cells).[207]

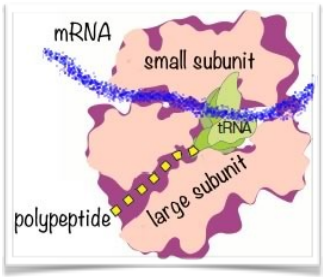
Now you might wonder if there are errors in polypeptide synthesis, as there are in nucleic acid synthesis. In fact there are. For example, if a base is skipped, the reading frame will be thrown off. Typically, this leads to a completely wrong sequence of amino acids added to the end of the polypeptide and generally quickly leads to a stop codon, which terminates translation, releasing a polypeptide that cannot fold correctly and is (generally) rapidly degraded.[208] Similarly, if the wrong amino acid is inserted at a particular position and it disrupts normal folding, the polypeptide could be

[207] see http://bionumbers.hms.harvard.edu/default.aspx

[208] Quality control by the ribosome following peptide bond formation: http://www.ncbi.nlm.nih.gov/pubmed/19092806

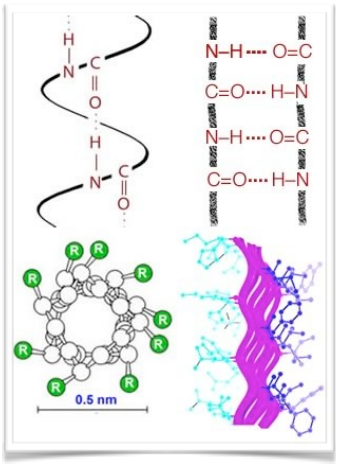
degraded. What limits the effects of mistakes during translation is that most proteins (unlike DNA molecules) have finite and relatively short half-lives; that is, the time an average polypeptide exists before it is degraded by various enzymes. Normally (but not always) this limits the damage that even an antimorphic polypeptide can do to the cell and organism.

Factors influencing polypeptide folding and structure: Polypeptides are synthesized, and they fold, in a vectorial, that is, directional manner. The polypeptide is synthesized in an N- to C- terminal direction and exits the ribosome through a tunnel approximately 10 nm long and 1.5 nm in diameter. This tunnel is narrow enough to block the folding of the newly synthesized polypeptide chain. As the polypeptide emerges from the tunnel, it encounters the crowded cytoplasmic environment; at the same time it begins to fold. As it folds, the polypeptide needs to avoid low affinity, non-specific, and non-physiologically significant interactions with other cellular components. These arise due to the fact that all molecules interact with each other via van der Waals interactions. If it is part of a multi-subunit protein, it must "find" its partner polypeptides, which again is a stochastic process. If the polypeptide does not fold correctly, it will not function correctly and may damage the cell. A number of degenerative neurological disorders are due, at least in part, to the accumulation of misfolded polypeptides (see below).



http://youtu.be/i8rGTYQ6oZ8

We can think of the folding process as a “drunken” walk across an energy landscape, with movements driven by thermal fluctuations and thermodynamic factors. The goal is to find the lowest point in the landscape, the energy minimum of the system. This is generally assumed to be the native or functional state of the polypeptide. That said, this state is not necessarily static, since the folded polypeptide (and the final protein) will be subject to thermal fluctuations; it is possible that it will move between various states with similar, but not identical stabilities. The problem of calculating the final folded state of a polypeptide is an extremely complex one. Generally two approaches are taken, in the first the structure of the protein is determined directly by X-ray crystallography or Nuclear Magnetic Resonance spectroscopy. In the second, if the structure of a homologous protein is known (and we will consider homologous proteins later on), it can be used as a framework to model the structure of a previously unsolved protein.

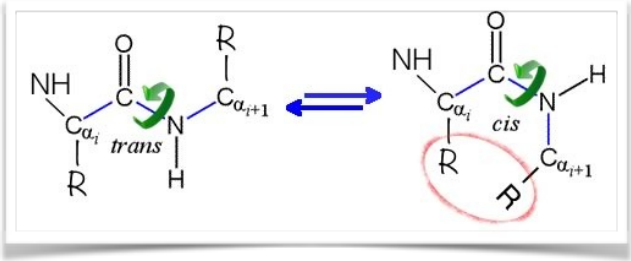


There are a number of constraints that influence the folding of a polypeptide. The first is the peptide bond itself. All polypeptides contain a string of peptide bonds. It is therefore not surprising that there are common patterns in polypeptide folding. The first of these common patterns to be recognized, the α-helix, was discovered by Linus Pauling and Robert Corey in 1951. This was followed shortly thereafter by their description of the β-sheet. The forces that drive the formation of the α-helix and the β-sheet will be familiar. They are the same forces that underlie water structure.

In an α-helix and a β-sheet, all of the possible H-bonds involving the peptide bond's donor and acceptor groups (–N–H : O=C– with “:” indicating a H-bond) are formed within the polypeptide. In the α-helix these H-bond interactions run parallel to the polypeptide chain. In the β-sheet they occur between

polypeptide strands. These strands can be within the same polypeptide and can run parallel or anti-parallel to one another. requiring one or more bends in the polypeptide. It is also possible to have β-sheet interactions between polypeptides located in different polypeptides. In an α-helix, the R-groups point outward from the helix axis. In β-sheets the R-groups point in an alternating manner either above or below the sheet. While all amino acids can take part in either α-helix or β-sheet structures, the imino acid proline cannot - the N-group coming off the α-carbon has no H, so its presence in a polypeptide chain leads to a break in the pattern of intrachain H-bonds.

Peptide bond rotation and proline: Although drawn as a single bond, the peptide bond behaves more like a double bond, or rather like a bond and a half. In the case of a single bond, there is free rotation around the bond axis in response to thermal motion. In contrast, rotation around a peptide bond requires more energy to move from the trans to the cis configuration and back again, that is, it is more difficult to rotate around the peptide bond. In addition, in the cis configuration the R groups of adjacent amino acids are on the same side of the polypeptide chain. If these R groups are large, they can bump into each other. If they get too close the repulsions between the outer electrons of each group make this arrangement less stable. This will usually lead to the polypeptide chain to prefer (at least locally) to be in the *trans* arrangement. In both α-helix and β-sheet configurations, the peptide bonds are in the *trans* configuration because the *cis* configuration disrupts their regular organization. However peptide bonds containing a proline residue have a different problem. The amino group is “locked” into a particular shape by the ring and therefore inherently destabilizes both α-helix and β-sheet structures (see above). Prolines are found in the cis configuration ~100 times as often as those between other amino acids. This *cis* configuration leads to a bend or kink in the polypeptide chain. The energy involved in the rotation around a proline bond is much higher than that of a standard peptide bond; so high, that there exist protein catalysts (peptidyl proline isomerases) that facilitate *cis-trans* rotations in such bonds. That said, the polypeptide chain folds as a unit, so increased stability elsewhere in the folded molecule can lead to an otherwise unfavorable local configuration elsewhere.



Hydrophobic R-groups: Many polypeptides and proteins exist primarily in an aqueous (water-based) environment. Yet, a number of their amino acid R-groups are hydrophobic. That means that their interactions with water will decrease the entropy of the system. Very much like the process that drives the assembly of lipids into micelles and bilayers, a typical polypeptide, with hydrophobic R groups along its length will, in aqueous solution, collapse onto itself so as to minimize the interactions of its hydrophobic residues with water. All else being equal minimizing their interaction with water will be thermodynamically favorable (since entropy will increase.) In practice this means that the first step in the folding of a polypeptide as it is synthesized is generally to move hydrophobic R-groups out of contact with water. This drives the collapse of the polypeptide into a compact and dynamic "molten

globule." In contrast where there are no (or few) hydrophobic R groups in the polypeptide, it will tend to adopt an elongated configuration. In contrast, if a protein comes to be embedded within a membrane (and we will briefly consider how this occurs later on), then the hydrophobic R-groups will be located on the surface of the folded polypeptide, so that they interact with the hydrophobic interior of the lipid bilayer. Hopefully this makes sense to you, thermodynamically.

The path to the native (that is, most stable) state is not necessarily a smooth or predetermined one. The folding polypeptide can get "stuck" in a local energy minimum; there may not be enough energy (derived from thermal collisions) for it to get out again. If a polypeptide gets stuck, there are active mechanisms to unfold it and let it try again to reach its native state. This process of partial unfolding is carried out by proteins known as chaperones. There are many types of protein chaperones; some interact with specific polypeptides as they are synthesized and attempt to keep them from getting into trouble, that is, folding in an unproductive way. Others can recognize inappropriately folded polypeptides and couple ATP hydrolysis with polypeptide unfolding, allowing the polypeptide a second (or third or …) chance to fold correctly. In the "simple" eukaryote, the yeast *Saccharomyces cerevisiae*, there are at least 63 distinct molecular chaperones [209]

chaperone video
http://youtu.be/b39698t750c

One class of chaperones are known as "heat shock proteins." The genes that encode these proteins are activated in response to increased temperature (as long as the increase is not so severe that it kills the cell immediately.) Given what you know about polypeptide/protein structure, you should be able to develop a plausible model by which to regulate the expression of heat shock genes. Heat shock proteins recognize unfolded polypeptides which are more likely to be present at higher temperatures. Heat-shock chaperones couple ATP hydrolysis reactions to unfold misfolded polypeptides. They then release the unfolded polypeptides giving them another chance to refold correctly. The chaperone does not directly control the behavior of the polypeptide. You might be asking now, how do chaperones recognize unfolded or abnormally folded proteins? Well unfolded proteins will tend to have hydrophobic regions exposed on the surface. the chaperones can recognize and interact with these regions and then help the polypeptide refold.

Heat shock proteins can be used to help an organism adapt. In classic experiments, when bacteria were grown at temperatures sufficient to turn on the expression of the genes that encode heat shock proteins, the bacteria had a higher survival rate when exposed to elevated temperatures compared to bacteria that had been grown continuously at lower temperature. Heat shock response-mediated survival at higher temperatures is an example of the ability of an organism to adapt to its environment - it is a physiological response. The presence of the heat shock system itself, however, is likely to be a selectable trait, encouraged by temperature variation in the environment. It is the result of evolutionary factors.

Acidic and basic R-groups: Some amino acid R-groups contain carboxylic acid or amino groups and so act as weak acids and bases. Depending on the pH of their environment these groups may be uncharged, positively charged, or negatively charged. Whether a group is charged or uncharged can have a dramatic effect on the structure, and therefore the activity, of a protein. By regulating pH, an

[209] An atlas of chaperone–protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2710862/

organism can modulate the activity of specific proteins. There are, in fact, compartments within eukaryotic cells that are maintained at low pH in part to regulate protein structure and activity. In particular, it is common for internal spaces of vesicles associated with endocytosis to become acidic (through the ATP-dependent pumping of H^+ across their membrane), which activates a number of enzymes involved in the hydrolysis of proteins and nucleic acids.

Subunits and prosthetic groups: Now you might find yourself asking yourself (a philosophically complex task), if most proteins are composed of multiple polypeptides, but polypeptides are synthesized individually, how are proteins assembled in a cytoplasm crowded with other proteins and molecules? This is a process that often involves specific chaperone proteins that bind to the newly synthesized polypeptide and either stabilize its folding, or hold it until it interacts with the other polypeptides it must interact with to form the final, functional protein. The absence of appropriate chaperones can make it difficult to assemble multisubunit proteins into functional proteins *in vitro*.

Many functional proteins also contain non-amino acid-based components, known generically as co-factors. A protein minus its cofactors is known as an apoprotein. Together with its cofactors, it is known as a holoprotein. Generally, without its cofactors, a protein is inactive and often unstable. Cofactors can range in complexity from a single metal ion to quite complex molecules, such as vitamin B12. The retinal group of bacteriorhodopsin and the heme group (with its central iron ion) are co-factors. In general, co-factors are synthesized by various anabolic pathways, and so they represent the activities of a number of genes. So a functional protein can be the direct product of a single gene, many genes, or (indirectly) entire metabolic pathways.

Questions to answer & to ponder

- How does entropy drive protein folding and assembly?
- Why does it matter that rotation around a peptide bond is constrained?
- How might changing the pH of a solution alter a protein's structure and activity?
- What happens to a typical protein if you place it in a hydrophobic solvent?
- What would be your prediction for the structure of a polypeptide if all of its R-groups were hydrophilic?
- How might a chaperone recognize a misfolded polypeptide?
- How would a chaperone facilitate the assembly of a protein composed of multiple polypeptides?
- Summarize the differences in structure between a protein that is soluble in the cytoplasm and one that is buried in the membrane.
- Why might proteins that require co-factors misfold in the absence of the co-factor?
- How might surface hydrophobic R-groups facilitate protein-protein interactions.
- Suggest a reason why cofactors would be necessary in biological systems (proteins)?
- Map the ways that a mutation in a gene encoding a chaperone influence a cell or organism?

Regulating protein localization

Translation of proteins occurs in the cytoplasm, where mature ribosomes are located. Generally, if no information is added, a newly synthesized polypeptide will remain in the cytoplasm. Yet even in the structurally simplest of cells, those of the bacteria and archaea, there is more than one place that a protein may need to be to function correctly: it can remain in the cytoplasm, it can be inserted into the plasma membrane or it may be secreted from the cell. Both membrane and secreted polypeptides must

be inserted into, or pass through, the plasma membrane.

Polypeptides destined for the membrane or for secretion are generally marked by a specific tag, known as a signal sequence. The signal sequence consists of a stretch of hydrophobic amino acids, often at the N-terminus of the polypeptide. As the signal sequence emerges from the ribosomal tunnel it interacts with a signal recognition particle (SRP) - a complex of polypeptides and a structural RNA. The binding of SRP to the signal sequence causes translation to pause. SRP acts as a chaperone for a subset of membrane proteins. The nascent mRNA/ribosome/nascent polypeptide/SRP will find (by diffusion), and attach to, a ribosome/SRP receptor complex on the cytoplasmic surface of the plasma membrane (in bacteria and archaea.) This ribosome/SRP receptor is associated with a polypeptide pore. When the ribosome/SRP complex docks with the receptor, translation resumes and the nascent polypeptide passes through a protein pore and so through the membrane. As the polypeptide emerges on the external, non-cytoplasmic face of the membrane, the signal sequence is generally removed by an enzyme, signal sequence peptidase. If the polypeptide is a membrane protein, it will remain within the membrane. If it is a secreted polypeptide, it will be released into the periplasmic space, that is the region outside of the cell's plasma membrane and inside its cell wall. Other mechanisms can lead to the release of the protein from the cell.

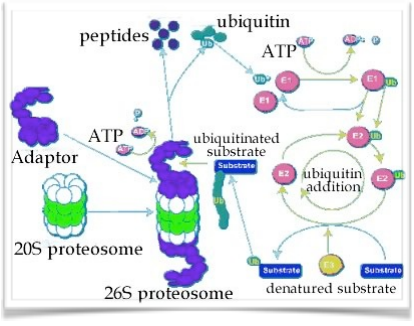
Eukaryotic cells are structurally and topologically more complex than bacterial and archaeal cells; there are more places for a newly synthesized protein to end up. While we will not discuss the details of those processes, one rule of thumb is worth keeping in mind. Generally, in the absence of added information, a newly synthesized polypeptide will end up in the cytoplasm. As in bacteria and archaea, a eukaryotic polypeptides destined for secretion or insertion into the cell's plasma membrane or internal membrane systems (that is the endoplasmic reticulum) are directed to their final location by a signal sequence/SRP system. Proteins that must function in the nucleus generally get there because they have a nuclear localization sequence, other proteins are actively excluded from the nucleus using a nuclear exclusion sequence (see above). Likewise, other localization signals and sequences are used to direct proteins to other intracellular compartments, including mitochondria and chloroplasts. While details of these targeting systems are beyond the scope of this course, you can assume that each specific targeting event requires signals, receptors, and various mechanisms that drive what are often thermodynamically unfavorable reactions.

Regulating protein activity

Proteins act through their interactions with other molecules. Catalytic proteins (enzymes) interact with substrate molecules; these interactions lower the activation energy of the reaction's rate limiting step, leading to an increase in the overall reaction rate. At the same time, cells and organisms are not static. They must regulate which proteins they produce, the final concentrations of those proteins within the cell (or organism), how active those proteins are, and where those proteins are located. It is primarily by altering proteins (and so indirectly gene expression) that cells (and organisms) adapt to changes in their environment.

A protein's activity can be regulated in a number of ways. The first and most obvious is to control the total number of protein molecules present within the system. Let us assume that once synthesized, a protein is fully active. With this simplifying assumption, the total concentration of a protein in a system [P_{sys}] is proportional to the rate of that protein's synthesis (dSynthesis/dt) minus the rate of that protein's degradation (dDegradation/dt). dt indicates per unit time. The combination of these two processes, synthesis and degradation, determines the protein's half-life. The degradation of proteins is mediated by a special class of enzymes (proteins) known as proteases. Proteases cleave peptide bonds via hydrolysis reactions. Proteases that cleave a polypeptide chain internally are known as endoproteases - they generate two polypeptides. Those that hydrolyze polypeptides from one end or the other, to release one or two amino acids at a time, are known as exoproteases. Proteases also can act more specifically, recognizing and removing a specific part of a protein in order to activate it or to inactivate it, or to control where it is found in a cell. For example, nuclear proteins become localized to the nucleus (typically) because they contain a nuclear localization sequence or they can be excluded because they contain a nuclear exclusion sequence. For these sequences to work they have to be able to interact with the transport machinery associated with the nuclear pores; but the protein may be folded so that they are hidden. Changes in a protein's structure can reveal or hide such targeting sequences, thereby altering the protein's distribution within the cell and its activity. For example, many proteins are originally synthesized in a longer, and inactive "pro-form". To become active the pro-peptide must be removed - it is cut by an endoprotease. This proteolytic processing activates the protein. Proteolytic processing is itself often regulated (see below).

