Goals & Principles of this text

Education is not the learning of facts, but the training of the mind to think." --Albert Einstein

Interactive: (no, really!) Most of us learn by doing and remember what we've done better than what we heard. The ubiquity of computers, smartphones & tablets means we can deliver *exercises* and *tasks* to students that are richer and more reflective of the content than word interactions (multiple choice questions, etc.). The goal of this eBook is to deliver truly interactive challenges to students tailored to each topic.

Approachable: Biology isn't 'hard', but it can be made inaccessible by walling it off with jargon or a commitment to 'cover everything' at the expense of delivering understanding. This text seeks to engage students by limiting technical terminology and always approaching topics from the perspectives of "how does this work", "why is it this way" and "why does this matter?"

Real: The idea of 'Introductory' here means 'from the ground up' as opposed to a shallow or "words first-meanings later" approach. This text represents the philosophy that it is actually easier to completely understand something and to be presented with the reasons why things are as they are than to be asked to memorize things that are not well understood. As such, it must cover fewer topics in greater depth than a more survey-oriented approach would allow.

Concept driven: This is not a text about the names of things or sharing long lists of steps, components, or actors. It's about how things work. I do not except the premise that students are 'not ready' to understand at the introductory level or its corollary that we should fill them up with terms for the understanding that will somehow sew it all together later. If it cannot be explained right now, it won't be introduced right now.

What's a 'concept'? In this text, the general rule will be *a principle essential to the operation of living things as we know them*. For example, the genetic material wouldn't 'work' if the pairings were flexible (allowed multiple partners), but if we had four different bases? So *what the bases do* is a principle; their names and structures may be historical accident. In the case of amino acids, it's critical that there be a diversity of 'building blocks' representing the major molecular properties; it's elegant that there are some 'specialists' (cysteine, proline)… But the fact that there are 20, the specific 20 we have, or the codons specifying each? Interesting details.

What's 'understanding'? Here, understanding will be defined as a working knowledge of how systems and machines actually work. The emphasis will be on enabling readers to explain Big Ideas in their own words and drawing from everyday experience rather

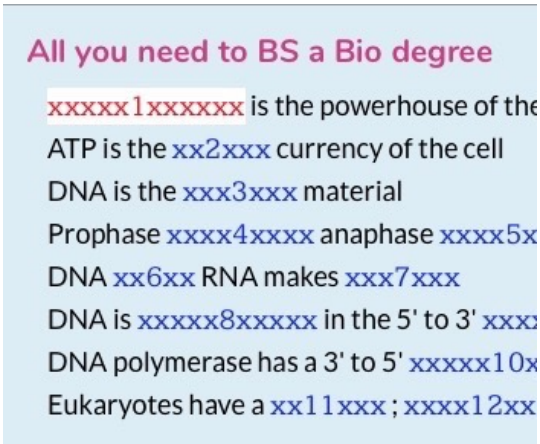
than quoting technically correct but remote and potentially meaningless key phrases.

Pedagogically sound: Everyone learns things when they perceive a value to the learning and when they can take new ideas and apply them. The driving principle behind the layout of this text is "foundations and dependencies": nothing is introduced until the underlying conceptual/material framework is in place (dependencies), and ideas are introduced when there's something to 'do' with them—they're the foundation for something useful or interesting.

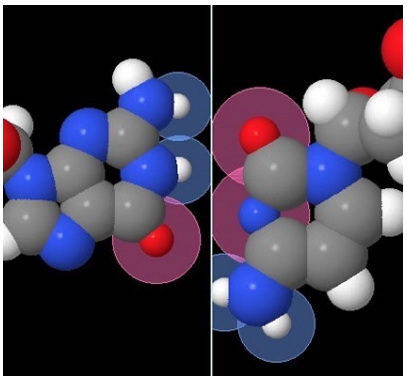
To read more about goals consistent with the above, laid out by U.S. Organizations
Bio2010
Vision and Change
AP Biology

Special approaches

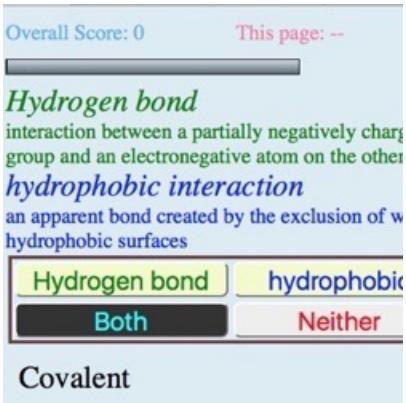
A number of elements are unique or foundational to this text and approach:



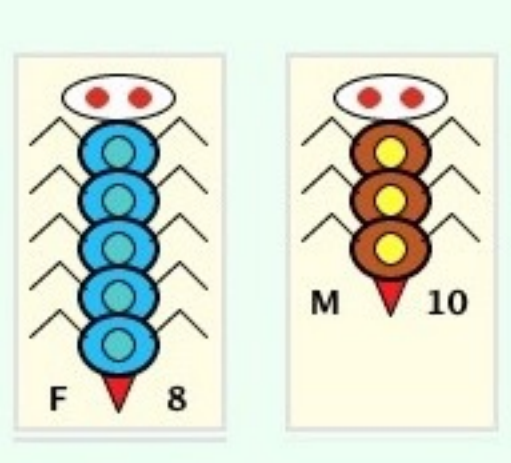
Topic summary tool: This is an interactive short text where the user is challenged to fill in the blanks. To prevent frustration and encourage interaction, words are progressively revealed if first choices aren't correct.



