

BIOFUNDAMENTALS



Klymkowsky & Cooper
University of Colorado Boulder &
Michigan State University

University of Colorado Boulder & Michigan
State University
Biofundamentals

Klymkowsky & Cooper

This open text is disseminated via the Open Education Resource (OER) LibreTexts Project (<https://LibreTexts.org>) and like the hundreds of other open texts available within this powerful platform, it is licensed to be freely used, adapted, and distributed. This book is openly licensed which allows you to make changes, save, and print this book as long as the applicable license is indicated at the bottom of each page.

Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



The LibreTexts mission is to unite students, faculty and scholars in a cooperative effort to develop an easy-to-use online platform for the construction, customization, and dissemination of OER content to reduce the burdens of unreasonable textbook costs to our students and society. The LibreTexts project is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning by developing an Open Access Resource environment. The project currently consists of 13 independently operating and interconnected libraries that are constantly being optimized by students, faculty, and outside experts to supplant conventional paper-based books. These free textbook alternatives are organized within a central environment that is both vertically (from advance to basic level) and horizontally (across different fields) integrated.

The LibreTexts libraries are [Powered by MindTouch®](#) and are supported by the Department of Education Open Textbook Pilot Project, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739. Unless otherwise noted, LibreTexts content is licensed by [CC BY-NC-SA 3.0](#).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org. More information on our activities can be found via Facebook (<https://facebook.com/Libretexts>), Twitter (<https://twitter.com/libretexts>), or our blog (<http://Blog.Libretexs.org>).



CSU The California State University
AFFORDABLE LEARNING
SOLUTIONS (AL\$)

UCDAVIS
UNIVERSITY OF CALIFORNIA

UCDAVIS
LIBRARY

MERLOT
Multimedia Educational Resource
for Learning and Online Teaching

This text was compiled on 01/30/2021

TABLE OF CONTENTS

Biofundamentals™ is an attempt to build a more conceptually coherent, rigorous and engaging introductory course in modern biology

1: UNDERSTANDING SCIENCE & THINKING SCIENTIFICALLY

- 1.0: INTRODUCTION
- 1.1: THE INTERCONNECTEDNESS OF SCIENCE
- 1.2: MODELS, HYPOTHESES, AND THEORIES
- 1.3: SCIENCE IS SOCIAL
- 1.4: TEACHING AND LEARNING SCIENCE
- 1.5: UNDERSTANDING SCIENTIFIC IDEAS

2: LIFE'S DIVERSITY AND ORIGINS

- 2.0: INTRODUCTION
- 2.1: WHAT IS LIFE, EXACTLY?
- 2.2: THE CELL THEORY AND THE CONTINUITY OF LIFE
- 2.3: THE ORGANIZATION OF ORGANISMS
- 2.4: SPONTANEOUS GENERATION AND THE ORIGIN OF LIFE
- 2.5: THE DEATH OF VITALISM
- 2.6: THINKING ABOUT LIFE'S ORIGINS
- 2.7: EXPERIMENTAL STUDIES ON THE ORIGINS OF LIFE
- 2.8: MAPPING THE HISTORY OF LIFE ON EARTH
- 2.9: FOSSIL EVIDENCE FOR THE HISTORY OF LIFE ON EARTH
- 2.10: LIFE'S IMPACT ON EARTH

3: EVOLUTIONARY MECHANISMS AND THE DIVERSITY OF LIFE

- 3.0: INTRODUCTION
- 3.1: ORGANIZING ORGANISMS (HIERARCHICALLY)
- 3.2: NATURAL AND UN-NATURAL GROUPS
- 3.3: FOSSILS AND FAMILY RELATIONSHIPS: INTRODUCING CLADISTICS
- 3.4: EVOLUTION THEORY'S CORE CONCEPTS
- 3.5: SO WHAT DO WE MEAN BY GENETIC FACTORS?
- 3.6: LIMITS ON POPULATIONS
- 3.7: THE CONCEPTUAL LEAP MADE BY DARWIN AND WALLACE
- 3.8: MUTATIONS AND THE ORIGINS OF GENOTYPE-BASED VARIATION
- 3.9: THE ORIGINS OF POLYMORPHISMS
- 3.10: A SHORT ASIDE ON THE GENOTYPE-PHENOTYPE RELATIONSHIP
- 3.11: VARIATION, SELECTION, AND SPECIATION
- 3.12: TYPES OF SIMPLE SELECTION
- 3.13: A SHORT NOTE ON PEDAGOGICAL WEIRDNESS
- 3.14: POPULATION SIZE, FOUNDER EFFECTS AND POPULATION BOTTLENECKS
- 3.15: POPULATION BOTTLENECKS
- 3.16: GENETIC DRIFT
- 3.17: GENE LINKAGE: ONE MORE COMPLICATION
- 3.18: A BRIEF REFLECTION ON THE COMPLEXITY OF PHENOTYPIC TRAITS
- 3.19: SPECIATION AND EXTINCTION
- 3.20: MECHANISMS OF SPECIATION
- 3.21: ISOLATING MECHANISMS
- 3.22: SYMPATRIC SPECIATION
- 3.23: SIGNS OF EVOLUTION: HOMOLOGY AND CONVERGENCE
- 3.24: THE LOSS OF TRAITS
- 3.25: SIGNS OF EVOLUTIONARY HISTORY
- 3.26: HOMOLOGIES PROVIDE EVIDENCE FOR A COMMON ANCESTOR
- 3.27: ANTI-EVOLUTION ARGUMENTS

4: SOCIAL EVOLUTION AND SEXUAL SELECTION

- 4.0: INTRODUCTION

- 4.1: SELECTING SOCIAL (COOPERATIVE) TRAITS
- 4.2: QUORUM SENSING
- 4.3: ACTIVE (ALTRUISTIC) CELL DEATH
- 4.4: INCLUSIVE FITNESS, KIN AND GROUP SELECTION, AND SOCIAL EVOLUTION
- 4.5: GROUP SELECTION
- 4.6: DEFENSE AGAINST SOCIAL CHEATERS
- 4.7: DRIVING THE EVOLUTIONARY APPEARANCE OF MULTICELLULAR ORGANISMS
- 4.8: ORIGINS AND IMPLICATIONS OF SEXUAL REPRODUCTION
- 4.9: SEXUAL DIMORPHISM
- 4.10: SEXUAL SELECTION
- 4.11: CURBING RUNAWAY SELECTION

5: MOLECULAR INTERACTIONS, THERMODYNAMICS & REACTION COUPLING

- 5.0: INTRODUCTION
- 5.1: A VERY LITTLE THERMODYNAMICS
- 5.2: REACTIONS: FAVORABLE, UNFAVORABLE, AND THEIR DYNAMICS
- 5.3: THINKING ENTROPICALLY (AND THERMODYNAMICALLY)
- 5.4: REACTION RATES
- 5.5: COUPLING REACTIONS
- 5.6: MOLECULES, LONDON DISPERSION FORCES, AND VAN DER WAALS INTERACTIONS
- 5.7: COVALENT BONDS
- 5.8: BOND STABILITY AND THERMAL MOTION (A NON-BIOLOGICAL MOMENT)
- 5.9: BOND POLARITY, INTER- AND INTRAMOLECULAR INTERACTIONS
- 5.10: THE IMPLICATIONS OF BOND POLARITY
- 5.11: INTERACTING WITH WATER

6: MEMBRANE BOUNDARIES AND CAPTURING ENERGY

- 6.0: INTRODUCTION
- 6.1: DEFINING THE CELL'S BOUNDARY
- 6.2: THE ORIGIN OF BIOLOGICAL MEMBRANES
- 6.3: TRANSPORT ACROSS MEMBRANES
- 6.4: TRANSPORT TO AND ACROSS THE MEMBRANE
- 6.5: CHANNELS AND CARRIERS
- 6.6: GENERATING GRADIENTS: USING COUPLED REACTIONS AND PUMPS
- 6.7: SIMPLE PHOTOTROPHS
- 6.8: CHEMO-OSMOSIS (AN OVERVIEW)
- 6.9: OXYGENIC PHOTOSYNTHESIS
- 6.10: CHEMOTROPHS
- 6.11: USING THE ENERGY STORED IN MEMBRANE GRADIENTS
- 6.12: OSMOSIS AND LIVING WITH AND WITHOUT A CELL WALL
- 6.13: AN EVOLUTIONARY SCENARIO FOR THE ORIGIN OF EUKARYOTIC CELLS
- 6.14: MAKING A COMPLETE EUKARYOTE

7: THE MOLECULAR NATURE OF HEREDITY

- 7.0: INTRODUCTION
- 7.1: DISCOVERING HOW NUCLEIC ACIDS STORE GENETIC INFORMATION
- 7.2: LOCATING HEREDITARY MATERIAL WITHIN THE CELL
- 7.3: IDENTIFYING DNA AS THE GENETIC MATERIAL
- 7.4: UNRAVELING NUCLEIC ACID STRUCTURE
- 7.5: DISCOVERING THE STRUCTURE OF DNA
- 7.6: DNA, SEQUENCES, AND INFORMATION
- 7.7: DISCOVERING RNA: STRUCTURE AND SOME FUNCTIONS
- 7.8: DNA REPLICATION
- 7.9: REPLICATION MACHINES
- 7.10: FURTHER REPLICATION COMPLEXITIES
- 7.11: MUTATIONS, DELETIONS, DUPLICATIONS & REPAIR
- 7.12: GENES AND ALLELES
- 7.13: MUTATIONS AND EVOLUTION
- 7.14: TRIPLET REPEAT DISEASES AND GENETIC ANTICIPATION

8: PEPTIDE BONDS, POLYPEPTIDES & PROTEINS

- 8.0: INTRODUCTION
- 8.1: POLYPEPTIDE AND PROTEIN STRUCTURE BASICS
- 8.2: AMINO ACID POLYMERS
- 8.3: MAKING A POLYPEPTIDE IN A BACTERIAL CELL
- 8.4: THE ORIGIN OF THE GENETIC CODE
- 8.5: PROTEIN SYNTHESIS: TRANSCRIPTION (DNA TO RNA)
- 8.6: PROTEIN SYNTHESIS: TRANSLATION (RNA TO POLYPEPTIDE)
- 8.7: THE TRANSLATION (POLYPEPTIDE SYNTHESIS) CYCLE
- 8.8: BURSTING SYNTHESIS AND ALARM GENERATION
- 8.9: GETTING MORE COMPLEX: GENE REGULATION IN EUKARYOTES
- 8.10: TURNING POLYPEPTIDES INTO PROTEINS
- 8.11: REGULATING PROTEIN LOCALIZATION
- 8.12: REGULATING PROTEIN ACTIVITY
- 8.13: ALLOSTERIC REGULATION
- 8.14: POST-TRANSLATIONAL REGULATION
- 8.15: DISEASES OF FOLDING AND MISFOLDING
- 8.16: WHY DO HARMFUL ALLELES PERSIST?

9: GENOMES, GENES & REGULATORY NETWORKS

- 9.0: INTRODUCTION
- 9.1: GENOMES AND THEIR ORGANIZATION
- 9.10: PARALOGOUS GENES AND GENE FAMILIES
- 9.11: PACKING DNA INTO A CELL
- 9.12: LOCATING INFORMATION WITHIN DNA
- 9.13: NETWORK INTERACTIONS
- 9.14: E. COLI AS A MODEL SYSTEM
- 9.15: FINAL THOUGHTS ON (MOLECULAR) NOISE
- 9.16: TYPES OF REGULATORY INTERACTIONS
- 9.2: A TECHNICAL ASIDE - TRANSPOSONS
- 9.3: GENES ALONG CHROMOSOMES
- 9.4: NATURALLY OCCURRING HORIZONTAL GENE TRANSFER MECHANISMS
- 9.5: TRANSFORMATION
- 9.6: CONJUGATION
- 9.7: TRANSDUCTION
- 9.8: SEXUAL REPRODUCTION
- 9.9: GENOME DYNAMICS

10: SOCIAL SYSTEMS

- 10.0: INTRODUCTION TO SOCIAL SYSTEMS
- 10.1: MICROBIAL COMMUNITIES
- 10.2: MAKING METAZOANS
- 10.3: STEPS TO METAZOANS MULTICELLULAR ANIMALS AND PLANTS
- 10.4: DIFFERENTIATION
- 10.5: STEM CELLS
- 10.6: CELLULAR DIFFERENTIATION AND GENOMIC INFORMATION

BACK MATTER

- [INDEX](#)
- [GLOSSARY](#)

CHAPTER OVERVIEW

1: UNDERSTANDING SCIENCE & THINKING SCIENTIFICALLY

Topic hierarchy

- 1.0: INTRODUCTION
- 1.1: THE INTERCONNECTEDNESS OF SCIENCE
- 1.2: MODELS, HYPOTHESES, AND THEORIES
- 1.3: SCIENCE IS SOCIAL
- 1.4: TEACHING AND LEARNING SCIENCE
- 1.5: UNDERSTANDING SCIENTIFIC IDEAS



1.0: Introduction

In which we consider what makes science a distinct, productive, and progressive way of understanding how the universe works and how science lets us identify what is possible and plausible from what is impossible. We consider the “rules” that distinguish a scientific approach to a particular problem from a non-scientific one.

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*¹⁰. In science, we do not accept that an observation or a conclusion is true just 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 phenomenon, 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 *de novo*. In 1665, Isaac Newton (1642–1727) performed a series of experiments that he interpreted as demonstrating that white light was not pure, but in fact was composed of light of different colors¹¹. 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 one prism so that passed through a second prism; a beam of white light emerged from the second prism. One could go on to show that the light emerging from the prism 1 lens prism 2 combination behaved the same as the original beam of white light by passing it through a third prism, which again produced a spectrum. In the second type of experiment, 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 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 nor 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 phenomenon, 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 reproduce, confirm, and extend his observations. If you have access to glass prisms and lenses, you can repeat Newton’s experiments yourself, and you will come to the same empirical conclusions; that is, you would observe the same phenomena that Newton did¹². In 1800, William Herschel (1738-1822) did just that. He used Newton’s experimental approach and discovered infrared (beyond red) light. Infrared light is invisible to us but its presence can be revealed by the fact that when absorbed by an object, say by a thermometer, it leads to an increase in the temperature of the object¹³. 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 light¹⁴. Subsequent researchers 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, that is, not composed of other types of light. His view was 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 the different colors of light 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 passes 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.

References

¹⁰ The Age of Reason: www.ushistory.org/paine/reason/singlehtml.htm

¹¹ Newton's Prism Experiments: <http://micro.magnet.fsu.edu/primer/j...pticsu/newton/> & http://youtu.be/R8VL4xm_3wk

¹² Infrared astronomy: coolcosmos.ipac.caltech.edu/c...discovery.html

¹³ There are some animals that can see infrared light: <https://sciencing.com/animals-can-se...t-6910261.html> and <http://news.sciencemag.org/plants-an...infrared-light>

¹⁴ Ritter discovers ultraviolet light: coolcosmos.ipac.caltech.edu/c...itter_bio.html

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

1.1: 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; 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.¹⁵ Alternatively, 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 biological processes 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 on these rules. It may well be that life, as it exists on Earth, is unique. The only way we will know otherwise is if we 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 or are based on untestable assumptions, something that some would argue is the case for string theory in physics, are no longer science, whether or not they are “true” in some other sense.¹⁶

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

1.2: Models, Hypotheses, and Theories

Tentative scientific models are commonly known as hypotheses. Such models are valuable in that they serve as a way to clearly articulate one's assumptions and their implications. 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 forms of the model 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, also known as the Principle of Parsimony, named after the medieval philosopher William of Occam (1287–1347). This rule states that all other things being equal, the simplest explanation is to be preferred. 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 phenomenon, 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, 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, or interact with the physical 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 talk by Murray Gell-Mann, is that simple, albeit often counter-intuitive rules can explain much of the Universe with remarkable precision.¹⁷ 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 science 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 many different observations is known as a theory – at least this is the meaning of the word in a rigorous 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 phenomenon, it is 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 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 and measures – that is people who have no vested interest in the outcome – and found to provide accurate descriptions of the phenomenon 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.

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

That said, theories are not static. New or more accurate observations that a theory cannot explain will inevitably drive the theory's revision or replacement. 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, a naturalistic explanation for why gravity exists and why it behaves the way it does. He relied, in fact, on a supernatural explanation.¹⁸ 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.¹⁹ So is general relativity true? Not necessarily, which is why scientists continue to test its predictions in increasingly extreme situations.

References

- ¹⁷ Beauty, truth and ... physics?
- ¹⁸ Want to read an interesting biography of Newton, check out “Isaac Newton” by James Gleick
- ¹⁹ [A good video on General Relativity](#)

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

1.3: 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.²⁰ Scientists present their observations, hypotheses, and conclusions 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 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: these include the Physicochemical Theory of Life, the Cell Theory, and the Theory of Evolution.²¹ 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 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 using multiple methods. 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.

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 of 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 100 million years, based on the rate of heat dissipation of a once molten object, the Earth.²² This was a time-span that seemed too short for a number of geological and evolutionary processes, and greatly troubled Charles Darwin. Somebody was wrong, or better put, their 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. The heat released by radioactive decay led to an increase the calculated age of the earth by more than ten to one hundred fold, to ~5 billion years, an age compatible with both evolutionary and geological processes.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

1.4: 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 new 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 that are nevertheless logically unavoidable. While it is now 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 have come to know that these ideas 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 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, the mystic Giordano Bruno (1548–1600) was burned at the stake for holding these and other ideas (some of which are similar to those currently being proposed by string theorists) and Galileo Galilei (1564–1642), known as the father of modern physics, 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.²³ In 1616, the Roman Catholic Church placed Galileo's book, which held that the sun was the center of the solar system, on the list of forbidden books where it remained until 1835.

The idea that we are standing on the surface of a planet that is rotating at ~1000 miles an hour and flying through space at ~67,000 miles per hour is difficult to reconcile with our everyday experience, yet science continues to generate even weirder ideas. Based on observations and logic, it appears that the Universe arose from "nothing" ~13.8 billion years ago.²⁴ 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 and leads to black holes. A range of biological observations indicate 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 appears to be 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 act together to generate consciousness and self-consciousness, contains ~86 billion (10^9) neurons as well as a similar number of non-neuronal (glial) cells. These cells are connected to one another through $\sim 1.5 \times 10^{14}$ connections, known as synapses.²⁵ How exactly such a system produces thoughts, ideas, dreams, feelings, and self-awareness remains obscure, but it appears that these are all emergent behaviors that arise from this staggeringly complex natural system. Scientific ideas, however weird, arise from the interactions between the physical world, our brains, and the social system of science that tests ideas based on their ability to explain and predict the behavior of the observable universe.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

1.5: Understanding Scientific Ideas

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 chemical reaction systems, which in turn requires an understanding of molecules, which rests upon an understanding of how atoms (matter) 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).²⁶ In it, he explains the complexity of understanding 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 the key observations and common rules upon which basic conclusions and working concepts 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 the way it does, based on various facts, observations, or explicit presumptions that logically support your claim. You also need to present your model to others, knowledgeable in the topic, to get their feedback, to answer (rather than ignore or disparage) 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 this 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 is 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 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?²⁷
- 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?
- If “science” concludes that free will is an illusion, would you accept it and behave like a robot?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

2: LIFE'S DIVERSITY AND ORIGINS

2.0: INTRODUCTION

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 that act in a coordinated (social) manner. We consider the origins of organisms, their basic properties, and their relationships to one another.

2.1: WHAT IS LIFE, EXACTLY?

2.2: THE CELL THEORY AND THE CONTINUITY OF LIFE

2.3: THE ORGANIZATION OF ORGANISMS

2.4: SPONTANEOUS GENERATION AND THE ORIGIN OF LIFE

2.5: THE DEATH OF VITALISM

2.6: THINKING ABOUT LIFE'S ORIGINS

2.7: EXPERIMENTAL STUDIES ON THE ORIGINS OF LIFE

2.8: MAPPING THE HISTORY OF LIFE ON EARTH

2.9: FOSSIL EVIDENCE FOR THE HISTORY OF LIFE ON EARTH

2.10: LIFE'S IMPACT ON EARTH



2.0: Introduction

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 that act in a coordinated (social) manner. We consider the origins of organisms, their basic properties, and their relationships to one another.

Biology is the science of organisms, how organisms function, behave, interact, adapt, 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 answers are not simple; they assume a working knowledge of thermodynamics, a complex topic that we address 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 some outcome that takes the input of energy to achieve. In the context of biological systems, work ranges from generating and maintaining molecular gradients and driving other unfavorable, that is energy-requiring reactions, such as the synthesis of a wide range of biomolecules, including nucleic acids, proteins, lipids, and carbohydrates, required for growth, reproduction, the generation of movement, and so on.

We will focus on what is known as free energy, which is energy available to make things happen. When a system is at equilibrium, its free energy is 0, which means that there are no macroscopic (visible) or net changes. 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 from the external world. Organisms are different from other non-equilibrium systems in that they contain a genetic (heritable) component. While other types of 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, no baby hurricanes. In contrast, each organism alive today arose from one or more pre-existing organisms (its parent(s)) 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, and makes organisms unique among physicochemical systems.

Biology has only a few overarching 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 both the diversity of organisms and how populations of organisms can change 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 all physical and chemical systems.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.1: 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 discover that the closer we look the more these different “types of life” are in fact all versions of a common underlying motif, they represent versions of a single 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 no reasonable doubt that all organisms are related, 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 and where we will discover alien life forms. 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, most of the methods used rely on the fact that all known organisms use DNA to encode their genetic information; these methods would not be expected to recognize dramatically different types of life; they certainly would not detect organisms that used a non-DNA method to encode genetic information. If we could generate living systems *de novo* in the laboratory we would have a better understanding of what functions are necessary for life and how to look for possible “non-standard” organisms using better methods. It might even lead to the discovery of alternative forms of life right here on Earth, assuming they exist.²⁸ 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 produces new organisms.²⁹

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, extinction) by accident. Over time, the probability of death for a single individual will approach one—that is, certainty.³⁰ (→) 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 of its 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,³¹ 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 of themselves and to export waste products into their environment.³² 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. The basic barrier or organisms appears to be a homologous structure—that is, it was present in and inherited from their common ancestor. 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 state and its dynamic structure. The boundary must be able to retain the valuable molecules 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. While we assume that you have at least a basic understanding of the laws of thermodynamics, we will review the core ideas in Chapter 5.

We see evidence of the non-equilibrium nature of organisms most obviously in their ability of move, but it is important for all aspects of the living state. In particular, organisms use energy, captured from their environment, to drive a wide range of thermodynamically unfavorable chemical reactions. These reactions are driven by coupling them to thermodynamically favorable reactions. An organism that reaches thermodynamic equilibrium is a dead organism.

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 the environment, a process associated with increased entropy. These non-living systems 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 observe them) be similar to those that form today. Since we understand (more or less) the conditions that produce tornados and hurricanes, we can predict, fairly reliably, the conditions that will lead to their appearance 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 into 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, modern tornados and hurricanes originate anew, they are not derived from parental storms.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.2: 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 was 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) together with 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 aspect of the Cell Theory 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, so the Cell Theory does not say anything about the how life originally began. 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 large body 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 the information present in DNA molecules, as well as careful comparisons of how cells are constructed at the molecular level suggest that there was a single common ancestor for all life 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 an uninterrupted 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 single-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 here 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 sexually reproducing multicellular organisms (including humans), 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 pre-existing 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 present in the living 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 than the organism itself; 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.3: The organization of organisms

Some organisms consist of a single cell, while 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 cannot survive outside the organism nor can it produce a new organism - it depends upon cooperation with the other cells of the organism. 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 underlies 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 (that is, the cells that die when the organism dies) are known as somatic cells.³³ All organisms in the modern world, and for apparently the last ~3.5-3.8 billion years, arise from a pre-existing organism or, in the case of sexually reproducing organisms, from the cooperation of two organisms, an example of social evolution that 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 the disruption of the social system. Cancer is the most obvious example of an anti-social behavior; in evolutionary terms, it can, initially, be rewarded (more copies of the cancerous cell are produced) but ultimately leads to the extinction of the cancer, and often the death of the organism.³⁴ This is because evolutionary mechanisms are not driven by long term outcomes, but only by immediate ones.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.4: 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 view held 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, together with the accumulation of data enables us to conclude 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.³⁵ 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. In contrast, all bets are off if the process is supernatural. 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. 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 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 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.

He who experiments increases knowledge. He who merely speculates piles error upon error.

- Arabic epigraph quoted by Francisco Redi.

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, a range of unicellular photosynthetic and non-photosynthetic eukaryotes.³⁶ 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 Spallanzani (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 obvious 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 air-borne microorganisms to reach the broth. Microbes in the air were trapped in the bended region of the flask's 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 served as what is known as a “positive control” experiment; it indicated that the heating of the broth had not destroyed some vital element needed for standard growth to occur. We carry out positive control experiments to test our assumptions; for example, if we are using a drug in a study, we first need to test to make sure that the drug we have is actually active. In Pasteur’s experiment, if the boiled broth could not support growth (after the flask was broken) we would not expect it to support spontaneous generation, and so the experiment would be meaningless. We will return to the description of a “negative control” experiment later.³⁷

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, or millions, of years to occur? If true, this objection would invalidate Pasteur’s conclusions. Clearly an experiment to address that particular objection has its own practical issues. Nevertheless, the results of various experiments on spontaneous generation have 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 due to 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, essentially everywhere, who are actively eating complex molecules to maintain their non-equilibrium state, to grow and, to 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 minimum time it could take from geological data using the time from 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 Darwin did →) that once organisms had appeared in a particular environment they would suppress any subsequent spontaneous generation events – they would have eaten the molecules needed for the process to occur. 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. Once life was present, only its descendants could survive.

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.

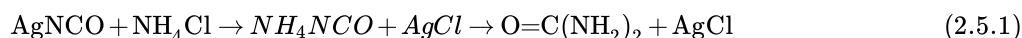
- Charles Darwin (1887).

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.5: 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.³⁸ 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:



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.

While simple, Wohler's *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, the synthesis of urea was a standard chemical process. 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. A huge amount of time and money is devoted to the industrial synthesis of a broad range of organic molecules.

Questions to answer and to ponder:

- Generate 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.6: 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 (we will discuss these processes throughout the book, and more specifically 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 to be wrong, often only a few years after such predictions were made! This speak to the power of science and the technologies based on science; for example, would an intelligent design creationist try to synthesize human proteins in bacteria?

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. More to the point panspermia 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 or the moons in our solar system, and we can retrieve and analyze it, it would be extremely informative, particularly if it were found that this extra-terrestrial life originated independently from life on earth (rather than being transferred from Earth through various astronomical impact events).³⁹

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.7: Experimental studies on the origins of life

One strategy to understanding how life might have arisen naturally 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 an early and influential example of this approach.⁴⁰ 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 other combinations of compounds, more likely to represent the environment of early Earth, with similar results: various biologically important organic molecules accumulate rapidly.⁴¹ 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 ~4.1 through ~3.9 billion years ago, meteorite impacts could have supplied substantial amounts of organic molecules.⁴² It therefore appears likely that early Earth was rich in organic molecules (which are, remember, carbon containing rather than life-derived molecules), the building blocks of life.

Given that the potential building blocks for life 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 (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.⁴³

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 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 accumulated in, 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:

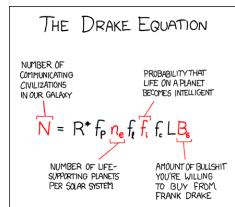
If we assume that spontaneous generation occurred in the distant past, why is it not occurring today? How could you tell if it were?