Controlling protein levels: Clearly the amount of a protein within a cell (or organism) is a function of the number of mRNAs encoding the protein, the rate that these mRNAs are recognized and translated, and the rate at which functional protein is formed, which in turn depends upon folding rates and their efficiency. It is generally the case that once translation begins, it continues at a more or less constant rate. In the bacterium *E. coli*, the rate of translation at 37ºC is about 15 amino acids per second. The translation of a polypeptide of 1500 amino acids therefore takes about 100 seconds. After translation, folding and, in multisubunit proteins, assembly, the protein will function (assuming that it is active) until it is degraded.



Many proteins within the cell are necessary all of the time. Such proteins are considered "constitutive." Protein degradation is particularly important for controlling the levels of "regulated" proteins, whose presence or concentration within the cell may lead to unwanted effects in certain situations. The regulated degradation of a protein typically begins when the protein is specifically marked for degradation. This is an active and highly regulated process, involving ATP hydrolysis and a multi-subunit complex known as the proteosome. The proteosome degrades the polypeptide into small peptides and amino acids that can be recycled. As a mechanism for regulating protein activity, however, degradation has a serious drawback, it is irreversible. Since both a protein's synthesis and degradation can be regulated, its half-life can be regulated.

Allosteric regulation

A reversible form of regulation is known as allosteric regulation, where regulatory molecules bind reversibly to the protein altering its conformation, which in turn alters the protein's activity and can alter its location within the cell and its half-life. Such allosteric effectors are not covalently attached to the protein and can act either positively or negatively. The nature of such factors is broad, they can be a small molecule or another protein. What is important is that the allosteric binding site is distinct from the enzyme's catalytic site. In fact allosteric means other site. Because allosteric regulators do not bind to the same site on the protein as the substrate, changing substrate concentration generally does not alter their effects.

Of course there are other types of regulation as well. A molecule may bind to and block the active site of an enzyme. If this binding is reversible, then increasing the amount of substrate can overcome the inhibition. An inhibitor of this type is known as a competitive inhibitor. In some cases, the inhibitor chemically reacts with the enzyme, forming a covalent bond. This type of inhibitor is essentially irreversible, so that increasing substrate concentration does not overcome inhibition. These are therefore known as non-competitive inhibitors. Allosteric effectors are also non-competitive, since they do not compete with substrate for binding to the active site. That said, binding of substrate could, in theory, change the affinity of the protein for its allosteric effectors, just as binding of the allosteric effector changes the binding affinity of the protein for the substrate.

Post-translational regulation

Proteins may be modified after synthesis - this process is known as post-translational modification. A number of post-translational modifications have been found to occur within cells. In general where a protein can be modified it can also be unmodified. The exception, of course, is when the modification involves protein degradation. The first, and most common type of modification we will consider involves the covalent addition of specific groups to specific amino acid side chains on the protein - these groups can range from phosphate groups (phosphorylation), an acetate group (acetylation), the attachment of lipid/hydrophobic groups (lipid modification), or carbohydrates (glycosylation). Such post-translational modifications are generally reversible, one enzyme adds the modifying group and another can remove it. For example, proteins are phosphorylated by enzymes known as protein kinases, while protein phosphotases remove these phosphate groups. Post-translational modifications act in much the same way as do allosteric effectors, they modify the structure and, in turn, the activity of the polypeptide to which they are attached. They can also modify a protein's interactions with other proteins, the protein's localization within the cell, or its stability.

Questions to answer & to ponder

- A protein binds an allosteric regulator - what happens to the protein?
- How is the post-translational modification of a protein like allosteric regulation? how is it different?
- Why are enzymes required for post-translational modification?
- Why is a negative allosteric regulator not considered a "competitive" inhibitor?
- Why do post-translational modifications (and their reversals) require energy?
- How does a signal sequence influence translation?
- How would a cell recover from the effects of an irreversible, non-competitive inhibitor?
- Why might a cell want a specific protein to have a short half-life?

- What would happen if you somehow put a signaling sequence at the beginning of a normally cytoplasmic polypeptide?
- Draw out the factors and their interactions that control the half-life, activity, and location of a particular protein within a biological system.

Diseases of folding and misfolding

If a functional protein is in its native (or natural) state, a dysfunctional misfolded protein is said to be denatured. It does not take much of a perturbation to unfold or denature most proteins. In fact, under normal conditions, proteins often become partially denatured spontaneously, normally these are either refolded (often with the help of chaperone proteins) or degraded (through the action of proteosomes and proteases). A number of diseases, however, arise from protein misfolding.

Kuru was among the first of these protein misfolding diseases to be identified. Beginning in the 1950s, D. Carleton Gadjusek (1923 – 2008)[210] studied a neurological disorder common among the Fore people of New Guinea. The symptoms of kuru, which means "trembling with fear", are similar to those of scrapie, a disease of sheep, and variant Creutzfeld-Jakob disease (vCJD) in humans. Among the Fore people, kuru was linked to the ritual eating of the dead. Since this practice has ended, the disease has disappeared. The cause of kuru, scrapie and vCJD appears to be the presence of an abnormal form of a normal protein, known as a prion. We can think of prions as a type of anti-chaperone. The idea of proteins as infectious agents was championed by Stan Prusiner, who was awarded the Nobel Prize in Medicine in 1997.[211]

The protein responsible for kuru and scrapie is known as PrPc. It normally exists in a largely α-helical form. There is a second, abnormal form of the protein, PrPsc for scrapie; whose structure is primarily of β-sheet. The two polypeptides have the same primary sequence. PrPsc acts as an anti-chaperone, catalyzing the transformation of PrPc into PrPsc. Once initiated, this leads to a chain reaction and the accumulation of PrPsc. As it accumulates, PrPsc assembles into rod-shaped aggregates that appear to damage cells. When this process occurs within the cells of the central nervous system it leads to severe neurological defects. There is no natural defense, since the protein responsible is a normal protein.

Disease transmission: When the Fore ate the brains of their beloved ancestors, they inadvertently introduced the PrPsc protein into their bodies. Genetic studies indicate that early humans evolved resistance to prion diseases, suggesting that cannibalism might have been an important selective factor during human evolution. Since cannibalism is not nearly as common today, how does anyone get such diseases in the modern world? There are rare cases of iatrogenic transmission, that is, where the disease is caused by faulty medical practice, for example through the use of contaminated surgical instruments or when diseased tissue is used for transplantation.

But where did people get the disease originally? Since the disease is caused by the formation of PrPsc, any event that leads to PrPsc formation could cause the disease. Normally, the formation of

---

[210] Carleton Gajdusek: http://www.theguardian.com/science/2009/feb/25/carleton-gajdusek-obituary

[211] Stanley Prusiner: 'A Nobel prize doesn't wipe the skepticism away': http://www.theguardian.com/science/2014/may/25/stanley-prusiner-neurologist-nobel-doesnt-wipe-scepticism-away and http://youtu.be/yzDQ8WgFB_U

PrPsc from PrPc is very rare. We all have PrPc but very few of us spontaneously develop kuru-like symptoms. There are, however, mutations in the gene that encodes PrPc that greatly enhance the frequency of the PrPc → PrPsc conversion. Such mutations may be inherited (genetic) or may occur during the life of an organism (sporadic). Fatal familial insomnia (FFI) is due to the inheritance of a mutation in the *PRNP* gene, which encodes PrPc. This mutation changes the normal aspartic acid at position 178 of the PrPc protein to an asparagine. When combined with a second mutation in the *PRNP* gene at position 129, the FFI mutation leads to Creutzfeld-Jacob disease (CJD). If one were to eat the brain of a person with FFI or CJD one might well develop a prion disease.

So why do PrPsc aggregates accumulate? To cut a peptide bond, a protease must position the target peptide bond within its catalytic active site. If the target protein's peptide bonds do not fit into the active site, they cannot be cut. Because of their structure, PrPsc aggregates are highly resistant to proteolysis. They gradually accumulate over many years, a fact that may explain the late onset of PrP-based diseases.

Why do harmful alleles persist?

At this point, you might well ask yourself, given the effectiveness of natural selection, why do alleles that produce severe diseases exist at all? There are a number of possible scenarios. One is that a new mutation arose spontaneously, either in the germ line of the organism's parents or early in the development of the organism itself, and that it will disappear from the population with the death of the organism. The prevalence of the disease will then reflect the rate at which such pathogenic mutations occur. The second, more complex reason involves the fact that many organisms carry two copies of each gene (they are diploid), and that carrying a single copy of the allele might either have no discernible effect on the organism's reproductive success or, in some cases, might even lead to an increase in reproductive success. In this case, the allele will be subject to positive selection, that is, it will increase in frequency. This increase will continue until the number of individuals carrying the allele reaches a point where the number of offspring with two copies of the mutant (pathogenic) allele becomes significant. These individuals (and the alleles they carry) are subject to strong negative selection. We will therefore arrive at a steady state population where the effects of positive selection (on individuals carrying one copy of the allele) will be balanced by effects of negative selection on individuals that carry two copies of the allele. You could model this behavior in an attempt to predict the steady state allele frequency by considering the sizes of the positive and negative effects and the probability that a mating will produce an organism with one (a heterozygote) or two (a homozygote) copies of the allele.

Generally the process of selection occurs gradually, over many (hundreds to thousands) of generations, but (of course) the rate depends on the strength of the positive and negative effects of a particular allele on reproductive success. As selection acts, and the population changes, the degree to which a particular trait influences reproductive success can also change. The effects of selection are themselves not static, but evolve. For example, a trait that is beneficial when rare may be less beneficial when common. New mutations that appear in the same or different genes can further influence the trait, and so how the population will change over time. For example, alleles that were

"neutral" or without effect in the presence of certain alleles at other genes (known as the genetic background) can have effects when moved into another genetic background. A (now) classic example of this effect was described by studies on the laboratory evolution of the bacterium *Escherichia coli*. A mutation with little apparent effect occurred in one lineage and its presence made possible the emergence of a new trait (the ability to use citrate for food) about 20,000 generations later.[212] We will return to how this works exactly toward the end of the course, but what is important here is that it is the organism (and its traits and all its alleles) that is "selected". Only in rare cases of extremely strong positive or negative selection, does it make sense to say that specific alleles are selected.

Questions to answer & to ponder

- How does the presence of PrPsc lead to the change in the structure of PrPc?
- Why is it, do you think, that FFI and CJD are late onset diseases?
- Which do you think would be more susceptible to proteolytic degradation, a compact or an extended polypeptide?

[212] Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli: http://www.pnas.org/content/105/23/7899.long

9. Genomes, genes, and regulatory networks

In which we consider the dynamics of genes and genomes, and how genome dynamics leads to families of genes and facilitates evolutionary change. We consider how DNA is organized within a cell and how its organization influences gene expression. Finally we consider the behavior of regulatory networks at the molecular level and the role of noise in producing interesting behaviors.

At this point we have introduced genes, DNA, and proteins, but we have left unresolved a number of important questions. These include how genomes are organized, how they evolve, how new genes and alleles are generated, and how they work together to produce the various behaviors that organisms display.[213] This includes trendy topics such as epigenetics (which is probably less interesting than most suppose) and the rather complex molecular and cellular level processes behind even the simplest behaviors. The details, where known–and often they are not–are beyond the scope of this course, but the basic themes are relatively straightforward, although it does takes some practice to master this type of thinking. The key is to keep calm and analyze on!

Genomes and their organization

Genomes are characterized by two complementary metrics, the number of base pairs of DNA and the number of genes present within this DNA. The number of base pairs is easier to measure, we can count them. This can, however, led to a mistake conclusion, namely that the number of base pairs of DNA within the genome of a particular species, organism, or even tissue within an organism is fixed and constant. In fact genomes are dynamic, something that we will return to shortly.

The genome of an organism (and generally the cells of which it is composed) consists of one or more DNA molecules. When we talk about genome size we are talking about the total number of base pairs present in all of these DNA molecules added together. The organism with the largest known genome is the plant *Paris japonica;* its genome is estimated to be ~150,000 x 10^6 (millions of) base pairs.[214] In contrast the (haploid) human genome consists of ~3,200 x 10^6 base pairs of DNA. The relatively small genome size of birds (~1,450 x 10^6 base pairs) is thought to be due to the smaller genome size of their dinosaurian ancestors.[215] That said there are interesting organisms that suggest that in some cases, natural selection can act to dramatically increase or decrease genome size without changing gene number. For example, the carnivorous bladderwort *Utricularia gibba*, has a genome of

---

[213] Gene Duplication: The Genomic Trade in Spare Parts: http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0020206

[214] A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. http://www.ncbi.nlm.nih.gov/pubmed/24299166

[215] Origin of avian genome size and structure in non-avian dinosaurs: http://www.ncbi.nlm.nih.gov/pubmed/17344851

~80 x 10^6 base pairs and ~28,000 genes, significantly fewer base pairs of DNA, but apparently more genes than humans.

Very much smaller genomes are found in prokaryotes, typically their genomes are a few millions of base pairs in length. The smallest genomes occur in organisms that are obligate parasites and endosymbionts. For example the bacterium *Mycoplasma genitalium*, the cause of non-gonococcal urethritis, contains ~0.58 x 10^6 base pairs of DNA, which encodes ~500 distinct genes. An even smaller genome is found in the obligate endosymbiont *Carsonella ruddii;* it has 159,662 (~0.16 x 10^6) base pairs of DNA encoding "182 ORFs (open reading frames or genes), 164 (90%) overlap with at least one of the two adjacent ORFs".[216] Eukaryotic mitochondria and chloroplasts, derived as they are from endosymbionts, have very small genomes. Typically mitochondrial genomes are ~16,000 base pairs in length and contain ~40 genes, while chloroplasts genomes are larger, ~120,000–170,000 base pairs in length, and ~100 genes. Most of the gene present in the original endosymbionts appear to have either been lost or transferred to the host cell's nucleus. This illustrates a theme that we will return to, namely that genomes are not static. In fact, it is their dynamic nature that makes significant evolutionary change possible.

An interesting question is what is the minimal number of genes that an organism needs. Here we have to look at free living organisms, rather than parasites or endosymbionts, since they can rely on genes within their hosts. A common approach is to use mutagenesis to generate non-functioning (amorphic) versions of genes. One can then count the number of essential genes within a genome, that is, genes whose functioning is absolutely required for life. One complication is that different sets of genes may be essential in different environments, but we will ignore that for now. In one such lethal mutagenesis study Lewis et al found that 382 of the genes in *Mycoplasma genitalium* are essential; of these ~28% had no known function.[217]

A technical aside: transposons

In their study, Lewis et al used what is known as a "mobile genetic element" or transposon to generate mutations. A transposon is a piece of DNA that can move (jump) from place to place in the genome.[218] The geneticist Barbara McClintock (1902 –1992) first identified transposons in the course of studies of maize (*Zea mays*).[219] There are two basic types of transposons. Type II transposons consist of DNA sequence that encodes proteins that enable it to excise itself from a larger (host) DNA molecule, and insert into another site within the host cell's genome. The second type (type I) can make copies of themselves, through an RNA intermediate, and this copy can be inserted into the host genome, leaving the original copy in place. Both types of transposon encode the proteins required to recognize the transposon sequence and mediated its movement or replication, and subsequent

[216] The 160-Kilobase Genome of the Bacterial Endosymbiont Carsonella: http://www.ncbi.nlm.nih.gov/pubmed/17038615

[217] Essential genes of a minimal bacterium: http://www.pnas.org/content/103/2/425.full

[218] Transposons: The Jumping Genes: http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518

[219] Barbara McClintock: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1983/mcclintock-bio.html

inserting into new sites. If the transposon sequence is inserted into a gene, it can create a null or amorphic mutation in that gene by disrupting the gene's regulatory or coding sequences. Transposons are only one of a class of DNA molecules that can act as molecular parasites, something neither Darwin nor the founders of genetics ever anticipated, but which makes sense from a molecular perspective, once the ability to replicate, cut, and join DNA molecules had evolved. These various activities are associated with the repair of mutations involving single and double stranded breaks in DNA, but apparently they also made DNA-based parasites possible. If a host cell infected with a transposon replicates, it also replicates the transposon sequence, which will be inherited by the offspring of the cell. This is a process known as vertical transmission, a topic we will return to shortly.

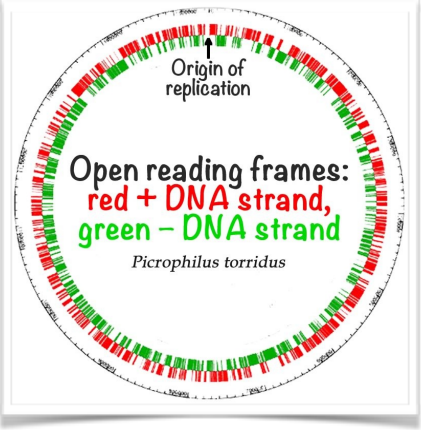
Because transposons do not normally encode essential functions, mutations can inhibit the various molecular components involved in their replication and jumping within a genome. They can be inactivated (killed) by random mutation, and there is no (immediate) selective advantage to maintaining them. If you remember back to our discussion of DNA, the human (and many other types of genomes), contain multiple copies of specific sequences. Subsequent analyses have revealed that these represent "dead" forms of transposons and related DNA-based molecular parasites. It is estimated that the human genome contains ~50,000 copies of the Alu type transposon, and that ~50% of the human genome consists of dead transposons. It is probably not too surprising then that there is movement within genomes during the course of an organism's life time.