Embedded 3D molecular display: Throughout the text, molecules from water to amino acids to proteins are displayed in interactive, 3D windows (on all platforms, including tablets). Students can generally select how to view the molecules and which properties to focus on.



'Gamified' vocabulary: Vocabulary may not be the primary teaching goal of Introductory Biology, but it can stand in the way of other goals. 'VocabuWary' game packets are provided for each chapter and provide multiple ways of thinking about each word as well as checking comprehension of similar words/ideas (transcription/translation; meiosis/mitosis)



Dedicated interactive activities: biology is a rich field; the Big Ideas often don't fit into multiple choice questions and similar verbal formats. Throughout this text you'll find activities that challenge the students to solve problems and demonstrate understanding of key principles.

Embedded review questions with instant feedback.

Experiments and how-we-know *integrated* as part of the text itself.

1. Why understanding biology matters

Life is innately interesting

While an understanding of biology is both useful and more important than it's ever been, life is also just plain interesting. Biology is the study of how we work as cells, organisms, and members of complex webs of food and energy. It offers us insight into the origins of the wonderful complex living things we now see from simpler ones that came before (and if we peer hard enough, perhaps even how life can arise from non-life). The study of life is in many ways the study of ourselves.

You often hear the concern that understanding something takes all the wonder out of it. One of the most dramatic disproofs of this worry is the focus of this book: the 'secrets' (or more correctly, the underlying mechanisms) of life. We'll build from the tiny upwards, starting with atoms, moving to molecules, and growing to cells, organisms, and ecosystems.

One amazing things about living systems is how few Big Ideas your really need to grasp to achieve a useful understanding. You'll come to see that this is actually inevitable—for life to have evolved from simple beginnings, the steps must have been 'achievable' and small… and the echoes of those beginnings are still present in the machines and processes of Life today.

In the first half of the text, you'll see how the business of cells is transacted and the goings on follow a few overriding principles: first, the laws of chemistry constrain what biological molecules can do (in the end, biological molecules are non-magical chemicals!); second, a major organizational principle of cellular machines is that they 'take care of themselves'—most often at the level of self-forming into functional shapes. This arises not because they're 'smart'—again, 'just' chemicals—but because each of the major molecules is a collection of 'beads on a string', and the interaction of water with the properties of the beads drive each string into its functional shape. Third, every molecule is a mindless machine—functions can't be 'thought out' or based on 'what is best for the cell'—everything is based on knee-jerk cause-and-effect. But oh, the wonderfully tuned, mindless machines you'll meet!

In the second half, you'll look at the many levels of webs of life--cycles of energy and materials, interactions between organisms, and how both chance and competition drive the changes in populations that we call evolution. We'll also look at the role humankind has played in shaping the world of today, and the choices and opportunities we face as we stand at multiple crossroads, picking our future.

Modern Medicine

In the past, medicines were discovered by chance or trial-and-error. While these 'methods' continue to make important contributions, drug and therapy discovery is increasingly about, testing hypotheses and knowledge-driven design. Our understandings of how diseases afflict us and how the molecules involved act provide avenues for designing approaches that save lives and make life better for millions. Recent developments promise almost unlimited opportunities to actually fix defective genes…or alter 'normal' ones. All of these abilities arise from our increasingly subtle understanding of biology.

One major path has been the conversion of information gained from curiosity about the biological world into tools for changing it. Characterization of self-defense mechanisms in bacteria laid the foundation for molecular cloning; understanding of how DNA is copied set up PCR, and the new kid on the block—the DNA re-writing "CRISPR-Cas" systems—is the result of detective work seeking to understand yet another bacterial self-defense system.

Knowing the code, we can design

Our understanding of how the code and machines of life work gives us powers both wonderful and frightening. More and more, we are in a position to re-design both molecular and macroscopic properties of living things… but we need to be well-informed and thoughtful lest the bugaboo from 'Jurassic Park' come back to haunt us: "Your scientists were so preoccupied with whether or not they could, they didn't stop to think if they should."

But some of the capabilities we now have are worthy of serious consideration. Should we modify mosquitoes so that they can no longer carry the malaria parasite? Or should we modify them out of existence? What is the dividing line between alleviating human suffering and curing disease versus vanity changes or eugenics? There are no clear answers here, and each of us needs to fully understand the question in order to formulate and express our views and interests.

Decisions: resources & the environment

"With great power comes great responsibility."

There's no question that we are an immensely powerful species. We've spread over all continents, and some believe it's now appropriate to assign a new epoch of life, the Anthropocene, to reflect our influence… which informed observers argue includes not only changes in global climate, but mass extinction and massive alteration and destruction of ecosystems that support fascinating forms of life… The changes contribute both materially and aesthetically to our own situation.

Both action and inaction represent choices, and in order to achieve the ends we desire, we need to understand the consequences of potential actions. There was a time when it seemed like introducing cane toads to guard sugarcane crops was a great idea; water lilies were viewed as a lovely, exotic addition to waterways. And sometimes, mere carelessness is enough; if you're familiar with fire ants, kudzu vine, zebra mussels… or rats, these have all hitchhiked with human travelers and set up housekeeping to very noticeable effect in their new domains.