In 1961, Frank Drake, a radio astronomer, proposed an equation to estimate the number of technological civilizations that exist within the observable Universe (N).⁴⁴ The equation is

$$N = R \times f_p \times n_e \times f_l \times f_i \times f_c \times L \quad (2.7.1)$$

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 = The number planets, per solar system, with an environment suitable for life.
- f_l = The fraction of suitable planets on which life actually appears.
- f_i = 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.



Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.8: 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 nebulae were from Earth. Edwin Hubble (1889-1953) and his co-workers were the first to realize that nebulae were in fact galaxies in their own right, each very much like our own Milky Way and each is 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).

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 of 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 $\sim 13.8 \pm 0.2$ 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 arrive at an estimated age of Earth and the other planets in the solar system as $\sim 4.5 \times 10^9$ years.

After Earth first formed, it was bombarded by extraterrestrial materials, including comets and asteroids. This bombardment began to subside around ~ 3.9 billion years ago and reached its current level by ~ 3.5 billion years ago.⁴⁵ 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.9: Fossil 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 the 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, which 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.

geologists 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. 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 isotope stability and the chemistry of atomic interactions. It uses the radioactive decay of elements with isotopes with long half-lives, such as ^{235}Ur (uranium) which decays into ^{207}Pb (lead) with a half-life of ~704 million years and ^{238}Ur which decays into ^{206}Pb with a half-life of ~4.47 billion 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 a rock, assuming that it originally contained Ur.

In order to use isotope abundance to accurately date rocks, it is critical that all of the atoms in a mineral measured stay there, that none are washed 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."⁴⁶ We know now, however, that this statement is not accurate; while very, very old, Earth had a beginning, it coalesced around ~5 billion years ago, and it will disappear when the sun expands and engulfs it in about ~5.5 billion years from now.⁴⁷

Now, 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.⁴⁸ Their presence in ancient rock implies that living organisms were present at the time the rock formed. Chemical fossils first appear in rocks that are between ~3.8 to $\sim 3.5 \times 10^9$ 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, there are what are known as 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 immobile 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, they can reveal the general shape of an organism or its ability to move or to move in a particular way. To move, an organism 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 mouth 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.

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 woodlands, 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

2.10: Life's impact on earth

Based on fossil evidence, the current model for life on Earth is that for a period of $\sim 2 \times 10^9$ (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 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 (Fe_2O_3) - that is, rust. This rust reaction removed large amounts of O_2 from the atmosphere, keeping levels of free O_2 low. The rusting of iron in the oceans is thought to be largely responsible for the massive banded iron deposits found around the world.⁴⁹ 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 O_2 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 $\sim 35\%$, almost twice the current level. It has been suggested that these high levels of atmospheric O_2 made the evolution of giant insects possible.⁵⁰

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; its appearance posed serious challenges and unique opportunities to, organisms. As we will see later on O_2 can be "detoxified" through reactions that lead to the formation of water; this type of thermodynamically favorable reaction appears to have been co-opted for a wide range of biological purposes. For example, through coupled reactions O_2 can be used to capture the maximum amount of energy from the breakdown of complex molecules (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 ~ 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 (i.e., animals), capable of moving along and through the mud on the ocean floor. About 0.6×10^9 years ago, new and more complex structural fossils begin to appear in the fossil record. Since the fossil record does not contain all organisms, we are left to speculate on what the earliest metazoans looked like. The first of these to appear in the fossil record are the so-called Ediacaran organisms, named after the geological formation in which their fossils were first found.⁵¹ Current hypotheses suggest they were immobile, like modern sponges but flatter; it remains unclear how or if they are related to later animals. By the beginning of the Cambrian age ($\sim 545 \times 10^6$ years ago), a wide variety of organisms had appeared within the fossil record, many clearly related to modern animals. 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 the presence of predators, and a need for a defensive response.

Viruses: Now, before we leave this chapter you might well ask, have we forgotten viruses? Well, no - viruses are often a critical 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 replicate 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. That said, the recent discovery of giant viruses, such as Mimivirus, suggests that something interesting is going on.⁵²

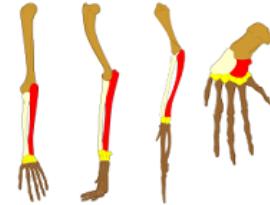
Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

3: EVOLUTIONARY MECHANISMS AND THE DIVERSITY OF LIFE

- 3.0: INTRODUCTION
- 3.1: ORGANIZING ORGANISMS (HIERARCHICALLY)
- 3.2: NATURAL AND UN-NATURAL GROUPS
- 3.3: FOSSILS AND FAMILY RELATIONSHIPS: INTRODUCING CLADISTICS
- 3.4: EVOLUTION THEORY'S CORE CONCEPTS
- 3.5: SO WHAT DO WE MEAN BY GENETIC FACTORS?
- 3.6: LIMITS ON POPULATIONS
- 3.7: THE CONCEPTUAL LEAP MADE BY DARWIN AND WALLACE
- 3.8: MUTATIONS AND THE ORIGINS OF GENOTYPE-BASED VARIATION
- 3.9: THE ORIGINS OF POLYMORPHISMS
- 3.10: A SHORT ASIDE ON THE GENOTYPE-PHENOTYPE RELATIONSHIP
- 3.11: VARIATION, SELECTION, AND SPECIATION
- 3.12: TYPES OF SIMPLE SELECTION
- 3.13: A SHORT NOTE ON PEDAGOGICAL WEIRDNESS
- 3.14: POPULATION SIZE, FOUNDER EFFECTS AND POPULATION BOTTLENECKS
- 3.15: POPULATION BOTTLENECKS
- 3.16: GENETIC DRIFT
- 3.17: GENE LINKAGE: ONE MORE COMPLICATION
- 3.18: A BRIEF REFLECTION ON THE COMPLEXITY OF PHENOTYPIC TRAITS
- 3.19: SPECIATION AND EXTINCTION
- 3.20: MECHANISMS OF SPECIATION
- 3.21: ISOLATING MECHANISMS
- 3.22: SYMPATRIC SPECIATION
- 3.23: SIGNS OF EVOLUTION: HOMOLOGY AND CONVERGENCE
- 3.24: THE LOSS OF TRAITS
- 3.25: SIGNS OF EVOLUTIONARY HISTORY
- 3.26: HOMOLOGIES PROVIDE EVIDENCE FOR A COMMON ANCESTOR
- 3.27: ANTI-EVOLUTION ARGUMENTS



3.0: Introduction

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

In medieval Europe there was a tradition of books known as bestiaries; these were illustrated catalogs of real 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.”⁵³ 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 the study 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 10,000,000⁵⁴. 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, existed in the past. So the obvious question is, why are there so many different types of organisms⁵⁵? Do they represent multiple independent creation events, and if so, how many such events have occurred?

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 of these was the fact that a number of organisms 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 infected ants to migrate to positions that favor fungal growth before killing the infected ant. Similarly, the nematode worm *Myrmeconema neotropicum* infects the ant *Cephalotes atratus*, leading to dramatic changes in the infected ant’s morphology and behavior.

The infected ant’s abdomen turns red and is held raised up, which makes it resemble a fruit and increases the likelihood of the infected ant being eaten by birds. The birds transport the worms, which survive in their digestive systems until they are excreted; they are then eaten by new ants to complete the worm’s life cycle⁵⁶. Perhaps the most famous example of this type of behavior occurs in wasps of the family *Ichneumonidae*. Female wasps deposit their fertilized eggs into the bodies of various types of caterpillars. The wasp eggs hatch out and produce larvae which then feed on the living caterpillar, consuming it from the inside out. Charles Darwin, in a letter to the American naturalist Asa Gray, remarked “*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 apparently cruel behaviors, Darwin and others sought alternative, morally neutral naturalistic processes that could both 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: 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 revealed that there were often many similarities between quite different types of organisms. A classic work, published in 1555, compared the skeletons of a human and a bird, both vertebrates⁵⁷. While many bones have different shape and relative sizes, what was most striking is how many bones are at least superficially similar between the two organisms. This 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 (a small terrestrial mammal), 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. So the scientific question is, what explains such similarities? Why build an organism that walks, runs, and climbs, such as humans, with a skeleton similar to that of an organism that flies (birds), swims (dugongs), or tunnels (moles). Are these anatomical similarities just flukes or do they imply something deeper about how organisms were initially formed?

References

53. <http://www.getty.edu/art/gettyguide/...?artobj=304109>
54. How many species are there on Earth and in the ocean? <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001127>
55. As a technical point, which we will return to, we will refer to each distinct type of organism as a species.
56. The Life of a Dead Ant: The Expression of an Adaptive Extended Phenotype: www.jstor.org/stable/10.1086/603640
57. Belon P (1555) L'Histoire de la Nature des Oyseaux. Paris, Guillaume Cavellat

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.1: Organizing organisms (hierarchically)

Carl Linnaeus (1707-1778) was the pioneer in taking the similarities between different types of organisms seriously. Based on such 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 how to decide 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 based his model only 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 explanation for why such similarities should exist in the first place, very much like Newton's law of gravitation did not explain why there was gravity, just how it behaved. So what are the features of a model? A model has to suggest observations or predict outcomes that have not yet been observed. It is the validity of these predictions that enable us to identify useful models. A model that makes no empirically validated predictions is not particularly useful, as least from a scientific perspective. 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 predictions and their confirmation, refutation, or revision, has been found to accurately explain a particular phenomenon can become promoted to a theory. We assume that the way the model works is the way the world works. 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.⁵⁸

Back to Linnaeus, whose classification system placed organisms of a particular type 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; it depended on which particular traits he felt to be important and useful to distinguish organisms of one species from those of another, perhaps quite, similar species. The choice of these key traits was subject to debate. Based on the perceived importance and presence of particular traits, organisms could be split into two or more types (species), or two types originally considered separate could be reclassified into a single species.

As we will see, the individual organisms that make 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 in nature. This criterion, reproductive compatibility, can be used to determine species distinctions on a more empirical basis, but it cannot be used with asexual 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 organisms 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 classifying. This can lead to conflicts between researchers that can be resolved by the collection of 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 the name in italics with the genus name capitalized, for example, *Homo sapiens*. Following on this pattern, one or more genera 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 group - all of the organisms on earth can be placed into a single unified phylogenetic "tree" or perhaps better put, bush – they are connected. That this should be the case is by no means obvious. This type of analysis could have produced multiple, disconnected classification schemes, but it did not.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.2: Natural and un-natural groups

It is worth reiterating that while a species can be seen as a natural group, the higher levels of classification may or may not reflect biologically significant information. Such higher-level classification is an artifact of the human need to make sense of the world; it also has the practical value of organizing information, much like the way books are organized in a library. We can be sure that we are reading the same book, and studying the same organism!

Genera and other higher-level classifications are generally based on a decision to consider one or more traits as more important than others. The assignment of a particular value to a trait can seem arbitrary. Let us consider, for example, the genus *Canis*, which includes wolves and coyotes and the genus *Vulpes*, which includes foxes. The distinction between these two groups is based on smaller size and flatter skulls in *Vulpes* compared to *Canis*. Now let us examine the genus *Felis*, the common house cat, and the genus *Panthera*, which includes tigers, lions, jaguars and leopards. These two genera are distinguished by cranial features and whether (*Panthera*) or not (*Felis*) they have the ability to roar. So what do we make of these distinctions, are they really sufficient to justify distinct groups, or should *Canis* and *Vulpes* (and *Felis* and *Panthera*) be merged together? Are the differences between these groups biologically meaningful? The answer is that often the basis for higher order classifications are not biologically meaningful. This common lack of biological significance is underscored by the fact that the higher order classification of an organism can change: a genus can become a family (and vice versa) or a species can be moved from one genera to another. 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 acknowledged. Looking at all bear-like organisms we recognize eight types.⁵⁹ We currently consider four of these, the brown bear (*Ursus arctos*), the Asiatic black bear (*Ursus thibetanus*), the American 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 types of bear-like organisms, the spectacled bear (*Tremarctos ornatus*), the sloth bear (*Melurus ursinus*), the sun bear (*Helarctos malayanus*), and the giant panda (*Ailuropoda melanoleuca*) in their own separate genus, because scientists consider these species more different from one another than are the members of the genus *Ursus*. The problem here is how big do these differences have to be to warrant a new genus?

So where does that leave us? Here the theory of evolution together with the cell (continuity of life) theory come together. We work on the assumption that the more closely related (evolutionarily) two species are, the more traits they will share and that the development of new, biologically significant trait is what distinguishes one group from another. Traits that underlie a rational classification scheme are known as synapomorphies (a technical term); basically these are traits that appeared in one or the other branch point of a family tree and serve to define that branch point, such that organisms on one branch are part of a "natural" group, distinct from those on the other branch (lineage). In just the same way that the distortion of space-time provided a reason for why there is a law of gravity, so the ancestral relationships between organisms provides a reason for why organisms can be arranged into a Linnaean hierarchy.

So the remaining question is, how do we determine ancestry when the ancestors lived, thousands, millions, or billions of years in the past. Since we cannot travel back in time, we have to deduce relationships from comparative studies of living and fossilized organisms. Here the biologist Willi Hennig played a key role.⁶⁰ He established rules for using shared, empirically measurable traits to reconstruct ancestral relationships, such that each group should have a single common ancestor. As we will discover later on, one of the traits now commonly used in modern studies is gene (DNA) sequence and genomic organization data, although even here there are plenty of situations where ambiguities remain, due to the very long times that separate ancestors and present day organisms.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.3: Fossils and family relationships: introducing cladistics

As mentioned previously, we continue to discover new fossils and new organisms. In most cases, fossils appear to represent organisms that lived many millions to hundreds of millions of years ago but which are now extinct. We can expect there are dramatic differences between the ability of different types of organisms to become fossilized.⁶¹ Perhaps the easiest organisms to fossilize are those with internal or external skeletons, yet it is estimated that between 85 to 97% of such organisms are not represented in the fossil record. A number of studies indicate that many other types of organisms have left no fossils whatsoever⁶² and that the number of organisms (at the genus level) that have been preserved as fossils may be less (often much less) than 5%.⁶³ For some categories of modern organisms, such as the wide range of microbes, essentially no informative fossils exist at all.

Once scientists recognized that fossils provide evidence for extinct organisms, the obvious question was, do extinct organisms fit into the same cladistic classification scheme as do living organisms 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 cladistic classification scheme developed for modern organisms, with a few possible exceptions, such as the Ediacaran organisms that lived very long ago and appear (perhaps) to be structurally distinct from all known living organisms.⁶⁴ The presumption, however, is that if we had samples of Ediacaran organisms for molecular analyses, we would find it that they would fall nicely into the same classification scheme as all other organisms do.⁶⁵ A similar example are the dinosaurs, which while extinct, are clearly descended from a specific type of reptile that also gave rise to modern birds, while 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 humanoids, that went extinct about 30,000 years ago. This is information that has been used to clarify their relationship to modern humans (*Homo sapiens*).⁶⁶ In fact, such data has been interpreted as evidence for interbreeding between these groups and has led for calls to reclassify Neanderthals and Denisovians as subspecies of *Homo sapiens*.

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

– John Maynard Smith

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.4: Evolution theory's core concepts

So what are the facts and inferences upon which the Theory of Evolution is based? Two of its foundational observations are deeply interrelated and based on empirical observations associated with plant and animal breeding and the characteristics 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 controlled matings between individuals often displayed phenotypes similar to those of their parents, indicating that 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. For example, beginning ~10,000 years ago plant breeders in Mesoamerica developed modern corn (maize) by the selective breeding of variants of the grass teosinte.⁶⁸ 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).⁶⁹ In all cases, the crafting of specific domesticated organisms followed the same pattern.

Organisms with desirable (or desired) traits 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 display 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 heritable, factors. Originally, what these heritable factors were was completely unclear, but we can refer to them as the organism's genotype, even though early plant and animal breeders would never have used that term.

This implies that different organisms have different genotypes and that different genotypes produce different phenotypes, but where genotypic differences came from was completely unclear to early plant and animal breeders. Were they imprinted on the organism in some way based on its experiences or were they the result of 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 still remains poorly resolved for many phenotypes.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.5: 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.⁷⁰



Figure 3.5.1: King Charles II, the last Habsburg king of Spain, displayed the "Habsburg lip" in reference to heredity of their Austro-Hungarian monarchy via successive inbreeding. King Charles suffered from both physical and mental disabilities; he could not chew his food and had a reduced intelligence. He was also impotent and had no heirs to the throne.

In the case of artificial selection, an important to keep in mind is that the various types of domesticated organisms produced are often dependent for their survival on their human creators (much like European royal families). This relieves them of the 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 short distances and so are extremely vulnerable to predators. Neither modern corn (*Zea mays*) or chihuahuas, one of the smallest breeds of dog developed by Mesoamerican breeders, would be expected to survive for long in the wild.⁷¹

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 -

Michael Crichton (1969) *The Andromeda Strain*

References

70. 'Imperial Stigmata' The Habsburg Lip, A Grotesque 'Mark' Of Royalty Through The Centuries!: <http://thesotericcuriosa.blogspot.com/2011/08/imperial-stigmata-habsburg-lip.html>
71. How DNA sequence divides chihuahua and great dane: www.theguardian.com/science/2012/jan/03/dna-chihuahua-great-dane

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.6: 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 reproduce periodically over the course of their lifespan. By the end of 500 years, a single pair of elephants could (theoretically) produce ~15,000,000 living descendants.⁷² Clearly if these 15,000,000 elephants paired up to form 7,500,000 breeding pairs, within another 500 years (1000 years altogether) there could be as many as $7.5 \times 10^6 \times 1.5 \times 10^7$ 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, asexual 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 at least one mistake in our logic.

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 before they die. This trait is known as superfecundity. But unlimited growth does not and cannot happen for very long - other factors must act to 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 total number of organisms remains roughly constant.

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 territory before it can successfully reproduce.⁷³ 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.⁷⁴ 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, in other words 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 its cause).

An immediate example of this type of behavior involves the human population. Once constrained by disease, war, and periodic famine, the introduction of better public health and sanitation measures, a more secure food supply, and reductions in infant mortality has led the human population to increase dramatically. Now, in many countries, populations appear to be heading to a new steady state, although exactly what that final population total level will be is unclear.⁷⁵ 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!⁷⁶ 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 itself can generate 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.

*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*

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.7: The conceptual leap made by Darwin and Wallace

Darwin and Wallace recognized the implications and significance of these key facts: the heritable 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 numbers of descendants. Over time, we would expect that the phenotypic variations associated with greater reproductive success (and the genotypes underlying these phenotypic differences) will increase in frequency within the population; over time they would replace those organisms with less reproductively successful phenotypes. Darwin termed this process natural selection, in analogy to process of artificial selection practiced 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 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 genotype) involved in producing one trait influences other traits – they are interdependent, after all they are aspects of a single organism. There are also non-genetic sources of variation. For example, there are molecular level fluctuations that occur at the cellular level; these can lead genetically identical cells to display different behaviors, that is, different phenotypes. Environmental factors and stresses also 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.8: 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 within a DNA molecule, which we will consider in detail in Chapter 7. Within this string of characters 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) as corresponding to **genes** and the other sequences as **intragenic** regions, that is, spaces between genes. We estimate that humans have ~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, over ~55,000,000 variations have been found 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 specific regions. In sexually reproducing organisms, like humans, there are typically 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 gametes: sometimes these are the same size. When gametes differ in size the smaller one is known as a sperm and the larger is known as an egg. Each gamete contains one copy of its own unique version of the genomic book and 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. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.9: The origins of polymorphisms

So what produces the genomic variations between individuals found within current populations? Are these processes still continuing to produce genotypic and phenotypic variations 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.⁷⁷ In addition, we are continually bombarded by radiation that can damage DNA (although not to worry, the radiation energy associated with cell phones, bluetooth, and wifi devices is too low 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 while remarkably accurate 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 ends up producing ~100-200 new mutations (changes) added to an individual's genome per generation.⁷⁸ 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 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 specify where a mutation will occur. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.10: 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.⁷⁹ 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), cystic fibrosis, 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.⁸⁰ 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.

The vast majority of traits 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 is not always possible), we find that each population can be characterized graphically 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:

In this case, the mean weight of the population is 180 pounds. It is common to recognize another characteristic of the population, 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.⁸¹ 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.

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 this graph. 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 relationship, however, is how many genes are involved in the genetic determination of height or how their effects are influenced by the environment and the environmental history that the offspring experience. For example, “human height has been increasing during the 19th century when comprehensive records began to be kept. The mean height of Dutchmen, for example, increased from 165cm in 1860 to a current 184cm, a spectacular increase that probably reflects improvements in health care and diet”, rather than changes in genes.⁸² Geneticists currently estimate that allelic differences at more than ~50 genes make significant contributions to the determination of height, while allelic

differences at hundreds other genes have smaller effects that contribute to differences in height.⁸³ 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 gigantism.

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 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 (i.e., the presence or absence of particular alleles of other genes). Environmental factors can also have significant effects on the phenotype associated with a particular allele or genotype.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.11: Variation, selection, and 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, or disappear altogether. The effects of specific alleles on an organism's reproductive success will, 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 effects 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. 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 selected for (or against) more weakly.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.12: 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 begin with three basic types of selection: stabilizing (or 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 effects on phenotype. 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, stronger to weaker, lighter to darker, or 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.

Stabilizing 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 sometimes, many more organisms than are needed to replace them when they die and that these organisms mate randomly with one another. Now we have to consider the factors that lead to the original population distribution: why is the mean value of the trait the value it is? What factors influence the observed 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 compared 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 inactive, 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 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 predict selection against extreme values of the trait. We can measure that degree of selection "pressure" by following the reproductive success of individuals with different values of the trait. We might 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 is 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 the impact of the trait on reproductive success will be relatively weak. **Directed selection:** Now imagine that the population's environment changes. It may now be the case that the phenotype of the mean is no longer the optimal phenotype in terms of reproductive success (the only factor that matters, evolutionarily); a smaller or a larger value may be more favorable. Under these conditions we would expect that, over time, the mean of the distribution would shift toward the phenotypic value associated with maximum reproductive success. Once reached, and assuming the environment stays constant, stabilizing 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 and too rapid 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 is 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 (this is not an issue in asexual populations). 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 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.13: 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 in 1908) 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 allele associated with the expressed (visible) phenotypic trait is said to be dominant to the other allele, which is termed recessive.⁸⁴ 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 trait 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 leave and none enter; 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.⁸⁵ 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 stabilizing selection?
- What would lead stabilizing selection to be replaced by directed or disruptive selection?
- Explain the caveats associated with assuming that you know why a trait was selected.
- How could phenotypic variation influence random mating?
- By looking at a population, how might you estimate the strength of selection with respect to a particular trait?

References

84. 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.
85. Hardy-Weinberg Equilibrium: <http://www.tiem.utk.edu/~gross/bioed...-weinberg.html>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.14: Population size, founder effects and population bottlenecks

When we think about evolutionary processes from a Hardy-Weinberg perspective, we ignore some extremely important factors that normally impact the populations. 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. Something similar happens when a large population is dramatically reduced in size for any of a number of reasons, a situation known as a population bottleneck (see below). In both founder effects and population bottlenecks, the small populations that result are more susceptible to the effects of random, non-selective effects, a process known as genetic drift. Together these process can produce a population with unique traits, traits not due to the effects of natural selection.

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 the isolation of, and evolutionary change within, small populations can cause a random jump from one place in the landscape to another; in the new position, new adaptations can be possible. In addition, a population invading a new environment will encounter a new set of organisms to compete and cooperate with. Similarly, a catastrophic environmental change will change the selective landscape, removing competitors, predators, pathogens, and cooperators, often favoring new adaptations and selecting against others. One effect of the major extinction events that have occurred during the evolution of life on Earth is that they provide a new adaptive context, a different and less densely populated playing field with fewer direct competitors. The expansion of various species of mammals that followed the extinction of the dinosaurs is an example of one such opportunity, associated with changes in selection pressures.

Founder effects: What happens when a small subpopulation becomes isolated, for whatever reason, from its parent population? The original (large) population will contain a number of genotypes and alleles. If it is in a stable environment the population will be governed primarily 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 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 a large number of times, a fair six-sided (cubical) die will be expected to produce the numbers 1, 2, 3, 4, 5, and 6 with equal probabilities. Each would appear $1/6^{\text{th}}$ of the time. But imagine that the number of rolls is limited and small. Would you expect to get each number appearing with equal probability? You can check your intuition using various on-line dice applets.⁸⁶ See how many throws are required to arrive at an equal $1/6^{\text{th}}$ probability distribution; the number is almost certainly much larger than you would guess. We can apply this to populations in the following way: imagine a population in which each individual carries one of six alleles or a particular gene and the percentage of each type is equal ($1/6^{\text{th}}$). The selection of any one individual from this population is like a throw of the die; there is an equal $1/6^{\text{th}}$ 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 individual rolls of die and equally likely to result in a $1/6^{\text{th}}$ chance to select any one of the six alleles. But producing a small subpopulation with $1/6^{\text{th}}$ 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 genetically 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 that the new subpopulation will be genetically different from the original population. So when a small group from a parent population invades or migrates into a new environment, it will very likely have a different genotypic profile compared to the parent population. This difference is not due to natural selection but rather to chance alone. Nevertheless, it will influence subsequent evolutionary events; the small subpopulation will likely respond in different ways to new mutations and environmental pressures based on which alleles are present within it.

The human species appears to have emerged in Africa ~200,000 years ago. The people living in Africa represent the parent population of *Homo sapiens* and genetic studies reveal 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*) ~1.2 million years ago.⁸⁷

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.15: 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 (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. Who survives the bottleneck can be random, due only to luck, or based on genetic factors (for example, leading to disease resistance).

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 all marine species and ~75% of land species became extinct.⁸⁸ If most species were affected, we would not be surprised if the surviving populations also 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. The Cretaceous-Tertiary event, which occurred ~65 million years ago, contributed to the extinction of the dinosaurs and led to the diversification of mammals (which had first appeared in the fossil record ~160 million years ago), particularly the placental mammals.

While surviving an asteroid impact (or other dramatic changes in climate) may be random, in other cases who survives a bottleneck is not. 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 significantly influence their chance of survival. In such a case, the effect of the bottleneck event would produce non-random changes in the distribution of genotypes (and alleles) in the post-bottleneck population – these selective effects could continue to influence the population in various ways. For example, a trait associated with pathogen resistance may also have negative phenotypic effects. After the pathogen associated bottleneck, mutations that mitigate the resistance trait's negative effects (and may have their own effects) would be selected. The end result is that traits that would not be selected in the absence of the pathogen, are selected. 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) and could make finding a mate more difficult.