Genes along chromosomes

Genomes are typically divided into chromosomes, which are distinct DNA molecules together with all of the other molecules that associate with them in the cell. These associated molecules, primarily proteins, are involved in organizing the DNA, recognizing genes and initiating or inhibiting their expression. An organism can have one chromosome or many. Each chromosome has a unique sequence and specific genes are organized in the same order along a particular chromosome. For example, your chromosome 4 will have the same genes in the same sequence along its length as those of all of the people you ever met. The difference is that you are likely to have different versions of those genes, different alleles. In this light, most macroscopic organisms are diploid (including humans), and so have two copies of each chromosome, with the exception of the chromosomes (X and Y) that determine sex. So you may have two different alleles for any particular gene. Most of these sequence differences will have absolutely no discernible effect on your molecular, physiological, or behavioral processes. However, some will have an effect, and these form the basis of genetic differences between organisms. That said, their effects will be influenced by the rest of your genome, so for most traits there is no simple link between genotype and phenotype.

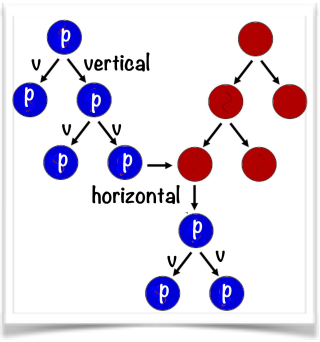
In humans, only ~5% of the total genomic DNA is involved in encoding polypeptides. The amount of DNA used to regulate gene expression is more difficult to estimate, but it is clear that lots of the genome (including the 50% that includes dead transposons) is not directly functional. That said, gene organization can be quite complex. We can see an example of this complexity by looking at organisms with more "streamlined" genomes. While humans have an estimated ~25,000 genes in ~3.2 x 10^9 base pairs of DNA (about 1 gene per 128,000 base pairs of DNA), the single circular chromosome of the bacterium *E. coli (K-12 strain)* contains 4,377 genes in 4,639,221 base pairs of DNA, of which

4,290 encode polypeptides and the rest RNAs.[220] That is about one gene per 1000 base pairs of DNA.



In prokaryotes and eukaryotes, genes can be on either strand of the DNA molecule, typically referred (rather arbitrarily) as the "+" and the "–" strands of the molecule. Given that the strands are anti-parallel, a gene on the + strand would run in the opposite direction from a gene on the - strand. We can illustrate this situation using the euryarchaea *Picrophilus torridus.* This archaea organism can grow under extreme conditions, around pH 0 and up to 65°C. Its genome is 1,545,900 base pairs of DNA in length and it encodes 1,535 polypeptides (open reading frames), distributed fairly equally on the + and – strands.[221]

While most prokaryotic genes are located within a single major chromosome, the situation is complicated by the presence of separate, smaller circular DNA molecules within the cell known as plasmids. In contrast to the organism's chromosome, plasmids can (generally) be gained or lost. That said, because plasmids contain genes, it is possible for an organism to become dependent upon or addicted to a plasmid. For example, a plasmid can carry a gene that makes its host resistant to certain antibiotics. Given that most antibiotics have their origins as molecules made by one organism to kill or inhibit the growth of others, if an organism is living in the presence of an antibiotic, losing a plasmid that contains the appropriate antibiotic resistance gene will be lethal. Alternatively, plasmids can act selfishly. For example, suppose a plasmid carries the genes encoding an "addiction module" (which we discussed previously.) When the plasmid is present, both toxin and anti-toxin are made. If, however, the plasmid is lost, the synthesis of the unstable anti-toxin ceases, while the stable toxin persists, becomes active (uninhibited), and kills the host. As you can begin to suspect, the ecological complexities of plasmids and their hosts are not simple.



Like the host chromosome plasmids, have their own "origin of replication" sequence required for DNA synthesis, and can therefore replicate independently. Plasmids can be transferred from cell to cell either when the cell divides (vertical transmission) or between "unrelated" cells through what is known as horizontal transmission. If you think back to Griffith's experiments on pneumonia, the ability of the DNA from dead S-type bacteria to transform R-type bacteria (and make them pathogenic) is an example of horizontal transmission.

Naturally occurring horizontal gene transfer mechanisms

Many horizontal transmission mechanisms are regulated by social and/or ecological interactions between organisms.[222] It is important to note that the mechanisms involved are quite complex, one

---

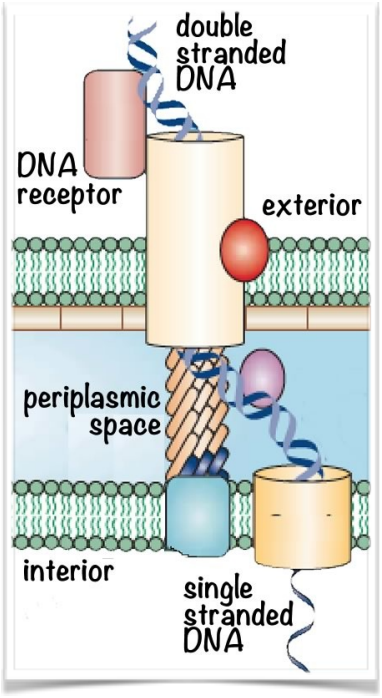
[220]Genome Sizes: http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/G/GenomeSizes.html

[221] Genome sequence of Picrophilus torridus and its implications for life around pH 0: http://www.pnas.org/content/101/24/9091.full

[222] DNA uptake during bacterial transformation: http://www.ncbi.nlm.nih.gov/pubmed/15083159

could easily imagine an entire course focused on this topic. So keep in mind that we are only introducing the broad features of these systems. Also, we want to be clear about the various mechanisms of DNA uptake. First it is worth noting that when organisms die their DNA can be eaten and become a source of carbon, nitrogen, and phosphorus. Alternatively, a nucleotide sequence of a DNA molecule could be integrated into another organism's genome, resulting in the acquisition of information developed (evolved) within another organismic lineage. The study of these natural DNA import systems has identified very specific mechanisms for DNA transfer. For example some organisms use a system that will preferentially import DNA molecules that are derived from organisms of the same or closely related types. You can probably even imagine how they do this – they recognize species specific "DNA uptake sequences." The various mechanisms of horizontal gene transfer, unsuspected until relatively recently, have had profound influences on evolutionary processes. It turns out that a population of organisms does not have to "invent" all of its own genes, but can adopt genes generated (by evolutionary mechanisms) by other organisms in other environments for other purposes. So the question is, what advantages might such information uptake systems convey, and (on the darker side), what dangers do they make possible?

Transformation



There are well established methods used in genetic engineering to enhance the ability of bacteria to take up plasmids from their environment.[223] We, however, will focus on the natural processes associated with the horizontal transfer of DNA molecules from the environment into a cell, or from cell to cell. The first of these processes is known as transformation. It is an active process that involves a number of components, encoded by genes that can be on or off depending upon environmental conditions. Consider a type of bacteria that can import DNA from its environment. If, however, the density of bacteria is low, then there will be little DNA to import, and it may not be worth the effort to express the genes and synthesize the proteins involved in the transformation machinery. In fact, bacteria can sense the density of organisms in their environment using a process called quorum sensing, which we will consider in more detail later. Bacteria use quorum sensing systems to synthesize the DNA uptake system when conditions warrant, apparently by activating a specific σ factor (see above). When present in a crowded environment, the quorum sensing system turns on the expression of the DNA update system and generate cells competent for transformation.

Here we outline the process in a Gram-negative bacteria (which are identified by how they stain with crystal violet) but a similar mechanism is used in Gram-positive bacteria.[224] Double-stranded DNA binds to the bacterial cell's surface through a variety of DNA receptors. In some cases these receptors

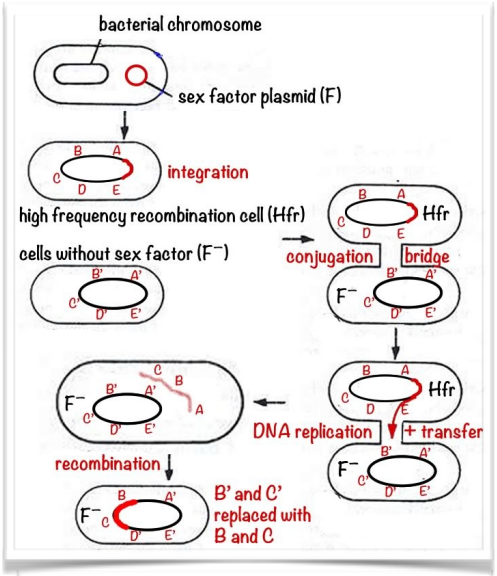
[223] Making Calcium Competent (bacterial) Cells: http://mcb.berkeley.edu/labs/krantz/protocols/calcium_comp_cells.pdf

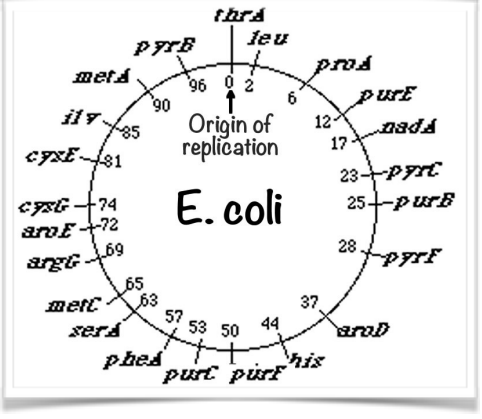
[224] Gram positive bacteria have a single membrane, the plasma membrane, surrounded by a think layer of protein and carbohydrate (peptidoglyan). http://en.wikipedia.org/wiki/Gram-positive_bacteria

bind specific DNA sequences, in others they bind DNA generically (that is any DNA sequence). As shown, Gram negative bacteria have two lipid membranes, an outer one and an inner (plasma) membrane, with a periplasmic space in between. In an ATP-hydrolysis coupled reaction, DNA bound to the exterior surface of the bacterium is moved, through a protein pore through the outer membrane and into the periplasmic space, where it is passed to the DNA channel protein, Here one strand is degraded by a nuclease while the other moves through the channel into the cytoplasm of the cell in a 5' to 3' direction. Once inside the cell, the DNA associates with specific single-stranded DNA binding proteins and, by a process known as homologous recombination, is inserted into the host genome.[225] While the molecular details of this process are best addressed elsewhere, what is key is that transformation enables a cell to decide whether or not to take up foreign DNA and to add those DNA sequences to its genome.

Conjugation and transduction

There are two other processes that can lead to horizontal gene transfer in bacteria: conjugation and transduction. In contrast to transformation, these processes "force" DNA into what may be a reluctant cell. In the process of conjugation, we can distinguish between two types of bacterial cells (of the same species). One contains a plasmid known as the sex factor (F) plasmid. These are known as a Hfr (high frequency recombination) cells. This plasmid contains the genes needed to transfer a copy of its DNA into a cell that lacks an F-plasmid, a so called F^- cell. Occasionally, the F-plasmid can integrate into the host cell chromosome and when this happens, the F-plasmid mediated system can transfer host cell genes (in addition to plasmid genes) into an F^- cell. To help make things a little simpler, we will refer to the Hfr cell as the DNA donor and F^- cells as the DNA recipients.





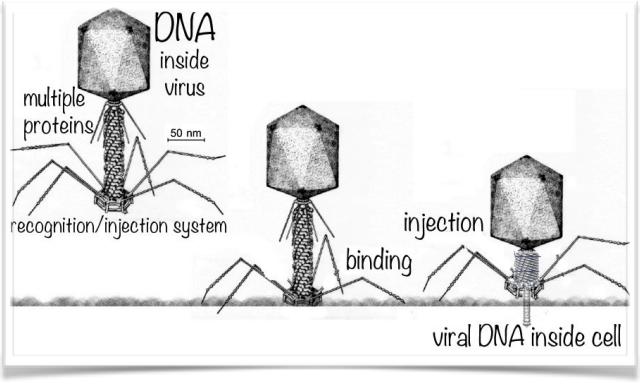
To initiate conjugation, the Hfr cell makes a physical bridge to the F– cell. A break in the donor DNA initiates a process by which single stranded DNA is synthesized and moved into the recipient (F^-) cell. The amount of DNA transported is determined largely by how long the transporter bridge remains intact. It takes about 100 minutes to transfer the entire donor chromosome from an Hfr to an F- cell. Once inside the F^- cell, the DNA is integrated into the recipient's chromosome, replacing the recipient's versions of the genes transferred (through a process of homologous recombination, similar to that used in transfection). Using Hfr strains with integrated F^- plasmids carrying different alleles of various genes, and by controlling the duration of conjugation (separating the cells by placing them in a kitchen

[225] Bacterial transformation: distribution, shared mechanisms and divergent control.: http://www.ncbi.nlm.nih.gov/pubmed/24509783

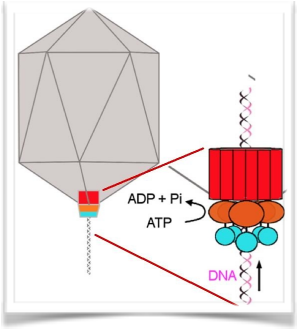
blender), experimenters were able to determined the order of genes along the chromosome. The result was the discovery that related organisms had the same genes arranged in the same order. The typical drawing of the circular bacterial chromosome is like a clock going from 0 to 100, with the genes placed in their respective positions, based on the time it takes to transfer them (in minutes). This is an example of synteny, that is the conservation of gene order along a chromosome.[226] We will return to synteny soon.

If the entire F-plasmid sequence is transferred, the original F^- cell becomes an Hfr cell. If the Hfr cell loses the F-plasmid sequence it will revert to a F^- state. The end result of the conjugation process is similar to that obtained in sexual reproduction in eukaryotes (see below), namely the original F^- cell now has a genome derived in part from itself and from the "donor" Hfr strain cell.

Transduction



The final form of horizontal gene transfer is one that involves the behavior of viruses. The structure and behavior of viruses is an extremely complex topic, the details being well beyond us here, but we can consider them generally as nucleic acid transport machines. Viruses are completely dependent for their replication on a host cell, they have no active metabolic processes and so are not really alive in any meaningful sense, although they can certainly be rendered non-infectious. The simplest viruses contain a nucleic acid genome and a protein-based transport and delivery system. We will consider a typical bacterial virus, known as a bacteriophage or bacteria eater, which uses a double stranded DNA molecule to encode its genetic information. The bacterial virus we consider here, the T4 bacteriophage, looks complex and it is (other viruses are much simpler). T4 phage (short for bacteriophage) have a ~169,000 base pair double-stranded DNA genome that encodes 289 polypeptides.[227] The assembled virus has an icosahedral head that contains the DNA molecule and a tail assembly that recognizes and binds to target cells. Once a suitable host is found, the tail domain attaches and contracts, like a syringe. The DNA emerges from the bacteriophage and enters the (now) infected cell. Genes within the phage genome are expressed leading to the replication of the phage DNA molecule and the fragmentation of the host cell's genome. The next round of infection involves the assembly of new phage heads, DNA is packed into these heads by a protein-based DNA pump, the pump is driven by coupling to an ATP hydrolysis complex.[228] In the course of packaging virus DNA, occasionally the system will make a mistake and package undigested host DNA. When such a phage particle infects another



[226] Synteny: http://en.wikipedia.org/wiki/Synteny

[227] http://en.wikipedia.org/wiki/Bacteriophage_T4

[228] The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces: http://www.ncbi.nlm.nih.gov/pubmed/19109896

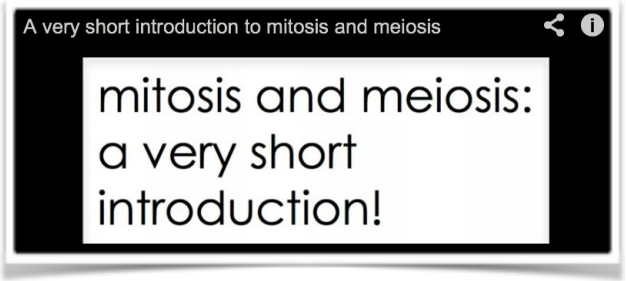
cell, it injects that cell with a DNA fragment derived from the previous host. Of course, this mispackaged DNA may not contain the genes the virus needs to make a new virus or to kill the host. The transferred DNA can be inserted into the newly infected host cell genome, with the end result being similar to that discussed previously for transformation and conjugation. DNA from one organism is delivered to another, horizontally rather than vertically.

Sexual reproduction

The other major mechanism for shuffling genes is through sexual reproduction. In contrast to prokaryotes, eukaryotes typically have multiple chromosomes. Chromosomes are composed of both single linear double-stranded DNA molecules and associated proteins, but for our purposes only the DNA molecules are important. Different chromosomes can be distinguished by the genes they contain, as well as the length of their DNA molecules. Typically the chromosomes of an organism are numbered from the largest to the smallest. Humans, for example, have 23 pairs of chromosomes. In humans the largest of these chromosomes, chromosome 1, contains about 250 million base pairs of DNA and over 2000 polypeptide-encoding genes, while the smaller chromosome 22 contains about 52 million based pairs of DNA and around 500 polypeptide encoding genes.[229]

In sexually reproducing organisms, somatic cells are typically diploid, that is, they contain two copies of each chromosome rather than one. The two copies of the same chromosomes are known as homologs of each other or homologous chromosomes. As we will now describe, one of these homologous chromosomes is inherited from the maternal parent and the other from the paternal parent. Aside from allelic differences the two homologous chromosomes are generally very similar, the exception are the so called sex chromosomes. While the sex of an organism can be determined in various ways in different types of organisms, in humans (and most mammals, birds and reptiles) the phenotypic sex of an individual is determined by which sex chromosomes their cell's contain. In humans the 23rd chromosome comes in two forms, known as X and Y. An XX individual typically develops into a female, while an XY individual develops into a male.