There are also potential positives, of course. We have created strains of rice known as 'golden rice' the produces more vitamin A precursor than natural rice, alleviating a nutritional deficiency that can lead to severe vision problems or blindness. We're contemplating multiple ways of combatting humankind greatest killer, malaria, ranging from eliminating the mosquitoes to rendering them inhospitable to the microbe that infects us. At the other end of the scale, there's potential to restore woolly mammoths, dodo birds… and who knows what else? It is your generation that will determine not only what we *can* do, but whether we proceed to do it.

Is it getting warm in here?

Few knowledgeable individuals dispute that the world is changing around us. One of the most notable and worrisome is that humankind's conversion of carbon-containing fuels into carbon dioxide (CO_2) is changing the atmosphere's heat-retention properties. A consequence of increased CO_2 is that more infrared radiation (essentially, warmth) that would have escaped into space is reflected back toward the Earth. This change in balance has temperatures climbing… and they'll continue to do so for some time. Understanding the problem and making difficult decisions about mitigating future effects will require an understanding of climate… but also biology.

2. Ways of knowing: the scientific approach

What you will learn

> *Science is simply common sense at its best--that is, rigidly accurate in observation, and merciless to fallacy in logic.*
>
> *--Thomas Huxley*

Solid thinking is a skill just as cooking or playing basketball or dancing. And like these, some of it comes naturally, but proficiency requires study, identifying a process and practice. Approaching questions and problems in a scientific way is not a reflex—in a real sense, scientific thinking was invented. The effectiveness of the method can be glimpsed by looking at advances in medicine, technology and understanding over the centuries. For these reasons, thoughtful study of the scientific approach and practice in its application are important for all of us.

While scientific thinking runs through every section of this text, in this chapter, we'll visit some of the early triumphs of scientific thinking in biology. The goal is to lay bare some of the key elements of this way of problem solving as well as to give you signposts to watch for as you explore our current understanding of life and its machines. To get the most benefit from these opportunities, ask yourself throughout "Would I have had this idea?" "What was the key question or insight that drove this discovery?". By challenging yourself not only to understand, but to expand your abilities you'll develop as a thinker and a scientist

Approaching questions scientifically is not everyone's reflex; indeed credit for first formally laying out the broad strokes of the approach is sometimes given to Francis Bacon, who lived in the 16th and 17th centuries. While picking a single individual or time oversimplifies, it's worth noting that rigorous observation and moving in an orderly way to conclusions about the world is not necessarily 'natural' nor our default way of doing things!

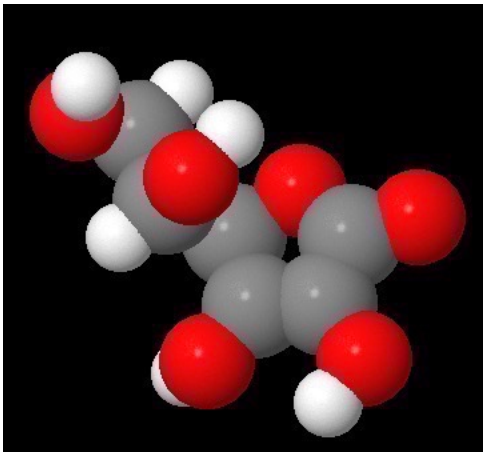
Scurvy and a barrel of limes

'**The Scientific Method**' can sound like an abstract tool that should be parked at the classroom door on you way out… but in reality, it's a well-honed tool for reliably figuring out how the world works. And it's been saving lives for a long time. Consider the case of scurvy, a debilitating disease that wrought havoc (in the form of rotting gums, bruising, muscle weakness) amongst sailors on long journeys… that is, until lime juice became standard issue on British ships.

The solution has roots in the decision by James Lind to perform side-by-side comparison (a.k.a. controlled experiment) of several potential treatments. His hypothesis was that scurvy was the result of bodily decay and that acids could prevent this. He therefore took a dozen sickened sailors and gave each pair a different treatment, including cider (containing acetic acid), sulfuric acid (much diluted, lest the sailors dissolve!)… and [two oranges and a lemon]. Before his supplies ran out, the last group was displaying a marked reversal of symptoms.

And thus, scurvy was immediately eradicated? Nope. As so often happens, advances require not only knowledge, but a willingness to hear and learn. It was decades before rations of lemon or lime juice were routine in the British navy.

Why were citrus-deprived sailors losing teeth? We'll look more closely at enzymes, the machines that do cellular chores, later. Suffice it to say that several of the machines involved in building collagen, a protein critical in joints and in connecting tissue, require vitamin C. Citrus fruits are one source of vitamin C, a molecule humans cannot make… but require. Without it, tissues weaken (loose teeth, bleeding gums…). By restoring it to the diet in the form of limes, disaster can be averted. A molecular model of vitamin C is shown below--we'll cover how to interpret this kind of molecule later, so don't worry if it doesn't make sense now! And no, there's no way *looking at it* can tell you what it's doing just yet!