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 non-sense (intragenic) parts of the genome at a constant rate over time (can you explain why?) The rate of the accumulation of neutral polymorphisms serves as a type of population-based biological clock. Its rate can be estimated, at least roughly, by comparing the genotypes of individuals of different populations whose time of separation can be accurately estimated (assuming of course that there has been no migrations between the populations). Such 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.⁸⁹ This bottleneck occurred around the time of the major migration of people out of Africa into Europe and Asia. Comparing genotypes, that is, neutral polymorphisms, between isolated populations enables us to estimate that aboriginal Australians reached Australia ~50,000 years ago, well before other human migrations⁹⁰ and that humans arrived in the Americas in multiple waves beginning ~15,000 to 16,000 years ago.⁹¹ The arrival of humans into a new environment has been linked to the extinction of a group of mammals known as the megafauna in those environments.⁹² The presence of humans changed the environmental pressures on these organisms around the world.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.16: Genetic drift

Genetic drift is an evolutionary phenomenon that is impossible in a strict Hardy-Weinberg world, yet it 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 Haplorrhini (from the Greek meaning “dry noses”) and the Strepsirrhini (meaning “wet noses”). The Strepsirrhini include the lemurs and lorises, while the Haplorrhini include the tarsiers and the anthropoids (monkeys, apes, and humans). One characteristic trait of the Haplorrhini 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.”*⁹³

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* gene is the root cause of vitamin C dependence in Haplorrhini by putting a working copy of the *gulo1* gene, for example derived from a 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 a preventative treatment for scurvy was discovered⁹⁴), no new, functional *gulo1* allele has appeared. Organisms do not always produce the alleles they need or that might be beneficial, such alleles must be selected from alleles already present in the population or that appear through mutation. In some cases, however, there may be no molecule pathway that can generate such an allele.

The mutant *gulo1* 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 lose 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 (that is, there was little or no selective pressure for the ability to make vitamin C). 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 functional 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, it is due to random effects on organismic survival, which can, in practice be difficult to distinguish from selective effects. In contrast, drift is unavoidable in sexually reproducing organisms. 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 the alleles present within the population, any particular gamete contains only a single 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. Moreover, not all gametes (something particularly true of sperm) become part of the next generation. 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 with respect to one another, 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. Of course new alleles continue to arise by mutation, but they are originally very infrequent, just one in the entire population, so unless they are strongly selected for they are likely to disappear from the population.⁹⁵ Drift can lead to some weird outcomes. 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 BBY 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.

This use of pre-existing variation, rather than the idea that an organism invents variations in its genome as they are required, was a key point in Darwin's view of evolutionary processes. The organism cannot create the alleles it might need, nor are there any known processes 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 to work with.⁹⁶ 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.17: 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 isolated 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, 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, genes 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 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 that 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 can lead to the fixation of mildly deleterious alleles, in neighboring regions. We refer to a position of a particular gene within the genome as a genetic locus (or the plural, loci). In Latin locus means ‘place’ (think location, which is derived from the same root). A particular genetic locus can be occupied by any of a number of distinct alleles (DNA sequences). As we will see, there are various mechanisms that can duplicate, delete, or move within the genome a region of DNA, creating (or eliminating) new genetic loci. The phenotype associated with an allele is influenced by its genetic locus, as well as the details of the rest of the genome.

It is worth noting the combination of non-adaptive, non-selective processes can lead to the appearance and maintenance of mildly non-adantageous traits within a population. Similarly, a trait that increases reproductive success, by increasing the number of surviving offspring, may be associated with other not-so- beneficial, and sometime seriously detrimental (to individuals) effects. The key is to remember that evolutionary mechanisms do not necessarily result in what is best for an individual organism but what in the end enhances net reproductive success. Evolutionary processes do not select for particular genes (we will consider how new genes appear later on) or new versions of genes but rather for those 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.

Of course, the situation gets more complex when evolutionary mechanisms generate organisms, like humans, who feel and can actively object to the outcomes of evolutionary processes. From the point of view of self-conscious organisms, evolution can appear cruel. This was one reason that Darwin preferred impersonal (naturalistic) mechanisms over the idea of a God responsible for what can appear to be gratuitously cruel aspects of creation.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.18: 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 (random) processes, like drift, linkage, and bottlenecks. These traits become established not because they improve reproductive success but simply because they happened to be fixed randomly within the population. If an allele is extremely deleterious independent of its environment, it will be expected to rapidly disappear from the population. Such strongly deleterious alleles are, most likely, the result of a new mutation that occurred within the affected individual or the germ line of its parents.

When we consider a deleterious allele we mean it in terms of its effects on reproductive success. An allele can harm the individual organism carrying it yet persist in the population because it improves reproductive success in some measurable way. Similarly, there are traits that can be seen as actively maladaptive, but which occur because they are linked mechanistically to some other 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 can be beneficial, others deleterious. As an example, a trait that dramatically increases the survival of the young, and so increases their potential reproductive success, but leads to senility and death 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 (think the effects of diet on the effects of the *gulo1* mutation). All of which is to say that when thinking about evolutionary mechanisms, do not assume that a particular trait exists independently of other traits, that it functions in the same way in all environments, or that the presence of a trait is evidence that it is beneficial.

*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.*

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.19: Speciation and extinction

As we have noted, an important observation that any useful biological theory needs to explain is why there are so many (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. At the same time, we know from the fossil record and from modern experiences that types of organisms can disappear – they can become extinct. What leads to the formation of a new species or the disappearance of existing ones?

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, which involves a certain degree of cooperation between individuals, we have to consider how far a particular organism (or its gametes) can travel. The range of some organisms is quite limited, whereas others can travel significant distances. Another factor to consider is how an organism makes its living - where does 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 (the past selection pressures acting within a particular environment) and its current behavior, 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, stabilizing 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 continue to compete directly with other species. This can lead to a number of outcomes. In one case, one species becomes much better than others at occupying a particular niche, driving the others 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) 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.⁹⁷ 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 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 an 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 as 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.20: Mechanisms of speciation

So now we consider the various mechanisms that can lead a species to give 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 generally no magic moment when one species changes into another, rather a new species emerges over time from a pre-existing species⁹⁸. Of course the situation is more complex in organisms that reproduce asexually, but we will ignore that for the moment. More generally, 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 interbreed with one another. If we assume that the environments inhabited by the subpopulations are distinct, and 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 period of time, the populations will diverge. Both selective and non-selective processes will drive this divergence, and will be influenced 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 Hawaiian islands⁹⁹. Derived from an ancestral population, these organisms have adapted to a number of highly specialized niches. These specializations give them a competitive edge with respect to one another 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 a little like a fungus that can only grow on one particular place on a particular type of beetle, as we 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 and effectively and competitively exploit the remaining niches, leading to its extinction. It is a sobering thought that current estimates are that greater than ~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 changes¹⁰⁰. In a complementary way, the migration of organisms into a new environment can produce a range of effects as the competition 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 geological 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 appreciate the fact 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 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 called 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).¹⁰¹

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, Through the Looking Glass

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.21: 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 display the behavior, even if it is in a primitive form, 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, could a long neck have evolved in a species that did not eat the leaves of 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, the song produced also changes¹⁰². 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 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 isolation¹⁰³. 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 fuse with one another, or embryos might not be viable - there are many possibilities.

References

¹⁰² Beaks, Adaptation, and Vocal Evolution in Darwin's Finches: <http://bioscience.oxfordjournals.org...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>

¹⁰³ Beak size matters for finches' song: news.nationalgeographic.com/n...ins_finch.html

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.22: 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 region. How could this possibly occur? What stops (or inhibits) the distinct sub-populations from inbreeding and reversing the effects of selection and nascent speciation? Recently a number of plausible mechanisms have been identified. One involves host selection¹⁰⁴. 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-specific niches. Through a process of disruptive selection (see above), organisms that live primarily on a 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 is known as reproductive isolation - as we will see this is a simple form of sexual selection¹⁰⁵. One population has become two distinct, reproductively independent populations, one species has become two.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.23: 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 their version of the trait 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 and 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. Given that other rodentia are vitamin C independent, we can assume that the common ancestor of the rodent/primate lineages 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 - 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 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, 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 structurally similar starting point. But most tetrapod vertebrates do not fly, and forelimbs have become adapted to many different functions. An analysis of tetrapod vertebrate wings indicates that each took a distinctly different approach 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 of evolutionary convergence consider teeth. 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; morphologically similar dagger-like teeth have evolved 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, like most mammals, appears to have lacked such dagger-like teeth. While teeth are a homologous feature of Smilodon and Thyacosmilus, elongated dagger-like teeth are analogous structures that resulted from the convergent evolution.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.24: The loss of traits

A major challenge when trying to determine a plausible relationship between organisms based on anatomy has been to distinguish homologous from convergent (analogous) traits. Homologous traits, known as synapomorphies, are the basis of placing organisms together within a common group. In contrast, convergent traits are independent solutions to a common problem, and so are irrelevant when it comes to defining evolutionary relationships. It is, however, also true that evolution can lead to the loss of traits; this can confuse or complicate the positioning of an organism in a classification scheme. It is worth noting that very often developing a particular trait, whether it is an enzyme or an eye, requires energy. If the trait does not contribute to an organism's reproductive success it will not be selected for; on the other hand, it is expensive to build, but has no useful function, its loss may be selected for. 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 are completely dependent on their hosts for many essential functions, this allows them to 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.25: 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 success¹⁰⁶. 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 communication¹⁰⁷.

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 can change in response to changing environmental factors, which alter selective pressures. In fact, lifespan itself is a selected trait, since it is the population not the individual that evolves¹⁰⁸.

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 aging¹⁰⁹. 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 such a blind spot. We have adapted to this retinal blind spot through the use of saccadic eye 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 of reproductive success associated with removing the blind spot. 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 stabilizing selection (as previously described). The eyes of different vertebrates differ in their acuity (basically how fine a pattern of objects they can resolve at what distance) and sensitivity (what levels and wavelengths of light they can perceive). Each species has eyes (and their connections to the brain) adapted for its ecological niche. For example, an eagle sees details at a distance four to five times as far as the typical human; why, because such visual acuity is useful in terms of the eagle's life-style, whereas such visual details could well be just a distraction for humans¹¹⁰.

References

- ¹⁰⁶ Evolutionary tinkering: virtuallaboratory.colorado.edu...nTinkering.pdf
- ¹⁰⁷ How the Hyoid Bone Changed History: <http://www.livescience.com/7468-hyo...d-history.html>
- ¹⁰⁸ 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: www.plosone.org/article/info%...l.pone.0000785
- ¹⁰⁹ <http://www.pbs.org/wgbh/nova/evoluti...-suggests.html>
- ¹¹⁰ <http://www.livescience.com/18658-hum...le-vision.html>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.26: 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 (DNA) sequencing, have made it possible to treat gene sequences and genomic organization as traits that can be compared. Detailed analyses of many 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 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 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

3.27: 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

4: SOCIAL EVOLUTION AND SEXUAL SELECTION

4.0: INTRODUCTION

In which we consider how organisms, including many 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 behaviors and morphologies of organisms.



4.1: SELECTING SOCIAL (COOPERATIVE) TRAITS

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.

4.2: QUORUM SENSING

Another type of community behavior at the unicellular level involves a behavior known as quorum sensing. This is a process by which organisms can sense the density of other organisms in their immediate environment. Each individual secretes a molecule, which they can also respond to; the organism's response to this molecule dependent on the secreted molecule's concentration and it is non-linear.

4.3: 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 surprisingly common in a wide range of systems. 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 programmed cell death process amounts to cellular suicide.

4.4: INCLUSIVE FITNESS, KIN AND GROUP SELECTION, AND SOCIAL EVOLUTION

4.5: GROUP SELECTION

4.6: DEFENSE AGAINST SOCIAL CHEATERS

4.7: DRIVING THE EVOLUTIONARY APPEARANCE OF MULTICELLULAR ORGANISMS

4.8: ORIGINS AND IMPLICATIONS OF SEXUAL REPRODUCTION

4.9: SEXUAL DIMORPHISM

4.10: SEXUAL SELECTION

4.11: CURBING RUNAWAY SELECTION

4.0: Introduction

In which we consider how organisms, including many 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 behaviors and morphologies of organisms.

The naturalist Ernst Mayr stressed the difference in thinking in biology compared to physics and chemistry. The history of an electron, an atom, or a molecule is totally irrelevant to its physical and chemical properties. Each carbon isotope atom, for example, is identical to all others - one could be replaced by another and you could never, in theory or in practice, be able to 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 ~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. In fact, both are old, both reflect a continuous history stretching back to the origin of life.

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, we find no clear line between the different types (that is, species) of organisms. When did humans (*Homo sapiens*) appear from pre-humans, or birds from their dinosaurian progenitors? The answer is necessarily arbitrary, since cellular (and organismic) 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 close and complex ways. A dramatic example of this behavior 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 effectively sterile) female workers, along with a few males to inseminate the queen¹¹¹. So what, exactly, is the organism, the social group or the individuals that make it up? From an evolutionary perspective, selection is occurring at a social level as well as the organismic level. Similarly, consider yourself and other multicellular organisms (animals and plants). Most of the cells in your body, known as somatic cells, do not directly contribute to the next generation, rather they cooperate to insure that a subset of cells, known as germ line cells, 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 produce 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 thumb¹¹². 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. This involves the differentiation of cells, one from the other, so that certain cells within the group become specialized to carry out specific roles. Reproducing the next generation is one such specialized cellular 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 earth¹¹³. Is this a violation of evolutionary theory or do we have to get a little more sophisticated in our thinking?

References

¹¹¹ An Introduction to Eusociality: <http://www.nature.com/scitable/knowledge-base/eusociality-15788128>

¹¹² A twelve-step program for evolving multicellularity and a division of labor: onlinelibrary.wiley.com/doi/10.1002/bies.20197/full

¹¹³ The Origins of Multicellularity: bcrc.bio.umass.edu/courses/f...default/files/bonner_multicellularity_1998.pdf

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.1: 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 or organism that expresses that trait. 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 a cost-benefit perspective. There are costs (let us call that term “c”) in terms of energy needed to produce a trait and risks associated with expressing the trait, and benefits (“b”) in terms of the trait’s effects on reproductive success. To be evolutionarily preferred (or selected), the benefit b must be greater than the cost c, that is $b > c$. 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 let us assume further that this same trait benefits neighboring organisms. 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 self-sacrificing behavior?

Let us consider an examples of this type of behavior, provided by social amoebae of the genus *Dictyostelium*¹¹⁴. These organisms have a complex life style that includes a stage in which unicellular amoeba-like organisms crawl around in the soil eating bacteria, growing, and dividing. In this phase of their life cycle, the cells 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 multicellular “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 differentiation¹¹⁵. All through the cellular aggregation and slug migration stages, the original amoeboid cells remain distinct. Upon differentiation ~20% of the cells in the slug differentiate into stalk cells, which can no longer divide, in fact they die. Before they die the stalk cells act together, through changes in shape, to lift the non-stalk cells above the soil, where they go on to form spores. The stalk cells sacrificed themselves so that other cells can form spores. These spores are specialized cells that can survive harsh conditions; they can be transported 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 vegetatively. The available 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 genetically identical cells may both form spores, both stalk cells, or one might become a stalk and one a spore cell¹¹⁶.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.2: Quorum Sensing

Another type of community behavior at the unicellular level involves a behavior known as quorum sensing. This is a process by which organisms can sense the density of other organisms in their immediate environment. Each individual secretes a molecule, which they can also respond to; the organism's response to this molecule dependent on the secreted molecule's concentration and it is non-linear. So what does a non-linear response look like? As the concentration of signaling molecules increases, there is a discrete concentration, known as the threshold concentration; below the threshold concentration the cells (or organisms) do not change their behavior in response to the secreted compound. When cells or organisms are present at a low density, the concentration of the signaling molecule never exceeds the threshold concentration. As the density of organisms per unit volume increases, however, the concentration of the molecule exceeds the threshold concentration and interesting things start to happen; there are changes in behavior, often associated with changes in gene expression (we will consider what that means exactly later on)¹¹⁷.

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 relationship with the squid *Euprymna scolopes*¹¹⁸. In these squid, *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 there are many steps in the colonization process, and its regulation is complex, we will just consider just a few to indicate how cooperative behaviors between the bacteria are critical. For the colonization of the squid's light organs the *V. fischeri* 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 (known as a 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 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 emit light through a reaction involving the luciferin molecule. This reaction system involves various coupled chemical reactions (we will consider in some detail the thermodynamics of such reactions in the next section of the course) and is catalyzed (that is, sped up) by the protein luciferase. The luciferase protein is encoded by one of the bacteria's genes (its original role has been proposed to be in the "detoxification of the deleterious oxygen derivatives"¹¹⁹). Given that bacteria are small, you can imagine that very little light would be emitted from a single bacterium. If there were 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 bacterium rises above a threshold concentration. This insures that when a biologically costly 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.3: 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 surprisingly common in a wide range of systems¹²⁰. 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 programmed cell death 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 well beyond the scope of this book (but extremely important)¹²¹. 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 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 leads 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 will increase their chances of their survival and successful reproduction. This strategy works because as the environment becomes hostile, not all cells 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 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 stable - 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. The anti-toxin molecule 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 cellular activity, including the synthesis of cellular components (such as the anti-toxin) slows or stops. Now can you predict what happens? The level of the stable toxin molecule within the cell remains high, decreasing only slowly, while the level of the unstable anti-toxin drops rapidly. As the level of the anti-toxin drops below the threshold level required to keep the toxin inactive, 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 machinery 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 kills itself before new viruses are synthesized, it also kills 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 kind of altruistic, self-sacrificing behavior we have been considering¹²².

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.4: Inclusive fitness, kin and group selection, and social evolution

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 reproduce¹²³. 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 individual’s differential reproductive success, how are such traits possible? W.D. Hamilton (1936–2000) provided the formal answer, expressed in the equation r times $b > c$ (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 (see above).

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 or organism’s neighbors are likely to be closely related. In fact, 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 and 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 individual as a result of their cooperative/altruistic behavior = 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 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 is, how is its behavior influenced by the degree of relationship to others) 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 if they are mechanistically linked to a trait that is highly valued (that is, selected for). 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.5: Group selection

A proposed alternative to inclusive fitness is 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. It is the reproductive success of the group, compared to other groups of the organism, that is the basis of selection. Groups that display cooperative and altruistic traits have a selective advantage over groups that do not. Again, the mathematical analysis is similar (and it has been claimed that mathematically group and kin selection are equivalent)¹²⁴. 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 masse and return to active life. They move and feed in a cooperative manner through the release of digestive enzymes, which because they are acting in a quorum mode, can reach high levels¹²⁵. A well-coordinated group is expected to have a significant reproductive advantage over more anarchic collection of individuals.

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 algae¹²⁶. Each of these examples represents a cooperative strategy by which organisms can gain a reproductive advantage over those that do not display the 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(s) that produces the most offspring, (i.e., the most reproductively successful groups) will come to dominate.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.6: Defense against social cheaters

Now an interesting question arises: within a social organization, such as a group of cooperating microbes or hunters¹²⁷, 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 various internal and external mechanisms. This makes cooperators recognizable and works to maintain a cooperative or altruistic trait even in the face of 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, particularly early on-set and childhood cancers, are diseases 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 and begin to divide in an uncontrolled manner, disrupting 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 apoptosis, a type of active cell death (see above). A cell carrying a mutation that enables it to grow in an uncontrolled and inappropriate manner will likely kill itself before it can produce significant damage¹²⁸. 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 in 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 of the immune system can recognize these changes and kill the mutant cell¹²⁹. Of course, given that tumors occur and kill people, we can assume that there are mutations that enable tumor cells to avoid such 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 social 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 social systems, from cooperating bacteria to complex societies, we see examples of intrinsic and extrinsic control.

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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.7: 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 organisms¹³⁰. 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 nigrofasciatus*); it can grow to more than 600µm 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 fields¹³¹. 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, including those with which we are most likely to be familiar with, 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 themselves or to avoid predation. For example, Borass et al.¹³², reported that the unicellular algae *Chlorella vulgaris* (5-6µm 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 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. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.8: Origins and implications of sexual reproduction

One type of social interaction that we have mentioned in passing is sex. Sexual reproduction involves a cooperative interaction between organisms of different mating types, something unnecessary in asexual reproduction. While we are used to two distinct sexes (male and female), this is not universal: many unicellular eukaryotes are characterized by an number of distinct mating types. Typically, sexual reproduction involves the fusion of specialized cells, known as gametes, of different mating types (or sexes). 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 an 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 together to form a single cell. The original cells are haploid, meaning that they have a single copy of their genome. When two haploid cells fuse, the resulting cell has two copies of the genetic material and is referred to as diploid. This diploid cell will then go through a series of events, known collectively as meiosis, that results in the production of four haploid cells. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.9: 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. An individual *Tetrahymena* cell can mate with another individual of a different mating type but not with an individual of the same mating type as itself. Mating involves fusion and so the identity of the parents is lost; the four cells that result are of one or the other of the original mating types.

In multicellular organisms, the parents do not themselves fuse with one another. Rather they produce cells, known as gametes, which do. Also, instead of two or more mating types, there are usually only 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 organisms produce. The larger fusing cell is termed the egg and an organism that produces eggs it is termed a female. The smaller fusing cell, which is often motile (while eggs are generally immotile), is termed a sperm and organisms that produce sperm are 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 for all individuals to originally develop as males; based on environmental cues, the largest of these males changes its sex to become female. Alternatively, one organism can produce both eggs and sperm; such an organism is known as a hermaphrodite.

The size difference between male and female gametes changes the reproductive stakes for the two sexes. Simply 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 significantly 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 supplies sperm and leaves. As you can imagine, there are many different reproductive strategies (many more than we can consider here), and they all have distinct implications. For example, a contributing factor in social evolution is that where raising offspring is particularly biologically expensive, cooperation between the sexes or within groups of organisms in child rearing can improve reproductive 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 can diverge dramatically from one another, and that such divergence has evolutionary and behavioral implications.

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 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 can occur (as often happens in pack animals) 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 and energy diverted from the male's search for a new mate; it will reduce the male's

overall reproductive success, and so is a behavior likely to be selected against. 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 are stable and (largely) exclusive. In some cases these pairs last for extremely long times; in others there is what has been called serial monogamy, 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 males¹³³.

It is common that while caring for their young, females are 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 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 al¹³⁴. Male *Pisaura mirabilis* 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 offer a worthless gift, an inedible object wrapped in silk. Females cannot initially tell that the gift is worthless but quickly terminate mating if they discover that 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 events¹³⁵. There is even evidence that in some organisms, such as the wild fowl *Gallus gallus*, females can bias against fertilization by certain 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 with¹³⁶. And so it goes, each reproductive strategy leads, over time, to counter measures¹³⁷. For example, in species in which a male guards a set of females (its harem), groups of males can work together to distract the guarding 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 world¹³⁸. 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 a similar time period. Moreover, the biological cost of generating offspring is substantially greater for the female, compared to the male¹³⁹. 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 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, i.e., 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 following mating with a more dominant male, a mating event likely to result in more robust offspring, that is, off-spring more likely to survive and reproduce¹⁴⁰. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.10: Sexual Selection

As we have already noted, it is not uncommon to see morphological and behavioral differences between the 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 potential¹⁴¹. 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 selection¹⁴². 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 exactly is happening is an interaction between costs, benefits, and specific behaviors is complex.

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 unambiguous features¹⁴³. To be a true sign of the male's robustness, this advertisement needs to be difficult to fake and so accurately 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 established and can defend such a territory, only the most robust males are reproductively successful. An alternative scenario involves males monopolizing females 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 can be counted on to help the female raise their joint offspring. In this situation, there is a significant reproductive advantage if females can accurately identify those males who will, in the future, display this type of reproductive loyalty¹⁴⁴. Under these conditions (the shared rearing of offspring with a committed male) females will be competing with other females for access to such loyal 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 health, loyalty, robustness, and fertility. The size and symmetry of a beetle's or an elk's antlers or a grasshopper's song communicate rather clearly their state of health¹⁴⁵. 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 logic¹⁴⁶.

It is critical that both females and males correctly read and/or respond to various traits, and this ability is likely to be selected for. For example, males that can read the traits of other males can determine whether they are likely to win a fight with another male; not being able to make such an accurate determination could result in crippling injuries. A trickier question is how does a female or a male determine whether a possible mate will be loyal? As with advertisements of overall robustness, we might expect that traits that are difficult or expensive to generate will play a key role. So how does one unambiguously signal one's propensity to loyalty and a 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 giver 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 tests the value of gift, a behavior we might term rational skepticism, as opposed to naive gullibility.

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, there can be situations where continuity a pregnancy could threaten the survival of the mother. In such cases, spontaneous abortion could save the female, who can go on and mate again. In a number of organisms, spontaneous abortion occurs in response to signs of reproductive distress in the fetus. Of course, spontaneous abortion is not in the interest of the offspring and we can expect that mechanisms will exist to maintain pregnancy, even if it risks the life of the mother, in part because the fetus and the mother, which related are not identical; there can be a conflict of interest between the two.

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 often disconcerting 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; behaviors that were not directly selected for but emerged in the course of the evolution of other traits, and that once present, play important roles in subsequent organismic behavior (and evolution).

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 non-committed 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).

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).

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

4.11: Curbing runaway selection

Sexual selection can lead to what has been termed (but is not really) 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 significant negative effects. All of which is to say that there will be both positive and negative selection for tail size, which will be influenced by the overall probability that a particular male mates successfully. Selection does not ever really run away, but settles down when the positive (in terms of sexual success) and negative (in terms 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 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 of generating antlers and benefits of effective 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 also 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 supported by Darwin or by scientifically-established evolutionary mechanisms, but rather by a number of pundits who used it to justify various political (that is, inherently non-scientific) positions, particularly arguing against social programs that helped the poor (often characterized as the unfit) 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 genetically inferior people should be killed, removed, or sterilized, before their "bad" traits overwhelmed a particular culture. Eugenics was a particularly influential ideology in the United States during the early part of the 20th century and inspired the genocidal programs of the Nazis in Germany. What is particularly odd about this evolutionary perspective is that it is actually anti-evolutionary, since if the unfit really were unfit, they could not possibly take over a population. In addition, it completely ignores the deeply social (cooperative) aspect of the human species.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

5: MOLECULAR INTERACTIONS, THERMODYNAMICS & REACTION COUPLING

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, and in particular the laws of thermodynamics.

5.0: INTRODUCTION

5.1: A VERY LITTLE THERMODYNAMICS

5.2: REACTIONS: FAVORABLE, UNFAVORABLE, AND THEIR DYNAMICS

5.3: THINKING ENTROPICALLY (AND THERMODYNAMICALLY)

5.4: REACTION RATES

5.5: COUPLING REACTIONS

5.6: MOLECULES, LONDON DISPERSION FORCES, AND VAN DER WAALS INTERACTIONS

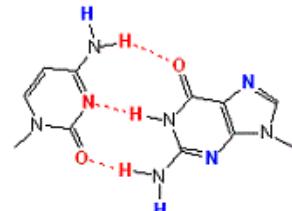
5.7: COVALENT BONDS

5.8: BOND STABILITY AND THERMAL MOTION (A NON-BIOLOGICAL MOMENT)

5.9: BOND POLARITY, INTER- AND INTRAMOLECULAR INTERACTIONS

5.10: THE IMPLICATIONS OF BOND POLARITY

5.11: INTERACTING WITH WATER



5.0: Introduction

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 and how these interactions and reactions determine the properties of substances and systems.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.1: 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, and in particular the laws of thermodynamics. What makes biological systems unique is that, unlike simpler physicochemical systems that 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 to be clear about what it means and implies 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 to 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 macroscopic and the molecular worlds 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, for example, could rust, which would complicate things. In our model the room is initially at a cosy 70 °F (~21 °C) and the bar of gold is at 200°C. What will happen? Can you generate a graph that describes how the system will behave over time? Our first task is to define the system – that is, that part of the universe in which we are interested. 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, its history. We could, if we wanted to, demonstrate quite convincingly that the system's history will have no influence on its future behavior – this is a critical difference between biological and simple physicochemical 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 energy will be transferred from the gold bar and the rest of the room and that the temperature of the gold bar will decrease over time; the behavior of the system has a temporal direction. 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, N°C, no further macroscopic change will occur. 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, constantly colliding with one another, and with molecules within the bar and the table. The molecules within the bar and the table are also vibrating. These collisions can change the velocities of the colliding molecules. (What happens if there was no air in the room? How would this change your graph of the behavior of the system?) 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.