The sexual reproductive cycle involves two distinct mechanisms of allele shuffling. The cells of the body that take an integral part in sexual reproduction (of course, the entire body generally takes part in sex, but we are trying to stay simple here) are known as germ line cells. A germ line cell is diploid, but through a process known as meiosis it can produces as many as four haploid cells, known as gametes. A first step in this process is the replication of the cell's DNA; each individual chromosome will be duplicated. Instead of separating from one another, these replicated DNA molecules remain attached through associated proteins, at a structure known as the centromere. In standard, asexual division (known as mitosis), each replicated chromosome interacts



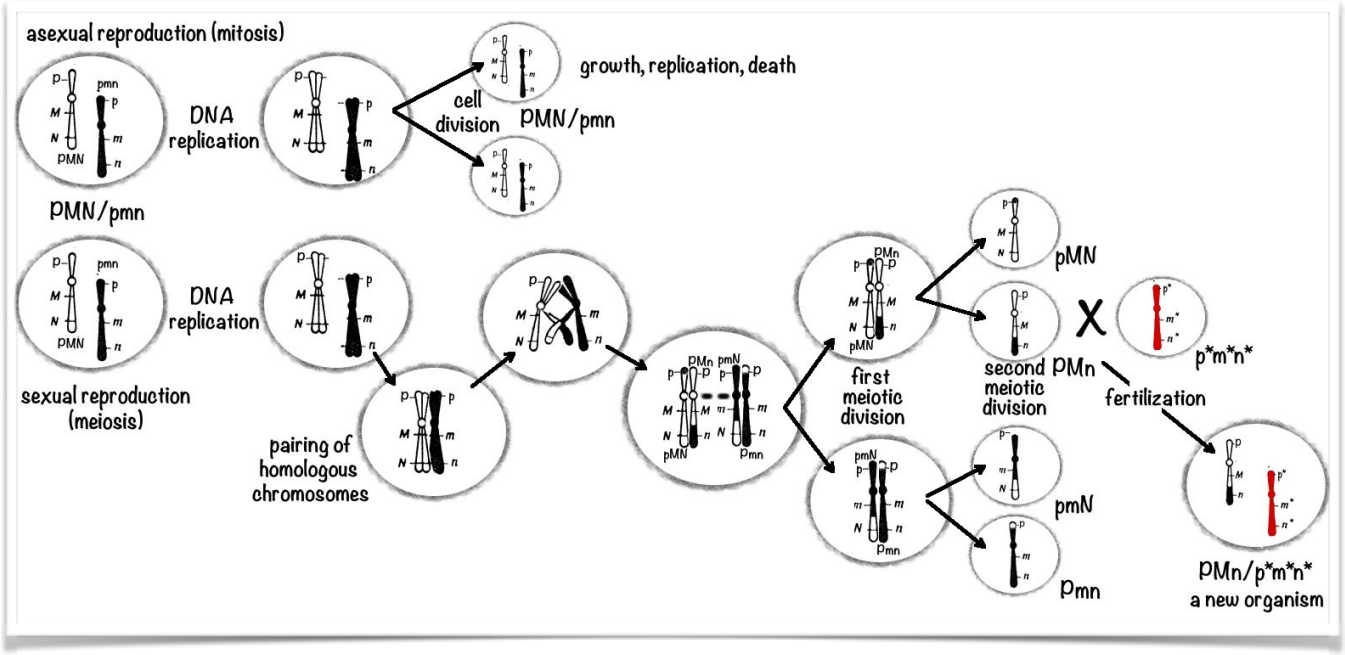
http://youtu.be/i4jTu7IN65k

[229] We are only discussing polypeptide-encoding genes because it remains unclear whether (and which) other transcribed regions are genes, or physiologically significant.

independently with a molecular machine (the mitotic spindle) whose role is to send one copy of each chromosome to each of the two daughter cells that will be formed. During mitosis (asexual reproduction) a diploid cell produces a diploid cell, and nothing about the genome has changed. The cells that are formed are fated to be part of the original organism.

In contrast, the purpose of meiosis is to produce a new organism that will have a genome distinct from that of either of its parents (even in the case of hermaphrodites, in which one organism acts as both mother and father!) To accomplish this, chromosomes are shuffled in various ways. First, remember that the diploid cell contains two sets of chromosomes, one set from the mother and a set from the father. In meiosis (sexual reproduction), the process diverges from mitosis after the chromosomes are duplicated. Instead of one copy of each chromosome (both maternal and paternal) being delivered to two daughter cells, the homologous duplicated chromosomes pair up with one another. This pairing is based on the fact that the DNA sequences along each homologous chromosome, while not identical, are extremely similar. They are syntenic, that is, they have the same genes in the same order. In contrast, the DNA of two different, that is, non-homologous chromosomes, say human chromosomes 1 and 8 have many sequence differences and contain different genes. Based on their sequence similarity, the replicated maternal and paternal homologous chromosomes line up with one another into a structure with four DNA strands. At this point, at positions more or less random along the length of the chromosome, there are double strand breaks in two adjacent DNA molecules. The DNA molecules can then be rejoined, either back to themselves (maternal to maternal, paternal to paternal) or with another DNA molecule (maternal to paternal, or paternal to maternal). Typically, multiple “crossing-over” events occur along the length of each set of paired, replicated homologous chromosomes. At the first meiotic division, the duplicated maternal and paternal chromosomes remain attached at their centromeres, but because of crossing over these will, in fact, be different from the original chromosomes. Each of the two daughter cells receives either the replicated maternal or paternal chromosome centromere region. Each of the organism’s chromosomes are segregated at random. For an organism with 23 different chromosomes, that generates 2^{23} possible different daughter cells. There is no DNA replication before the second meiotic division. During this division, the two daughter cells each receive a copy of one and only one homologous chromosome. The four cells that are generated by meiosis are known as gametes (or at least are potential gametes) and they are haploid. In the human, they each contain one and only one copy of each of the 23 chromosomes.

But let us take a closer look at the chromosomes in these gametes, compared to those in the cells from which they were derived. Our original cell (organism)(on the left of the diagram on the next page) was derived from the fusion of two haploid gametes. These haploid gametes each contained one full set of chromosomes, but those chromosomes differed from one another in the details of their nucleotide sequences, specifically which alleles they contain. There will be nucleotide differences at specific positions (known as single nucleotide polymorphisms or SNPs - pronounced snips), small insertions and deletions of nucleotide sequences, and various other structural variants. For our purposes, we will consider only one single chromosome set, but remember there are often multiple chromosomes (23 pairs in human). In our example, the chromosomes inherited from one parent had alleles P, M, and N, while the chromosome from the other parent had alleles p, m, and n. Barring new



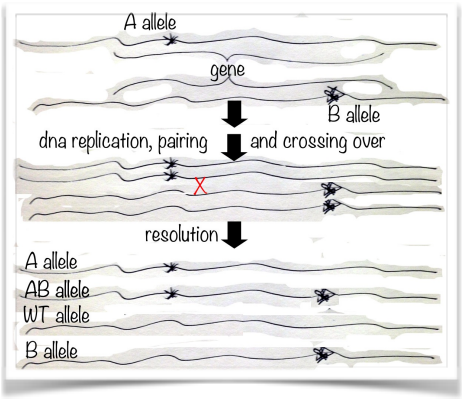
mutations, all of the cells in the body will have the same set of alleles at these genetic positions, and all cells will contain chromosomes similar to the parent PMN and pmn chromosomes (top panel).

Now let us consider what happens when this PMN/pmn organism is about to reproduce. It will begin meiosis (bottom panel). The processes of homolog pairing and crossing over will generate new combinations of alleles: the four haploid cells formed have pMN, PMn, pmN, and Pmn genotypes. All of these are different from the PMN and pmn parental chromosomes. At fertilization one of these haploid cells will fuse with a haploid cell from another organism, to produce a unique individual. While we have considered only two (or three, if you include the p*, m*, and n* alleles) at three genes, two unrelated individuals will differ from each other by 3 to 12 million DNA differences. Most phenotypes are influenced to a greater or lesser extent by the entire genotype, new combinations of alleles will generate new phenotypes.

Meiotic recombination arising from crossing over has two other important outcomes. First consider what happens when a new allele arises by mutation on a chromosome. If the allele has a strongly selected, either positive or negative, phenotype, then organisms that inherit that allele will be selected for (or against). But remember that the allele sits on a chromosome, and is linked to neighboring genes (and their alleles). Without recombination, the entire chromosome would be selected as a unit. In the short term this is still the case, but recombination allows alleles of neighboring genes to disconnect from one another eventually. When the probability of a recombination event between two genes is 50% or greater, the genes appear to behave as if they are on different chromosomes, they become “unlinked.” Linkage distances are calculated in terms of centimorgans, named after the Nobel prize winning geneticist Thomas Hunt Morgan (1866-1945.) A centimorgan corresponds to a 1% chance of a crossing over event between two genes (or specific sites in a chromosome). In humans, a centimorgan corresponds to about 1 million base pairs of DNA, so two genes/alleles/sites along a chromosome separated by more than ~50 million base pairs would be separated by 50 centimorgans, and so would appear unlinked. That is, a crossing over event between the two originally linked alleles

would be expected to occur 50% of the time, which is the same probability that a gamete would inherit one but not the other allele if the genes were located on different chromosomes.

In addition to shuffling alleles, meiotic crossing over (recombination) can create new alleles. Consider the situation in which the two alleles of a particular gene in a diploid organism are different from one another (see figure.) Let us assume that each allele contains a distinct sequence difference (as marked). If, during meiosis, a crossing over event takes place between these sites, it can result in an allele that contains both molecular sequences (AB), and one with neither (indicated as WT), in addition to the original A and B allele chromosomes.



Genome dynamics

Up to now, aside from the insertion of "external" DNA and the recombination events of meiosis we have considered the genome, once inherited by a cell, to be static, but it has become increasingly apparent that genomes are more dynamic than previously thought. For example, consider the number of new mutations (SNPs and such) that arise in each generation. This can be estimated based on the number of times a DNA molecule is replicated between the formation of a new organism (the fusion of haploid cells during fertilization) and the ability of that organism to form new haploid cells (about 400 replication events in a human male, fewer in a female), and the error rate of DNA replication ($\sim 1 \times 10^{-10}$ per nucleotide per division.) Since each diploid cell contains $\sim 6 \times 10^{-9}$ nucleotides, one can expect about 1 new mutation for every two rounds of DNA replication. It has been estimated that, compared with the chromosomes our parents supplied us, we each have between 60 to 100 new mutations in our chromosomes. Given that less than ~5% of our DNA encodes gene products, only of few of these new mutations are likely to influence gene expression or the gene's encoded. Even in the coding region, the redundancy of codons means that many SNPs will not lead to functionally significant alterations in the behavior of gene products. That said, even apparently "neutral" mutations do lead to changes in genotype that can have effects on phenotype, and so evolutionary impacts. As we have already discussed, in small populations alleles with mild effects on reproductive success may or may not be retained in the population. They tend to be lost by genetic drift since they are originally present in a very low percentage of the population.

In addition to the point mutations that arise from mistakes in DNA replication, a whole other type of genomic variation has been uncovered in the course of genome sequencing studies. These are known as "structural variants." They include small (between 1 to 1000 base pair) sequence insertions or deletions (known as InDels), the flipping of the orientation of a DNA region, and a distinct class known as copy number variations (CNV). As noted previously, about 50% of the human genome (and similar levels in other eukaryotic genomes) is composed of various virus-like sequences. Most of these have been degraded by mutation, but some remain active. For example, there are ~100 potentially active L1

type transposons (known as LINE elements) in the human (your) genome.[230] These 6000 base pair long DNA regions encode genes involved in making and moving a copy of themselves to another position in the genome. Some genomic variants have no direct phenotypic effects. For example a region of a chromosome can be "flipped" around; as long as no regulatory or coding sequences are disrupted, there may be no effect on phenotype. That said, large flips or the movements of regions of DNA molecules between chromosomes can have effects on chromosome pairing during meiosis. It has been estimated that each person contains about 2000 "structural variants".[231]

An important point with all types of new variants is that if they occur in the soma, that is in cells that do not give rise to the haploid cells (gametes) involved in reproduction, they will be lost when the host organism dies. At this point, there is no evidence of horizontal gene transfer between somatic cells. Moreover, if a mutation disrupts an essential function, the affected cell will die, to be replaced by surrounding normal cells. Finally, as we have discussed before and will discuss later on, multicellular organisms are social systems. Mutations, such as those that give rise to cancer, can be seen as cheating the evolutionary (cooperative) bargain that multicellular organisms are based on. It is often the case that organisms have both internal and social policing systems. Mutant cells often actively kill themselves (through apoptosis) or, particularly in organisms with an immune system, they will be actively identified and killed.

Paralogous genes and gene families

As noted previously genome dynamics plays a critical role in facilitating evolutionary change, particularly in the context of multicellular organisms.[232] When a region of DNA is duplicated, the genes in the duplicated region may come to be regulated differently, and they can be mutated in various ways while the other copy of the gene continues to carry out the gene's original function. This provides a permissive context in which mutations can alter what might have been a gene product's off-target or as it is sometime called, promiscuous activities.[233] While typically much less efficient than the gene product's primary role, they can have physiologically significant effects.

The two versions of a duplicated gene are said to be paralogs of each other. In any gene duplication event, the two duplicated genes can have a number of fates. For example, both genes could be conserved, providing added protection against mutational inactivation. The presence of two copies of a gene often leads to an increase the amount of gene product generated, which may provide a selective advantage. For example, in the course of cancer treatment, gene duplication may be selected for because increased copies of genes may encode gene products involved in the detoxification of, or

---

[230] Natural mutagenesis of human genomes by endogenous retrotransposons: http://www.ncbi.nlm.nih.gov/pubmed/20603005

[231] Child Development and Structural Variation in the Human Genome: http://www.ncbi.nlm.nih.gov/pubmed/23311762

[232] Ohno's dilemma: evolution of new genes under continuous selection: http://www.ncbi.nlm.nih.gov/pubmed/17942681 and Copy-number changes in evolution: rates, fitness effects and adaptive significance: http://www.ncbi.nlm.nih.gov/pubmed/24368910

[233] Enzyme promiscuity: a mechanistic and evolutionary perspective: http://www.ncbi.nlm.nih.gov/pubmed/20235827 and Network Context and Selection in the Evolution to Enzyme Specificity: http://www.ncbi.nlm.nih.gov/pubmed/22936779

resistance to an anti-cancer drug.[234] It is possible that both genes retain their original function, but are expressed at different levels and at different times in different cell types. One gene's activity may be lost through mutation, in which case we are back to where we started. Alternatively, one gene can evolve to carry out a new, but important functional role, so that conservative selection acts to preserve both versions of the gene.

Such gene duplication processes can generate families of evolutionarily related genes. In the analysis of gene families, we make a distinction between genes that are orthologs of each other and those that are paralogs. Orthologous (or homologous) genes are found in different organisms, but are derived from a single common ancestral gene present in the common ancestor of those organisms. Paralogous genes are genes present in a particular organism that are related to each other through a gene duplication event. A particular paralog in one organism can be orthologous to a gene in another organism, or it could have arisen independently in an ancestor, through a gene duplication event.

Detailed comparisons of nucleotide sequence can distinguish between the two. The further in the past that a gene duplication event is thought to occur, the more mutational noise can obscure the relationship between the duplicated genes. Remember, when looking at DNA there are only four possible bases at each position. A mutation can change a base from A to G, and a second mutation from G back to A. If this occurs, we cannot be completely sure as to the number of mutations that separate two genes, since it could be 0, 2 or a greater number. We can only generate estimates of probable relationships. Since many multigene families appear to have their origins in organisms that lived hundreds of millions of years ago, the older the common ancestor, the more obscure the relationship can be. The exceptions involve genes that are extremely highly conserved, which basically means that their sequences are constrained by the sequence of their gene product. In this case most mutations produce a lethal phenotype, meaning that the cell or organism with that mutation dies or fails to reproduce. These genes evolve very slowly. In contrast, gene/gene products with less rigid constraints (and this includes most genes/gene products) evolve much faster, which can make relationships between genes found in distantly related organisms more tentative. Also, while functional similarities are often seen as evidence for evolutionary homology, it is worth considering the possibility, particularly in highly diverged genes/gene products, of convergent evolution. As with wings, the number of ways to carry out a particular molecular level function may be limited.

Questions to answer & to ponder:

- Make a diagram that illustrates how genes can "overlap".
- Make a diagram and analyze the effects of flipping a region of a chromosome around (180°) or moving it from one chromosome to another, on gene expression.
- Consider the effects of such rearrangements on chromosome pairing during meiosis.
- Think about eukaryotic gene structure; explain how a transposon could insert within a gene without negatively influencing gene function. Is such a thing possible?
- What factors might drive the evolution of overlapping genes?
- Explain why parasites and endosymbionts can survive with so few genes.
- How does sexual reproduction increase the genetic diversity within a population?
- Speculate on what selective factors might favor sexual over asexual reproduction.