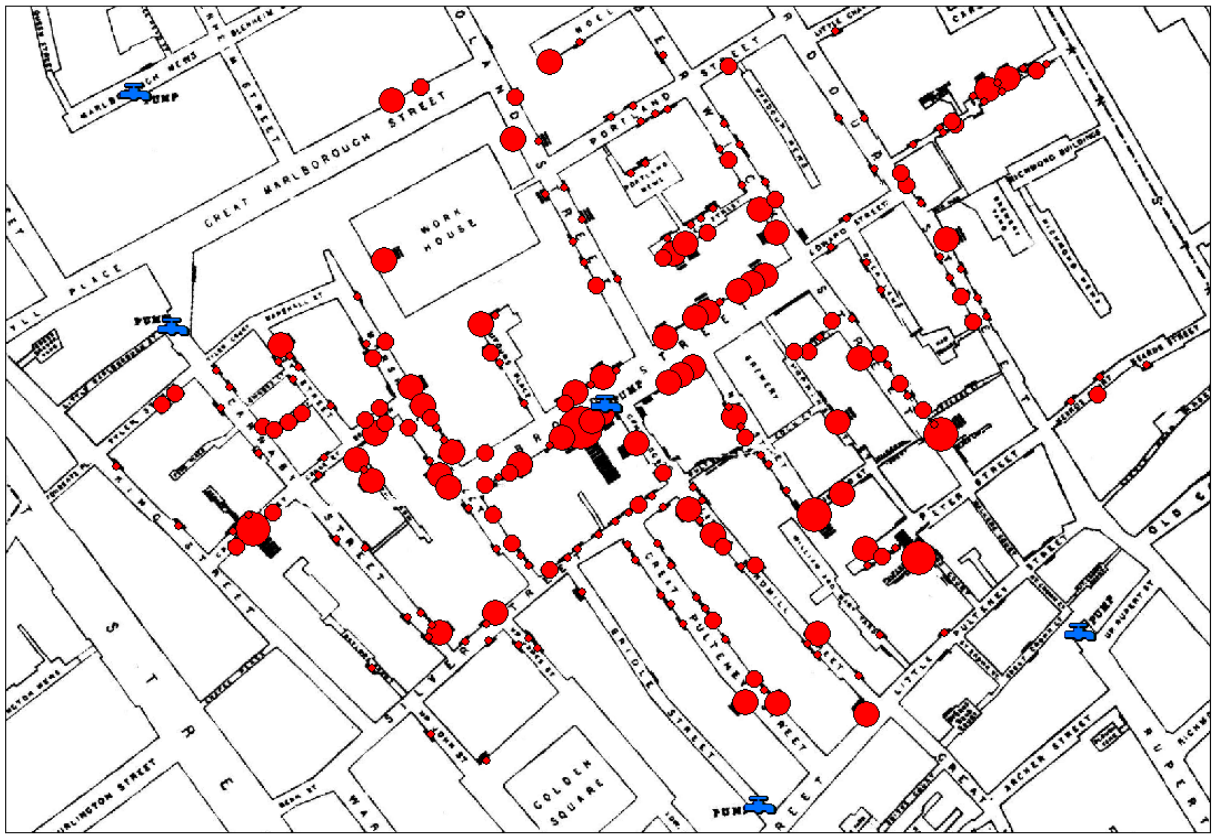
Vitamin C

Did you get it? **Interactive review**

Pump marks the spot: Snow and cholera

 One of the hallmarks of a good theory is that it simplifies ways of knowing the universe. Instead of having to associate unique observations with each organism or phenomenon, we seek to discover patterns… and then to understand why those patterns hold. One great example of this method in practice lies in a cholera epidemic of London in 1854.

First, a little background: at the time one of the primary theories of disease was the 'miasma' theory, which held that disease was caused by fogs or gases (the word malaria, for instance, comes from 'mala aria' or 'bad air'). This theory offered limited pathways to protect from or cure disease. However, Dr. John Snow took a calculated approach to analyzing the disease: he mapped out deaths occurring from cholera. And what a map it proved to be!



Source: http://blog.rtwilson.com/wp-content/uploads/2012/01/SnowMap_Points.png

Take a moment--what do you see? What do you think? This is one starting point for a sound scientific hypothesis: an observation that either requires or suggests an explanation. Snow's thinking? Unless the disease was being distributed by a tornado (hint: England doesn't have them), an explanation besides 'creeping evil fog' was required.

The map rather creates an impression of a target with a bulls-eye, doesn't it? FYI, the blue symbols scattered around the maps are public water pumps—do you notice that there's one right at the heart of the bulls-eye? From this pattern in the data, Snow put forth the reasoned conjecture (hypothesis) that it was the well (the specific blue symbol near the densest cluster of red 'death spots'), not the air, that was the source of the disease. He then designed a simple test of his hypothesis: he removed the pump handle, rendering the pump inoperable and thus requiring the locals to go elsewhere for their water. Poof! The local cholera outbreak came to an end shortly thereafter.