And this is what is important about a system at equilibrium: 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 imaginary closed room system, a non-equilibrium 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 bar and the room by illuminating the bar and removing heat from the room as a whole. A temperature difference between the bar and the room could then (in theory) produce what is known as a heat engine, which can do work (that is, produce macroscopic change.) As long as we continue to heat one block and remove heat from the rest of the system, we can continue to do work, that is, 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 state of suspended animation for decades. What is critical to note, however, is 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. The organism can be reanimated when returned to normal conditions¹⁴⁷. Cryptobiosis is an genetically-based adaptation that takes energy to produce and energy is used to emerge from stasis. While the behavior of tardigrads is extreme, many organisms display a range of adaptive behaviors that enable them to survive hostile environmental conditions.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.2: Reactions: favorable, unfavorable, and their dynamics

As we will see, biological systems are extremely complex; 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 \times 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:

$$e \text{ (energy)} = m \text{ (mass)} \times c^2 \text{ (c = speed of light)}$$

To illustrate this principle, 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 you 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 that 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 the 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 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 molecules¹⁴⁸.

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 be used to do work. 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.

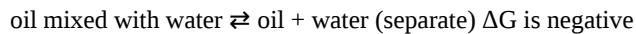
Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.3: Thinking entropically (and thermodynamically)

We certainly are in no position to teach you (rigorously) the basics of chemistry and chemical reactions (or physics for that matter), but we can provide a short refresher that focuses on the key points we will be using over and over again¹⁴⁹. The first law of thermodynamics is that while forms of energy may change, that is, can be converted 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, 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 is 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 system¹⁵⁰. 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:



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 above) and the separation of different types of molecules in a liquid (the oil-water example). 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).

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.

$$K_{eq} = k_f / k_r \tag{5.3.1}$$

$$K_f(\text{reactants}) = K_r(\text{products}) \tag{5.3.2}$$

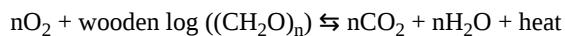
While the concentration of reactants and products of a reaction at equilibrium remains constant it is a mistake to think that the system is static. If we were to peer into the system at the molecular level we would find that, at equilibrium, reactants are continuing to form products and products are rearranging to form reactants at similar rates¹⁵¹. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.4: 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 with which we are concerned. To know the reaction's rate we need to know the reaction kinetics for the specific system with which we are dealing. 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 (O₂) the reaction:



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? Or more generally why is the world so annoyingly complex?

The answer lies in the details of the reaction, how exactly the reactants are converted into the products. For simplicity let us consider another non-chemical but rather widespread type of reaction. In this reaction system, there is a barrier between two compartments, specifically the barrier membrane that separates the inside from the 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 $\text{A}_{\text{outside}} \rightleftharpoons \text{A}_{\text{inside}}$ (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 $[\text{A}_{\text{outside}}]$, with the square brackets indicating concentration) is much greater than $[\text{A}_{\text{inside}}]$. At any moment in time, the number of collisions between $\text{A}_{\text{outside}}$ 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 $\text{A}_{\text{outside}}$ to A_{inside} . The real question is how large this net flux will be, which 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. In our simplified view of the subject, the difference between the intermediate with the highest free energy ($\Delta G_{\text{transition}}$) and the free energy of the reactants ($\Delta G_{\text{reactants}}$) corresponds roughly to the rate limiting step in the reaction and reflects the reaction's activation energy.

For a reaction to move from reactants ($\text{A}_{\text{outside}}$) 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 them 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 (or Maxwell-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 can cross the boundary layer, that is, move from outside to inside. If not, it will probably lose 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 ($\Delta G_{\text{reaction}}$) 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 $\text{A}_{\text{outside}} \rightleftharpoons \text{A}_{\text{inside}}$ 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 $\Delta G_{\text{reaction}}$ and either a large or small $\Delta G_{\text{transition}}$. So assuming that there is enough energy in the

system, and $\Delta G_{\text{transition}}$ 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 $\Delta G_{\text{transition}}$ is high enough, however the $A_{\text{outside}} \rightleftharpoons A_{\text{inside}}$ reaction will not occur to any significant extent.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.5: 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 to a region of lower concentration will be favored. So how exactly can we build the big molecules, such as DNA and proteins, and concentration gradients 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.

Unfavorable reactions occur when 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 predictions¹⁵².

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



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 to the system. The answer is that the reaction moves 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; the A + B reaction would continue in an unbalanced manner until the level of C (and D) increased and C + D to A + B reaction would balance 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 balanced the A + B to C + D reaction. This type of behavior arises directly from the fact that at equilibrium reaction systems are not static, but dynamic (at the molecular level) – things are still occurring, they are just balanced 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 return to equilibrium over time.

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 + 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 reproduction¹⁵³.

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 does the energy released by the fire come 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?

References

152. http://en.wikipeCreate Linkdia.org/wiki/Le_Chatelier's_principle

153. haha.nu/science/the-amazing-human-body/

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.6: Molecules, London Dispersion Forces, and van der Waals interactions

We have briefly (admittedly absurdly briefly) defined what energy is and begun to consider how it can be transformed from one form to another. Now we need to consider what we mean by matter, which implies an understanding of the atomic organization of the molecules that compose matter. As you hopefully know by now, all matter is composed of atoms. The internal structure of atoms is the subject of quantum physics and we will not go into it any depth. Suffice to say that each atom consists of a tiny positively charged nucleus and cloud of negatively charged electrons¹⁵⁴. Typically atoms and molecules, which after all are collections of atoms, interact with one another through a number of different types of interactions. The first are known as van der Waals interactions, which are mediated by London Dispersion Forces (LDF). These forces arise from the fact that the relatively light negatively-charged electrons are in continual movement, compared to the relatively massive and stationary positively-charged nuclei. Because charges on the protons and electrons are equal in magnitude the atom is electrically neutral, but because the electrons are moving, at any one moment, an observer outside of the atom or molecule will experience a small fluctuating electrical field.

As two molecules approach one another, their fluctuating electric fields begin to interact, this interaction generates an attractive LDF, named after its discoverer Fritz Wolfgang London (1900–1954). This force varies as $\sim 1/R^6$ where R is the distance between the molecules; this relationship means that LDFs act only over very short distances, typically less than 1 nanometer ($1 \text{ nm} = 10^{-9}\text{m}$). As a frame of reference, a carbon atom has a radius of $\sim 0.07 \text{ nm}$. 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 (the van der Waals radius of a carbon atom is $\sim 0.17 \text{ nm}$). If they move closer than this distance, the 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 molecules¹⁵⁵. This repulsive interaction keeps atoms from fusing together and is one reason why molecule can form.

Each atom and molecule has its own characteristic van der Waals radius, although since most molecules are not spherical, it is better to refer to a molecule's van der Waals surface. This surface is the closest distance that two molecules can approach one another before repulsion kicks in and drives them back away from one another. It is common to see molecules displayed in terms of their van der Waals surfaces. Every molecule generates LDFs when it approaches another, so van der Waals interactions are universal. The one exception involves pairs of small, similarly charged "ionic" molecules, that is molecules with permanent net positive or negative charge, approach each other. The strength of their electrostatic repulsion will be greater than their LDFs.

The strength of the van der Waals interactions between molecules is determined primarily by their shapes. The greater the surface complementarity, the stronger the interaction. Compare the interaction between two monoatomic Noble atoms, such as helium, neon or argon, and two molecules with more complex shapes. 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 LDFs between them is greater resulting a stronger van der Waals interaction.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.7: Covalent bonds

In the case of 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 system¹⁵⁶. This sharing of electrons produces what is known as a covalent bond. Covalent bonds are ~20 to 50 times stronger than van der Waals interactions. What exactly does that mean? Basically, it takes much more energy to break these bonds. While the bonded form of atoms in a molecule is always more stable than the unbounded form, it may not be stable enough to withstand the energy delivered through collisions with neighboring molecules. Different bonds between different atoms in different molecular contexts differ in terms of bond stability; the bond energy refers the energy needed to break a particular bond. A molecule is stable if the bond energies associated with bonded atoms within the molecule are high enough to survive the energy delivered to the molecule through either collisions with neighboring molecules or the absorption of energy (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-methylpropane¹⁵⁷. As molecules become larger, as is the case with many biologically important molecules, it can become impossible to appreciate their underlying organization 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. To a first order approximation, a molecule's properties are based on its shape, which is dictated by how the various atoms within the molecule are connected to one another. These geometries are imposed by each atom's quantum mechanical properties and (particularly as molecules get larger, as they so often do in biological systems) the interactions between different parts of the molecule. Some atoms, common to biological systems, such as hydrogen (H), can form only a single covalent bond. Others can make two (oxygen (O) and sulfur (S)), three (nitrogen (N)), four (carbon (C)), or five (phosphorus (P)) bonds.

In addition to smaller molecules, biological systems contain a number of distinct types of extremely large molecules, composed of thousands of atoms; these are known as macromolecules. Such macromolecules are not rigid; they can often fold back on themselves leading to intramolecular interactions. There are also interactions between molecules. The strength and specificity of these interactions can vary dramatically and even small changes in molecular structure can have dramatic effects.

Molecules and molecular interactions are dynamic. Collisions with other molecules can lead to parts of a molecule rotating around a single bond¹⁵⁸. 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 one another and 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 to constrain molecular geometry. As we will see later on with the peptide bond, which occurs between a carbon (C) and a nitrogen (N) atom in polypeptide chain, is an example of such a resonance structure. Similarly, the ring structures found in the various "bases" present in nucleic acids produces flat structures that can pack one top of another. These various geometric complexities combine to make predicting a particular molecule's three dimensional structure increasingly difficult as its size increases.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.8: Bond Stability and Thermal Motion (a non-biological moment)

Molecules do not exist out of context. In the real, or at least the biological world they do not sit alone in a vacuum. Most biologically-relevant 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, -273.15 °C, -459.67 °F). At this, biologically irrelevant temperature, molecular movements are minimal, but not apparently absent all together¹⁵⁹. 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 a system is a measure of the average kinetic energy of the molecules within it. The average kinetic energy is:

$$E_k = \frac{1}{2}(\text{average mass}) \times (\text{average velocity})^2 \quad (5.8.1)$$

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 all of the different molecules has one value. This is not to say that all molecules have the same kinetic energy, they certainly do not; each forms a distribution that is characterized by its average energy, this distribution is known as the Boltzmann (or Maxwell-Boltzmann) distribution. The higher the temperature, the more molecules will have a higher kinetic energy.

In a gas we can largely overlook the attractive intermolecular interactions between molecules 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 decreases. As the average kinetic energy gets low enough, the molecules will form a liquid. In a liquid, the movement of molecules is not enough to completely 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 in equilibrium. What we will find, if we think about it, is that there is a reaction going on, that reaction is:



At the particular temperature, the liquid phase is favored, although there will be some molecules in the system's gaseous phase. 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 this equilibrium state, that is, the relative amounts of molecules in the gaseous versus the liquid states will change. The equilibrium is dynamics, in that different molecules may be in gaseous or the liquid states, even though the level of molecules will be steady.

In a liquid, while molecules associate with one another, they can still move with respect to one another. 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 container, whose shape is independent of what it contains. In a solid the molecules are tightly associated with one another and so do not translocate with respect to one another (although they can rotate and jiggle in various ways). Solids do not flow. The cell, or more specifically, the cytoplasm, acts primarily as a liquid and many biological processes take place in the liquid phase: this has a number of implications. First molecules, even very large macromolecules, move with respect to one another. Driven by thermal motions, molecules will move in a Brownian manner, a behavior known as a random walk.

Thermal motion will influence whether and how 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; A and B interact to form a complex, AB. Assume that this complex is held together by van der Waals interactions. In an aqueous solution, the A:B complex is colliding with water molecules. These water molecules have various energies (from low to high), as described by the Boltzmann distribution. There is a probability that in any unit of time, one or more of these collisions will deliver energy greater than the interaction energy hold them together; this will lead to the disassociation of the AB complex into separate A and B molecules. Assume we start with a population of 100% AB complexes, the time it takes for 50% of these molecules to dissociate into A and B is 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 remain intact. Why?

Because we cannot predict exactly which collisions will provide sufficient energy to disassociate a particular AB complex¹⁶⁰. 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 population it is not possible to predict which alleles will be lost and which fixed, or if and when fixation will occur. A hallmark of a stochastic process is that they are best understood in terms of probabilities.

Stochastic processes are particularly important within biological systems because, generally, cells are small and may 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 that protein and (usually) only one or two copies of the gene (that is, the DNA molecule) present, we will find that whether or not a copy of the protein is bound to a specific region of the DNA is a stochastic process¹⁶¹. If there are enough cells, then the group average will be predictable, but the behavior of any one cell will not be. 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 stochastic property of cells can play important roles 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.

Questions to answer & to ponder:

- Explain why the Boltzmann distributions is not symmetrical around the highest point.
- 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? How might changes in temperature influence molecular shape (particularly in a macromolecule)?
- Why are some liquids more viscous (thicker) than others? Draw a picture of your model.
- In considering generating a graph that describes radioactive decay or the dissociation of a complex (like the AB complex discussed above) as a function of time, why does population size matter?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.9: 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 papers¹⁶². Each type of atom (element) has a characteristic electronegativity, a measure of how tightly it 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 the two atoms and the bond is said to be non-polar (meaning without direction). 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 be shared un-equally. On average, there will be more electrons more of the time around the more electronegative atom and fewer around the less electronegative atom. This leads to partially negatively and positively-charged regions to the bonded atoms - the bond has a direction. 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 termed 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; like covalent bonds they have a directionality to them – 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 atom interacting with an O or an N atom, 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 than it does to break a covalent bond involving a H atom.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.10: 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 considering a pure sample that contains extremely large numbers 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 delivered by collision with other molecules. The molecules are said to be a gaseous state and the transition from liquid to gas is the boiling point. Similarly, starting with a liquid, when we reduce the temperature, the interactions between molecules become longer lasting until such a temperature is reached that the energy transferred through collisions is no longer sufficient to disrupt the interactions between molecules¹⁶³. 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: solid to liquid and liquid to gas.

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 and the ability to form H-bonds. For example, neither CH₄ (methane) and Ne (neon) contain polar bonds and cannot form intra-molecular H-bond-type electrostatic interactions. In contrast NH₃ (ammonia), H₂O (water), and FH (hydrogen fluoride) have three, two and one polar bonds, respectively, and can take part in one or more intra-molecular 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 the presence of polar bonds influences these properties.

In particular water stands out as dramatically different from the rest of the molecules, with significantly higher (> 70°C) melting and boiling point than its neighbors. So why is water different? 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 its 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 neighboring 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. Even if one is broken, however, the 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 water 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. Hydrophobic molecules interact with other molecules, including water molecules, only through van der Waals interactions. Molecules that can make H-bonds with water are termed **hydrophilic**. As molecules 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 greater detail in the next chapter.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

5.11: Interacting with Water

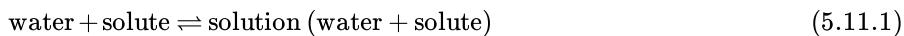
We can get an idea of the hydrophilic, hydrophobic/hydroapathetic, and amphipathic nature of molecules through their behaviors when we try to dissolve them in water. Molecules like sugars (carbohydrates), alcohols, and most amino acids are primarily hydrophilic. They dissolve readily in water. Molecules like fats are highly hydrophobic (hydroapathetic), and they do not dissolve significantly in water. 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, if 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. 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 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: In a liquid water molecules will typically be found in a state that maximizes the number of H-bond-type electrostatic interactions present. And because these interactions have a distinct, roughly tetragonal geometry, their presence constrains the possible orientations of molecules with respect to one another. This constraint is captured when water freezes; it is the basis for ice crystal formation, why the density of water increases before freezing, and why ice floats in liquid water¹⁶⁴. In the absence of the hydrophobic 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 to a much smaller number of configurations that result in maximizing 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; they 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.

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 $\Delta G = \Delta H - T\Delta S$, if ΔS is negative, then $-T\Delta S$ will be positive. The ΔG of a thermodynamically favorable reaction is, by definition, negative. This implies that the reaction:



will be 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 interior?
- Under what conditions might entropic effects influence the interactions between two solute molecules?

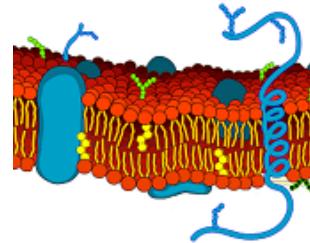
Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

6: MEMBRANE BOUNDARIES AND CAPTURING ENERGY

- 6.0: INTRODUCTION
- 6.1: DEFINING THE CELL'S BOUNDARY
- 6.2: THE ORIGIN OF BIOLOGICAL MEMBRANES
- 6.3: TRANSPORT ACROSS MEMBRANES
- 6.4: TRANSPORT TO AND ACROSS THE MEMBRANE
- 6.5: CHANNELS AND CARRIERS
- 6.6: GENERATING GRADIENTS: USING COUPLED REACTIONS AND PUMPS
- 6.7: SIMPLE PHOTOTROPHS
- 6.8: CHEMO-OSMOSIS (AN OVERVIEW)
- 6.9: OXYGENIC PHOTOSYNTHESIS
- 6.10: CHEMOTROPHS
- 6.11: USING THE ENERGY STORED IN MEMBRANE GRADIENTS
- 6.12: OSMOSIS AND LIVING WITH AND WITHOUT A CELL WALL
- 6.13: AN EVOLUTIONARY SCENARIO FOR THE ORIGIN OF EUKARYOTIC CELLS
- 6.14: MAKING A COMPLETE EUKARYOTE



6.0: Introduction

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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.1: 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?

When a new cell is formed its plasma membrane is derived from the plasma membrane of the progenitor 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. Structurally diverse molecules, such as cholesterol and phospholipids, are both considered lipids. 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**¹⁶⁵.

All lipids have two distinct domains: a hydrophilic (circled in red in this figure domain characterized by polar regions and one or more hydrophobic/hydroapathetic domains that are usually 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 aqueous 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 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); while the details are beyond our scope here you might be able to generate a plausible mechanism. A number of distinct mechanisms are used to insert molecules into membranes, but they all 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 generic classes, 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 its 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 with one another, 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 these bilayers 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, proteins embedded within the membrane will be immobile; if is in the liquid state, these proteins will move by diffusion, that is, by thermally driven movement, within the plane of the membrane. In addition, since lipids and proteins are closely associated with one another in the membrane, the physical state of the membrane can influence the activity of embedded proteins, 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 archaea¹⁶⁶. 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 temperatures¹⁶⁷. 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 the archaea 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) and hydrophilic molecules through¹⁶⁸. Given the importance of membrane integrity, you will (perhaps) not be surprised to find “double-headed” lipids in organisms that live at high temperatures, known as 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 °C¹⁶⁹. 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 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 here¹⁷⁰.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.2: 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, from 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 leaky¹⁷¹. 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:

- Is the universe at equilibrium? If not when will it get to equilibrium?
- 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?
- Draw what "double-headed" lipids look like in the context of a bilayer membrane.
- 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?
- 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?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.3: 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 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 ($> 60 \text{ mg/ml}$) 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.

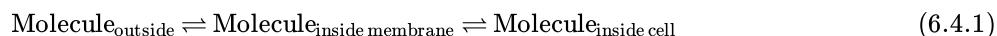
A lipid bilayer membrane poses an interesting barrier to the movement of molecules. First for larger molecules, particles or other organisms, it acts as a physical barrier. Typically when larger molecules, particles (viruses), and other organisms enter a cell, they are actually engulfed by the membrane, 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 exocytosis (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 subsequent 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).

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.4: 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”. Which types of carriers, channels, and pumps are present will determine what types of molecules move through the cell’s 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:



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 each 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 during the transport reaction. 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 across the membrane 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 as well as the concentrations of the molecules.

If a molecule is hydrophobic (non-polar) it will be more soluble in the hydrophobic environment that exists in the central region 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. You can start by drawing the steps involved in "flipping" a lipid molecule's orientation with a membrane.

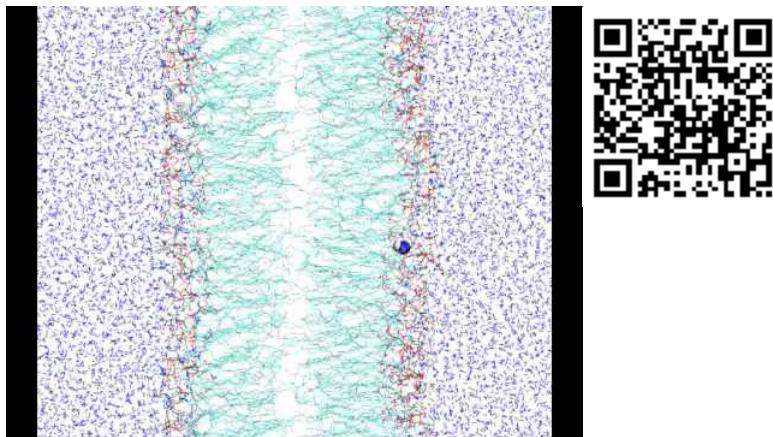
Let us begin with water itself, which is small and uncharged. When a water molecule begins to leave the aqueous phase and enter the hydrophobic (central) region of the membrane, there are no H-bonds to take the place of those that are lost, no strong molecular handshakes; the result is that often the molecule is “pulled back” into the water phase. 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 $\text{Water}_{\text{outside}} \rightleftharpoons \text{Water}_{\text{inside}}$ reaction is low enough that water can pass through a membrane (in both directions) at a reasonable rate.

Small non-polar molecules, such as O₂ and CO₂, 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 involved in passing through the membrane. However now we begin to see changes in the free energies of the molecules on the inside and outside of the cell. For example, in organisms that depend upon O₂ (obligate aerobes), the O₂ outside of the cell comes from the air; it is ultimately generated by plants that release O₂ as a waste product. Once O₂ 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 O₂ outside the cell will be greater than the concentration of O₂ inside the cell. That means that the free energy of O₂ outside will be greater than the free energy of O₂ inside. The reaction O₂ _{outside} \rightleftharpoons O₂ _{inside} is now thermodynamically favorable and there will be a net flux of O₂ 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:



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.



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

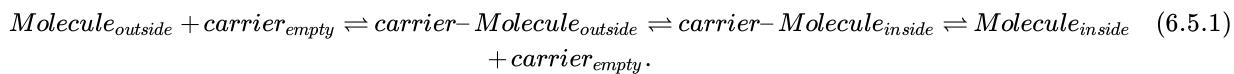
Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.5: 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 features 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 can be quite high for highly hydrophilic and larger molecules, we will need a catalyst to reduce it to that the reaction can proceed. There are two generic types of membrane permeability catalysts: carriers and channels.

Carrier proteins are membrane proteins that shuttle back and forth across the membrane. They 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. Both the movements of carrier and cargo across the membrane, and the release of transported molecules, are driven by thermal motion (collisions with other molecules), so no other energy source is necessary. We can write this class of reactions as:



There are many different types of carrier molecules and each type of carrier has preferred cargo molecules. Related molecules may be bound and transported, but with much less specificity (and so at a much lower rate). 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 membrane carrier is a class of antibiotics, known generically as ionophores, that carry ions across membranes. They kill cells by disrupting the normal ion balance across the cell's membrane and within the cytoplasm, which in turn disrupts normal metabolic activity¹⁷². 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 ~10⁵ times more effectively than it binds Na⁺ ions. Together with the bound ion, the valinomycin molecule shuttles 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, K⁺ will tend to bind to the valinomycin molecule, whereas on the side where [K⁺] is low, the K⁺–valinomycin complex will dissociate (in response to collisions with other molecules), breaking the valinomycin–K⁺ interaction. Where there is K⁺ concentration gradient, 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 and contain an aqueous 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 significantly lower activation energy than would be associated with moving through the lipid part of the membrane in the absence of the channel. Channels are generally highly selective in terms of which particles will pass through them. For example, there are channels in which 10,000 potassium ions will pass through for every one sodium ion.

Often the properties of these channels 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. Channels cannot, however, determine in which direction an ion will move - that is determined by the ion gradient across the membrane. The transition between open and closed states can be regulated through a number of processes, including the reversible binding of small molecules to the protein and various other molecular changes (which we will consider when we talk about proteins).

Another method of channel control depends on the fact that channel proteins are i) embedded within a membrane and ii) contain 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⁺ concentration¹⁷³. The ion gradient in turn can

produce an electrical field 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?
- Draw a picture of valinomycin's position and movements within a typical membrane. What drives the movement of valinomycin in the membrane and what factors lead to a net flux in K^+ movement?
- 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? Explain your reasoning.

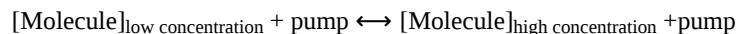
Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.6: Generating gradients: using coupled reactions and pumps

Both carriers and channels 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 transport specific types molecules across a membrane and against their concentration gradient. We will call these types of molecule pumps and write the reaction it is involved in as:

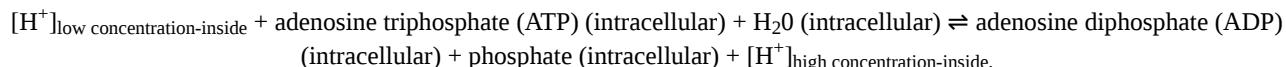


As you might already suspect this is a thermodynamically unfavorable reaction. Like a familiar macroscopic pump, it will require the input of energy. We will have to “plug in” our molecular pump into some source of energy. What energy sources are available to biological systems? Basically we have two 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; the absorbance of light (energy) is 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 that reaction is 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 will 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 orientation. 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). One of the most common ones involve 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, drives the synthesis of ATP. The movement of H⁺ from the side of the membrane with relatively high [H⁺] to that of relatively low [H⁺] is coupled to the ATP synthesis through the membrane bound ATP synthase enzyme:



This reaction can run in reverse, in which case ATP is hydrolyzed to form ADP and phosphate, and H⁺ is moved against its concentration gradient, that is, from a region of low concentration to a region of higher concentration.



In general, 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.7: Simple Phototrophs

Phototrophs are organisms that capture particles of light (photons) and transform their electromagnetic energy into energy stored in unstable molecules, such as ATP and carbohydrates. Phototrophs eat light. 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 considerations, a particular molecule can only absorb photons of specific wavelengths (energies). Because of this property, we can identify molecules at great distances based on the photons they absorb or emit, 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 if at all. To capture this energy, organisms have evolved the ability to synthesize molecules, known as pigments to capture, and therefore allow organisms to use (absorb) visible light. The color we see for a typical pigment is the color of the light that is it does not absorb but rather that it reflects. 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 a chemical system, is provided by the archaea *Halobacterium halobium*¹⁷⁴. Halobacteria are extreme halophiles (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 A¹⁷⁵. Together the two, opsin + retinal, form the functional bacteriorhodopsin protein.