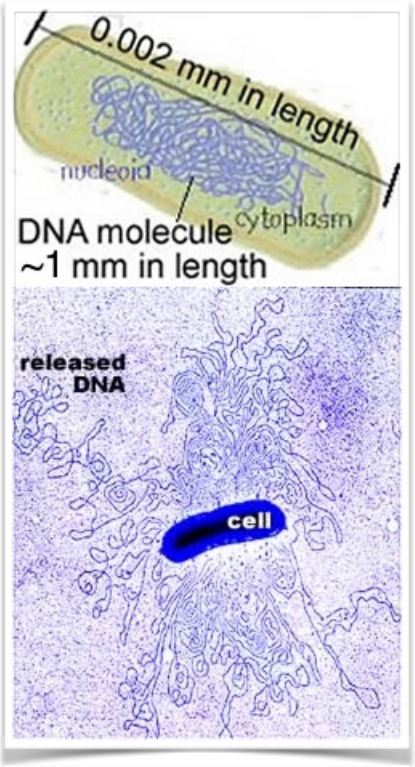
---

[234] Dihydrofolate reductase amplification and sensitization to methotrexate of methotrexate-resistant colon cancer cells: http://www.ncbi.nlm.nih.gov/pubmed/19190117

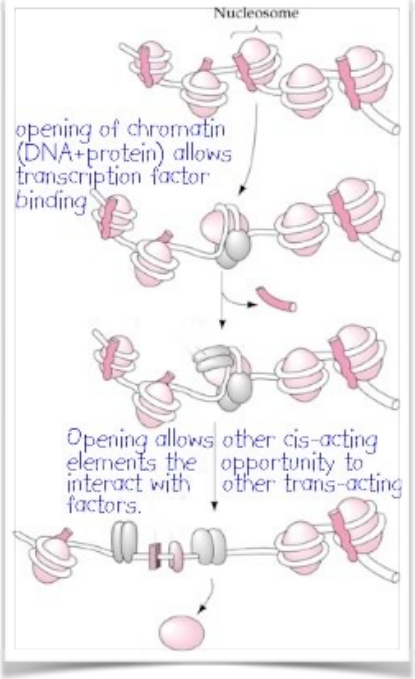
- Provide an explanation for the persistence of duplicated genes. What forces would act to remove them?
- What type of event would lead to total genome duplication?
- Why are some genes lost after genome duplication?

Packing DNA into a cell

An important part of our approach to biology is to think concretely about the molecules we are considering. No where is this more important than with DNA. DNA molecules are very long and cells, even the largest cells, are (generally) quite small. For example, a typical bacterium is roughly cylindrical and around 2 µm in length and about 1 µm in circumference. Based on the structure of DNA, each base pair is about 0.34 nm in length. A kilobase (that is, 10^3 base pairs) of DNA is therefore about 0.34 μm in length. A bacterium, like *E. coli*, has ~ 3 x 10^6 base pairs of DNA – that is a DNA molecule almost a millimeter in length, or about 500 times the length of the bacterial cell in which it finds itself. That implies that at the very least the DNA has to be folded back on itself at least 250 times. A human cell has about 6000 times more DNA, that is a total length of greater than 2 meters (per cell), which has to fit into a nucleus of approximately 10 μm in diameter. In both cases, the DNA has to be folded and packaged in ways that allow it to fit and yet still be accessible to the various proteins involved in the regulation of gene expression and the replication of DNA. To accomplish this, the DNA molecule is associated with specific proteins and the resulting DNA:protein complex is known as chromatin.



The study of how DNA is regulated is the general topic of epigenetics (on top of genetics), while genetics refers to the genetic information itself. If you consider a particular gene (based on our previous discussions) you will realize that to be expressed, transcription factor proteins must be able to find (by diffusion) and bind to specific regions (defined by their sequences) of the DNA in the gene's regulatory region(s). But the way the DNA is organized into chromatin, particularly in eukaryotic cells, can dramatically influence the ability of transcription factors to interact with and bind to their regulatory sequences. For example, if a gene's regulatory regions are inaccessible to protein binding because of the structure of the chromatin, the gene will be "off" (unexpressed) even if the transcription factors that would normally turn it on are present and active. As with essentially all biological systems, the interactions between DNA and various proteins can be regulated.



Different types of cells can often have their DNA organized differently through the differential expression and activity of genes involved in opening up (making accessible) or closing down (making inaccessible) regions of DNA. Accessible, transcriptionally active regions of DNA are known as euchromatin while DNA packaged so that the DNA is inaccessible is known as heterochromatin. A particularly dramatic example of this process occurs in female mammals. The X chromosome contains about 1100 genes that play important roles in both males and females.[235] But the level of gene expression is influenced by the number of copies of a particular gene. While various mechanisms can often compensate for differences in gene copy number, this is not always the case. For example, there are genes in which the mutational inactivation of one of the two copies leads to a distinct phenotype, a situation known as haploinsufficiency. This raises issues for genes located on the X chromosome, since XX organisms have two copies of these genes, while XY organisms have only a single copy.[236] While one could imagine a mechanism that increased expression of genes on the male's single X chromosome, the actual mechanism used is to inhibit expression of genes on one of the female's two X chromosomes. In each XX cell, one of the two X chromosomes is packed into a heterochromatic state, more or less permanently. It is known as a Barr body. The decision as to which X chromosome is "inactivated" is made in the early embryo, and appears to be stochastic - that means that it is equally likely that in any particular cell, either the X chromosome inherited from the mother or the X chromosome inherited from the father may be inactivated (made heterochromatic). Importantly, once made this choice is inherited, the offspring of a cell will maintain the active/inactivated states of the X chromosomes of its parental cell. Once the inactivation event occurs it is inherited vertically.[237] The result is that XX females are epigenetic mosaics, they are made of clones of cells in which either one or the other of their X chromosomes have been inactivated. Many epigenetic events can persist through DNA replication and cell division, so these states can be inherited through the soma. A question remains whether epigenetic states can be transmitted through meiosis and into the next generation.[238] Typically most epigenetic information is reset during the process of embryonic development.

Locating information within DNA

So given that a gene exists within a genome, for it to be useful there have to be mechanisms by which it can be recognized and transcribed.[239] This is accomplished through a two-component system. The first part of this system are specific nucleotide sequences. These regulatory sequences provide a molecular address that can be used to identify the specific region and the specific strand of the DNA to be transcribed. The regulatory region of a gene can be simple and relatively short or long and complex. In some human genes, the gene's regulatory region is spread over thousands of base-pairs of DNA,

---

[235] Human Genome Project: Chromosome X: http://www.sanger.ac.uk/about/history/hgp/chrx.html

[236] The Y chromosome is not that serious an issue, since its ~50 genes are primarily involved in producing the male phenotype.

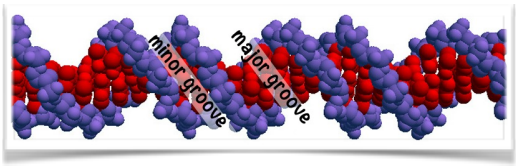
[237] X Chromosome: X Inactivation: http://www.nature.com/scitable/topicpage/x-chromosome-x-inactivation-323

[238] Identification of genes preventing transgenerational transmission of stress-induced epigenetic states: http://www.ncbi.nlm.nih.gov/pubmed/24912148

[239] As an aside, are many transcribed DNA sequences that do not appear to encode a polypeptide or regulatory RNAs. It is not clear whether this transcription is an error, due to molecular level noise or whether such RNAs play a physiological role..

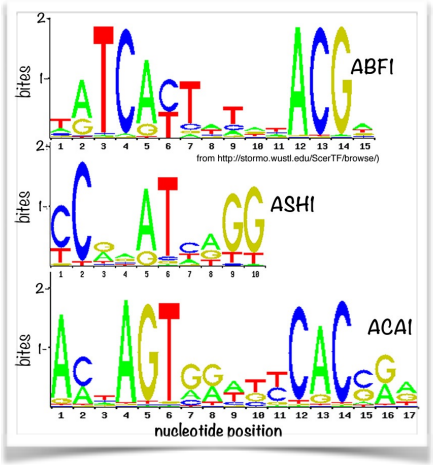
located “up-stream”, "down-stream" and within the coding region.[240] This is possible because DNA can fold back on itself.

In eukaryotes, the proteins that bind to regulatory sequences are known as transcription factors - they function similarly to the sigma (σ) factors of prokaryotes. In early genetic studies, two types of mutations were found that influence the activity of a gene. “cis” mutations were mapped to or near the gene, and include mutations in the gene’s regulatory sequences. “trans” mutations mapped at other (distant) sites, and they turn out to influence genes that encoded the transcription factor proteins involved in the target gene’s regulation. Transcription regulating proteins can act either positively to recruit and activate DNA-dependent, RNA polymerase or negatively, to block RNA polymerase binding and activity. Genes that efficiently recruit and activate RNA polymerase will make many copies of the associated RNA, and are said to be highly expressed. Generally, high levels of mRNA will lead to high levels of the encoded polypeptide. Mutations in the genes encoding transcription factors can influence the expression of many genes, while mutations in a gene’s regulatory sequence will influence its expression, unless of course the gene encodes a transcription factor or its activity influences the regulatory circuitry of the cell.



Transcription regulatory proteins recognize specific DNA sequences by interacting with the surfaces of base pairs visible in the major or minor grooves of the DNA helix. There are a number of different types of transcription factors that are members of various gene families.[241] A particular transcription factor’s binding affinity to a particular regulatory site will be influenced by the DNA sequence as well as the binding of other proteins in the molecular neighborhood. Different DNA sequences will bind transcription factors with different affinities. We can compare affinities of different proteins for different binding sites by using an assay in which short DNA molecules containing a particular nucleotide sequence are mixed in a 1:1 molar ratio, that is, equal numbers of protein and DNA molecules:

$$DNA_{sequence} + protein \leftrightarrows DNA{:}protein.$$



After the binding reaction has reached equilibrium, we can measure the percentage of the DNA bound to the protein. If the protein binds with high affinity the value is close to 100%, and close to 0% if it binds with low affinity. In this way we can empirically determine the relative binding specificities (binding affinity for a particular sequence) of various proteins, assuming that we can generate DNA molecules of specific length and sequence (which we can) and purify proteins that remain properly folded in a native rather than denatured or inactive configuration, which may or may not be simple.[242] What we discover is that transcription factors do not recognize unique nucleotide sequences, but rather have a range of affinities for related

[240] Regulatory regions located far from the gene’s transcribed region are known as enhancer elements.

[241] Determining the specificity of protein-DNA interactions: http://www.ncbi.nlm.nih.gov/pubmed/20877328

[242] Of course we are assuming that physiologically significant aspect of protein binding involves only the DNA, rather than DNA in the context of chromatin, and ignores the effects of other proteins, but it is a good initial assumption.

sequences. This binding preference is characteristic of each transcription factor protein; it involves both the length of the DNA sequence recognized and the pattern of nucleotides within that sequence. A simple approach to this problem considers the binding information present at each nucleotide position as independent of all others in the binding sequence, which is certainly not accurate but close enough for most situations. This data is often presented as a "sequence logo".[243] In such a plot, we indicate the amount of binding information at each position along the length of the binding site. Where there is no preference, that is, where any of the four nucleotides is acceptable, the information present at that site is 0. Where either of two nucleotides are acceptable, the information is 1, and where only one particular nucleotide is acceptable, the information content is 2. Different transcription factor proteins produce different preference plots. As you might predict, mutations in a transcription factor binding site can have dramatically different effects. At sites containing no specific information (0), a mutation will have no effect, whereas in sites of high information (2), any change from the preferred nucleotide will likely produce a severe effect on binding affinity.

This is not to say that proteins cannot be extremely specific in their binding to nucleic acid sequences. For example, there is a class of proteins, known as restriction endonucleases and site specific DNA modification enzymes (methylases) that bind to unique nucleotide sequences. For example the restriction endonuclease EcoR1 binds to (and cleaves) the nucleotide sequence GAATTC, change any one of these bases and there is no significant binding and no cleavage. So the fact that transcription factor's binding specificities are more flexible suggests that there is a reason for such flexibility, although exactly what that reason is remains conjectural.

An important point to take away is that most transcription factor proteins also bind to generic DNA sequences with low affinity. This "non-sequence specific" binding is transient and such protein:DNA interactions are rapidly broken by thermal motion. That said, since there are huge numbers of such non-sequence specific binding sites within a cell's DNA, most of the time transcription factors are found transiently associated with DNA (illustrated in the PhET applet:http://phet.colorado.edu/en/simulation/gene-expression-basics).

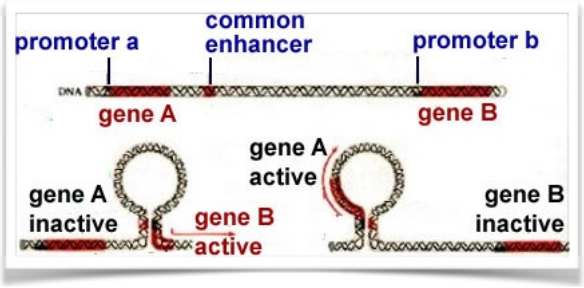
To be effective in recruiting RNA polymerases and other proteins to specific sites along a DNA molecule, the binding of a protein to a specific DNA sequence must be relatively long lasting. A common approach to achieving this outcome is for the transcription factor to be multivalent, that is, so that it binds to multiple (typically two) sequence elements. This has the effect that if the transcription factor dissociates from one binding site, it remains tethered to the other; since it is held close to the DNA it is more likely to rebind to its original site. In contrast, a protein with a single binding site is more likely to diffuse away. A related behavior involving the low affinity binding of proteins to DNA is that it leads to one-dimensional diffusion along the length of the bound DNA molecule (illustrated in the **PhET applet**). This enables a transcription factor protein to bind to DNA and then move back and forth along the DNA molecule until it interacts, and binds to, a high affinity site (or until it dissociates completely.)

[243] Sequence logos: a new way to display consensus sequences: http://www.ncbi.nlm.nih.gov/pubmed/2172928

This type of “facilitated target search” behavior can greatly reduce the time it takes for a protein to find a high affinity binding site among millions of low affinity sites present in the genome.[244]

While prokaryotic (bacterial) genes are normally regulated by a specific σ factor (see above), more complicated eukaryotic genes, particularly those in multicellular organisms, have a number of different cell types. These generally use distinct sets of transcription factors and regulatory sequences to regulate the time and level of gene expression. Not only do these proteins bind to DNA, they can interact with one another. For example, we can imagine that the binding affinity of a particular transcription factor will be influenced by the presence of another transcription factor already bound to an adjacent or overlapping site on the DNA. Similarly the structure of a protein can change when it is bound to DNA, and such a change can lead to interactions with DNA:protein complexes located at more distant sites, known as enhancers. Such regulatory elements, can be part of multiple various regulatory systems.

For example, consider the following situation. Two genes share a common enhancer, depending upon which interaction occurs, gene a or gene b but not both could be active. The end result is that combinations of transcription factors are involved in turning on and off gene expression. In some cases, the same protein can act either positively or negatively, depending upon the specific gene regulatory sequences and the context of other bound factors. Here it is worth noting that the organization of regulatory and coding sequences in DNA imposes directionality on the system. A transcription factor bound to DNA in one orientation or at one position may block the binding of other proteins (or RNA polymerase), while bound to another site it might stabilize protein (RNA polymerase) binding. Similarly, DNA binding proteins can interact with other proteins to control chromatin configurations that can allow or block accessibility to regulatory sequences. While it is common to see a particular transcription factor protein labelled a either a transcriptional activator or repressor, in reality the activity of a protein will often reflect the specific gene context and its interactions with various accessory factors, all of which can influence gene expression.



The place where RNA polymerase starts transcribing RNA is known as the transcription start site. Where it falls off the DNA, and so stops transcribing RNA, is known as the transcription termination site. As transcription initiates, the RNA polymerase moves away from the transcription start site. Once the RNA polymerase complex moves far enough away (clears the start site), there is room for another polymerase complex to associate with the DNA, through interactions with transcription factors. Assuming that the regulatory region and its associated factors remains intact, the time to load a new polymerase will be relatively faster than the time it takes to build up a new regulatory complex from scratch. This is one reason that transcription is often found to occur in bursts, a number of RNAs are synthesized from a particular gene in a short time period, followed by a period of transcriptional silence. A similar bursting behavior is observed in protein synthesis.

---

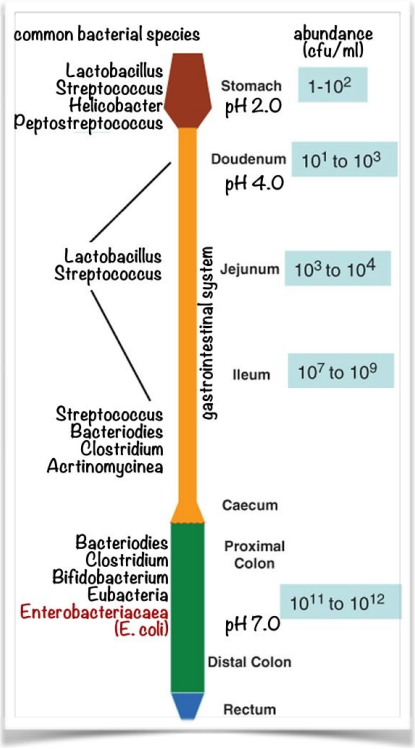
[244] Physics of protein-DNA interactions: mechanisms of facilitated target search: https://www.ncbi.nlm.nih.gov/pubmed/21113556

Network interactions

As we come to analyze the regulation of genes, we recognize that they represent an interaction network. A defining feature of all biological systems, from the molecular to the ecological and evolutionary, are interaction networks - generally organized in a hierarchical and bidirectional manner. So what exactly does that mean? Most obviously, at the macroscopic level, the behavior of ecosystems depends upon the interactions of organisms with one another. As we move down the size scale the behavior of individual organisms is based on the interactions between tissues formed during the process of embryonic development and maturation. At the same time the behavior of organisms is influenced by their environment. Similarly, the behavior of tissues and organs is based on the behavior of cells and their interactions with each other. Their behaviors are influenced by their environment, including the state of the organism as a whole. The behavior of individual cells is influenced by the activity of genes, which in turn in are influenced by the interactions between cells (and the extracellular environment) around them. The molecular level behavior of biological systems occurs within cellular, tissue, organismic, social, and ecological contexts that influence and are influenced by each other. And all of these interactions (and the processes that underlie a particular biological system) are the result of evolutionary mechanisms and historical situations (past adaptation and non-adaptive events.)