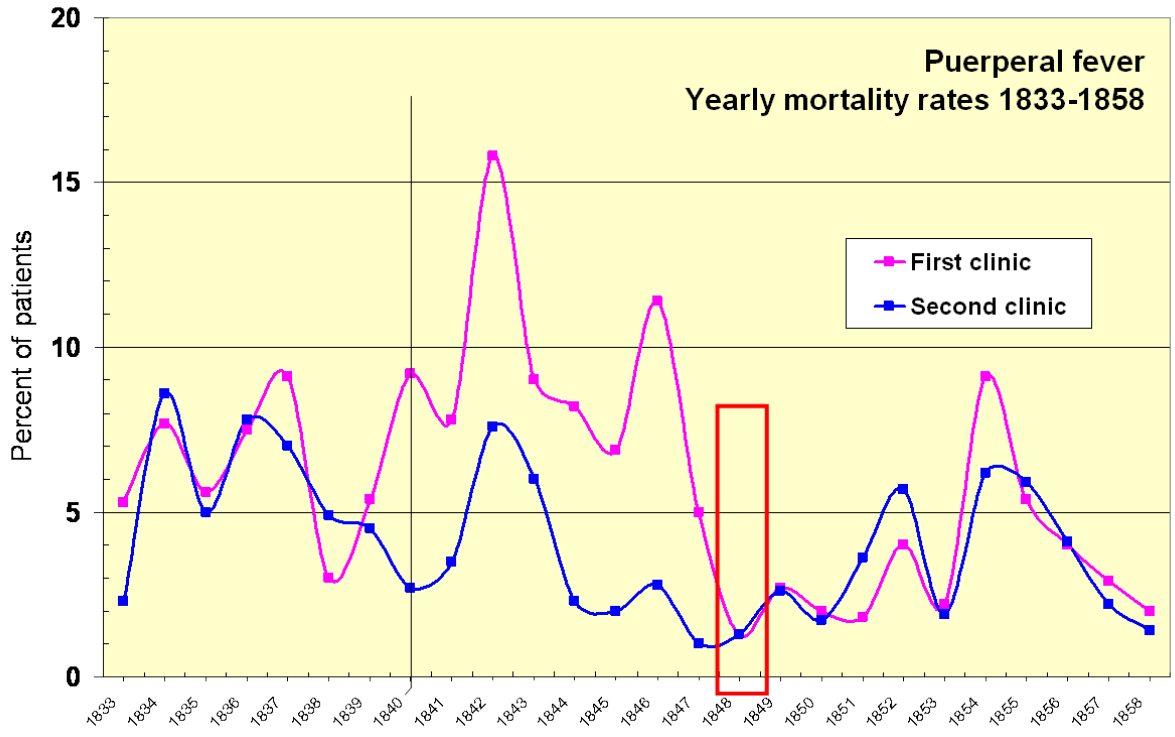
While the initial data was not experimentally generated (only observed), there is one more striking piece to the puzzle. There was one location that defied the odds and largely escaped the scourge of cholera during the outbreak. That place? The Broad Street Brewery… where

the workers consumed beer during their shifts…. Beer that was prepared in a process that included a step in which the water is boiled, thereby killing the cholera bacteria. In essence, this represents a negative control—an experiment in which the suspected cause is removed, with the expectation that the outcome will not be observed.

Did you get it? **Interactive review**

Surgeons must wash hands before childbirth

Propagating the species has always been dangerous work (for the women involved, that is), and this was particularly true in teaching hospitals in the 1900s. Enter one Ignaz Semmelweis… and another pattern in data. In this case, there were two clinics set up providing free birthing assistance… but the 'benefits' to patients were profoundly different; notably, in one ('First clinic') they died a lot:



[Image source: https://en.wikipedia.org/wiki/Historical_mortality_rates_of_puerperal_fever#/media/File:Yearly_mortality_rates_1833-1858.png]

The two clinics admitted patients on alternate days… but surely *that* couldn't be it. The fact that their work times were interleaved also ruled out issues of time-of-day, season of the year, etc. One particularly noteworthy fact: mortality between the clinics hadn't always been different; notice that prior to 1840, the death rates were largely similar. So… what changed? Prior to 1840, midwives and medical students worked at both clinics. In 1840 they were separated, with midwives attending second clinic and medical students first. Surely the training of medical students wasn't rendering them lethal! …Or was it? Semmelweis dug deeper, and found that there was also a seasonality to the increased mortality at First clinic. Mortality was peaking during periods when medical students were often working in the morgue prior to doing maternity work. Continuing to search for patterns, Semmelweis found

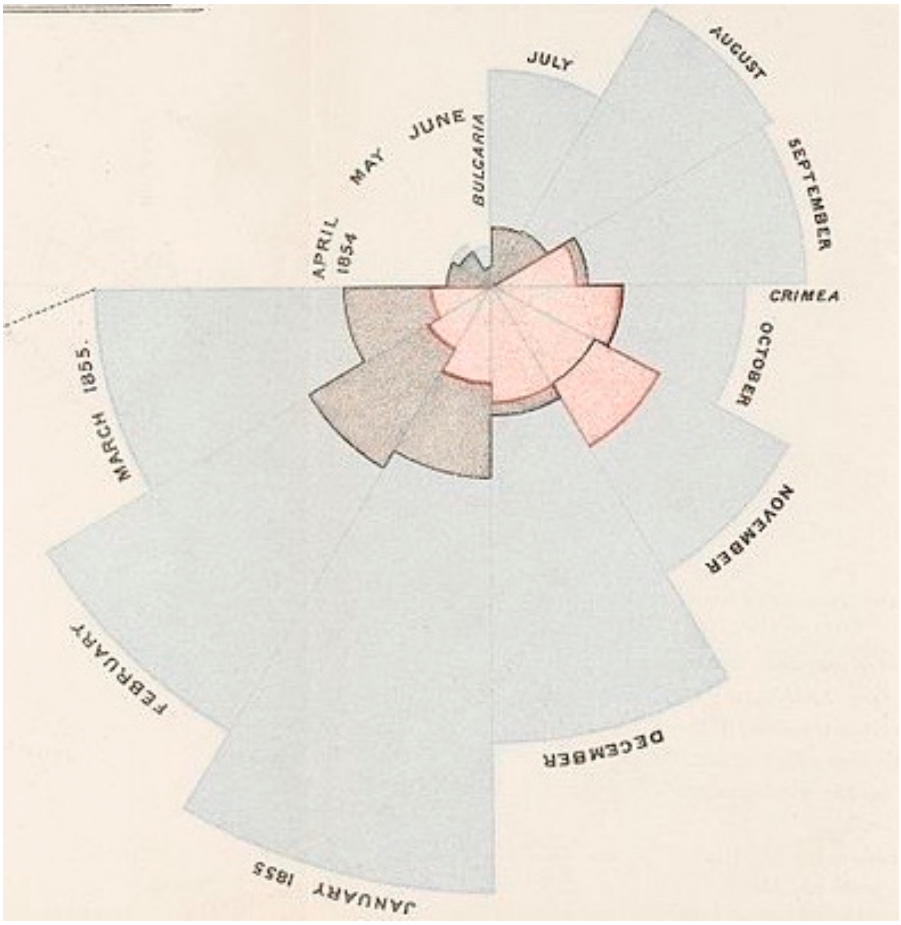
that deaths struck most heavily amongst those women that underwent extended dilation… which would increase the likelihood of extended contact with the physician's hands.