Because its electrons are located in extended molecular orbitals with energy gaps between them that are of the same order as the energy of visible light, absorbing of a photon of visible light moves an electron from a lower to a higher energy molecular orbital. Such extended molecular orbitals are associated with molecular regions that are often drawn as containing alternating single and double bonds between carbons; these are known as conjugated π orbital systems. Conjugated π systems are responsible for the absorption of light by pigments such as chlorophyll and heme (the pigment that makes blood red). 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 bacteriorhodopsin 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 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 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):

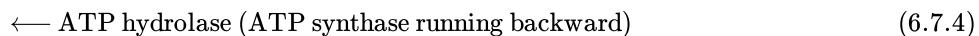
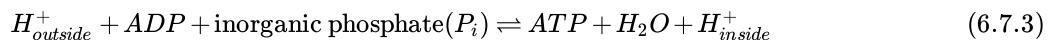


H⁺ is always present in water from the autoionization ([H⁺] = 1 × 10⁻⁷ for neutral water at room temperature) and it is these H⁺s that move.

In addition to the chemical gradient that forms when H⁺ ions are pumped out of the cell by the bacteriorhodopsin + light + water reaction, an electrical field is also established. There are excess positive charges outside of the cell (from H⁺ being moved there) and excess negative 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 across the membrane. Now, an H⁺ ion outside of the cell will experience two distinct forces, those associated with the electric field and those arising from the concentration gradient. If there is a way across the membrane, the [H⁺]

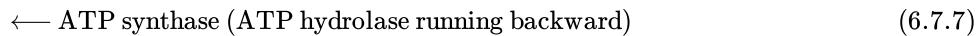
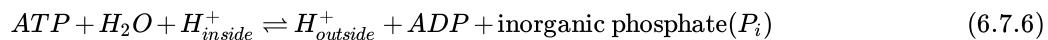
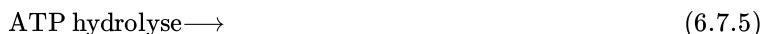
gradient will lead to the movement of H^+ ions back into the cell. Similarly the electrical field will also drive the positively charged H^+ back into the cell. The formation of the $[H^+]$ gradient basically generates a battery, a source of energy, into which we can plug in our pump.

So how does the pump tap into this battery? The answer is through a second membrane protein, an enzyme known as the H^+ - driven ATP synthase. H^+ ions move through the ATP synthase molecule in what is a thermodynamically favorable ($\Delta G < 0$) reaction. The ATP synthase couples this favorable movement to an unfavorable chemical reaction, a condensation reaction:



This reaction will continue as long as light is absorbed. Bacteriorhodopsin acts to generate a H^+ gradient and the H^+ gradient persists. That means that even after the light goes off (that is, night time) the H^+ gradient persists until H^+ ions have moved through the ATP synthase. ATP synthesis continues until the H^+ gradient no longer has the energy required to drive the ATP synthesis reaction. The net result is that the cell uses light to generate ATP, which is stored for later use. ATP acts as a type of chemical battery, in contrast to the electrochemical battery of the H^+ gradient.

An interesting feature of the ATP synthase molecule is that as H^+ ions move through it (driven by the electrochemical power of the H^+ gradient), a region of molecule rotates. It rotates in one direction when it drives the synthesis of ATP and in the opposite direction to couple ATP hydrolysis to the pumping of H^+ ions against their concentration gradient. In this form it is better called an ATP hydrolase:



Because the enzyme rotates when it hydrolyzes ATP, it is rather easy to imagine how the energy released through this reaction could be coupled, through the use of an attached propeller or paddle-like extension, to cellular or fluid movement.

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? When does ATP production stop and why?
- 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?
- Indicate how ATP hydrolysis or tapping into the H^+ gradient could lead to cell movement.

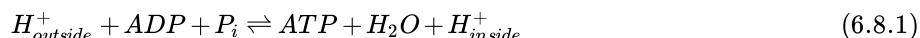
Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

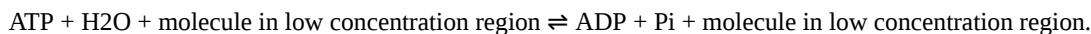
6.8: Chemo-osmosis (an overview)

One of the most surprising discoveries in biology was the wide spread, almost universal use of H⁺ gradients to generate ATP. What was originally known as the chemiosmotic hypothesis was produced by the eccentric British scientist, Peter Mitchell (1920–1992)¹⁷⁶. 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 might have a similar evolutionary root, that they are homologous? Basically, it is the observation that in both light- and chemical-based processes captured energy is transferred through the movement of electrons through a membrane-embedded “electron transport chain”. An electron transport chain involves a series of membrane and associated proteins and a series of reduction-oxidation or redox reactions (see below) during which electrons move from a high energy donor to a lower energy acceptor. Some of the energy difference between the two is used to move H⁺ ions across a membrane, generating a H⁺ concentration gradient. Subsequently the thermodynamically favorable movement of H⁺ down this concentration gradient (across the membrane) is used to drive ATP synthesis, a thermodynamically unfavorable process. ATP synthesis itself involves the rotating ATP synthase. The reaction can be written:



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 (ATP hydrolase) that can pump H⁺ (or other molecules) against its concentration gradient. Such pumping ATPases establishes most biologically important molecular gradients across membranes. In such a reaction:



The most important difference between phototrophs and chemotrophs is how high energy electrons enter the electron transport chain.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.9: Oxygenic Photosynthesis

Compared to the salt loving archaea Halobium with its purple bacteriorhodopin-rich membranes, photosynthetic cyanobacteria (which are true bacteria), green algae, and higher plants (both eukaryotes) use more complex molecular systems through which 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 bonding structure (drawn as a series of alternating 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 function¹⁷⁷.

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 often compete with one another for light; increasing the efficiency of the system through which an organism captures light can provide the organism with a selective advantage.

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 set 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 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.

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 they absorbed a photon) 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 reduction-oxidation (redox) reaction. Where an electron is within a molecule's electron orbital system determines the amount of energy present in the molecule. It therefore 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 (decreased) and the molecule is said to have been "oxidized"¹⁷⁸. Since electrons, like energy, are neither created nor destroyed in biological systems (remember, no nuclear reactions are occurring), 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 from the reduced molecule can be used to drive various types of thermodynamically unfavorable metabolic reactions.

Given the conservation of matter and energy in biological systems, if electrons are leaving the photosynthetic system (in the non-cyclic process) they must be replaced. So where could they be coming 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:



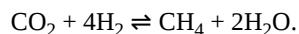
The four electrons, derived from two molecules of water, pass to the reaction center, while the $4H^+$ s contribute to the proton gradient across the membrane¹⁷⁹. O_2 is a waste product of this reaction. Over millions of years, the photosynthetic release of O_2 changed the Earth's atmosphere from containing essentially 0% molecular oxygen to the current ~21% level at sea level. Because O_2 is highly reactive, this transformation is thought to have been a major driver of subsequent evolutionary change. However, there remain 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; its level is based on how much O_2 is released into the atmosphere by oxygenic photosynthesis and how much is removed by various reactions, such as the decomposition of plant materials. When large amounts of plant materials are buried before they can decay, such as occurred with the formation of coal beds during the Carboniferous period (from ~360 to 299 million years ago), the level of atmospheric O_2 increased dramatically, up to an estimated ~35%. It is speculated that such high levels of atmospheric molecular oxygen made it possible for organisms without lungs (like insects) to grow to gigantic sizes¹⁸⁰.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.10: Chemotrophs

Organisms that are not phototrophic capture energy from other sources, specifically by transforming thermodynamically unstable molecules into more stable species. Such organisms are known generically as chemotrophs. They can be divided into various groups, depending upon the types of food molecules (energy sources) they use: these include organotrophs, which use carbon-containing molecules (you yourself are an organotroph) and lithotrophs or rock eaters, which use various inorganic molecules. In the case of organisms that can “eat” H₂, the electrons that result are delivered, along with accompanying H⁺ ions, to CO₂ to form methane (CH₄) following the reaction:



Such organisms are referred to as methanogens (methane-producers)¹⁸¹. In the modern world methanogens (typically archaea) are found in environments with low levels of O₂, such as your gut. In many cases reactions of this type can occur only in the absence of O₂. In fact O₂ 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 originally arose in the absence of molecular O₂ adapted as significant levels of O₂ began to appear in 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 employ can accept. If only electrons with high amounts of energy can be captured, which is often the case, then inevitably large amounts of energy are left behind. On the other hand, the lower the amount of energy that an electron acceptor can accept, the more energy can be extracted and captured from the original “food” molecules and the less energy is left behind. Molecular oxygen is unique in its ability to accept low energy electrons. For example, consider an organotroph that eats carbohydrates (molecules of the general composition [C₆H₁₀O₅]_n), a class of molecules that includes sugars, starches, and wood, a process known as glycolysis, from the Greek words meaning sweet (glyco) and splitting (lysis). In the absence of O₂, that is under anaerobic conditions, the end product of the breakdown of a carbohydrate leaves ~94% of the theoretical amount of energy present in the original carbohydrate molecule remaining in molecules that cannot be broken down further, at least by most organisms. These are molecules such as ethanol (C₂H₆O). However, when O₂ is present, carbohydrates can be broken down more completely into CO₂ and H₂O, a process known as respiration. In such O₂ using (aerobic) organisms, the energy released by the formation of CO₂ and H₂O is stored in energetic electrons and used to generate a membrane-associated H⁺ based electrochemical gradient that in turn drives ATP synthesis, through a membrane-based ATP synthase. In an environment that contains molecular oxygen, organisms that can use O₂ as an electron acceptor have a distinct advantage; instead of secreting energy rich molecules, like ethanol, they release the energy poor (stable) molecules CO₂ and H₂O.

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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.11: Using the energy stored in membrane gradients

The energy captured by organisms 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. Such processes involve what is known as coupled transport¹⁸². They rely on membrane-bound proteins that enable 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**, 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 an energetically unfavorable reaction, the movement of an ion against its concentration gradient. Without these pumps, and the chemical reactions that drive them, the membrane battery would quickly run down. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.12: 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 $[H_2O]_{\text{out}}$ will necessarily be higher than the concentration of water inside the cell $[H_2O]_{\text{in}}$. This concentration gradient in solvent leads to the net movement of water into the cells¹⁸³. 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 Nobel prize in chemistry, was that the membrane also contains water channels, known as aquaporins¹⁸⁴. 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. In addition to water, aquaporin-type proteins can also facilitate the movement of other small uncharged molecules across a membrane.

The difference or gradient in the concentrations of water across the cell membrane, together with the presence of aquaporins, leads to a system that is capable of doing work. The water gradient, can lift a fraction of the solution against the force of gravity, something involved in having plants stand up straight¹⁸⁵. How is this possible? If we think of a particular molecule in solution, it will move 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 is, without a preferred direction [this video¹⁸⁶ 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 $[H_2O]_{\text{outside}}$ 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 built into the wall as it is 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:

- Make a graph of the water concentration across a typical cellular membrane for an organism living in fresh water; explain what factors influenced your drawing.
- Look at this video: <https://www.youtube.com/watch?v=VctA...ature=youtu.be>. How could you use reverse osmosis to purify water?
- 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?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.13: 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 rampant. These are also, dare we remind you, events that took place billions of years ago. But such obstacles do not mean we cannot draw interesting, albeit speculative conclusions – there is relevant data present in each organisms' genetic data (its genotype), the structure of its cells, and their ecological interactions. This is data that can inform and constrain our speculations.

Animal cells do not have a rigid cell wall; its absence 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 perform such functions. 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 back 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 the 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 might conclude that organisms in the eukaryotic lineage never had a cell wall, rather than that they had one once and subsequently lost it. In this scenario, the development of eukaryotic cell walls by fungi and plants represents an example of convergent evolution and that these structures are analogous (rather than homologous) to the cell walls of prokaryotes (bacteria and archaea).

But now a complexities 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? One approach is to actively pump the water that flows into them back out using an organelle known as a contractile vacuole. Water accumulates within the contractile vacuole, a membrane-bounded structure within the cell; as the water accumulates the contractile vacuole inflates. To expel the water, the vacuole connects with the plasma membrane and is squeezed out by the contraction of a cytomuscular system. 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 such as cytomuscular system was originally involved in predation, that is, enabling the cell to move its membranes, 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 during cell division; distinctly different systems than is used prokaryotes¹⁸⁷. So which came first, different cell division mechanisms, which led to differences in membrane behavior, with one leading to a predatory active membrane and the other that led to a passive membrane, perhaps favoring the formation of a cell wall?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

6.14: Making a complete eukaryote

Up to this point we have touched on only a few of the ways that prokaryotes (bacteria and archaea) differ from eukaryote. The major ones include 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 further later on) and the relative locations of chemo-osmotic and photosynthetic systems in 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 derived from the plasma membrane. In contrast, in eukaryotes (plants, animals, fungi, protozoa, and other types of organisms) 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 located within 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 was engulfed by the ancestral pro-eukaryotic cell, with the host cell's membrane surrounding the engulfed cells plasma membrane. A more detailed molecular analysis reveals that the mitochondrial and chloroplast electron transport systems, as well as the ATP synthase proteins, more closely resemble those found in two distinct types of bacteria, rather than in 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 γ -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 cyanobacteria¹⁸⁸. In contrast, many of the nuclear systems found in eukaryotes appear more similar to systems found in archaea. How do we make sense of these observations?

Clearly when a eukaryotic cell divides it must have also replicated its mitochondria and chloroplasts, otherwise they would eventually be lost through dilution. 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 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. In addition, it was discovered that mitochondria and chloroplasts contained circular DNA molecules organized in a manner similar to the DNA molecules 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*¹⁸⁹ 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, initially to detoxify molecular oxygen, and then to utilize molecular oxygen as an electron acceptor so as to maximize the energy that could be derived from the break down of complex molecules. All eukaryotes (including us) are descended from this mitochondria-containing eukaryotic ancestor, which appeared around 2 billion years ago. The second endosymbiotic event in eukaryotic evolution occurred when a cyanobacteria-like bacterium formed an 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 a mutually beneficial interaction (symbiosis)¹⁹⁰. For example, a number of insects have intracellular bacterial parasites and some pathogens and parasites live inside human cells¹⁹¹. In some cases,

even these parasites can have parasites. Consider the mealybug *Planococcus citri*, a multicellular eukaryote; this organism contains cells known as bacteriocytes. Within these cells are *Tremblaya princeps* type β -proteobacteria. Surprisingly, within these *Tremblaya* bacterial cells, which lie within the mealybug cells, live *Moranella endobia*-type γ -proteobacteria¹⁹². 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 endosymbiotic relationship with eukaryotic green algal cells, to form a “secondary” endosymbiont. Similarly, secondary endosymbionts have been engulfed by yet another eukaryote, to form a tertiary endosymbiont¹⁹³. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

7: THE MOLECULAR NATURE OF HEREDITY

Topic hierarchy

- 7.0: INTRODUCTION
- 7.1: DISCOVERING HOW NUCLEIC ACIDS STORE GENETIC INFORMATION
- 7.2: LOCATING HEREDITARY MATERIAL WITHIN THE CELL
- 7.3: IDENTIFYING DNA AS THE GENETIC MATERIAL
- 7.4: UNRAVELING NUCLEIC ACID STRUCTURE
- 7.5: DISCOVERING THE STRUCTURE OF DNA
- 7.6: DNA, SEQUENCES, AND INFORMATION
- 7.7: DISCOVERING RNA: STRUCTURE AND SOME FUNCTIONS
- 7.8: DNA REPLICATION
- 7.9: REPLICATION MACHINES
- 7.10: FURTHER REPLICATION COMPLEXITIES
- 7.11: MUTATIONS, DELETIONS, DUPLICATIONS & REPAIR
- 7.12: GENES AND ALLELES
- 7.13: MUTATIONS AND EVOLUTION
- 7.14: TRIPLET REPEAT DISEASES AND GENETIC ANTICIPATION



7.0: Introduction

In which we discover how the physical basis of inheritance, DNA, was identified and learn about the factors that influence how DNA 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 within 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 or accurate understanding of genetic mechanisms. While it was very clear, based on the experiences of plant and animal breeders, that organisms varied with respect to one another and that part of that variation was inherited, the mechanism by which genetic information is 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 mechanisms¹⁹⁴. For example, some thought that evolutionary variation was generated by a type of inner drive or logic within the organism or even the species. 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 (1744–1829) suggested that inheritance somehow reflected the desires of the parent¹⁹⁵. This would have predicted a type of "internally directed" evolution. In contrast Darwin's model, based on random variations in the genetic material, 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 in 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 into his theory? The simplest explanation would be that Darwin was unaware of Mendel's work - in fact, the implications of Mendel's work were largely 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 quite specialized and could be attributed to 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 versions 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 variation at a single genetic locus (position or gene). The vast majority of traits, however, 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 multiple 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 present, can vary dramatically depending upon the rest of an organism's genotype. Finally, in an attempt to establish the general validity of his conclusions Mendel examined the behavior of a number of other plants, including hawkweed. Unfortunately, hawkweed uses a specialized, asexual reproductive strategy, known as apomixis, during which Mendel's laws are not followed¹⁹⁶. This did not help reassure Mendel or others that his genetic laws were universal or useful. Subsequent work, published in 1900, led to the recognition of the general validity of Mendel's basic conclusions¹⁹⁷.

Mendel deduced that there are stable hereditary "factors" - which became known as genes - and that these genes are present as discrete objects within an organism. Each gene can exist in a number of different forms, known as alleles. In many cases specific alleles (versions 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 or intolerant 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 LCT gene; the LCT gene encodes the enzyme lactase, which must be expressed for an organism to digest lactose¹⁹⁸. When a cell divides, its genes must be replicated so that each daughter cell receives a full set of genes (a genome). The exact set of alleles a cell 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 are linked together in a physical way, but that this linkage is not permanent - that is, processes exist that can shuffle linked genes (or rather the alleles of genes).

In sexually reproducing (as opposed to asexual or clonal) organisms, like the peas that Mendel originally worked with, two copies of each gene are present in each somatic (body) cell. Such cells are said to be diploid. During sexual reproduction, specialized cells (known as germ cells) are produced; these cells contain only a single copy of each gene and are referred to as haploid (although monoploid might 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. In a population there are typically a number of different alleles for each particular gene, and many thousands of different genes. An important feature of sexual reproduction is that the new organism reflects a unique combination of alleles inherited from the two parents. This increases the genetic variation within the population, which enables the population (as opposed to specific individuals) to deal with a range of environmental factors, including pathogens, predators, prey, and competitors. It leaves unresolved, however, exactly how genetic information is replicated and how new alleles form, how information is encoded, regulated, and utilized at the molecular, cellular, and organismic levels.

References

194. http://en.Wikipedia.org/wiki/The_eclipse_of_Darwinism
195. It is perhaps worth reading Evolution in Four Dimensions (reviewed here: www.ncbi.nlm.nih.gov/pmc/articles/PMC1265888/) which reflects on the factors that influence selection.
196. Apomixis in hawkweed: Mendel's experimental nemesis: <http://www.ncbi.nlm.nih.gov/pubmed/21335438>
197. https://en.Wikipedia.org/wiki/Gregor...endel.27s_work
198. <http://www.hhmi.org/biointeractive/m...es-and-culture>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.1: 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 are composed of cells, and that 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; because of resolution limits associated with available microscopes, these studies were restricted to larger eukaryotic cells. One such component of eukaryotic cells was the nucleus. At this point it is worth remembering that most cells do not contain pigments. Under these early microscopes, 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. Cross-linking stops these molecules from moving with respect to one another. Alternatively, the cell could be treated with organic solvents such as alcohols; this leads to the local 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 methods¹⁹⁹ and electron microscopes, such crude fixation methods became 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-regions 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 ions) and eosin, which leaves the cytoplasm pink and the nucleus dark blue²⁰⁰. 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) organisms²⁰¹. 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 divides, losing its (typically) roughly spherical shape which was replaced by discrete stained strands, known as chromosomes (colored bodies). In 1887 Edouard van Beneden reported that the number of chromosomes in a somatic (diploid) cell was constant for each species and that different species had different numbers of chromosomes. Within a particular species the individual chromosomes can be recognized based on their distinctive sizes and shapes. For example, in the somatic cells of the fruit fly *Drosophila melanogaster* there are two copies of each of 4 chromosomes. 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 eggs), each cell received one and only one copy of each chromosome. This strongly suggested that Mendel's genetic factors were associated with chromosomes²⁰². 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 a particular chromosome. These observations provided a physical explanation for the observation 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 nuclein²⁰³. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.2: Locating hereditary material within the cell

Further evidence suggesting that hereditary information was probably localized in the nucleus emerged from transplantation experiments carried out in the 1930's by Joachim Hammerling 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. Through microdissection, it is 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. You can remove and transplant regions of one organism (cell) to another. Finally, different species of *Acetabularia* have morphologically distinct "caps" that regrow faithfully following amputation. In his experiments, he removed the head and stalk regions from one individual, leaving a "holdfast" region that was much smaller but, importantly, contained the nucleus. He then transplanted large regions of anuclear stalk derived from an organism 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's morphology was located within the region of the cell that contained the nucleus, rather than dispersed throughout the cytoplasm. It's just a short step from these experimental results to the conjecture that all genetic information is located within the nucleus.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.3: Identifying DNA as the genetic material

The exact location, and the molecular level mechanisms behind the storage and transmission of the genetic information, still needed 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) found that exposing fruit flies to X-rays (a highly energetic form of light) generated mutations that could be passed from generation to generation. This suggested that genetic information was stored in a chemical form and that that information could be altered through interactions with radiation. Once altered, the information was again stable.

The second piece of 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*. These bacteria cause bacterial pneumonia and, when introduced, killed infected mice. Griffith grew these bacteria in the laboratory. This is known as culturing the bacteria. We say that bacteria grown in culture have been grown *in vitro* or in glass (although in modern labs, they are often grown in plastic), 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 salt water algae) containing various nutrients. Typically, a liquid culture of bacteria is diluted and spread on these plates, with individual and isolated bacteria coming to rest on the agar surface. Individual bacteria bind to the plate independently of, and separated from, one another. Bacteria are asexual and so each bacterium can grow up into a colony, a clone of the original bacterium that landed on the plate. The disease-causing strain of *S. pneumoniae* grew up into smooth or S-type colonies, due to the fact that the bacteria secrete a slimy mucus-like substance. Griffith found that mice injected with S strain *S. pneumoniae* 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 rather than some stable chemical toxin.

During extended cultivation *in vitro*, however, cultures of S strain bacteria sometimes gave rise to rough (R) colonies; R colonies were not smooth and shiny but rather rough in appearance. This appeared to be a genetic change since once isolated, R-type strains produced R-type colonies, a process that could be repeated many, many times. More importantly, mice injected with R strain *S. pneumoniae* did not get sick. A confusing complexity emerged however; mice co-injected with the living R strain of *S. pneumoniae* (which did not cause disease) and dead S strain *S. pneumoniae* (which also did not cause the disease) did, in fact, get sick and died! Griffith was able to isolate and culture *S. pneumoniae* from these dying mice and found that, when grown *in vitro*, they produced smooth colonies. He termed these 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 strain²⁰⁴. Unfortunately Fred Griffith died in 1941 during the bombing of London, which put an abrupt 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 strains of *S. pneumoniae* into S strains. Their approach was to grind up cells and isolate their various components, such as proteins, nucleic acids, carbohydrates, and lipids. They then digested these extracts with various enzymes (reaction specific catalysis) and ask whether the transforming principle remained 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, enzymes that 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 (absorption spectra of DNA versus protein). Subsequent studies confirmed this conclusion. Furthermore DNA isolated from R strain bacteria was not able to produce S-strain from R strain bacteria, whereas DNA from S strain bacteria could transform R strains into S 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, presumably, 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. This is a distinctly different process than the movement of genetic information from a parent to an off-spring, which is known as vertical gene transfer. Various forms of horizontal gene transfer occur 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 transfer²⁰⁵. 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 gene transfer is possible primarily because all organisms share the same basic system for encoding, reading and replicating genetic information. The hereditary machinery is homologous among existing organisms.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.4: 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, 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 a hydroxyl group attached to the 2' carbon of the ribose, 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 linkage of nucleotides together to form nucleic acid polymers.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.5: 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 nucleotides 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 of G to C were always equal to 1, no matter where the DNA came from. Knowing these rules, James Watson (b 1928) 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)²⁰⁶. Franklin got these data by pulling DNA molecules into oriented strands, fibers of many molecules aligned parallel to one another. By passing a beam of X-rays through these fibers she was able to obtain a diffraction pattern. This pattern is based on the structure of DNA molecules and 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 allowed Watson and Crick to make conclusions about the structure of a DNA molecule.

To understand this process, let us consider the chemical nature of a nucleotide and a nucleotide polymer like DNA. First the nucleotide bases in DNA (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 number of O-H groups, and a 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 insight, led to a direct explanation for why Chargaff's rules were universal; they recognized that pairs of nucleotide bases, in the 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 involving its 5' phosphate group attached to the 3' hydroxyl from of the existing strand. 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 interact with one another in a highly specific way. An A can form two hydrogen bonding interactions with a T on the opposite (anti-parallel) strand, while an G can form three hydrogen bonding interactions with a C. A key feature of this arrangement is that the lengths of the A::T and G::C base pairs are almost identical. The hydrophobic surfaces of the bases are stacked on top of each other, while the hydrophilic sugar and phosphate groups are in contact with the surrounding aqueous solution. The possible repulsion between negatively charged phosphate groups is neutralized (or shielded) by the presence of positively charged 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. This is the usual form of DNA in a cell. However, under different salt conditions, DNA can form two other double helical forms, known as A and Z. While the A and B forms of DNA are "right-handed" helices, the Z-form of DNA is a left-handed helix. We will not concern ourselves with these other forms of DNA, leaving that more advanced courses.

As soon as the Watson-Crick model of DNA structure 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, by convention 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 can be found, at least theoretically, in a DNA molecule. If information were encoded in the sequence of nucleotides along a DNA strand, 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 memory drive is not altered by whether it contains photos of your friends, a song, a video, or a textbook. At the same time, the double-stranded nature of the DNA molecule's structure and complementary nature of base pairing (A to T and G to C) 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 two original strands as templates. This model of DNA replication is facilitated by the fact that the two strands of the parental DNA molecule are held together by weak hydrogen bonding interactions, so no chemical reaction is required to separate them, no covalent bonds need to be broken. In fact, at

physiological temperatures DNA molecules often open up over short stretches and then close, a process known as DNA breathing²⁰⁷. 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 any information within the molecule is, in fact, 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).

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.6: DNA, sequences, and information

We can now assume that somehow the sequence of nucleotides in a DNA molecule encodes information but exactly what kind(s) of information are 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 how a single stranded DNA molecule does. Since the two strands of double stranded DNA molecules (often written dsDNA) are attached only by hydrogen bonds, increasing the temperature of the system will lead to their separation into two single stranded molecules (ssDNA). 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. What we find is that the temperature at which 50% of dsDNA molecules have separated into ssDNAs 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 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.

It quickly became clear that things are 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 can look at how fast ssDNA (the result of a melting experiment) reforms dsDNA. The speed of these "reannealing reactions" depends 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. 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 that contained similar nucleotide sequences. Because the single strands of these fragments can associate 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 is composed of various families of repeated sequences and that unique sequences amount 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), a 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 this imply about the presence of single and double stranded DNA in an organism?
- 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, 1mM MgCl₂ and some minor ions). Predict what will happen (what is thermodynamically favorable) if you place DNA into distilled water (that is, in the absence of dissolved salts.)