Not withstanding the complexity of biological systems, we can approach them at various levels through a systems perspective. At each level, there are objects that interact with one another in various ways to produce various behaviors. To analyze a system at the molecular, cellular, tissue, organismic, social, or ecological level we have to define (and understand and appreciate) the nature of the objects that are interacting, how they interact with one another, and what the results of those interactions are.

There are many ways to illustrate this way of thinking but we think that it is important to get concrete by looking at a (relatively) simple and well understood system by considering how it behaves at the molecular, cellular, and social levels. Our model system will be the bacterium *E. coli* and some of its behaviors, in particular how it behaves in isolation and in social groups and how it metabolizes the milk sugar lactose.[245] Together these illustrate a number of common regulatory principles that apply more or less universally to biological systems at all levels of organization.



E. coli as a model system:

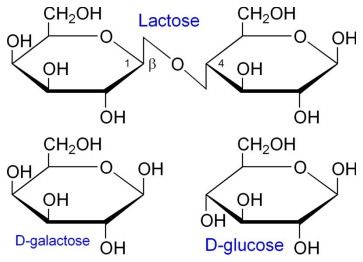
Every surface of your body, including your gastrointestinal tract, which runs from your mouth to your rectum, harbors a flourishing microbial ecosystem. Your gastrointestinal ecosystem includes a number of distinct environments, ranging from the mouth and esophagus, through the stomach, into the small and large intestine

[245] The Lac Operon: A Short History of a Genetic Paradigm http://books.google.com.et/books/about/The_Lac_Operon.html?id=ppRmC9-a6JQC

and the colon.[246] In addition to differences in pH there are also changes in O_2 levels between these environments. Near the mouth and esophagus, O_2 levels are high, and microbes can use aerobic (O_2 dependent) respiration to extract energy from food. Moving through the system, O_2 levels become extremely low and anaerobic (without O_2) mechanisms are necessary. At different position along the length of the gastrointestinal track, microbes with different ecological preferences and adaptabilities are found. One issue associated with characterizing the exact complexity of the populations of microbes in various locations is that it is often the case that these organisms are dependent upon one another for growth, and so when isolated as individuals they do not grow. The standard way to count bacteria is to grow them on a plate of growth medium in the lab. The sample is diluted so that single bacteria land (in isolation from one another) on the plate. As they grow and divide, each bacterium forms a macroscopic colony. We count the number of these "colony forming units" (CFUs) present in the original volume as a measure of the number of individual bacteria present. Bacteria that do not colonies under the assay conditions will appear to be absent from the population. But as we have just mentioned some bacteria are totally dependent on each other and therefore do not grow in isolation. In fact only less that 5% of the microbes present in a typical sample have been grown in the lab. To get a better estimate of the number of different types of organisms present in a sample scientists use DNA sequence analyses; this approach makes it possible to identify without having to grow them.[247] It reveals the true complexity of the microbial ecosystems living on and within us, a microbial ecosystem (known as the microbiome) that plays an important role in health.[248]

Here we focus on one well known, but relatively minor member of this microbial community, *Escherichia coli. E. coli* is a member of the Enterobacteriaceae family of bacteria and is found in the colon of birds and mammals.[249] *E. coli* is what is known as a facultative aerobe, it can survive in both an anaerobic environment as well as an aerobic one. This flexibility, as well as its generally non-fastidious growth requirement make it easy to grow in the laboratory. The laboratory strain of *E. coli* generally used, known as K12, is non-pathogenic–it does not cause disease in humans. There are other strains of *E. coli*, such as *E. coli* O157:H7 that are pathogenic. *E. coli* O157:H7 contains 1,387 genes not found in the *E. coli* K12. Scientists estimate that the two strains diverged from a common ancestor ~4 million years ago. The details of what makes *E. coli* O157:H7 pathogenic and *E. coli* K12 not is a fascinating topic but beyond our scope.

Adaptive behavior and gene networks (the lac response): Lactose is a disaccharide composed of D-galactose and D-glucose. It is synthesized, biologically, exclusively by female mammals. Mammals use lactose in milk as a source of calories (energy) for infants, one reason (it is thought) is that lactose is not easily digested by most microbes. The lactose synthesis system is derived from an evolutionary modification of an



[246] The gut microbiome: scourge, sentinel or spectator?: http://www.journaloforalmicrobiology.net/index.php/jom/rt/printerFriendly/9367/19922

[247] Application of sequence-based methods in human microbial ecology: http://www.ncbi.nlm.nih.gov/pubmed/16461883

[248] The human microbiome: our second genome: http://www.ncbi.nlm.nih.gov/pubmed/22703178

[249] The Evolutionary Ecology of Escherichia coli: http://www.americanscientist.org/issues/feature/the-evolutionary-ecology-of-escherichia-coli/1

ancestral gene that encodes the enzyme lysozyme. Through duplication and mutation, a gene encoding the protein α-lactoalbumin was generated. α-lactoalbumin is expressed only in mammary glands, where it forms a complex with a ubiquitously expressed protein, galactosyltransferase, to form lactose synthase.[250]

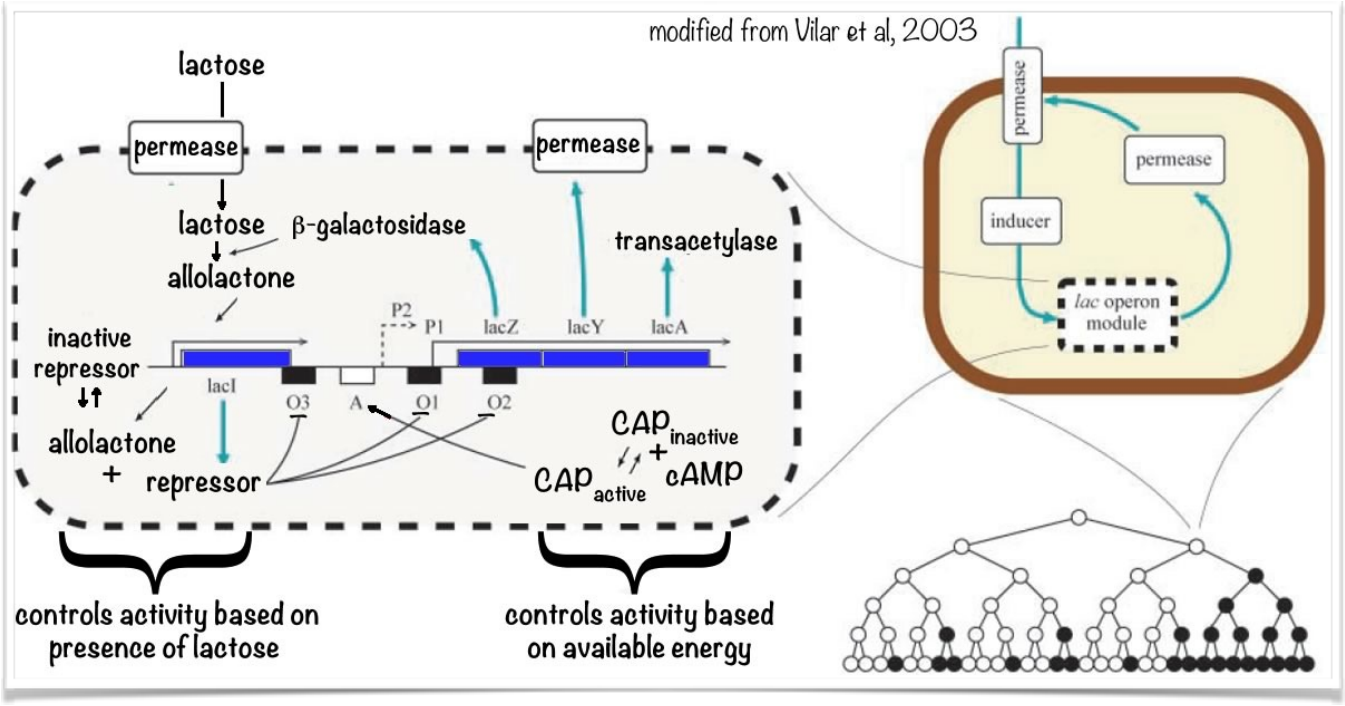
E. coli is capable of metabolizing lactose, but only when there are no better (easier) sugars to eat. If glucose or other compounds are present in the environment the genes required to metabolize lactose are turned off. Two genes are required for *E. coli* to metabolize lactose. The first encodes lactose permease. Lactose, being large and highly hydrophilic, cannot pass through *E. coli*'s plasma membrane. Lactose permease is a membrane protein that allows lactose to enter the cell, moving down its concentration gradient. The second gene encodes the enzyme β-galactosidase, which catalyzes the hydrolysis lactose into D-galactose and D-glucose. Both of these sugars can be metabolized by proteins expressed constitutively (that is, all of the time) in the cell. So how exactly does this system work? How are the lactose utilization genes turned off in the absence of lactose and how are they turned on when lactose is present and energy is needed. The answers illustrate general principles of the interaction networks controlling gene expression.

In *E. coli*, like many bacteria, multiple genes are organized into what are known as operons. In an operon, a single regulatory region controls the expression of multiple genes. It is also common that multiple genes involved in a single metabolic pathway are located in the same operon (the same region of the DNA). One powerful approach to the study of genes is to look for relevant mutant phenotypes. As we said, wild type (that is, normal) *E. coli* can grow on lactose as their sole energy sources. So an obvious phenotype to look for would be mutants of *E. coli* that cannot grow on lactose. To make the screen for such mutations more relevant, we will check to make sure that the mutants can grow on glucose. Why? Because we are not really interested (in this case) in mutations in genes that disrupt other aspects of metabolism, for example the ability to use glucose or synthesize proteins, but rather seek to identify genes specifically involved in the metabolism of lactose. This type of analysis revealed a number of distinct classes of mutations. Some mutations led to an inability to respond to lactose, while others led to the de-repression, that is expression of the genes lactose permease and β-galactosidase, even when lactose is absent. By mapping where these mutations are in the genome of *E. coli*, and a number of other experiments, the following model was generated (figure on next page).

The genes encoding lactose permease and β-galactosidase are part of an operon, known as the *lac* operon. This operon is regulated by two distinct factors. The first is the product of a constitutively active gene, *lacI*, which encodes a polypeptide that assembles into a tetrameric protein that acts as a transcriptional repressor; there are about 10 lac repressor proteins present per cell. The lac repressor protein binds to sites in the promoter of the *lac* operon. When bound to these sites the repressor protein blocks the transcription of the *lac* operon. It appears that the lac repressor's binding sites within the *lac* operon promoter are the only functionally significant binding sites in the *E. coli* genome (although perhaps we have not looked carefully enough). The *lac* operon's second regulatory element is known as the activator site. It can bind the catabolyte activator protein (or CAP), which is encoded by another gene. The DNA binding activity of CAP is regulated by the binding of a co-factor,

[250] Molecular divergence of lysozymes and alpha-lactalbumin: http://www.ncbi.nlm.nih.gov/pubmed/9307874

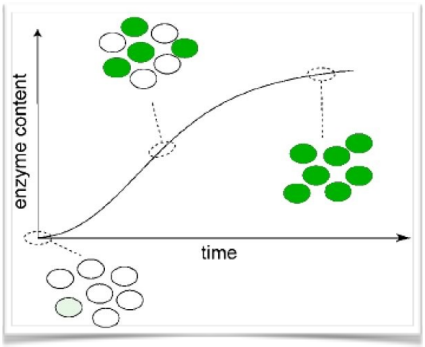
cyclic adenosine monophosphate or cAMP. cAMP accumulates in the cell when nutrients, specifically free energy delivering nutrients (like glucose) are low. Its presence serves as a signal that the cell needs energy. In the absence of cAMP, CAP does not bind to or activate expression of the *lac* operon, but when cAMP is present (that is, when energy is needed), the CAP-cAMP protein is active and binds to a site (known as the activator or A site) in the *lac* operon promoter, where it recruits and activates RNA polymerase, leading to the synthesis of lactose permease and β-galactosidase RNAs and proteins. However, if energy levels are low (and cAMP levels are high), the *lac* operon will be inactive in the absence of lactose because of the binding of the lac repressor protein to sites (labelled 0_1, 0_2, and 0_3) in *lac* operon.



So what happens when lactose appears in the cell's environment? Well, obviously nothing, since the cells are expressing the lac repressor; no lactose permease is present and lactose cannot enter the cell without it. But that assumes that, at the molecular level, the system works perfectly and deterministically. However, this is not the case. The system is stochastic, that is, it is subject to the effects of random processes–it is noisy and probabilistic. Given the small number of lac repressor molecules per cell (~10), there is a small but significant chance that, at random, the lac operon of a particular cell will be free of bound repressor (you could, if you were mathematically inclined, calculate this probability based on the binding constant of the lac repressor for its site in the *lac* promoter, about 1 x 10^{-9} M, and the concentration of the lac repressor protein in the cell, about 50 x 10^{-9} M). Under conditions in which CAP is active, periodically a cell will express the genes in the *lac* operon even though no lactose is present within the cell, because no repressor molecules are bound to the operon. We can use this information to predict what will happen to *lac* operon expression and the ability of *E. coli* cells to utilize lactose as a function of time following the addition of lactose.[251] When lactose is added those cells that have, because of stochastic events, expressed the lactose permease (a small

[251] Modeling network dynamics: the lac operon, a case study: http://www.ncbi.nlm.nih.gov/pubmed/12743100

percentage of the total cell population), will allow lactose to enter the cell. The noisy expression of the *lac* operon will also result in a small number of β-galactosidase molecules. β-galactosidase catalyzes two reaction. One reaction involves the hydrolysis of lactose into D-galactosidase and D-glucose. Their subsequent breakdown into CO_2 and H_2O is a thermodynamically favorable reaction which drives cellular metabolism. The second and more interesting reaction catalyzed by β-galactosidase, from a regulatory perspective, is the isomerization of lactose to form allolactone. It is, in fact, allolactone (not lactose) that binds to and inhibits the activity of the lac repressor protein. In the presence of allolactone, the repressor no longer inhibits *lac* operon expression, and there is a dramatic (~1000 fold) increase in the rate of expression of lactose permease and β-galactosidase. The cell goes from essentially no expression of the *lac* operon to full expression, and with full expression, becomes able to metabolize lactose, that is convert it into D-galactose and D-glucose, at its maximal rate. What is surprising then is that shortly after the addition of lactose, we will find that some cells in the culture are metabolizing lactose at the maximal rate, while others will not be metabolizing it at all. Only with time will more and more cell's turn on their copy of the lac operon, driven by the noisy (lactose independent) expression of the operon. Once "on", the presence of allolactone in the cell will keep the lac repressor protein in an inactive (unable to bind DNA) state and allow expression of the lac operon.



So even though all of the *E. coli* present in a particular culture may be genetically identical they can express different phenotypes. Shortly after the addition of lactose, some cells allow lactose to enter and then actively break it down, while other cells are unable to either import or metabolize lactose. The culture will be heterogeneous. That said, if we wait long enough, each cell will go through (with a certain probability per unit time) the transition to the ability to import and utilize lactose. In the presence of lactose this transition is stable and eventually all cells in the culture will be actively metabolizing lactose.

What happens if lactose disappears from the environment; what determines how long it takes for the cells to return to the state in which the the *lac* operon is no longer expressed? The answer is determined by the effects of cell division and regulatory processes. In the absence of lactose, there is no allolactone, so the lac repressor protein returns to its active state, which acts to inhibit the expression of the *lac* operon. Second, since they are no longer being synthesized lactose permease and β-galactosidase proteins will be degraded by proteases at a certain rate. Their concentrations in the cell will fall. Finally, and again because their synthesis has stopped, with each cell division the concentration of the lactose permease and β-galactosidase proteins will decrease by at least 50%. With time the proteins are diluted and degraded; the cells return to their initial state, that is, with the *lac* operon off due to the action of the lac repressor.

Final thoughts on (molecular) noise

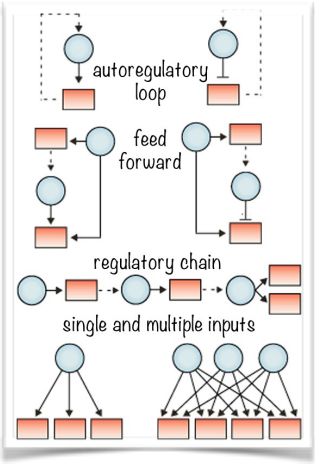
When we think about such stochastic behaviors, we can readily identify a few obvious sources of molecular level noise. First, there are generally only one or two copies of a particular gene within a cell. The probability that those genes are able to recruit and activate RNA polymerase is determined by the frequency of productive collisions between regulatory sequences and relevant transcription factors.