Putting the pieces together, Semmelweis proposed that students and physicians were transferring lethal agents from the cadavers to the women giving birth. This hypothesis lead to a prediction: thorough cleansing prior to contact with women giving birth should reverse the excess deaths in First clinic. Semmelweis therefore instituted a simple chlorine hand washing procedure in the middle of 1847, with the dramatic result shown in the red box above.

Alas, as with the case of scurvy preventive measures, the clear demonstration was not sufficient to change behaviors, largely because the surgeons of the time were unwilling to consider themselves as part of the problem. There's an obvious lesson here in listening for what the data is saying as opposed to hearing from it what we wish—or ignoring it altogether. For his part, Semmelweis was committed to a mental institution, and died shortly thereafter from a beating.

Did you get it? **Interactive review**

It's not the bullets that'll kill you, it's the bugs



***Needs to be updated to show the 'after' graph to scale reflecting sanitary changes

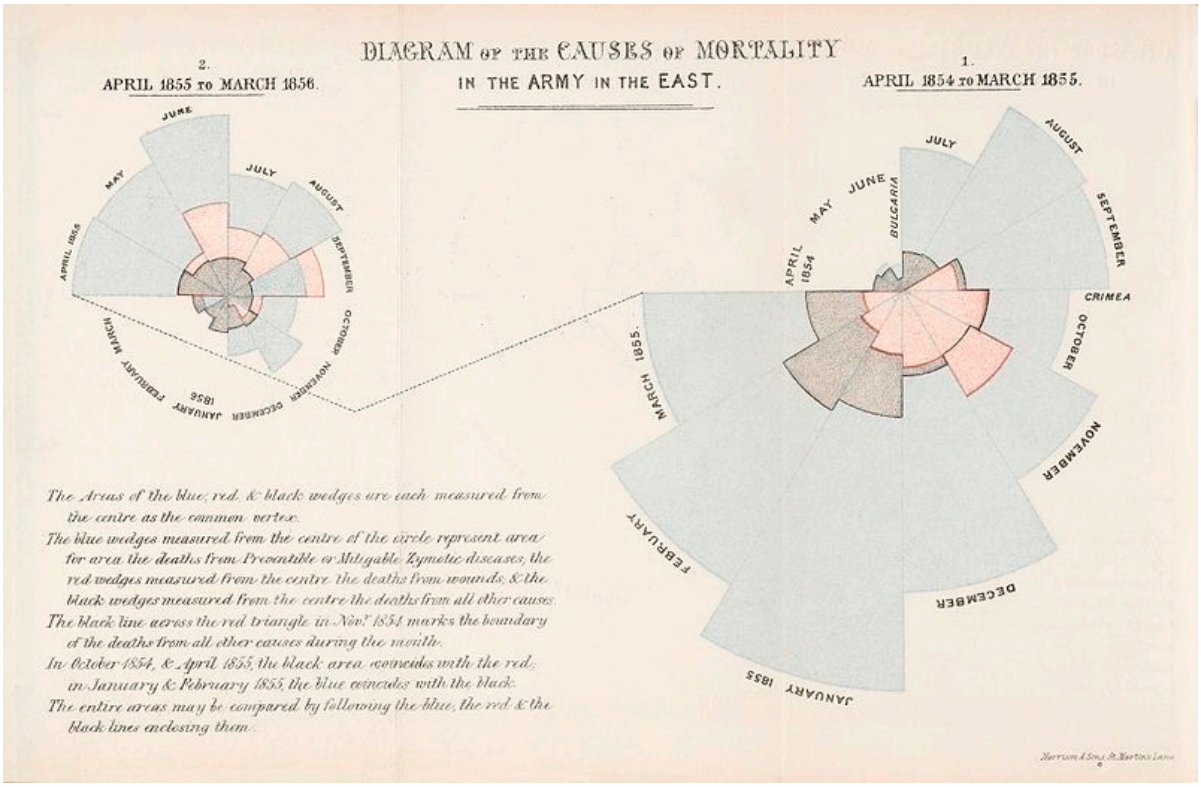
Science is more than collecting data--it's easy to have important information staring you in the face and never see it. During the Crimean war, Florence Nightingale worked in military hospitals and witnessed the suffering and incredible mortality rates there. She was determined to do something about the issues, but first she needed to understand them… and then to convince others that action was needed.

She carefully collected and categorized information on the deaths of British soldiers. In examining this information, she made several critical findings: first, soldiers weren't dying on battlefields at nearly the rates they were dying in the 'hospitals' of the time. Even in peacetime she found that the healthy young men in the army were dying at twice the rate of the rest of the citizenry.