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.7: 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 significant difference when the RNA strand interacts with DNA, since the number of hydrogen bonding interactions are the same. When RNA is 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 that mediates the synthesis of polypeptides, is a complex of structural and catalytic RNAs (known as ribosomal or rRNAs) and proteins. Transfer RNAs (tRNAs) are a integral component of the protein synthesis system. RNAs, in combination with proteins, also play a number of regulatory functions including recognizing and regulating the synthesis and subsequent 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 at least a subset of 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 throughout their evolution. The catalytic activity of the ubiquitous ribosome, which is involved in protein synthesis in all known organisms, is based on a ribozyme, a RNA-based catalyst.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

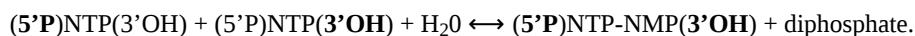
7.8: DNA Replication

Once it was proposed, the double-helical structure of DNA immediately suggested a simple mechanism for the accurate duplication of the information stored in DNA. Each strand contains all of the information necessary to specify the sequence of the complementary strand. The process begins when a dsDNA molecule opens to produce two single-stranded regions. Where DNA is naked, that is, not associated with other molecules (proteins), the opening of the two strands can occur easily. Normally, the single strands simply reassociate with 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 (conserved) features of DNA synthesis. The enzymes (DNA-dependent, DNA polymerases) that catalyze the synthesis of new DNA strands cannot start synthesis on their own, they have to add nucleotides to an existing nucleic acid polymer. In contrast, the catalysts that synthesize RNA (DNA-dependent, RNA polymerases) do not require a pre-existing nucleic acid strand, they can start the synthesis of new RNA strand, based on complementary DNA sequence, *de novo*. Both DNA and RNA polymerases link the 5' end of a nucleotide triphosphate molecule to the pre-existing 3' end of a nucleic acid molecule. This polymerization reaction is said to proceed in the 5' to 3' direction. As we will see later on, the molecules involved in DNA replication and RNA synthesis rely on signals within the DNA that are recognized by proteins and which determine where synthesis starts and stops, and when nucleic acid replication occurs, but for now let us assume that some process has determined where replication starts. We begin our discussion with DNA replication.

The first step in DNA replication is to locally open up the dsDNA molecule. A specialized RNA-dependent, DNA polymerase, known as primase, collides with, binds to, and synthesizes a short RNA molecule, known as a primer. Because the two strands of the DNA molecule point in opposite directions (they are anti-parallel), one primase complex associates with each DNA strand, and two primers are generated, one on each strand. Once these RNA primers are in place, the DNA-dependent, DNA polymerases replaces the primase and begins to catalyze the nucleotide-addition reaction; which nucleotide is added is determined by which nucleotide is present in the existing DNA strand. The nucleotide addition reaction involves various nucleotides colliding with the DNA-primer-polymerase complex; only the appropriate nucleotide, complementary to the nucleotide residue in the existing DNA strand is bound and used in the reaction.

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, generated through the reaction:



During the reaction the terminal diphosphate of the incoming NTP is released (a thermodynamically favorable reaction) and the nucleotide mono-phosphate is added to the existing polymer through the formation of a phosphodiester [-C-O-P-O-C] bond. This reaction creates a new 3' OH end for the polymer that can, in turn, react with another NTP. 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 further, exposing more single-stranded DNA. Keep in mind that this process is moving, through independent complexes, 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 iii) that eventually an extending DNA polymerase will run into the RNA primer part of an "upstream" molecule. However, there is a complexity: RNA regions are not found in "mature" DNA molecules, so there must be mechanisms that removes them. This is due to the fact that the DNA polymerase complex contains more than one catalytic activity. When the DNA polymerase complex reaches the upstream nucleic acid chain it runs into this RNA containing region; an RNA exonuclease activity removes the RNA nucleotides and 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. For a dynamic look at the process check out this video²⁰⁸ which is nice, but "flat" (to reduce the complexity of the process) and fails to start at the beginning of the process.

Evolutionary considerations: At this point you might well ask yourself, why (for heavens sake) is the process of DNA replication so complex. 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 addition of a nucleotide to the 5' end of an RNA molecule (the 5' capping reaction associated with mRNA synthesis) that we will briefly consider later on. But, such activities are simply 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 of all organisms and extremely difficult or impossible to change through evolutionary mechanisms, or simply worth the effort (in terms of its effects on reproductive success). 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 appears to function in all known organisms, so for practical purposes, we have to remember some of the key details involved, these include the direction of polymer synthesis and the need (in the case of DNA) of an RNA primer.

References

208. DNA replication video: <http://bcove.me/x3ukmq4x>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.9: 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 the actual machinery involved is complex. In part this 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 the processes of 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. Various proteins then bind to the DNA to locally denature (unwind and separate) and block the single strands from reannealing. This leads to the formation of a replication bubble. Multiprotein complexes, known as a replication forks, then assembles on the two DNA strands. 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 the form of 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. While a discussion of the exact mechanisms involved is beyond our scope here, it is also critical that DNA is complete before a cell attempts to divide.

DNA 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 as mistakes. 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 mistake in base pairing as a structural defect in the DNA molecule. When a mismatched base pair is formed and recognized, the DNA polymerase stops forward synthesis, reverses its direction, and removes the region of the DNA containing the mismatched base pair using a “DNA exonuclease” activity. It then resynthesizes the region, (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 in fact obeys reasonably straightforward 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 off a nucleotide located at the 5' end of the molecule, a 3' exonuclease, cuts off a nucleotide located at the molecule's 3' end. An intact 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 without interruption. But the interactions between nucleic acid chains are based on weak H-bonding interactions, and the enzymes involved in the DNA replication 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 molecule 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: <http://youtu.be/QMhi9dxWaM8>). The DNA polymerase complex is held onto the DNA by a doughnut shaped protein, known as a sliding clamp, that 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 loader²⁰⁹. 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, that is acting as an unloader.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.10: Further replication complexities

There are important differences between DNA replication in prokaryotes and eukaryotes. The DNA molecules found in eukaryotic nuclei are linear molecules, with free ends, known as telomeres. This leads to problems replicating the ends of the DNA molecules, a problem solved by the enzyme complex telomerase (discussed briefly below)²¹⁰. In contrast, the DNA molecules found in bacteria and archaea are circular; there are no free ends²¹¹. This creates a topological complexity. After replication, the two double-stranded DNA 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 what is known as the supercoiling of the DNA molecule. Left unresolved, supercoiling and knotting would inhibit DNA synthesis and the separation of replicated strands (perhaps you can explain why)²¹². These topological issues are resolved by enzymes known as topoisomerases. There are two types of topoisomerases. 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. 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 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 chromosome ends. This is solved by a catalytic system composed of proteins and RNA known as telomerase²¹³. The eukaryotic DNA replication enzyme complex is slower (about 1/20th as fast) as prokaryotic systems. While a bacterial cell can replicate its circular $\sim 3 \times 10^6$ base pair chromosome in ~ 1500 seconds using a single origin of replication, the replication of the billions of base pairs of eukaryotic DNAs involves the use of multiple (many) origins of replication, scattered along the length of each chromosome.

So what happens when replication forks collide with one another? In the case of a circular DNA molecule, with its single origin 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, with their multiple linear chromosomes, is much more complex and involves a more complex molecular machines that we will return to, although only superficially, later.

Questions to answer & to ponder:

- During DNA/RNA synthesis what is the average ratio of productive to unproductive interactions between nucleotides and the polymerase?
- Where would genetic variation come from if DNA were totally stable and DNA replication was error-free?
- Draw a diagram to explain how the DNA polymerase might recognize a mismatched base pair.
- Why do you need to denature (melt) the DNA double-helix to copy it?
- How would DNA replication change if H-bonds were as strong as 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 DNA replication video: <http://bcove.me/x3ukmq4x>
- 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 and the organism?
- Explain how the absence of the clamp would influence DNA replication?

References

210. more on telomerase: blogs.scientificamerican.com/...as-too-little/
211. The mitochondria and chloroplasts of eukaryotic cells also contain circular DNA molecules, another homology with their ancestral bacterial parents. ,
212. see this video on DNA supercoiling and topoisomerases: <http://youtu.be/EYGrElVyHnU>
213. <http://en.Wikipedia.org/wiki/Telomerase>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.11: Mutations, Deletions, Duplications & Repair

While DNA is used as the universal genetic material of organisms, 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 ~13°C, half of the phosphodiester bonds in a DNA sample will break after ~520 years²¹⁴. 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 will be replaced by an AU base pair. But, uracil is not normally found in DNA and its presence will be recognized by an enzyme that severs the bond between the uracil moiety and the deoxyribose group²¹⁵. 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 base²¹⁶. A DNA-dependent DNA polymerase binds to the open DNA and uses the existing strand as a primer and 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 DNA²¹⁷. Other hydrolysis reactions (depurination: the loss of a cytosine or thymine group and depyrimidination: the loss of an adenine or guanine group) lead to the removal of a base from the DNA. The rates of these reactions increases at acidic pH, which is probably one reason that the cytoplasm is not acidic. How frequent are such events? A human body contains ~10¹⁴ cells. Each cell contains about ~10⁹ base pairs of DNA. Each cell (whether it is dividing or not) undergoes ~10,000 base loss events per day or ~10¹⁸ 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 ~65,000,000 years ago) has disappeared from the earth, making it impossible to clone (or resurrect) a true dinosaur²¹⁸. In addition 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 pathogens²¹⁹.

References

214. Here is the paper from which statement is derived: <http://www.nature.com/news/dna-has-a...f-life-1.11555>
215. uracil-DNA-N-glycosidase
216. absent purine/absent pyrimidine endonuclease <http://omim.org/entry/300773>
217. Human DNA Repair Genes: <http://www.sciencemag.org/content/291/5507/1284.full>. video with lots of misspelled words:<http://youtu.be/g4khROaOO6c>
218. DNA has a 521-year half-life: <http://www.nature.com/news/dna-has-a...f-life-1.11555>
219. Dietary carcinogens, environmental pollution, and cancer: some misconception: <http://link.springer.com/article/10.1007/BF02988534>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.12: Genes and alleles

Up to now we have been considering genes as abstract entities and mentioning, only in passing, what they actually are. We can think about genes encoding traits, but this is perhaps the most incorrect possible view of what genes are and what they do. A gene is a region of DNA that encode a gene product, either an mRNA that itself encodes a polypeptide or a “non-coding” RNA that functions as an RNA. The gene also includes the sequences required for its proper expression, that is, when and where the gene is active, when RNAs are made from it. While we will not consider in any significant detail, it is worth noting that genes can be 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 through the process of RNA splicing²²⁰. How differences in gene sequence influence the activity and role(s) 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 a cell or 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 different alleles of a gene correspond to genes with different sequences. Two alleles of the same gene can differ from one another at as little as a single nucleotide position or at many positions. 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, particularly since different regulatory regions may be used in the different cell types present within a multicellular organism. A gene's regulatory regions can span many thousands of kilobases of DNA and be located upstream, downstream, or within the gene's 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 on, 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 that are the products of multiple genes 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.13: Mutations and evolution

There are often multiple alleles of a particular gene in the population and they all may be equally normal, that is have similar effects on reproductive success and in terms of the phenotypes they produce. If there is no significant selective advantage between them, their relative frequencies within a population will drift. At the same time, the phenotype associated with a particular allele can be influenced by which alleles are present at other genetic loci, known as the genetic background. Since most traits are the results of hundreds or thousands of genes functioning together, 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 any one locus may make only a minor contribution to the disease phenotype. On top of that, environmental and developmental differences can outweigh genetic influence on phenotype. Such genetic background effects can lead to a particular allele producing a disease in one person and not another²²¹.

Mutations are the ultimate source of genetic variation – without them evolution would not occur. Mutations can lead to a number of effects, in particular, they can create new activities. At the same time these changes may reduce the original (and necessary) 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 (overly) negatively influence the gene’s primary and essential function. The evolution of ancillary functions may be severely constrained or blocked altogether.

This problem can be circumvented based on the fact that the genome is not static. There are molecular level 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, which are mutations, occur continuously during embryonic development. The end result is that while most of the cells in your body have very similar genomes (perhaps consisting of single base pair changes that arose during DNA replication), some have genomes with different arrangements of DNA. These differences can include deletions, duplications, and translocations (moving a region of DNA from one place to another in the genome). Not all cells in your body will have exactly the same genome²²².

In the case above illustrated in the figure, imagine that the essential but multifunctional gene is duplicated. Now one copy can continue to carry out its essential function, while the second is free to change. While many mutations will negatively effect 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 its original (and essential) function. We see evidence of this type of process throughout the biological world. When a gene is duplicated, the two copies are known as paralogs. Such paralogs often evolve independently.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

7.14: Triplet repeat diseases and genetic anticipation

While they are essential for evolution, defects in DNA synthesis and genomic rearrangements more often lead to genetic (that is inherited) diseases than to any benefit to an individual. You can explore the known genetic diseases by using the web based On-line Mendelian Inheritance in Man (OMIM) database²²³. 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 dystrophy. 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, they can lead to regions of a repeated amino acids. For example, expansion of a domain of CAGs in the gene encoding the polypeptide Huntingtin causes the neurological disorder Huntington's chorea.

Fragile X: This DNA replication defect is the leading form of autism of known cause (most forms of autism have no known cause); ~6% of 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)²²⁴. In the unaffected population, the FMR-1 gene contains between 6 to 50 copies of a CGG repeat. Individuals with between 6 to 50 repeats are phenotypically normal. Those with 50 to 200 repeats carry what is known as a pre-mutation; 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 activity²²⁵.

Other DNA Defects: 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 prematurely²²⁶.

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 role within the organism, typically by encoding a gene product (which we will consider next) and/or the information needed to insure its correct "expression", that is, where and when the information in the gene is used. 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 amorphous mutation in a gene in the lab, whereas organisms that carry that allele may well be at a significant reproductive disadvantage in the real world. 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 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.

Contributors and Attributions

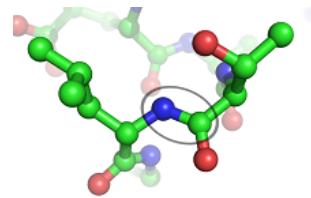
- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

8: PEPTIDE BONDS, POLYPEPTIDES & PROTEINS

Topic hierarchy

- 8.0: INTRODUCTION
- 8.1: POLYPEPTIDE AND PROTEIN STRUCTURE BASICS
- 8.2: AMINO ACID POLYMERS
- 8.3: MAKING A POLYPEPTIDE IN A BACTERIAL CELL
- 8.4: THE ORIGIN OF THE GENETIC CODE
- 8.5: PROTEIN SYNTHESIS: TRANSCRIPTION (DNA TO RNA)
- 8.6: PROTEIN SYNTHESIS: TRANSLATION (RNA TO POLYPEPTIDE)
- 8.7: THE TRANSLATION (POLYPEPTIDE SYNTHESIS) CYCLE
- 8.8: BURSTING SYNTHESIS AND ALARM GENERATION
- 8.9: GETTING MORE COMPLEX: GENE REGULATION IN EUKARYOTES
- 8.10: TURNING POLYPEPTIDES INTO PROTEINS
- 8.11: REGULATING PROTEIN LOCALIZATION
- 8.12: REGULATING PROTEIN ACTIVITY
- 8.13: ALLOSTERIC REGULATION
- 8.14: POST-TRANSLATIONAL REGULATION
- 8.15: DISEASES OF FOLDING AND MISFOLDING
- 8.16: WHY DO HARMFUL ALLELES PERSIST?



8.0: Introduction

In which we consider the nature of proteins, how they are synthesized, how they are assembled,, how they get to where they need to go within the cell and the organism, 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)²²⁷. After an analysis of a number of different substances, he proposed that all proteins 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 different proteins 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 systems²²⁸. As you can see, Mulder’s molecular formula was 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 proteins 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 reaction 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 (-NH₂) 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 polymer 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.1: 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 of subunits, and that each subunit is an unbranched polymer with a specific amino acid 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, and they can be composed of a number of distinct polypeptides each encoded by distinct gene. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.2: Amino acid polymers

As you might remember from chemistry carbon atoms (C) typically form four bonds. 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 (-NH₂), 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 forms, known as enantiomeric stereoisomers. Enantiomers are mirror images of one another and are referred to as the L- and D- forms. Only L-type amino acids are found in proteins, even though there is no obvious chemical reason that proteins could not have also been made using both types of amino acids or using only D-amino acids²²⁹. It appears that the universal use of L-type amino acids in the polypeptides found in biological systems is 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 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 sequentially (and 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; the formation of a peptide bond is associated with the release of a water molecule. When complete, the 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 the very short (5-10) to very long (many hundreds to thousands) amino acids in length. For example, the protein Titin²³⁰ (found in muscle cells) can be more than 30,000 amino acids in length. Because there is no theoretical constraint on which amino acids occurs at a particular position within a polypeptide, there is an enormous universe of possible polypeptides that can exist. In the case of a 100 amino acid long polypeptide, there are 20^{100} possible different polypeptides that could, in theory, be formed.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.3: Making a polypeptide in a bacterial cell

A bacterial cell synthesizes thousands of different polypeptides. The sequence of these polypeptides (the exact amino acids from N- to C-terminal) is 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 the information present in nucleic acid-sequence translated into polypeptide sequence.

To address the first question let us think 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 molecule²³¹. 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 (or 22) different amino acids. As pointed out by the physicist George Gamow (1904–1968)²³² 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 amino acids, a sequence two nucleotides in length could encode (4^2) or 16 different amino acids (not enough), while a sequence of three nucleotide (4^3) could encode 64 different amino acids (more than enough)²³³. Although the actual coding scheme that Gamow proposed was wrong, his thinking about the coding capacity of DNA influenced those who set out to experimentally determine 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. The 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 for by more than one codon. In addition there are three codons, UAA, UAG and UGA, that do not encode any amino acid but are used to mark the end of a polypeptide, they encode “stops” or periods.

The region of the nucleic acid that encodes a polypeptide begins with what is known as the “start” codon and continues until one of the three stop codons is reached. A sequence defined by in-frame start and stop codons (with some number of codons between them) is known as an open reading frame or an ORF. At this point it is important to point out explicitly, while the information encoding a polypeptide is present in the DNA, this information is not used directly to specify the polypeptide sequence. Rather, the process is indirect. The information in the DNA is first copied into an RNA molecule (known as a messenger RNA) and it is this RNA molecule that directs polypeptide synthesis. The process of using information within DNA to direct the synthesis of an RNA molecule is known as **transcription** because both DNA and RNA use the same language, nucleotide sequences. In contrast polypeptides are written in a different language, amino acid sequences. For this reason the process of RNA-directed polypeptide synthesis is known as **translation**.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.4: The origin of the genetic code

There are a number of hypotheses as to how the genetic code originated. 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, each using a different genetic code. The common genetic code found in all existing organisms 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. It is not clear which model reflects what actually happened. 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 code²³⁴. Since the genetic code does not appear to be predetermined, the 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.5: Protein synthesis: transcription (DNA to RNA)

Having introduced you to the genetic code and mRNA, 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; we are looking for a (relatively) short region of DNA within millions (in prokaryotes) or billions (in eukaryotes) of base pairs of DNA. 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 a region(s) of a larger DNA molecule. Part of the gene's sequence is known as its regulatory region; this region of DNA 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". So what is expressed? This is the part of the gene's sequence that is used to direct the synthesis of an RNA molecule, known as the transcribed region or a transcript. Within the transcribed region is the region of the RNA that actually encodes the polypeptide, through the process of translation - this is known as the coding region. The regions of the RNA that are not translated are known as untranslated regions (UTRs). Typically the coding region of an RNA molecule is located between a 5' UTR and 3' UTR.

Once a gene's regulatory region is identified (by the binding of specific type of protein - see below), a DNA-dependent, RNA polymerase binds to the protein-DNA complex and the synthesis of an mRNA molecule begins. As a general simplification, we will say that a gene is expressed when the RNA that its transcribed region encodes is synthesized (note: while regulatory regions are generally not transcribed, they are still part of the gene). We can postpone further complexities to later on (and to subsequent classes). It is important to recognize that an organism as "simple" as a bacterium contains thousands of genes and that different sets of genes are used in different environments and situations, and in different combinations 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 specific genes involved in rapid growth and division in order to prepare itself (through the expression of other genes that turn on) to survive in a more hostile environment. Our goal is not to have you accurate predictions about the behavior of an organism in a particular situation, 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 consider, although at a rather elementary level, a few of the regulatory processes 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. 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. Often groups of genes share regulatory sequences recognized by specific transcription factors. As we will see this makes it possible to regulate groups of particular 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 transcription factors. Which of these genes are expressed determines which transcription factor proteins are present and which genes are actively expressed. Of course, the gene encoding a specific transcription factor is itself regulated. Transcription factors can act positively or negatively, which means that they can lead to the activation of transcription or its inhibition. In addition the activity of a particular transcription factors can be regulated (a topic we will return to later on in this chapter).

For a transcription factor to regulate a specific gene, either positively or negatively, it must be able to bind to specific sites on the DNA. Whether or not a gene is expressed (whether it is "on" or "off") depends upon which transcription factors are expressed, are active, and can interact productively with the DNA-dependent, RNA polymerase (RNA polymerase). Inactivation of a transcription factor can involve a number of mechanisms, including its destruction, modification, or interactions with other proteins, so that it can no longer interacts productively with either its target DNA sequence or the RNA polymerase. Once a transcription factor is active, it can diffuse through out the cell and (in prokaryotic cells that do not have

barrier control interactions with DNA) can bind to its target DNA sequences. Now an RNA polymerase can bind to the DNA-transcription factor complex, an interactions that leads to the activation of the RNA polymerase and the initiation of RNA synthesis, using one DNA strand to direct RNA synthesis. Once RNA polymerase has been activated, it will move away from the transcription factor-DNA complex. The DNA bound transcription factor can then bind another polymerase or the transcription factor can release from the DNA (in response to molecular level collisions), which will diffuse away, interact with other regulatory factors, or rebind to other sites in the DNA. Clearly the number of copies of the transcription factor and its interaction partners and DNA binding sites will impact the behavior of the system.

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 within the DNA lead to the termination of transcription and the release of the RNA polymerase. Once released, the RNA polymerase returns to its inactive state. It can act on another gene if the RNA polymerase interacts with a transcription factor bound to its promoter. Since multiple types transcription 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 and activities of specific transcription factors and their regulatory proteins, together with the binding affinities of particular transcription factors for specific DNA sequences (compared to their general low-affinity binding to DNA in general).

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.6: 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 polypeptide²³⁵. 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?

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 genes that encode tRNAs (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, for which there are no tRNAs. 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 (in the mRNA-ribosome complex) 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: there is a phenylalanine-tRNA synthetase and a proline-tRNA synthetase, etc. An amino acyl tRNA synthetase binds the appropriate tRNA and the appropriate amino acid and, through a reaction coupled to a thermodynamically favorable nucleotide triphosphate hydrolysis 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 long²³⁶. 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 $\sim 3 \times 10^6$ 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 rRNA lies at the heart of the ribosome and catalyzes the transfer of an amino acid bound to a tRNA to the carboxylic acid end of the growing polypeptide chain. RNA based catalysis is a conserved feature of polypeptide synthesis and appears to represent an evolutionarily homologous trait.

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.

The cytoplasm of cells is packed with ribosomes. In a rapidly growing bacterial cell, $\sim 25\%$ 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 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.7: 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 (we think) are unnecessary details that obscure a basic understanding. Here we will remind you of two recurring themes. The first is the need 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 occur before a productive (correct) one, since there are more than 20 different amino acid/tRNA molecules bouncing around in the cytoplasm. The stochastic aspects of the peptide synthesis process are rarely illustrated.

The first step in polypeptide synthesis is the synthesis of the specific mRNA that encodes the polypeptide. (1) The mRNA contains a sequence²³⁷ 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)²³⁸. This interaction defines the beginning of the polypeptide as well as the coding region's reading frame. (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 sequentially. With each new amino acid added, the ribosome moves along the mRNA (or the mRNA moves through the ribosome). 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 is threaded through a molecular tunnel within 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 UAG²³⁹. Since there are no tRNAs for these codon, the ribosome pauses, waiting for a charged tRNA that 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 that could degrade it (ribonucleases), that is, break it down into nucleotides. Upon its release the mRNA may interact with a new small ribosomelsubunit, 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 RNA stability is regulated by the bonding of specific proteins to nucleotide sequences within the mRNA. The relationship between mRNA synthesis and degradation will determine the half-life of a population of mRNA molecules within the cell, the steady state concentration of the mRNA in the cell, and indirectly, the level of polypeptide present.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.8: 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.; the assembly of this initial complex involves the a number of components, and takes time to occur, but once formed can persist. While this complex exists (that is, before it dissociates) multiple ribosomes can interact with the mRNA, each synthesizing a polypeptide. This leads to a behavior known as translational bursting, in which multiple polypeptides are synthesized in a short period of time from a single RNA. Once the translation initiation complex dissociates, it takes time (more time than just colliding with another small ribosome subunit) before it forms again. This leads to bursts of new polypeptide synthesis followed by periods when no new polypeptides are made. A similar process, transcriptional bursting, is observed with the synthesis of mRNAs. Since the number of mRNA molecules encoding a particularly polypeptide can be small (less than 10 per cell in some cases), the combination of transcriptional and translational bursting can lead to noisy protein synthesis.

The translation system is dynamic and a major consumer of energy within the cell²⁴⁰. When a cell, particularly a bacterial cell, is starving, it does not have the energy to generate amino acid charged tRNAs. The result is that uncharged tRNAs accumulate. Since uncharged tRNAs fit into the 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 (see²⁴¹) that can lead to adaptive changes in the cell that enable it to survive for long periods in a “dormant” state²⁴².

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, which binds to and renders the toxin molecule inactive. The key feature of the toxin-anti-toxin system is that the toxin molecule is stable, it has a long-half life. The half-life of a molecule is the time it takes for 50% of the molecules present within a population at a particular time to be degraded (or to 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 the death of the cell. Death leads to the release of the cell's nutrients, nutrients that can be used by its neighbors. A similar process can occur if a virus infects a cell. If an infected cell kills itself before the virus can replicate, the virus is destroyed and the cell's neighbors (who are likely to be its relatives) survive.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.9: Getting more complex: gene regulation in eukaryotes

At this point, we will not take very 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 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 punctuated by nuclear pores, macromolecular complexes (protein machines). While molecules of molecular weight less than ~40,000 daltons can generally pass through the nuclear pore, larger molecules must be actively transported, that is, in a process that is coupled to a thermodynamically favorable reaction, in this case the hydrolysis of guanosine triphosphate (GTP) rather than 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, located withinin polypeptides. These are recognized by proteins (receptors) associated with the pore complex. A protein with an active nuclear localization sequence (NLS) will be found in the nucleus while a protein with an active nuclear exclusion sequence (NES) will be in the cytoplasm. By controlling NLS and NES activity a protein can come to accumulate in a regulated manner in either the nucleus or the cytoplasm. As we will see later on the nuclear envelope breakdowns occurs during cell division (mitosis) in many but not all eukaryotes. Tears in the nuclear envelope have also been found to occur when migrating cells trying to squeeze through small openings²⁴³. Once the integrity of the nuclear envelop is re-established, proteins with NLS and NES sequences move back to their appropriate location within the cell.