Cells are small, and the numbers of different transcription factors can vary quite dramatically. Some are present in reasonably high numbers (~250,000 per cell) while others (like the lac repressor) may be present in less than 10 copies per cell. The probability that particular molecules interact will be controlled by diffusion, binding, and kinetic energies (temperature). This will dramatically influence the probability that a particular gene regulated by a particular transcription factor is active or not.

A related process arises from the fact that in some cases, the generation of an active promoter complex may lead to a temporary stable state that has a higher probability of productive interaction. For example, if a complex of proteins binds to a gene's promoter, and this complex is stabilized through their mutual interactions, the result can be bursts of transcript synthesis. A similar situation can apply to the assembly of a ribosome/mRNA complex, again leading to bursts of polypeptide synthesis. Such bursting RNA and polypeptide synthesis effects have been observed and, in certain cases, are physiologically significant.[252] For example, a group of genetically identical *E. coli* cells containing genes encoding various fluorescent proteins display dramatically different levels of expression due to such noisy processes (see PhET gene expression applet). These variations mean that a single genotype can produce multiple phenotypes.

Types of regulatory interactions

A comprehensive analysis of the interactions between 106 transcription factors and regulatory sequences in the baker's yeast *Saccharomyces cerevisiae* revealed the presence of a number of common regulatory motifs.[253] These include:



- **Autoregulatory loops**: A transcription factor binds to sequences that regulate its own transcription. Such interactions can be positive (amplifying) or negative (squelching).
- **Feed forward interactions**: A transcription factor regulates the expression of a second transcription factor; the two transcription factors then cooperate to regulate the expression of a third gene.
- **Regulatory chains**: A transcription factor binds to the regulatory sequences in another gene and induces expression of a second transcription factor, which in turn binds to regulatory sequences in a third gene, etc. The chain ends with the production of some non-transcription factor products.
- **Single and multiple input modules**: A transcription factor binds to sequences in a number of genes, regulating their coordinated expression (σ factors works this way). In most cases, sets of target genes are regulated by sets of transcription factors that bind in concert.

[252] **A single molecule view of gene expression**: http://www.ncbi.nlm.nih.gov/pubmed/19819144

[253] Transcriptional regulatory networks in Saccharomyces cerevisiae: http://www.ncbi.nlm.nih.gov/pubmed/12399584

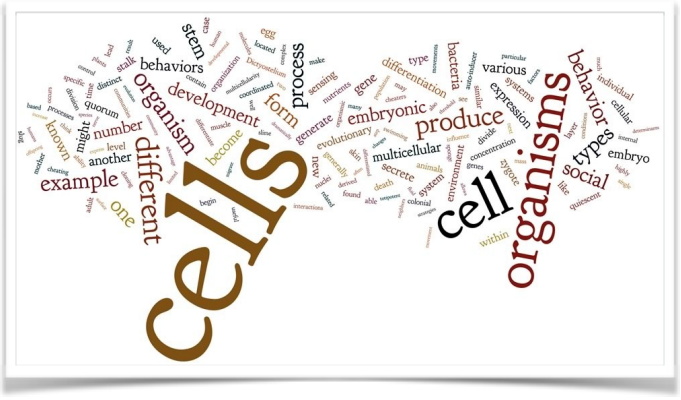
In each case the activity of a protein involved in an interaction network can, like the *lac* repressor, be regulated through interactions with other proteins, allosteric factors, and post-translational modifications. It is through such interactions that signals from inside and outside the cell can control patterns of gene expression leading to maintenance of the homeostatic state or various adaptations.

Questions to answer & to ponder:

- Make a model for how a transcription factor determines which DNA strand will be transcribed.
- Make a model for how one could increase the specificity of the regulation of a gene.
- Describe the possible effects of mutations that alter the DNA-binding specificity of a transcription factor or a DNA sequence normally recognized by that transcription factor.
- Consider a particular gene, what factors are likely to influence the length of its regulatory region?
- What factors might drive the evolution of overlapping genes?
- How could you tell which X chromosome was inactivated in a particular cell of a female person?
- How would you design a regulatory network to produce a steady level of product?
- How can transcription factor proteins be regulated?
- How does regulating the intracellular localization of a transcription factor alter gene expression?
- What kinds of mutation would permanently inactivate a gene?

10. Social systems:

We end up by considering the dynamics of social systems, from bacterial quorum sensing to the development of an embryo.



Thinking about biology, we have to adopt a systems perspective. At each level of biological organization we can identify the objects that interact, how they interact, and the outcomes of their interactions. At the molecular level it is common to focus on the interactions between proteins and DNA (genes) that control gene expression (such as we have discussed in the context of the lac operon). These molecular level interactions play an important role in determining how cells behave. Interactions between cells influence the behaviors of the interacting cell, as well as the overall behavior(s) of biological communities and multicellular organisms. Interactions between organisms, ranging from mutual dependencies to host-pathogen and predator-prey interactions, underlie social and ecological systems. Interaction systems are complex. For example, interactions between cells will influence both lower (molecular level) and higher (organismic and social) systems. Moreover systems change over time and will respond to environmental perturbations in various, often unexpected ways. Systems thinking provides an analytical context to consider biological systems at all levels, from the gene to the ecosystem.

Microbial communities

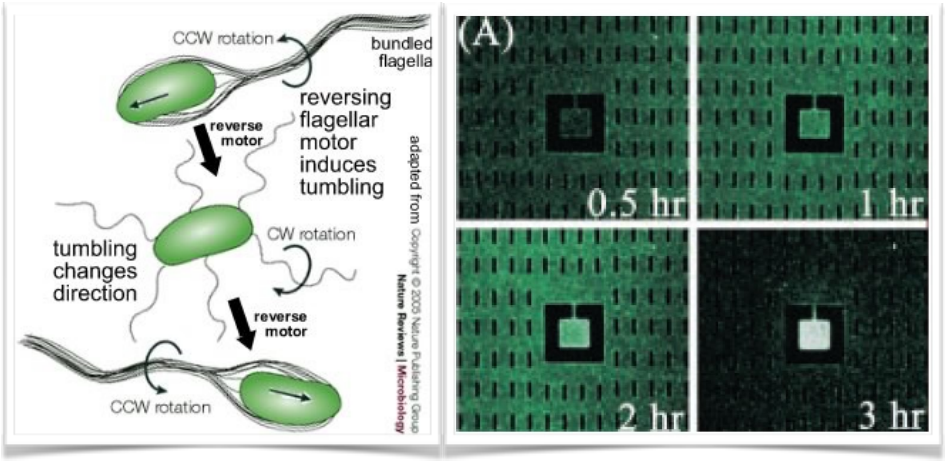
The organisms within a particular community are often critically dependent upon one another. Some organisms will secrete nutrients that are needed by others for their survival. Our own need for vitamins obtained from our diet reflects this interdependence. Some organisms secrete toxins to control the growth of others. Some will secrete molecules that influence the behaviors of other organisms (including themselves). There are complex molecular level conversations going on between the organisms in a ecosystem and the cells within an organism. Organisms are not independent, their behaviors are altered by their environment and they in turn, alter their environment.

An example of how even the simplest organisms can cooperate is an effect known as quorum sensing (which we have mentioned previously.) A bacterium of a particular species can secrete factors that are useful, for example, in the digestion of food into soluble nutrient molecules that it can ingest. But when growing in sparse situations (few organisms per unit volume or area), such a strategy is not efficient. For example if organisms are at low density, expensive to produce secreted molecules are more likely to diffuse away, and so be useless to the organism that produced them. However if there are large numbers of the organisms present, then the process becomes more efficient, the concentration of the secreted molecules will increase dramatically, reaching useful levels. By cooperating with their neighbors to produce a mutually beneficial behavior, each individual benefits.

How might this type of cooperation work? In bacteria a common strategy is for individuals to produce and secrete small (energetically inexpensive) molecules known as auto-inducers. They also

produce a cellular receptor for this same auto-inducer molecule. The auto-inducer-receptor pair enables organisms of the same type to recognize each other. The system works because the level of auto-inducer produced by a single bacterium is not sufficient to activate its receptors; only when the density of auto-inducer-secreting bacteria reaches a threshold level does the concentration of auto-inducer rise to a level high enough to activate the receptors. Activation of the auto-inducer-receptor generates a signal that in turn influences the bacterium's behavior (and generally gene expression).[254] One obvious behavior could be the secretion of digestive enzymes, but there are a number of others. For example, some types of bacteria (including *E. coli*) use quorum sensing to control cell migration. Over time individual cells migrate using their swimming system. One such system relies on flagellar (rotary) motors (driven by electrochemical gradients) to move the cell forward. In the absence of such a gradient, the motor reverses, this causes the cell to tumble and change direction. When moving up a gradient of attractant (or down a gradient of repulsant) tumbling is suppressed; the end result is directed movement.

This type of behavior has been illustrated dramatically by using *E. coli* that contain a plasmid that encodes the Green Fluorescent Protein (GFP). When illuminated with blue light, a cell expressing GFP enables glows green![255] When GFP-expressing *E. coli* are cultured in a maze-like environment with a central "chamber" with a single opening, the secreted attractant will accumulate to high concentrations within this space. Over a three hour period the bacteria will swim in a directed manner up the attractant concentration gradient into the chamber.[256] At this point quorum sensing behaviors will be activated. For example in situations where nutrients become scarce, a quorum sensing controlled behavior can lead some of the cells in the population to die, a process known as programmed cell death, releasing their nutrients for their neighbors to use. This can be seen as a type of altruism, since it helps the neighbors, who are likely to be relatives of the sacrificing cell.[257] Another type of behavior occurs under condition of stress, a subpopulation of cells will form slow or non-growing cells, known as quiescent or "persister" cells, while the rest of the population continues to grow.[258] If the environment turns seriously hostile, the persisters have a much higher probability of survival than do the actively growing cells. If conditions



[254] Bacterial quorum-sensing network architectures: http://www.ncbi.nlm.nih.gov/pubmed/19686078

[255] The original green fluorescent protein evolved in jelly fish Aequorea victoria, it is one of a multigene family of fluorescent proteins: see GFP-like Proteins as Ubiquitous Metazoan Superfamily: Evolution of Functional Features and Structural Complexity: http://www.ncbi.nlm.nih.gov/pubmed/14963095.

[256] Motion to Form a Quorum: http://www.ncbi.nlm.nih.gov/pubmed/12855801

[257] Programmed cell death in bacteria and implications for antibiotic therapy: http://www.ncbi.nlm.nih.gov/pubmed/23684151

[258] "Persisters": Survival at the Cellular Level: http://www.ncbi.nlm.nih.gov/pubmed/21829345

improve the persisters can reverse their behavior and reestablish an actively growing population. On the other hand, if the conditions never get hostile, the growing cells have an evolutionary advantage over cells that go quiescent. This implies the presence of a system can produce persisters when they might be useful. The ability of an organism to produce quiescent persister state helps insure the survival of the population within a wider range of environments than would be expected in a population that cannot produce persisters. This is a example of group selection. A similar behavior has been found to occur within populations of cancer cells.[259] Persister cells can survive therapeutic treatments and re-emerge later. We have already seen, in the context of the *lac* operon, how an initially uniform population of organisms can produce distinct phenotypes through stochastic processes; similar random events play an important role in the determination of cell fates in many social situations.

An important evolutionary question involves what to do with the emergence of social cheaters? First, what exactly do we mean by a social cheater? In the context of quorum sensing, suppose an individual does not make the auto-inducer, but continues to make its receptor. It gains the benefits of communicating with other bacteria, but minimizes its contribution. It might well gain an advantage in that the energy used to make the auto-inducer could instead be used for growth and reproduction. There are limits to cheating, however. If enough members become cheaters the quorum sensing system will fail because not enough members of the community secrete the auto-inducer. There are other more pro-active strategies that can be used to suppress cheaters. It may be that the production of the auto-inducer is a by-product of an essential reaction. In this case, loss of the ability to produce the auto-inducer could itself lead to death. A second approach is more pro-active. For example, many bacterial species synthesize toxins to which they themselves are immune, but which kill cells of related species. It could be that toxin immunity could be coupled to auto-inducer expression. Social cooperation between cells can provide benefits, but also opens up the system to selfish cheaters.[260] Cancer, and the mechanisms to suppress it, is a particularly prominent example of cheater and anti-cheater behaviors.

Making metazoans

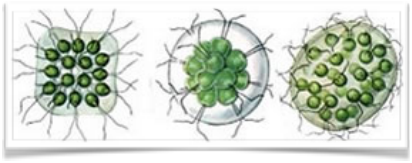
As we think about biological communities we begin our movement from biofilms and other ecologies to discrete systems, that is, what we think of as organisms. First, let us make it clear, a biofilm or microbial mat is not an organism, it is more correctly termed an ecological system or community, composed of distinct organisms, each of which gives rise to organisms genetically related to their parent(s). While horizontal gene transfer between organisms may occur to various extents, the idea of distinct organisms is still valid.

The next obvious level of organization is what we will call a colony. In colonial organisms individual cells are attached to one another, generally through the extracellular materials they secrete. They gain advantages associated with larger size (for example, they may be able to swim faster or being too big to swallow) but these advantages are constrained by the fact that the individual cells

[259] Evolution of cooperation among tumor cells: http://www.ncbi.nlm.nih.gov/pubmed/16938860

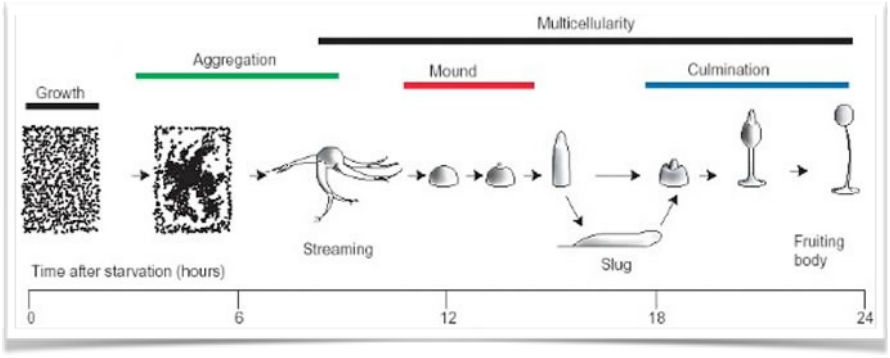
[260] Safeguards for cell cooperation in mouse embryogenesis shown by genome-wide cheater screen:http://www.ncbi.nlm.nih.gov/pubmed/24030493

retain their individuality. For example in colonial forms of algae there is no central coordination between the beating of neighboring cells. Moreover, in a pure colonial organism, each cell within the colony retains its ability to reproduce independently, either sexually or asexually. Previously we introduced the terms soma for the cells of the body that reproduce asexually and are responsible for the growth and repair of the organism, and the germ line, that is, the cells that are responsible for producing the next generation of organisms. In a purely colonial organism, all cells are potential germ cells. There is no central system for coordinating behavior.



So we might ask, what is the path from individual cells to integrated multicellular organisms? In general we think that the earliest step is likely to have been colonial organization. Some organisms can be used as part of a modern bestiary to illustrate various behaviors on the way to multicellular organisms.[261] This is not to claim that any represent real ancestors, all are modern organisms, well adapted to their current environment and the result of their own evolutionary history. Never the less, they have dealt with various aspects of multicellular coordination and differentiation in interesting ways.

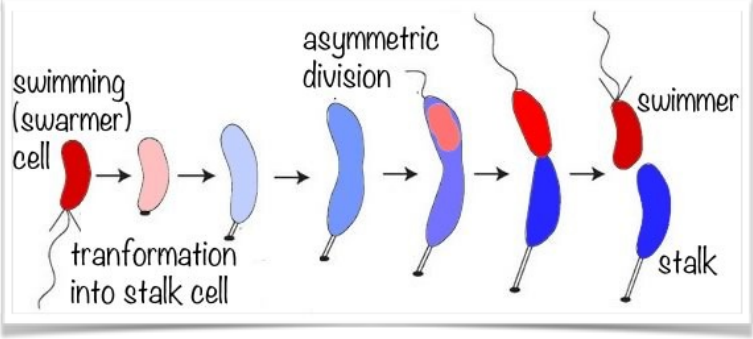
Consider the eukaryotic slime mold *Dictyostelium discoideum*. Cellular slime molds live in soil and eat bacteria - they are unicellular predators. Most of the time they are small, amoeba-like, haploid cells. Upon starvation they can undergo a dramatic aggregation process. Aggregation is triggered by the release, from individual cells, of pulses of cyclic adenosine monophosphate (cAMP); a process analogous to quorum sensing in bacteria (see above). The result is that individual cells begin to migrate up the cAMP concentration gradient, where they interact with and adhere to one another. Groups of cells produce more cAMP, and the end result are cellular aggregates, known as slugs, that contain between 10,000 to 100,000 discrete cells. Slugs migrate in a coordinated manner. Eventually the slug will stop its migration and begin a process of differentiation. Some of the cells of the slug differentiate to form stalk cells; the coordinated elongation of these stalk cells lifts the rest of the slug "body" into the air. The non-stalk cells differentiate to form spores, cells like the quiescent persisters we mentioned above. When released into the air, the spores are widely dispersed and, if they land in an appropriate environment, can go on to form single celled amoebae.