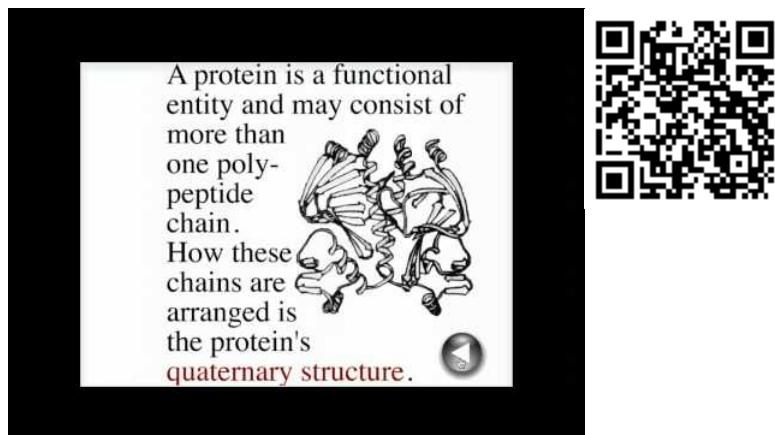
Aside from those within mitochondria and chloroplasts, the DNA molecules of eukaryotic cells are located within the nucleus. One difference between eukaryotic and bacterial genes is that the transcribed region of eukaryotic genes often contains what are known as intervening sequences or introns; introns involve sequences that do not encode a polypeptide. After an RNA is synthesized introns are removed enzymatically, resulting in a shorter mRNA. As a point of interest, which sequences are removed can be regulated, this can produce multiple different mRNAs from the same gene, mRNAs that encode somewhat (and often functionally) different polypeptides. In addition to removing introns, the mRNA is further modified (processed) at both its 5' and 3' ends. Only after RNA processing has occurred is the now "mature" mRNA exported out of the nucleus, through a nuclear pore, into the cytoplas where it can interact with ribosomes. One further difference from bacteria is that the interaction between a mature mRNA and 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) transcription and translation are distinct processes in eukaryotes. This makes the generation of multiple, functionally distinct RNAs (through mRNA processing) from a single gene possible and leads to significantly greater complexity from only a relatively small increase in the number of genes.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.10: 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 with other polypeptides. In a protein composed of multiple polypeptides, we need to consider how it comes to interact with those other polypeptides (often termed subunits). As we think about polypeptide structure it is common to see the terms primary, secondary, tertiary, and quaternary structure. The primary structure of a polypeptide is the sequence of amino acids along the 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 combine and are arranged to form a functional protein. In a protein that consists of a single polypeptide and no co-factors, 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 way in which a gene can play a role in a number of different processes and be involved in the generation of 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 productively bind only the amino-acyl-tRNA that the mRNA specifies, that is the tRNA with the right anticodon. This enables its attached amino acid to interact productively, leading to the addition of the amino acid to C-terminus of 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)²⁴⁴.

Now you might wonder whether there are errors in polypeptide synthesis as there are in nucleic acid synthesis. In fact there are. For example, if a base is skipped by the ribosomal system, the reading frame will be thrown off. Typically, this leads to a completely different sequence of amino acids added to the end of the polypeptide (down-stream of the skip) and generally

quickly leads to a stop codon, which terminates translation, leading to the release of a polypeptide that cannot fold correctly and is (generally) rapidly degraded²⁴⁵. Similarly, if the wrong amino acid is inserted at a particular position and it disrupts normal folding, the polypeptide could be 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 a mis-translated 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. Synthesis occurs in an N- to C-terminal direction and the newly synthesized polypeptide exits the ribosome through a 10 nm long and 1.5 nm in diameter tunnel. This tunnel is narrow enough to block the folding of the newly synthesized polypeptide chain. As the polypeptide emerges from the tunnel it begins to fold. At the same time it encounters the crowded cytoplasmic environment; the newly synthesized polypeptide needs to avoid low affinity, non-specific, and non-physiologically significant interactions with other cellular components²⁴⁶. If the polypeptide is part of a multi-subunit protein, it must also "find" its correct partner polypeptides, which again is a stochastic process. If the polypeptide does not fold correctly, it will not function correctly and may even damage the cell or the organism. A number of degenerative neurological disorders are due, at least in part, to the accumulation of misfolded polypeptides (see below).

We can think of the folding process as a "drunken" walk across an energy landscape, with movements driven by intermolecular interactions and collisions with other molecules. The successful goal of this process 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 native 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 challenge to calculating the final folded state of a polypeptide is that it is a extremely complex problem. Generally two approaches are taken to characterizing the structure of a functional protein. 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.

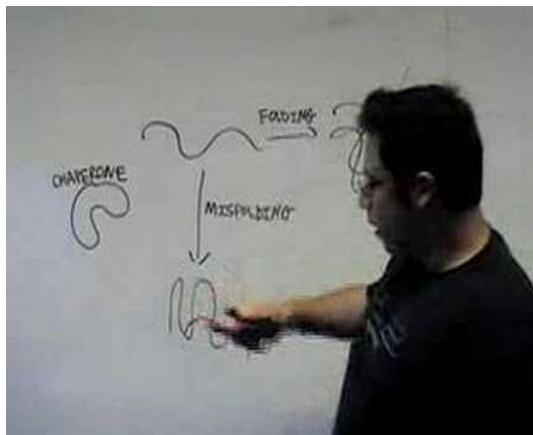
In an α -helix and a β -sheet, all of the possible H-bonds involving the peptide bond's donor and acceptor groups ($-N-H : O=C-$ with ":" indicating a H-bond) are formed within the polypeptide. In the α -helix these H-bond interactions run parallel to the polypeptide chain. In the β -sheet they occur between polypeptide chains. The interacting strands within a β -sheet can run parallel or anti-parallel to one another, and can in occur within a single polypeptide chain or between different polypeptide chains. 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. It is worth noting that some polypeptides can adopt functionally different structures: for example in one form (PrPC) the prion protein contain a high level of α -helix (42%) and essentially no β -sheet (3%), while an alternative form (PrPSc), associated with the disease scrapiecontains high levels of β -sheet (43%) and 30% α -helix (see below)²⁴⁷.

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 molecular collisions. 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 because it involves the partial breakage of the 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 both large they can bump into each other. If they get too close they will repel each other. The result is that usually the polypeptide chain will 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.

Peptide bonds involving 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). In addition, peptides bonds involving prolines are found in the cis configuration \sim 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 peptide bond involving a proline is much higher than that of a standard peptide bond; so high, in fact, that there are protein catalysts, peptidyl proline isomerasers, that facilitate the cis-trans rotation.

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, by leading to the organization of water molecules around the hydrophobic group, a thermodynamically unfavorable situation. This is 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, tend to collapse onto itself so as to minimize (although not always completely eliminate) the interactions of its hydrophobic residues with water. In practice this means that the first step in the folding of a newly synthesized polypeptide is, generally to collapse the polypeptide so that the majority of its hydrophobic R groups are located internally, out of contact with water. In contrast, where there are no (or few) hydrophobic R groups in the polypeptide, the polypeptide will tend to adopt an extended configuration. On the other hand, if a protein comes to be embedded within a membrane (we will consider how this occurs later on), then the hydrophobic R-groups will tend to be located on the surface of the folded polypeptide that interacts with the hydrophobic interior of the lipid bilayer. Hopefully this makes sense to you, thermodynamically.

The path to the native (that is, most stable, functional) 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, structurally, there are active mechanisms to unfold it and let the process leading to the native state proceed again. This process of partial unfolding is carried out by proteins known as chaperones. An important point to recognize; chaperones do not determine the native state of a polypeptide. 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, through coupling to ATP hydrolysis, catalyze the unfolding of the polypeptide, 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 chaperones²⁴⁸.



chaperone video <http://youtu.be/b39698t750c>

One class of chaperones are known as “heat shock proteins.” The genes that encode these proteins are activated (expressed) in response to increased temperature (as long as the increase is not so severe that it kills the cell immediately.) At these higher temperatures, the native protein can unfold and misfold, it can denature. 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. Once expressed, heat shock proteins recognize denatured polypeptides, couple ATP hydrolysis reactions to unfold them, and then release them giving them another chance to refold correctly.

Heat shock proteins 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 re-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.

By now you might be asking yourself, how do chaperones recognize unfolded or abnormally folded proteins? Well unfolded proteins will tend to have hydrophobic amino acid side chains exposed on their surface. Because of that they will also tend to aggregate. Chaperones recognize and interact with surface hydrophobic regions.

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 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 the internal regions of vesicles associated with endocytosis to become acidic (through the ATP-dependent pumping of H⁺ across their membrane), which in turn activates a number of enzymes (located within the vesicle) involved in the hydrolysis of proteins and nucleic acids.

Subunits and prosthetic groups: Now you might find yourself asking yourself, 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 a newly synthesized polypeptide and either stabilizes its folding, or hold it until it interacts with the other polypeptides 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 vitaminB12. 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. At the same time, the formation of protein can be dependent upon chaperones, which are themselves projects of other genes.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.11: Regulating protein localization

The synthesis 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 located 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 complex will find (by diffusion), and attach to, a ribosome/SRP receptor complex on the cytoplasmic surface of the plasma membrane (in bacteria and archaea) or a cytoplasmic facing membrane (in eukaryotes). 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 the protein pore and so enters into or passes 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 fold and remain within the membrane. If it is a secreted polypeptide, it will be released into the periplasmic space, that is the region topologically outside of the cytoplasm (either within a vesicle or other side of the plasma membrane). Other mechanisms can lead to the release of the protein from the cell.

Because 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 and Golgi apparatus) 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 receptors 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.12: 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 (which in turn influences 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, and the total protein activity in a system [P_{sys}] is proportional to the rate of that protein's synthesis ($d\text{Synthesis}/dt$) minus the rate of that protein's degradation ($d\text{Degradation}/dt$), with dt indicating synthesis or degradation per unit time. The combination of these two processes, synthesis and degradation, determines the protein's half-life. Since both a protein's synthesis and degradation can be regulated, its half-life can be regulated.

The degradation of proteins is mediated by a special class of enzymes (proteins) known as proteases. Proteases cleave peptide bonds via hydrolysis (adding water) 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 can also act more specifically, recognizing and removing specific parts of a protein in order to activate or 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 NLS or NES sequences, thereby altering the protein's distribution within the cell and therefore its activity. As an example, a transcription factor located in the cytoplasm is inactive, but it becomes active when it enters the nucleus. Similarly, many proteins are originally synthesized in a longer and inactive "pro-form". When the pro-peptide is removed, cut away by an endoprotease, the processed protein becomes active. 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 37°C 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 termed "constitutive" or house-keeping proteins. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.13: Allosteric regulation

A reversible form of regulation is known as allosteric regulation, where a regulatory molecule binds reversibly to the protein altering its conformation, which in turn alters the protein's structure, its location within the cell, its activity, and its half-life. Such allosteric effectors are not covalently attached to the protein and their interactions are reversible, influence by thermal factors and concentration. Allosteric regulators 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 over-come 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.14: Post-translational regulation

Proteins may be modified after their synthesis, folding, and assembly - 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 that modification can be reversed. The exception, of course, is when the modification involves protein degradation or proteolytic processing. There are many different types of post-translational modification, and we will consider them only generically. In general they involve the formation of a covalent bond linking a specific chemical group to specific amino acid side chains on the protein - these groups can range from a 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 phosphatases remove such 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.15: 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 many 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 proteasomes 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)²⁴⁹ 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 (mentioned above). We can think of prions as a type of anti-chaperone. The idea of proteins as infectious agents was championed by Stan Prusiner (b. 1942), who was awarded the Nobel Prize in Medicine in 1997²⁵⁰.

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 contains high levels of β -sheet. The two polypeptides have the same primary sequence. PrPsc acts to catalyze the transformation of PrPc into PrPsc. Once initiated, this reaction 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 neuronal cell death and dysfunction, and 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 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 very 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 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 \rightarrow PrPsc conversion. Such mutations may be inherited (genetic) or may occur during the life of an organism (sporadic). Fatal familial insomnia (FFI)²⁵¹ is due to the inheritance of a mutation in the PRNP gene, which encodes PrPc. This mutation replaces the normal aspartic acid at position 178 of the PrPc protein with an asparagine. When combined with a second mutation in the PRNP gene at position 129, the FFI mutation leads to Creutzfeld-Jacob disease (CJD)²⁵². 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 (an enzyme that cuts peptide bonds) 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

8.16: 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 together with the rate at which individuals carrying them are eliminated (before they have off-spring). 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 homozygous individuals (and the alleles they carry) then become subject to strong negative selection. We 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 emerged from studies on the 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) ~20,000 generations later²⁵³. We will return to how this works exactly toward the end of the next chapter, but what is important here is that it is the organism (and its traits and all its alleles) that is "selected". Only in cases of 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?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

9: GENOMES, GENES & REGULATORY NETWORKS

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. We will touch on such 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 biological 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 take some practice to master this type of thinking. The key is to keep calm and analyze on!

```
CATGACGTCGGACAAACCC  
CTGGGGCTTTACTGATGTCA  
ATTTTCTGAAAGTTACAGAC  
TTTTTCGACGTGTCAGGAC  
CGATAAAATTCAACTACTGG  
CCCTGGGTGTTCTATGATAA  
ATCCAAGCGCCCCGCTAATT  
AGCCCAGTCGAAGGGTCTG  
GGTTAAGGCGTGTGATCGAC
```

Topic hierarchy

9.0: INTRODUCTION

9.1: GENOMES AND THEIR ORGANIZATION

9.10: PARALOGOUS GENES AND GENE FAMILIES

9.11: PACKING DNA INTO A CELL

9.12: LOCATING INFORMATION WITHIN DNA

9.13: NETWORK INTERACTIONS

9.14: E. COLI AS A MODEL SYSTEM

9.15: FINAL THOUGHTS ON (MOLECULAR) NOISE

9.16: TYPES OF REGULATORY INTERACTIONS

9.2: A TECHNICAL ASIDE - TRANSPOSONS

A transposon is a piece of DNA that can move (jump) from place to place in the genome. Type II transposons consist of a 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 - it jumps from place to place in the genome. Type I transposons make copies of themselves, through an RNA intermediate to generate a complementary DNA that can then be inserted into the host genome.

9.3: GENES ALONG CHROMOSOMES

9.4: NATURALLY OCCURRING HORIZONTAL GENE TRANSFER MECHANISMS

9.5: TRANSFORMATION

Transformation 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.

9.6: CONJUGATION

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.

9.7: TRANSDUCTION

9.8: SEXUAL REPRODUCTION

9.9: GENOME DYNAMICS

9.0: Introduction

In which we consider the dynamics of genes and gene expression, 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 molecular level 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 display²⁵⁴. We will touch on such 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 biological 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 take some practice to master this type of thinking. The key is to keep calm and analyze on!

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.1: 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, lead to a mistaken 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 one of the largest known genomes is the plant *Paris japonica*; its genome is estimated to be $\sim 150,000 \times 10^6$ (millions of) base pairs²⁵⁵. In contrast the (haploid) human genome consists of $\sim 3,200 \times 10^6$ base pairs of DNA. The relatively small genome size of birds ($\sim 1,450 \times 10^6$ base pairs) is thought to be due to the smaller genome size of their dinosaurian ancestors²⁵⁶. 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 $\sim 80 \times 10^6$ base pairs and $\sim 28,000$ 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 $\sim 0.58 \times 10^6$ 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 ($\sim 0.16 \times 10^6$) base pairs of DNA encoding "182 ORFs (open reading frames or genes), 164 (90%) overlap with at least one of the two adjacent ORFs"²⁵⁷. Eukaryotic mitochondria and chloroplasts, derived from endosymbionts, have very small genomes. Typically mitochondrial genomes are $\sim 16,000$ base pairs in length and contain ~ 40 genes, while chloroplasts genomes are larger, $\sim 120,000\text{--}170,000$ base pairs in length, and encode ~ 100 genes. Most of the genes 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 $\sim 28\%$ had no (as yet) known function²⁵⁸.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.10: Paralogous genes and gene families

As noted previously genome dynamics play a critical role in facilitating evolutionary change, particularly in the context of multicellular organisms²⁷⁵. 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 activities²⁷⁶. 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 (a form of copy number variation) can be selected for because increased copies of genes may encode gene products involved in the detoxification of, or resistance to an anti-cancer drug²⁷⁷. It is possible that both genes retain their original functions, 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 sequences 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 or highly disadvantageous 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 more rapidly, which can make determining the relationships between genes found in distantly related organisms more tentative and speculative. Also, while functional similarities are often seen as evidence for evolutionary homology, it is worth considering the possibility, particularly in highly divergent genes and 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

- Using a diagram, consider the effects of flipping a chromosomal region around (180°) on gene expression.
- Consider the effects of such rearrangements on chromosome pairing during meiosis.
- How does sexual reproduction increase the genetic diversity within a population?
- Speculate on what selective factors might favor sexual over asexual reproduction (and visa versa).
- 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 often lost after genome duplication?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.11: Packing DNA into a cell

An important part of our approach to biology is to think concretely about the molecules we are considering. Nowhere is this more important than with DNA. DNA molecules are very long and cells, even the largest cells, are (generally) small. For example, a typical bacterium is roughly cylindrical and around $2\mu\text{m}$ in length and about $1\mu\text{m}$ 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 $\sim 3 \times 10^6$ base pairs of DNA – that is a DNA molecule almost a millimeter in length, or about 500 times the length of the 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 ~ 6000 times more DNA, that is a total length of greater than 2 meters (per cell), and this length of DNA has to fit into a nucleus that is $\sim 10\ \mu\text{m}$ 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; 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, the sequence of DNA molecules. 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 ~ 1100 genes that play important roles in both males and females²⁷⁸. But the level of gene expression is (generally) influenced by the number of copies of a particular gene. While various mechanisms can 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 one²⁷⁹. 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 packed away ("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 vertically²⁸⁰. 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 generation²⁸¹. Typically most epigenetic information is reset during the process of embryonic development.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.12: Locating information within DNA

So given that genes exists within a genome, for them to be useful there needs to be mechanisms by which specific genes can be recognized and expressed (transcribed)²⁸². Recognizing genes involves a two-component system consisting of regulatory nucleotide sequences that provide a molecular address that identifies a specific region of a DNA molecule as well as which specific strand of the DNA should be transcribed. The second component are proteins that recognize (and bind to) specific DNA sequences. The regulatory region of a gene can be simple and relatively short or long and complex. In some human genes, the regulatory region is spread over thousands of base-pairs of DNA, located "up-stream" or "down-stream" of and within the coding region²⁸³. This is possible because, over a long distance and in association with proteins DNA can fold back on itself.

The proteins that bind to regulatory sequences are known as transcription factors²⁸⁴. In early genetic studies, two general types of mutations were found that could influence the activity of a gene. "cis" mutations are located near the gene's coding (transcribed) region – these are mutations that alter the regulatory regions of a gene's. "trans" mutations mapped at other (distant) sites and they turn out to alter the genes that encode the transcription factors involved in the target gene's regulation. Transcription factors can act either positively to recruit and activate DNA-dependent, RNA polymerase or negatively, to block RNA polymerase binding and activity. It is also possible that post-translational modifications and the binding of allosteric factors can alter the activity of transcription factor. The pattern of transcription factor binding within a regulatory region can also influence whether a gene is expressed or not.

Genes that efficiently recruit and activate RNA polymerase will make many copies of the transcribed RNA and are said to be highly expressed. Generally, high levels of mRNA will lead to high levels of the encoded polypeptide. A mutation in a genes encoding a transcription factor can influence the expression of many genes, while mutations in a gene's regulatory sequence will directly effect only its own 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 edges of base pairs accessible through the major or minor grooves of the DNA helix. There are a number of different types of transcription factors, within structurally distinct DNA bonding domains; they can be grouped in various (presumably evolutionarily related)families²⁸⁵. The binding affinity of a particular transcription factor to a particular regulatory sequence will be influenced by the DNA sequence as well as the binding of other proteins in the molecular neighborhood. 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:



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 simple²⁸⁶. What we discover is that transcription factors do not recognize unique nucleotide sequences, but rather have a range of affinities for related 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"²⁸⁷. 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, and can lead to a dramatic change in gene expression.

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 from this discussion 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.

To be effective in recruiting a functional RNA polymerase complex 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 before rebinding can occur. 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 molecule²⁸⁸. 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.) 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 genome²⁸⁹.

As the conditions in which an organism lives get more complex, the more dynamic gene expression needs to be. This is particularly the case in multicellular eukaryotes, where different cell types need to express different genes, or different versions (splice variants) of genes. One approach is to have different gene regulatory regions, that bind different sets of transcription factors. Such regulatory factors not only bind to DNA, they interact with one another. 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 context, that is, the specific gene regulatory sequences accessible, the other transcription factors expressed, and their various post-translational modifications. 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 facilitate 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 often reflects the specific gene and its interactions with various accessory factors, all of which can influence gene expression.

The exact 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. As mentioned above, a similar bursting behavior is observed in protein synthesis.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.13: Network interactions

As we come to analyze the regulation of gene expression, we recognize that they represent an interaction network. A common feature of all biological systems, from the molecular to the ecological and evolutionary, are networks of interactions. These are 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 cells and tissues formed during the process of embryonic development and maturation. Since many of these interactions have a stochastic nature, chance also plays a role. At the same time there are regulatory interactions and feed-back loops that can act to suppress some of these stochastic effects and serve to make biological behaviors more predictable. All of these interactions (and the processes that underlie particular biological systems) are the result of evolutionary mechanisms and historical situations, including past adaptations and non-adaptive events.

Notwithstanding 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 specific behaviors. To analyze a system at the molecular, cellular, tissue, organismic, social, or ecological level we need to define (and understand and appreciate) the nature of the interacting objects, how they interact with one another, and what emerges from such interactions. 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 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 lactose²⁹⁰. Together these illustrate a number of common regulatory principles that apply more or less universally to biological systems at all levels of organization.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.14: *E. coli* as a model system

Every surface of your body harbors a flourishing microbial ecosystem. This is particularly true of the gastrointestinal system, which runs from your mouth and esophagus (with a detour to the nose), through the stomach, into the small and large intestine and the colon²⁹¹. Each of these regions supports its own unique microbial community (known as a microbiome). These environments differ in terms of a number properties, including differences in pH and O₂ levels. Near the mouth and esophagus O₂ levels are high and microbes can use aerobic (O₂ dependent) respiration to extract energy from food. Moving through the system O₂ levels decrease until anaerobic (without O₂) mechanisms are necessary. At different position along the length of the gastrointestinal track microbes with different ecological preferences and adaptabilities are found.

One challenge associated with characterizing the exact complexity of the microbiome present at various locations is that often the organisms present are dependent upon one another for growth; when isolated from one another they do not grow. The standard way to count bacteria is to grow them in the lab using plates of growth media. Samples are diluted so that single bacteria land (in isolation from one another) on the plate. When they grow and divide, they form macroscopic colonies and it is possible to count the number of “colony forming units” (CFUs) per original sample volume. This provides a measure of the number of individual bacteria present. If an organism cannot not form a colony under the assay conditions, it will appear to be absent from the population. But as we have just mentioned some bacteria are totally dependent on others and therefore do not grow in isolation. To avoid this issue, newer molecular methods use DNA sequence analyses to identify which organisms are present without having to grow them²⁹². The result of this type of analysis reveals the true complexity of the microbial ecosystems living on and within us²⁹³.

For our purposes, we will focus on one well known, but relatively minor member of this microbial community, *Escherichia coli*²⁹⁴. *E. coli* is a member of the Enterobacteriaceae family of bacteria and is found in the colon of birds and mammals²⁹⁵. *E. coli* is what is known as a facultative aerobe, it can survive in both an anaerobic and an aerobic environment. This flexibility, as well as *E. coli*'s generally non-fastidious nutrient requirements make it easy to grow in the laboratory. Moreover, the commonly used laboratory strain of *E. coli*, known as K12, does not cause disease in humans. That said, there are other strains of *E. coli*, such as *E. coli* O157:H7 that is pathogenic (disease-causing). *E. coli* O157:H7 contains 1,387 genes not found in the *E. coli* K12. It is estimated that the two *E. coli* strains diverged from a common ancestor ~4 million years ago. The details of what makes *E. coli* O157:H7 pathogenic is a fascinating topic, but beyond our scope here²⁹⁶.

Adaptive behavior and gene networks (the lac response): Lactose is a disaccharide (a sugar) 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 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 the protein lactose synthase²⁹⁷.

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 the *E. coli* cell membrane. Lactose permease is a membrane protein that allows lactose to enter the cell, moving down its concentration gradient. The second gene involved in lactose utilization encodes the enzyme β-galactosidase, which splits lactose into D-galactose and D-glucose, both of which can be metabolized by proteins expressed constitutively (that is, all of the time) within 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 in bacteria that multiple genes involved in a single metabolic pathway are located in the same operon (the same region of the DNA). A 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 to understand lactose utilization, we can look mutant *E. coli* that cannot grow on lactose²⁹⁸. To make the screen for such mutations more relevant, we first check to make sure that the mutant can grow on glucose. Why? Because we are not

really interested (in this case) in mutations in genes that disrupt standard metabolism, for example the ability to use glucose; we seek to understand the genes involved in the specific process of lactose metabolism. Such an analysis revealed a number of distinct classes of mutations: some led to an inability to respond to the presence of lactose in the medium, others led to the de-repression, that is the constant expression of two genes involved in the ability to metabolize lactose, lactose permease and β -galactosidase. In these mutant strains both genes were expressed where or not lactose was present. By mapping (using the Hfr system, see above) where these mutations are in the genome of *E. coli*, and a number of other experiments, the following model was generated.

The genes encoding lactose permease (*lacY*) and β -galactosidase (*lacZ*) 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. In a typical cell there are ~ 10 lac repressor proteins present. The lac repressor protein binds to sites in the promoter of the *lac* operon. When bound to these sites the repressor protein blocks transcription (expression) of the *lac* operon. The repressor's binding sites within the *lac* operon promoter appear to be its only functionally significant binding sites in the entire *E. coli* genome. The second regulatory element in the system is known as the activator site. It can bind the catabolite activator protein (or CAP), which is encoded by a gene located outside of the *lac* operon. The DNA binding activity of CAP is regulated by the binding of a co-factor, cyclic adenosine monophosphate (cAMP). cAMP accumulates in the cell when nutrients, specifically free energy delivering nutrients (like glucose) are low. Its presence acts 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 in its presence (that is, when energy is needed), the CAP-cAMP protein is active, binds to a site in the *lac* operon promoter, recruits and activates RNA polymerase, leading to the synthesis of lactose permease and β -galactosidase RNAs and proteins. However, even if energy levels are low (and cAMP levels are high), the *lac* operon is inactive in the absence of lactose because of the binding of the lac repressor protein to sites (labelled O₁, O₂, and O₃) in *lac* the regulatory region of the operon.

So what happens when lactose appears in the cell's environment? Well, obviously nothing, since the cells are expressing the lac repressor, so no lactose permease is present, and lactose cannot enter the cell without it. But that prediction assumes that, at the molecular level, the system works perfectly and deterministically. This is not the case, however, the system is stochastic, that 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. If this occurs under conditions in which CAP is active then, if lactose is present, we see the effect of positive feedback loop²⁹⁹. When lactose is added, those cells that have, by chance, expressed both the lactose permease and of β -galactosidase (a small percentage of the total cell population) will respond: lactose will enter these cells (since the permease is present) and, since β -galactosidase is also present, it will be converted to allolactone (a reaction catalyzed by β -galactosidase). Allolactone 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 further increase (~ 1000 fold) in the rate of expression of lactose permease and β -galactosidase. β -galactosidase also catalyzes the hydrolysis of lactose into D-galactosidase and D-glucose, which are then used to drive cellular metabolism. Through this process, the cell goes from essentially no expression of the *lac* operon to full expression, and with full expression becomes able to metabolize lactose. At the same time, those cells that did not (by chance) express lactose permease and β -galactosidase will not be able to metabolizing lactose at all. So even though all of the *E. coli* cells present in a culture may be genetically identical, they can express different phenotypes due to the stochastic nature of gene expression. An example of such behavior is presented in the PhET gene expression basics applet³⁰⁰. In the case of the lac system, over time the noisy nature of gene expression lead to more and more cells activating their copy of the lac operon. Once "on", as long as lactose is present in the system, its entry into the cell and its conversion into allolactone will keep the lac repressor protein in an inactive state and allow continued expression of the lac operon.

What happens if lactose disappears from the environment, what determines how long it takes for the cells to return to the state in which they no longer express the *lac* operon? The answer is determined by the effects of cell division and regulatory processes. In the absence of lactose the allolactone concentration falls and the lac repressor protein returns to its active state and inhibits expression of the *lac* operon. No new lactose permease and β -galactosidase will be synthesized and their concentrations will fall due to degradation. At the same time, and again because their synthesis has stopped, with each cell division the concentration of the lactose permease and β -galactosidase will decreases by $\sim 50\%$. With time the proteins will be diluted (and degraded) and so the cells return to the initial state, that is, with the *lac* operon off and no copies of either lactose permease or β -galactosidase present.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.15: Final thoughts on (molecular) noise

When we think about the stochastic behaviors of cells, we can identify a few reasonably obvious sources of noise. First, there are generally only one or two copies of a particular gene within a cell, and the probability that those genes are in the state to be able to recruit and activate RNA polymerase is determined by the frequency of productive collisions between regulatory sequences and relevant transcription factors together with the dissociation rate. Cells are small, and the numbers of different transcription factors can vary quite dramatically. Some transcription factors are present in 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. This will influence the probability that a particular gene regulated by a particular transcription factor is active or not. Once on, transcriptional and translational bursting will produce gene products that can alter the state of the cell such that secondary, down-stream changes occur in gene expression. and other cellular processes. These changes may (like the lac operon-lactose system) be reversible once the stimulus (lactose) is removed or they many be more or less irreversible, as occurs during cellular differentiation and embryonic development³⁰¹.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.16: 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 motifs³⁰². 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 work this way). In most cases, sets of target genes are regulated by sets of transcription factors that bind in concert.

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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.2: A Technical Aside - Transposons

In their study, Lewis et al used what is known as a “mobile genetic element” or transposons to generate mutations. A transposon is a piece of DNA that can move (jump) from place to place in the genome²⁵⁹. The geneticist (and Nobel prize winner) Barbara McClintock (1902–1992) first identified transposons in the course of studies of maize (*Zea mays*)²⁶⁰. There are two basic types of transposons. Type II transposons consist of a 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 - it jumps from place to place in the genome. The second, type I, transposon can make copies of themselves, through an RNA intermediate, and this copy can be used to generate a complementary DNA that can then 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 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 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 or apparent) 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, since some transposons are still active.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.3: 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 (or at least the vast majority) of the people you will ever meet. The difference is that you are likely to have different versions of those genes, different alleles. In this light, most macroscopic organisms (including humans) are diploid and so have two copies of each chromosome, with the exception of the chromosomes that determine sex, the X and Y chromosomes in humans. 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 $\sim 3.2 \times 10^9$ 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 RNAs²⁶¹. That is about one gene per 1000 base pairs of DNA.

In prokaryotes and eukaryotes, genes can be located on either strand of the DNA molecule, typically referred (arbitrarily) as the “+” and the “-“ strands of the molecule. Given that the strands are anti-parallel, a gene on the “+” strand runs in the opposite direction from a gene on the “-“ strand. We can illustrate this situation using the euryarchaeon *Picrophilus torridus*. This archaeal 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 – strands²⁶².

While most prokaryotic genes are located within a single major circular chromosome, the situation is complicated by the presence of separate, smaller circular DNA molecules 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 the initiation of DNA synthesis; this enables them to replicate independently of the main chromosome. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.4: Naturally occurring horizontal gene transfer mechanisms

Many horizontal transmission mechanisms are regulated by social and/or ecological interactions between organisms²⁶³. It is important to note that the mechanisms involved are complex, one could easily imagine an entire course focused on this topic, and we are introducing only the broad features of these systems. Also, we want to be clear about the various mechanisms of DNA uptake. First recognize that when organisms die their DNA can be eaten, and become a source of energy, as well as carbon, nitrogen, and phosphorus, by their neighbors. Alternatively, a nucleotide sequence of a DNA molecule can be integrated into another organism's genome, resulting in the acquisition of whatever information developed (evolved) within another organismic lineage. The study of these natural DNA import systems has identified very specific mechanisms for DNA transfer. 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 must have receptor systems that can 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?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.5: Transformation

There are well established methods used in genetic engineering to enhance the ability of bacteria to take up plasmids from their environment²⁶⁴. 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 transcription factor(see above). When present in a crowded environment, the quorum sensing system turns on the expression of the DNA uptake system and generate cells competent for transformation.

Here we outline the process in one type of bacteria but a functionally similar mechanisms are used in other bacterial species and Archaea. Double-stranded DNA binds to the bacterial cell's surface through a variety of DNA receptors. In some cases these receptors 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 recombination, is inserted into the host genome²⁶⁵. While the molecular details of this process and functionally similar processes 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.6: Conjugation

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 (insert) itself 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 known as homologous recombination). 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 blender), experimenters were able to determine 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 chromosome²⁶⁶. 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 reverts 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.

References

²⁶⁶ Synteny: <http://en.Wikipedia.org/wiki/Synteny>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.7: Transduction

The final form of horizontal gene transfer is one that involves viruses. The structure and behavior of viruses is a complex topic, the details of which are beyond us here, but it is not unreasonable to consider viruses 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 infectious (and can be denatured by various treatments). 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 polypeptides²⁶⁷. The assembled virus has an icosahedral protein head that contains its DNA molecule and a tail assembly that recognizes and binds to target cells. Once a suitable host is found (based on tail binding to cellular surface molecules), the tail domain attaches and contracts, like a syringe, punching a hole through the cell's external wall and plasma membrane. The DNA emerges from the bacteriophage and enters the cytoplasm, infecting the 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 phage DNA encodes the proteins of the phage will be assembled of new phage heads. DNA is packed into these heads by a protein-based DNA pump, a pump driven by coupling to an ATP hydrolysis complex²⁶⁸. In the course of packaging virus DNA will, occasionally, make a mistake and package a fragment of host DNA. When such a phage particle infects another 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.

Because the movement of DNA is so common, particularly in the microbial world, a number of defense mechanisms have evolved. These include the restriction / DNA modification systems used widely for genetic engineering, and the CRISPR-CAS9 system, which enable cells to recognize and destroy foreign DNA.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.8: Sexual reproduction

In eukaryotes, the major mechanism for shuffling genes is sexual reproduction. Because sex (usually) requires that organisms cooperate sex represents an example of social evolution. So what are the benefits of sexual reproduction²⁶⁹?

Sexual reproduction involves the alternation of diploid (two copies of each chromosome) and haploid (one copy of each chromosome) cell types. In a diploid cell, the two copies of a particular gene can be different, the the organisms is said to be heterozygous for that gene. Different alleles can (but do not necessarily do) encode functionally distinct gene products and can be associated with different phenotypes (discussed in greater detail in the future two-semester version). We can then characterize different alleles with respect to one another. Assume that allele-1 is associated with phenotype-1 and allele-2 is associated with phenotype-2. An organism heterozygous for allele-1 and allele-2 may express one of four phenotypes; it may express phenotype-1, phenotype-2, a new phenotype (phenotype 3), or no distinctive phenotype. If phenotype-1 is found, allele-1 is said to be dominant to allele-2 (which is said to be recessive), if phenotype-2 is found, allele-2 is dominant to allele-1 (which is said to be recessive); if phenotype 3 is found, the two alleles are said to be co-dominant. If both alleles of a gene are the same, the organism is said to be homozygous for that gene. If the phenotype associated with a particular allele is observed only in homozygous individuals, the allele is said to be recessive.

The process of sexual reproduction has two obvious impacts on a population's genetic variation. The process of homologous recombination, which occurs during the formation of haploid cells, can generate new alleles not present in the original cell. In addition, the generation of haploid cells can reveal the presence of recessive alleles. When the present an isolated recessive allele results in a lethal or highly deleterious phenotype, the haploid cells that contain such an allele will be eliminated, removing the allele from the population.

As we will see below, the process of recombination between homologous chromosome leads to mutations when cells attempt to divide. Avoiding such mutations is thought to be involved in driving the evolution of the linear chromosomes in eukaryotes. Eukaryotes also typically have multiple chromosomes. Each chromosome contains single (linear) double-stranded DNA molecule. Different chromosomes can be distinguished by the genes they contain, as well as the overall 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 ~250 million base pairs of DNA and over 2000 polypeptide-encoding genes, while the smaller chromosome 22 contains ~52 million based pairs of DNA and around 500 polypeptide encoding genes²⁷⁰.

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. Two homologous chromosomes are generally similar, with the exception of allelic differences in different genes (also known as genetic position or genetic loci). The exception to this rule are the so called sex chromosomes. While different types of organisms determine an individual's sex using different mechanisms, 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.

When a eukaryotic cell reproduces asexually, it goes through two linked processes known as mitosis (the segregation of chromosomes) and cytokinesis (the generation of two daughter cells). The sexual reproductive cycle involves a modified form of mitosis known as meiosis. It involves two distinct mechanisms not found in mitosis that lead to 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 the process known as meiosis, produces as many as four haploid cells, known as gametes. The first step in this process is the replication of the cell's DNA; each individual chromosome is duplicated, that is the DNA molecule it contains are copied to produce two DNA molecules. 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 (mitotic) division each replicated chromosome interacts independently with a molecular machine, known as the mitotic spindle, that acts to send one copy of each chromosome to each

of the two daughter cells. During mitosis a diploid cell produces a diploid cell, and nothing about the genome has changed. The cells that are formed become part of the original organism.

In contrast the purpose of meiosis is to produce a new organism, with 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, and in contrast to mitosis, the (now) duplicated homologous (material and paternal) chromosomes form a structure containing of four DNA strands, sometimes called a tetrad. This pairing is based on the fact that the DNA sequences along each homologous chromosome, while not identical, are extremely similar: they are syntenic.²⁷¹ At this point, at positions more or less random along the length of the chromosome, there are double strand breaks in two of the 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, generally involve different pairs of DNA molecules.

At the first meiotic division, the duplicated maternal and paternal chromosomes remain attached at their centromeres, but if you follow any one DNA molecule from beginning to end you will find that because of crossing over each will, in fact, be different from either the maternal or paternal chromosomes. Each of the two resulting daughter cells will receive either the replicated maternal or paternal chromosome centromere region, but which chromosome (defined by their centromeres, material or paternal) they inherit is a random process. For an organism with 23 different chromosomes, that produces 2^{23} possible different daughter cells. There is no DNA replication before the second meiotic division. During the second meiotic division the two daughter cells each receive a copy of one and only one homologous chromosome (one DNA molecule). The four cells that are generated by meiosis are known as gametes (or at least are potential gametes) and they are haploid..

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 the alleles they contain. There will be nucleotide differences at specific positions (known as single nucleotide polymorphisms or SNPs - pronounced snips) and small insertions and deletions of nucleotide sequences. For our purposes, we will consider only one single chromosome set, but remember there are generally 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 (an unrealistic assumption), 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 parental PMN and pmn chromosomes (top panel).

Now let us consider what happens when this PMN/pmn organism is about to reproduce (sexually). It will begin meiosis (bottom panel). The processes of homolog pairing and crossing over will generate new combinations of alleles: the four haploid cells illustrated here 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 between 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 new 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, on average, but recombination allows alleles of neighboring genes to disconnect from one another. 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 ~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. 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

9.9: 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 a gene coding region or its expression. Even when they occur in a gene’s coding region, the redundancy of codons means that many SNPs will not lead to functionally significant alterations in the gene products. That said, even apparently “neutral” mutations can lead to changes in genotype that can have effects on phenotype, and so evolutionary impacts. As we have already discussed, in small populations genetic drift can influence whether new alleles (with non-lethal effects) are retained in 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 flipping of the orientation of a DNA region (inversion) and sequence insertions or deletions, known as copy number variations).²⁷²

As noted previously~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) genome²⁷³. These 6000 base pair long DNA regions contain genes that encode proteins 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 obvious 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”²⁷⁴.

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 can be actively identified and killed.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

CHAPTER OVERVIEW

10: SOCIAL SYSTEMS

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.



10.0: INTRODUCTION TO SOCIAL SYSTEMS

At the molecular level it is common to focus on the interactions between proteins and DNA (genes) that control gene expression. These molecular level interactions play an important role in determining how cells behave. Interactions between cells influence the behaviors of the interacting cells, 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,

10.1: MICROBIAL COMMUNITIES

Some organisms secrete toxins to control the growth of others. Some secrete molecules that influence the behaviors of other organisms (including themselves). There are complex molecular level conversations going on between the organisms within an 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.

10.2: MAKING METAZOANS

As we move from quorum sensing between organisms of the same type, we come to what are known as biofilms. These are microbial communities, such as the plaque that forms on your teeth, and consist of a number of different types of often co-dependent organisms. While horizontal gene transfer may occur between these different organisms, they remain distinct and give rise to organisms genetically related to their parent(s).

10.3: STEPS TO METAZOANS MULTICELLULAR ANIMALS AND PLANTS

10.4: DIFFERENTIATION

10.5: STEM CELLS

10.6: CELLULAR DIFFERENTIATION AND GENOMIC INFORMATION

10.0: Introduction to Social Systems

We end up (for now) 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 considered 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 cells, 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

10.1: 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, molecules obtained from our diet, reflects this interdependence. Some organisms secrete toxins to control the growth of others. Some secrete molecules that influence the behaviors of other organisms (including themselves). There are complex molecular level conversations going on between the organisms within an 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. If organisms at low density in a aqueous environment the expensive to produce secreted molecules are more likely to diffuse away and so be effectively useless to the organism that produced them. However if organisms are present at high local densities, then the process becomes more efficient, the concentration of the secreted molecules increases 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 (relatively energetically inexpensive) molecules known as auto-inducers. They also produce a cellular receptor specific for this auto-inducer. The auto-inducer-receptor system enables organisms of the same type to recognize each other's presence. 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 increase 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 (including gene expression)³⁰³. 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 reverse causing 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 contains a plasmid that encodes the Green Fluorescent Protein (GFP), a protein that when illuminated with blue light glows green³⁰⁴! When GFP-expressing *E. coli* are cultured in a maze-like environment with a central "chamber" with a single opening, the secreted attractant accumulates to high concentrations within this space. Over a three hour period the bacteria swim in a directed manner up the attractant concentration gradient into the chamber³⁰⁵. At this point quorum sensing-controlled behaviors are 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 cell³⁰⁶. Another type of behavior that can occur under condition of stress is that a subpopulation of cells, known as quiescent or "persister" cells, grow slowly or not at all, while the rest of the population continues to grow³⁰⁷. If the environment turns seriously hostile the persisters have a much higher probability of survival than do the actively growing cells. If conditions 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, and suppress their formation when not. 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 for 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 cells³⁰⁸. 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.

Social cooperation between cells can provide benefits, but also opens up the system to selfish behaviors, essentially cheating³⁰⁹. This raises the evolutionary question what can be done to surpass 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. The cheater gains the benefits of communicating with other bacteria, but minimizes its contribution to the process (it does not use energy to synthesize the autoinducer). 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 inherent limits to cheating, however. If enough members of a group (population) become cheaters the quorum sensing system will fail because not enough members of the community secrete the auto-inducer. Assuming that the social behavior is critical for the survival of the population, a group with too many cheaters may die out (become extinct).

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 (the organisms become addicted to the autoinducer). Many bacterial species synthesize toxins to which they themselves are immune, but which kill cells of related species. It can be that toxin immunity is coupled to auto-inducer expression. In higher organisms there can be the development of the ability to recognize cheaters by their behavior in certain situations. We can think of cancer (tumor formation) as a form of social cheating with a multicellular organism, and their are mechanisms that are to suppress it (although we will not have to time to explore them to any significant extent).

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

10.2: Making Metazoans

As we move from quorum sensing between organisms of the same type, we come to what are known as biofilms. These are microbial communities, such as the plaque that forms on your teeth, and consist of a number of different types of often co-dependent organisms. While horizontal gene transfer may occur between these different organisms, they remain distinct and give rise to organisms genetically related to their parent(s).

The next obvious level of organization is what we will call a discrete colonial organism. In such 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 become too big to easily swallow) but these advantages are constrained by the fact that the individual cells retain their individuality. 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, 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 next step in the evolution of a truly integrated multicellular organism, one with a soma and a germ line? It appears that there have been a number of paths to multicellularity: for example animals, plants, and fungi all appear to have independent strategies (lineages). We can use modern organisms a modern bestiary to illustrate these strategies³¹⁰. 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.



Figure 10.2.1: *Dictyostelium* showing aggregation behavior. Image from Wikimedia Commons (“*Dictyostelium Aggregation*” by Bruno in Columbus – Own work. Licensed under Public domain via Wikimedia Commons – http://commons.wikimedia.org/wiki/File:Dictyostelium_Aggregation.JPG)

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 leading to the formation of cellular aggregates known as slugs that contain between 10,000 to 100,000 discrete cells. Slugs migrate in a coordinated manner. Eventually a slug will stop its migration and begin a process of known as differentiation, a process involves changes in gene expression and cellular behavior. 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 considered migrating bacterial farmers³¹¹. 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 challenge 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 sensing³¹².

An organism that displays a distinct type of differentiation behavior is the bacterium *Caulobacter crescentus*. Under certain conditions it will produce a distinct type of cell, a stalk cell, which attaches to a surface. It can 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. crescentus* has established two different cell types designed to exploit two distinct environments.

References

³¹⁰ The medieval bestiary: <http://bestiary.ca>

³¹¹ Small molecules mediate bacterial farming by social amoebae. <http://www.ncbi.nlm.nih.gov/pubmed/23975931>

³¹² Kin Recognition Protects Cooperators against Cheaters: <http://www.ncbi.nlm.nih.gov/pubmed/23910661>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

10.3: 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 first occurred ~1 billion years ago. In true multicellular organisms (as opposed to colonial organisms), different cells become highly specialized. Most cells (somatic cells) are relieved of the need to produce a new organism; that task is taken up by specialized cells (germ cells). 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 eukaryotes³¹³. 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 sister group are the choanoflagellates³¹⁴. Choanoflagellates have cells that are characterized by a single flagellum surrounded by a distinctive collar structure³¹⁵. Choanoflagellates exist in both unicellular and simple colonial forms.

Sponges (porifera) are among the simplest of the metazoans (multicellular animals). Fossils of extinct sponges, such as the Archaeocyathids, have been found in Cambrian rock over 500 million years old. Earlier sponge-like organisms have been found in even older Pre-cambrian rock. Sponges contain only a few different types of cells. These include the cells that form the outer layer of the organism (pinococytes) and those (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 choanoflagellates (and subsequent genomic analyses) that led to the hypothesis that choanoflagellates and animals are sister groups³¹⁶.

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 (anchored) and benthic, or floating, lifestyles³¹⁷. 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 behave as 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 waste is 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 muscle 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 poison³¹⁸. Hydrozoans are complex enough to be true predators.

Questions to answer & ponder:

- What types of social signals do humans 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 isolation?
- 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?
- What types of evidence suggest that choanoflagellates and sponges are related?
- Why is the presence of highly specialized cells considered 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?
- Does coordinated movement require a brain? Does having a brain equal self-awareness?

References

³¹³ Multicellularity arose several times in the evolution of eukaryotes: <http://www.ncbi.nlm.nih.gov/pubmed/23315654>

³¹⁴ http://www.nytimes.com/2010/12/14/science/ures.html?_r=0

³¹⁵ Introduction to the Choanoflagellata: <http://www.ucmp.berkeley.edu/protista/choanos.html>

³¹⁶ The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans: <http://www.ncbi.nlm.nih.gov/pubmed/18273011>

³¹⁷ The live cycle of jellyfish: http://youtu.be/oHiVA9J_YIM

³¹⁸ How do jellyfish sting: <http://youtu.be/HyIwa7W-ZV8>

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

10.4: Differentiation

Complex multicellular organisms undergo a process known as embryonic development³¹⁹. Development begins with the fusion of a haploid sperm and egg (produced through meiosis) to form a new diploid organism (known as zygote). This cell then divides (by mitosis) to produce an embryo that 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 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 produce—they become committed to one or another specific fate. In part this fate restriction is due to the fact that as cells divide, different cells come to have different neighbors and they experience different environments, leading to the expression of different genes. The question becomes, what determines what types of cells an embryonic (increasingly differentiated) cell produce?

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 specific developmental directions based on changes in gene expression. At the same time, there are changing interactions between cells and cells come to inhabit different environments. Together asymmetries and interactions lead cells to express different genes and to adopt specific fates. There are many different types of embryonic development, since this stage of an organism's life cycle is as subject to the effects of evolutionary pressures as any other (although it is easy to concentrate our attention 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, these new organisms are highly vulnerable to predation. In such organisms, early embryonic development generally 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 (SEP marks sperm entry point in the figure early frog development). Maternal and fertilization-driven asymmetries are stabilized by the rapid cycles of DNA replication and cell division, with growth being dependent upon the utilization of maternally supplied nutrients. As distinct cells are formed, they begin to become different from one another as i) they inherit different determinants, ii) the presence of these determinants leads to changes in gene expression, and iii) cells secrete and respond to different factors that drive their differentiation further 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 (generally) dramatically slower. In such developmental systems, it is not the asymmetries associated with the oocyte and fertilized egg that are critical, but rather the asymmetries that arise during embryonic development. As the zygote divides, a major factor that drives the differentiation is whether a cell comes to 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 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.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

10.5: 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 divide further and 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 comes 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 remain 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 differentiate 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 that enable them to circumvent the effects of senescence.

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

10.6: 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 1960s³²⁰. 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 (taken from differentiated cells) 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) of a small number (typically) four molecules to induce differentiated somatic cells to become pluripotent cells. These “induced pluripotent stem cells” (iPSCs) behave much like embryonic stem cells. The hope is that iPSCs 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:

- How might asymmetries be generated in the zygote?
- 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?
- 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 ethical objections to human cloning? What if the clone were designed to lack a brain, and destined to be used for "spare parts"?

Contributors and Attributions

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.

Index

B

Biofilms

[10.2: Making Metazoans](#)

Biofundamentals

- 1: Understanding Science & Thinking Scientifically
 - 1.1: The Interconnectedness of Science
 - 1.2: Models, Hypotheses, and Theories
 - 1.3: Science is Social
 - 1.4: Teaching and Learning Science
 - 1.5: Understanding Scientific Ideas
 - 2: Life's Diversity and Origins
 - 2.1: What is life, exactly?
 - 2.2: The cell theory and the continuity of life
 - 2.3: The organization of organisms
 - 2.4: Spontaneous generation and the origin of life
 - 2.5: The Death of Vitalism
 - 2.6: Thinking about life's origins
 - 2.7: Experimental studies on the origins of life
 - 2.8: Mapping the history of life on earth
 - 2.9: Fossil evidence for the history of life on earth
 - 2.10: Life's impact on earth
 - 3: Evolutionary Mechanisms and the Diversity of Life
 - 3.1: Organizing organisms (hierarchically)
 - 3.2: Natural and un-natural groups
 - 3.3: Fossils and family relationships: introducing cladistics
 - 3.4: Evolution theory's core concepts
 - 3.5: So what do we mean by genetic factors?
 - 3.6: Limits on populations
 - 3.7: The conceptual leap made by Darwin and Wallace
 - 3.8: Mutations and the origins of genotype-based variation
 - 3.9: The origins of polymorphisms
 - 3.10: A short aside on the genotype-phenotype relationship
 - 3.11: Variation, selection, and speciation
 - 3.12: Types of simple selection
 - 3.13: A short note on pedagogical weirdness
 - 3.14: Population size, founder effects and population bottlenecks
 - 3.15: Population bottlenecks
 - 3.16: Genetic drift
 - 3.17: Gene linkage: one more complication
 - 3.18: A brief reflection on the complexity of phenotypic traits
 - 3.19: Speciation and extinction
 - 3.20: Mechanisms of speciation
 - 3.21: Isolating mechanisms
 - 3.22: Sympatric speciation
 - 3.23: Signs of evolution: homology and convergence
 - 3.24: The loss of traits
 - 3.25: Signs of evolutionary history
 - 3.26: Homologies provide evidence for a common ancestor
 - 3.27: Anti-evolution arguments
 - 4: Social Evolution and Sexual Selection
 - 4.1: Selecting social (cooperative) traits
 - 4.2: Quorum Sensing
 - 4.3: Active (altruistic) cell death
 - 4.4: Inclusive fitness, kin and group selection, and social evolution
 - 4.5: Group selection
 - 4.6: Defense against social cheaters
 - 4.7: Driving the evolutionary appearance of multicellular organisms
 - 4.8: Origins and implications of sexual reproduction
 - 4.9: Sexual dimorphism
 - 4.10: Sexual Selection
 - 4.11: Curbing runaway selection
 - 5: Molecular Interactions, Thermodynamics & Reaction Coupling
 - 5.1: A very little thermodynamics
 - 5.2: Reactions: favorable, unfavorable, and their dynamics
 - 5.3: Thinking entropically (and thermodynamically)
 - 5.4: Reaction rates
 - 5.5: Coupling Reactions
 - 5.6: Molecules, London Dispersion Forces, and van der Waals interactions

5.7: Covalent bonds

- 5.8: Bond Stability and Thermal Motion (a non-biological moment)
- 5.9: Bond polarity, inter- and intramolecular interactions
- 5.10: The implications of bond polarity
- 5.11: Interacting with Water
- 6: Membrane Boundaries and Capturing Energy
 - 6.1: Defining the Cell's Boundary
 - 6.2: The origin of biological membranes
 - 6.3: Transport across membranes
 - 6.4: Transport to and across the membrane
 - 6.5: Channels and carriers
 - 6.6: Generating gradients: using coupled reactions and pumps
- 7: The Molecular Nature of Heredity
- 8: Peptide Bonds, Polypeptides & Proteins
- 9: Genomes, Genes & Regulatory networks
- 10: Social Systems

C

Caulobacter crescentus

[10.2: Making Metazoans](#)

conjugation

[9.6: Conjugation](#)

D

Dictyostelium

[10.2: Making Metazoans](#)

differentiation

[10.4: Differentiation](#)

E

embryonic development

[10.4: Differentiation](#)

G

genetic drift

[3.16: Genetic drift](#)

M

Metazoans

[10.2: Making Metazoans](#)

P

programmed cell death

[4.3: Active \(altruistic\) cell death](#)

Q

quorum sensing

[4.2: Quorum Sensing](#)

S

Stem Cells

[10.5: Stem Cells](#)

T

transformation

[9.5: Transformation](#)

transposons

[9.2: A Technical Aside - Transposons](#)

Glossary

Lass="lt-bio-39877">Sample Word 1 | ass="lt-bio-39877">Sample Word 1 Word 1]
[Source: " class="lt-bio-39877">Sample