By now you may be able to generate a plausible scenario to explain exactly how the self-sacrificing behavior of stalk cells is possible. The answer lies in inclusive fitness. The purpose of the slug and stalk are to enable *Dictyostelium* cells to escape a hostile environment and colonize new, more hospitable environments. In fact, in a number of cases the spores carry with them bacteria that inoculate their new environments; these are bacteria that the amoeba can eat. The slime mold could be

[261] The medieval bestiary: http://bestiary.ca

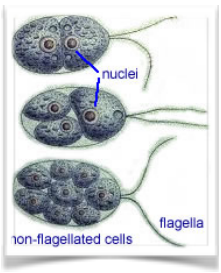
considered migrating bacterial farmers.[262] Since individual *Dictyostelium* amoeboid cells can not migrate far, most of the cells in any particular region, that is the cells that combine to form a slug, are likely to be closely related to one another - they are part of a clone. The sacrifice of the stalk cells is more than made up for by the increased chance that the spore cells will survive and produce lots of offspring. Of course there is a danger that some cells will diverge (through mutation) and cheat the system. That is, they will avoid becoming stalk cells. Such cheating has been observed in wild type *Dictyostelium* and cheating is a challenged faced by all multicellular systems. There are a number of strategies that are used to suppress cheaters, generally they are similar to those exploited in the context of quorum sensing.[263]



An organism that displays a distinct type of differentiation behavior is the bacterium *Caulobacter crescentus*. Under certain conditions it will produce stalk cells. These cells attach to a surface and divide to produce swimming cells that can migrate away and colonize new surfaces. The stalk cell can continue to produce swimming cells, and the swimming cells can settle down and transform into stalk cells. *C. crecentus* has established two different cell types designed to exploit two distinct environments.

Steps to metazoans multicellular animals and plants

As we think about how organisms can increase in complexity, there are really only a few strategies available. One way is to generate very complex unicellular organisms. This strategy is limited, however, and organisms of this type are generally small, only a few hundred micrometers in length. The alternative path to complexity is through multicellularity, which appears to have occurred around 1 billion years ago. In true multicellular organisms (as opposed to colonial organisms), different cells become highly specialized. Most cells are relieved of the need to produce a new organism; that task is taken up by specialized cells in the germ line. As noted above, this allows for the formation of cells with very limited, but highly useful abilities.



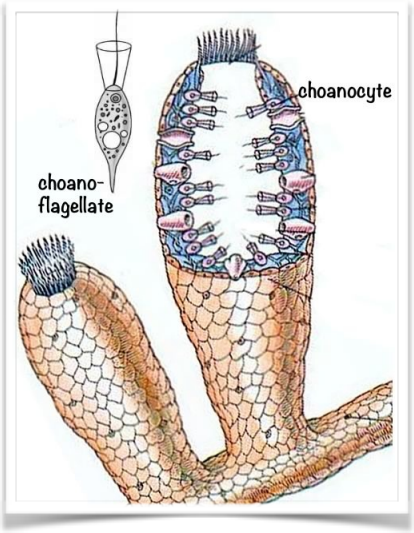
To get a better idea of the evolutionary history of multicellularity it is helpful to look in detail at the organization, both cellular and genomic, of current organisms. It has been estimated that multicellularity arose multiple times among the eukaryotes.[264] To begin to understand the steps in the process it is useful to consider those unicellular organisms most closely related to a particular metazoan lineage (known as a sister group). We can then speculate on the various steps between the unicellular and multicellular forms. In the case of the animals, it appears that their (our) unicellular

[262] **Small molecules mediate bacterial farming by social amoebae**. http://www.ncbi.nlm.nih.gov/pubmed/23975931

[263] Kin Recognition Protects Cooperators against Cheaters: http://www.ncbi.nlm.nih.gov/pubmed/23910661

[264] Multicellularity arose several times in the evolution of eukaryotes: http://www.ncbi.nlm.nih.gov/pubmed/23315654

sister group are the choanoflagellates.[265] Choanoflagellates have cells that are characterized by a single flagellum surrounded by a distinctive collar structure.[266] Choanoflagellates exist in both unicellular and simple colonial forms.



Sponges (porifera) are among the simplest of the metazoans. Fossils of extinct sponges, such as the Archaeocyathids, are found in Cambrian rock that is over 500 million years old. Earlier sponge-like organisms have been found in even older Precambrian rock. Sponges contain only a few different types of cells. These include the cells that form the outer layer of the organism (pinococytes) and the cells (porocytes) that form the pores in the organism's outer layer. The skeletal system of the sponge, the spicules, are produced by sclerocytes. A distinct type of cell (archaeocytes) function in digestion, gamete production, tissue repair and regeneration. Sponges also include cells, known as choanocytes, that move fluid through the body. It is the striking resemblance of these cells to the unicellular choanaflagellates (and subsequent genomic analyses) that led to the hypothesis that choanoflagellates and animals are sister groups.[267]

The next level of metazoan complexity is represented by hydra and related organisms, the hydrozoa, which include jellyfish. Some of these organisms alternate between a sessile and benthic, or floating, lifestyles.[268] The hydrozoa contain more distinct cell types than the porifera. The most dramatic difference is their ability to produce coordinated movements associated with swimming and predation. While sponges are passive sieves, the hydrozoa have a single distinct mouth, an internal stomach-like cavity, and motile arms specialized to capture prey. Their mouth also serves as their anus, through which wastes are released.

Hydrozoan movements are coordinated by a network of cells, known as a nerve net, that acts to regulate contractile muscle cells. Together the nerve net and muscles cells generate coordinated movements, even though there is no central brain (which in its simplest form is just a dense mass of nerve cells). A hydra can display movements complicated enough to capture and engulf small fish. Stinging cells, nematocysts, are located in the “arms". Triggered by touch, they explode outward, embedding themselves in prey and delivering a paralyzing poison.[269] Hydrozoans are complex enough to be true predators.

---

[265] http://www.nytimes.com/2010/12/14/science/14creatures.html?_r=0

[266] Introduction to the Choanoflagellata: http://www.ucmp.berkeley.edu/protista/choanos.html

[267] The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans: http://www.ncbi.nlm.nih.gov/pubmed/18273011

[268] The live cycle of jellyfish: http://youtu.be/oHiVA9J_YIM

[269] How do jellyfish sting: http://youtu.be/HyIwa7W-ZV8

Questions to answer & ponder:

- What types of signals do human send and receive?
- How would changes in the affinity of an auto-inducer receptor influence the behavior of an organism?
- Why might an organism grow well in a biofilm but not in an isolated monoculture?
- In the case of a cellular slime mold, what is the advantage of multicellularity?
- Why do *Dictyostelium* stalk cells "sacrifice themselves" for fruiting body cells?
- Does coordinated movement require a brain?
- Does having a brain equal self-awareness?
- What types of evidence suggest that choanoflagellates and sponges are related?
- Why is the presence of highly specialized cells evidence for common ancestry?
- In terms of cell types and functions, how do a hydra and a sponge differ from one another?
- What kind of evidence, in modern organisms, might lead you to conclude that the last common ancestor of plants and animals had flagella?
- What are the advantages of a closed gut versus a sieve?

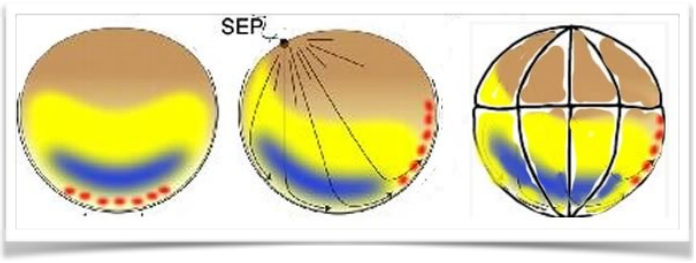
Differentiation

Complex organisms, from worms to humans, undergo a process known as embryonic development. This process begins with the fusion of a haploid sperm and a haploid egg (produced through meiosis) to form a new diploid organism. This cell then divides (by mitosis) to produce the embryo which develops into an adult. Cell division leads to embryonic cells that begin to behave differently from one another. For example, while the original diploid cell generated by fertilization (the zygote) is totipotent - that is, it can generate all of the cells found in the adult, the cells formed during development become more and more restricted with respect to the types of progeny that they can produce–they become committed to one or another specific fate. In part this is due to the fact that as cells divide, different ells come to have different neighbors and they experience different environments, leading to the expression of different genes. The question now becomes, what determines what types of cells does an embryonic cell produce in the adult?

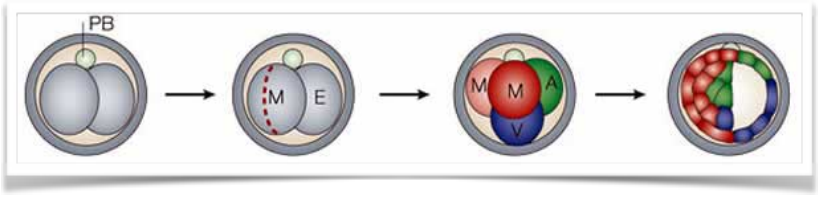
There are two basic, and interacting, processes that drive embryonic development. During the formation of the egg and following fertilization, cytoplasmic determinants (which may be proteins, RNAs, or metabolic products) can become localized to, or active in, specific regions of the egg, and later to specific regions of the embryo. The presence of these cytoplasmic determinants drives the cell that contains them in a specific developmental direction. This developmental direction is based on changes in gene expression. The second set of processes involved in embryonic development are the changing interactions between cells. These involve adhesive interactions and intercellular signals. They can direct a cell to adopt specific fates. There are many different types of embryonic development, since this stage of an organisms life cycle is as subject to the effects of evolutionary pressures as any other (although it is easy to concentrate our attentions on adult forms and behaviors). The study of these processes, known as embryology, is beyond our scope here, but we can outline a few common themes.

If fertilized eggs develop outside of the body of the mother and without parental protection, then these new organisms are highly vulnerable to predation. In such organisms, early embryonic development proceeds rapidly. The eggs are large and contain all of the nutrients required for development to proceed up to the point where the new organism can feed on its own. To facilitate such

rapid development, the egg is essentially pre-organized, that is, it is highly asymmetric, with specific factors that can influence gene expression, either directly or indirectly, positioned in various regions of the egg. Entry of the sperm (the male gamete), which itself is an inherently asymmetric process, can also lead to reorganization of the cytoplasm. Maternal and fertilization-driven asymmetries are stabilized by the rapid cycles of DNA replication and cell division, with growth being dependent upon the transformation of maternally supplied nutrients. As distinct cells are formed, they begin to become different from one another because i) they inherit different determinants, ii) the presence of these determinants leads to changes in gene expression, and iii) cells will secrete and respond to different factors, that further drive their differentiation into different cell types, with different behaviors based on differences in gene expression.



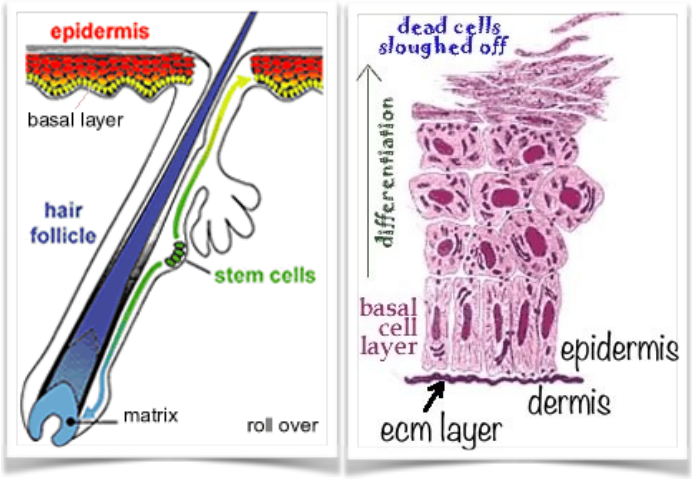
On the other hand, in a number of organisms, and specifically mammals, embryonic development occurs within the mother, so there is no compelling need to stockpile nutrients within the egg and the rate of development is dramatically slower. In such developmental systems, it is not the asymmetries associated with the oocyte and fertilized egg that are critical, but rather the geometries of the cells within the developing embryo. As the zygote divides, a major factor that drives the differentiation of cells is whether they lie on the surface of the embryo or within the interior. In mammals, the cells on the exterior form the trophectoderm, which goes on to form extraembryonic tissues, in particular the membranous tissues that surround the embryo and become part of the placenta, the interface between the embryo and the mother. Cells within the interior form the inner cell mass that produces to the embryo proper. Changes in gene expression will lead to changes in the ability to produce and respond to inductive signals, which will in turn influence cell behavior and gene expression. Through this process, the cells of the inner cell mass come to form the various tissues and organs of the organism; that is, skin, muscle, nerve, hair, bone, blood, etc. It is easy to tell a muscle cell from a neuron from a bone cell from a skin cell by the set of genes they express, the proteins they contain, their shapes (morphology), their internal organization, and their behaviors.



Stem cells

Stem cells are cells that continue to divide in the adult, but they divide in a very particular manner. At each division cycle, one daughter cell remains a stem cell, while the other goes on to differentiate. In part, this is due to the environment in which the stem cell finds itself, which is known as the stem cell niche. For example, in mammals, the stem cells that lead to the continuous regeneration of the skin and hair are located in a region of the hair follicle, known as the bulge. These cells divide rarely, with one daughter migrating away from the bulge and the other remaining in place. The migrating daughter cell will come to colonize the basal layer of the epidermis, where it continues to divide a number of times. Again, this is a stem cell-like division; the cells that remain attached to the

extracellular matrix layer remaining stem cell like, while those that leave the “basal cell layer” begin the process of differentiation that leads, eventually, to their death (you are constantly shedding dead skin cells.) In normal skin the process of cell birth and death is balanced. Hyperplasia occurs when cell birth occurs more frequently than cell death. Typically the non-stem cell products of a stem cell division are committed to differentiation and have a finite proliferative life span - they can divide only a limited number of times before they senesce (that is, stop dividing). Terminally differentiated cells no longer divide. The process of cellular senescence is thought to be an internal defense mechanism against cancer; often cancer cells accumulate mutations than enable them to circumvent the effects of senescence.



Cellular differentiation and genomic information

An important question that was asked by early developmental biologists was, is cellular differentiation due to the loss of genetic information. Is the genetic complement of a neuron different from a skin cell or a muscle cell? This question was first approached by Briggs and King in the 1950s through nuclear transfer experiments in frogs. These experiments were extended by Gurdon and McKinnell in the early 1960s. They were able to generate adult frogs via nuclear transfer using embryonic cells. The process was inefficient however - only a small percentage of transferred nuclei supported normal embryonic development. Nevertheless, these experiments suggested that it was the regulation rather than the loss of genetic information that was important in embryonic differentiation.

In 1996 Wilmut et al used a similar method to clone the first mammal, the sheep Dolly. Since then many different species of mammal have been cloned, and there is serious debate about the cloning of humans. In 2004, cloned mice were derived from the nuclei of olfactory neurons using a method similar to that used by Gurdon. These neurons came from a genetically engineered mouse that expressed GFP (see above). A hybrid gene contained the coding sequence for GFP and a regulatory sequence that led to its expression in most cell types of the mouse. Neuronal nuclei were transplanted into an oocyte from which the original nucleus had been removed (an enucleated oocyte). Blastula derived from these cells were then used to generate totipotent embryonic stem cells. It was the nuclei from these cells that were transplanted into enucleated eggs. The resulting embryos were able to develop into full grown and fluorescent mice, proving that neuronal nuclei retained all of the information required to generate a complete adult animal.

The process of cloning from somatic cells is inefficient – many attempts have to be performed, each using an egg, to generate an embryo that is apparently normal (most embryos produced this way were abnormal). At the same time, there are strong ethical concerns about the entire process of reproductive cloning. For example the types of cells used, embryonic stem cells, are derived from the

inner cell mass of mouse or human embryos. Embryonic stem cells can be cultured in vitro and under certain conditions can be induced to differentiate into various cell types. Since the generation of totipotent human embryonic stem cells involves the destruction of a human embryo, it raises a number of ethical issues.

Current research attempts to avoid these issues by focussing on optimizing the process by which somatic nuclei can be reprogrammed to support totipotent and pluripotent development. In this scenario, somatic cells from a patient are treated with genes (or more recently gene products) for a small number (typically) four molecules to induce differentiated somatic cells to become pluripotent cells. These “induced pluripotent stem” (iPS) cells behave much like embryonic stem cells. The hope is that a iPS cells derived from a patient could be used to generate tissues or even organs that could be transplanted back into the patient, and so reverse and repair disease-associated damage.

Questions to answer & to ponder:

- Are the advantage(s) of multicellularity the same for plants versus animals ?
- How might asymmetries be generated in the zygote?
- Why do differentiated cells express different genes than do undifferentiated cells?
- How could two cells that express the same set of transcription factors, express different genes?
- In terms of transcription factors and chromatin packing, why is it difficult to reverse differentiation?
- What is the primary characteristic of a stem cell?
- Why might the organism want to reduce the number of stem cells it contains?
- Based on your understanding of the control of gene expression, outline the steps required to reprogram a nucleus so that it might be able to support embryonic development.
- What is necessary for cells to become different from one another - for example how do muscle cells and skin cells come to be different from one another?
- What are the main objections to human cloning? What if the clone were designed to lack a brain, and destined to be used for "spare parts"?
- How would a clone be different from a twin?
- How do we "check" whether our reading of another's emotions are correct.?
- Would different types of social groups have different types of morality?
- Does social evolution explain morality?
- Is the next step in evolution the evolution of eusocial humans? speculate (please)