For Nightingale, as with researchers today, knowledge alone was not enough. In order to bring about action, she had to convince others. In order "to affect thro' the Eyes what we fail to convey to the public through their word-proof ears." she created visualizations of her data that dramatically made her points. In the figure above, the innermost (orange) sections

represent deaths from wounds in battle, while the dramatically larger 'blue' (outermost) sections are deaths from disease.

The combination of Nightingale's careful investigations with her ability to persuade others through clear and dramatic representations saved thousands of lives during her lifetime and established a model for scientists that is still inspiring today. The graphic below adds data from April 1865 onwards, after some of Nightingale's recommendations on sanitation had been incorporated--quite a difference!



A modern take on Nightingale's work and relevance: "Florence Nightingale: Joy of Stats (3/6)":

https://youtu.be/yhX0OR1_Vfc

Did you get it? **Interactive review** **(not yet ready)**

[Darwin and Natural Selection]

This section incomplete

> *My reflection, when I first made myself master of the central idea of the "Origin" was "How extremely stupid not to have thought of that!" --T. H. Huxley on Darwin's "Origin of Species"*

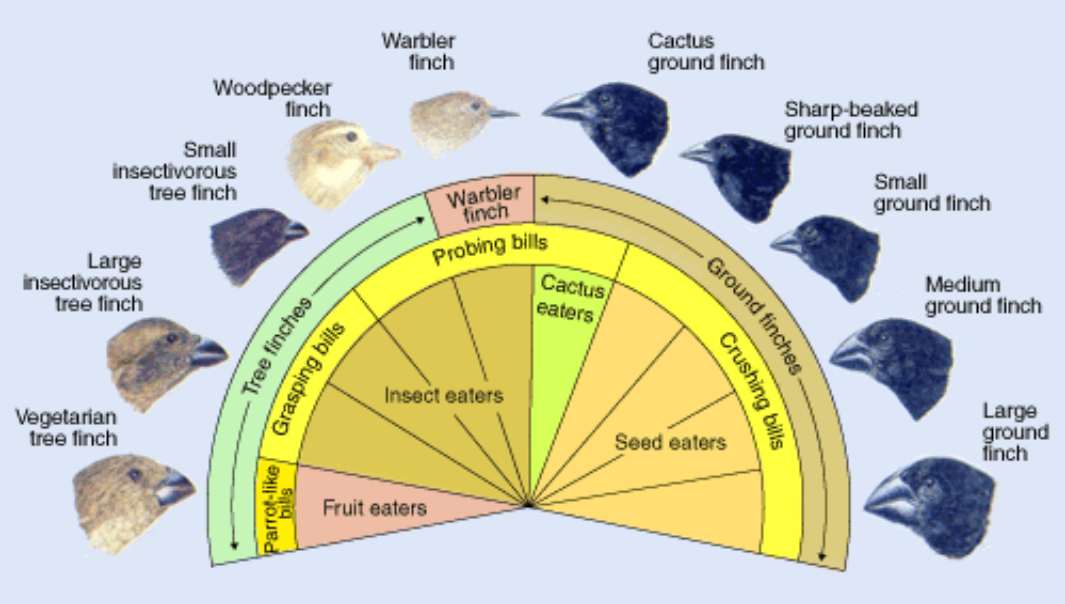


Image source: http://www.pbs.org/wgbh/evolution/library/01/6/image_pop/l_016_02.html

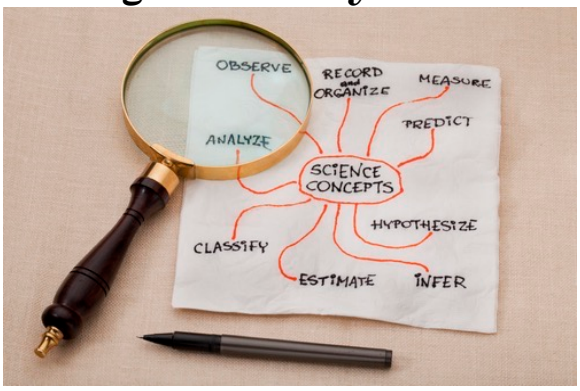
Darwin's name is synonymous with evolution not because he was the first or sole observer of ongoing change in nature, but because he found the *sense* (the 'because') behind it and compiled copious examples into a coherent theory. For all the hoopla, the underlying ideas and logic are straightforward and inescapable. The following exercise lets you work through them yourself.

A video mini-lecture of evolution

Rather than deliver a lot of words, this linked interactive exercise enables you to build the logic and assemble the key ideas yourself.

(now defined simply as the change in frequency of one or more gene variants over time)

Getting into the rhythm



If you've been following the italicized terms, you've (hopefully) noticed that several of them keep recurring--there's a certain rhythm to 'scientific' investigations and ways coming to understand things. While the path can vary, the common components are:

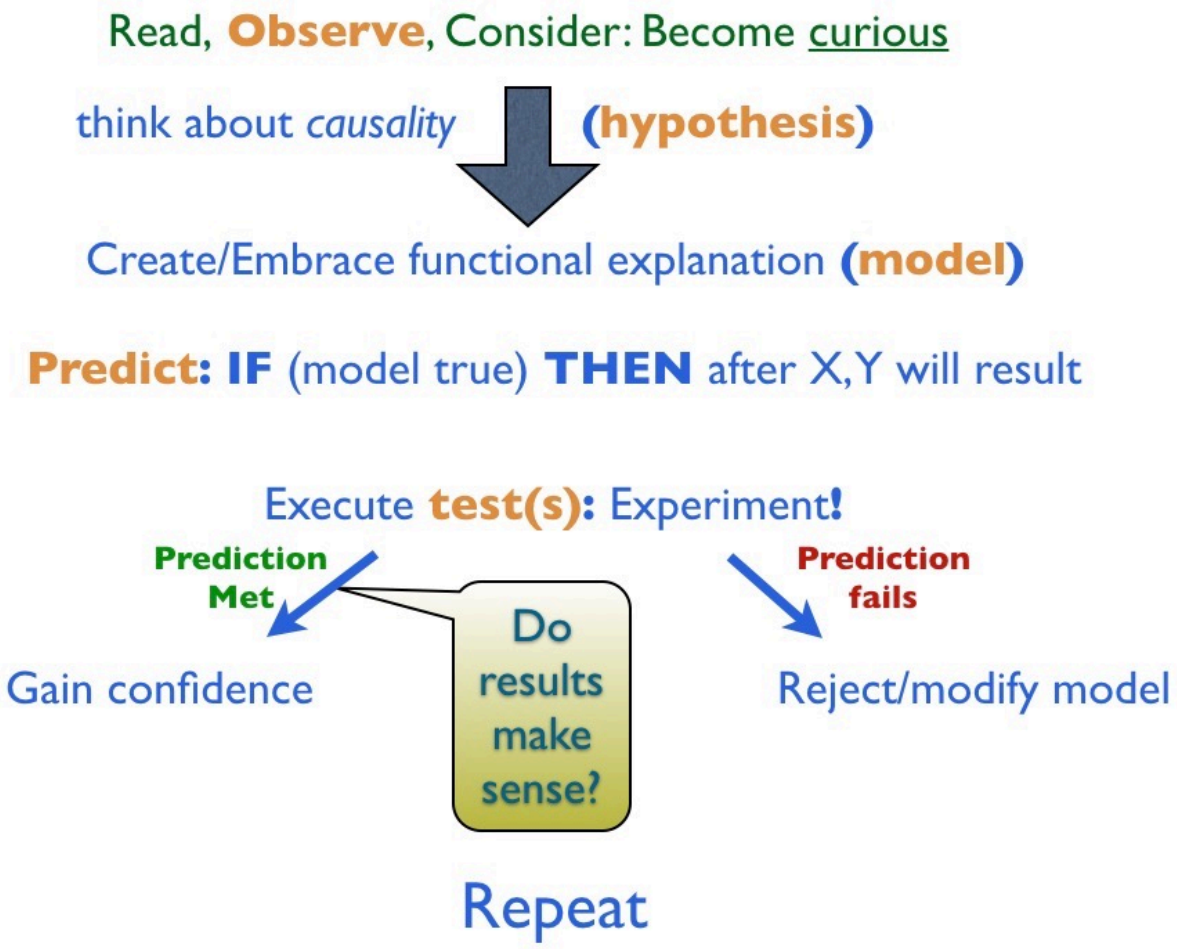
1) An *observation*: either something that doesn't fit into our existing worldview or a way of tying things together that simplifies the rules we have for the world, or highlights relationships we hadn't detected. Semmelweis and Snow saw dead people; Lind saw sailors falling apart at sea.
2) A *hypothesis*: this is really just putting the idea *and its justification* into words. Good hypotheses tend to have a <u>because</u> component--not just saying that the world IS a certain way, but providing a suggestion as to why this might be so. Lind suspected diet; Snow thought the problem was bad water, and Semmelweis looked to the behavior of surgeons. You'll often see *model* as a key component. A model is a concrete way of presenting the hypothesis--a description of the underlying reality.
3) Hindsight, as they say, is twenty-twenty. There are lots of ways to explain what has already happened; the test of whether an explanation is 'right' is whether it works frontwards as well. A *prediction* is thus the hallmark of a good model--if the model has sufficient <u>explanatory</u> <u>power</u>, then it should be able to go beyond what is already known and speak to what is knowable. Lind tried feeding some sailors limes; Snow stole a pump handle; Semmelweis implemented hand washing.
4) Of course, the whole point of the prediction is to find out if it's correct--to perform a *test*. Note that things get a little dicey in terms of terms: it's really pretty hard to separate the test from the prediction; one is the formulation; the other is the execution.
5) Repeat as necessary. The glory of science isn't that it attracts people who are always right the first time; it's that the rules of the game tend to improve the answers as time goes on. As more predictions are made and tests performed, the model may require refinement. Of course, it can even be replaced by a description that is more accurate or accounts for more observations (has more explanatory power).

No one should be lead to believe that there isn't a critical role for inspiration, insight, and creativity in science. As Albert von Szent-Györgyi (the man who purified… vitamin C!) said:

Discovery consists of seeing what everybody has seen and thinking what nobody has thought.

The one who asks the right question or recognizes the pattern or notices the 'interest' or aberrant observation is the first step in making a discovery. And testing a challenging hypothesis requires both knowledge of the tools available and insight into how they can be applied to a given question… and as you'll see in later chapters, sometimes the tools themselves need to be invented!

Word alert: besides the terms above, it's worth making sure that some others are defined. Whereas a *hypothesis* is a proposed explanation for an observation or group of related phenomena, a *theory* is a hypothesis that has survived the test of time--and the test of tests! A *law* is a rule that holds true over and over again… but isn't necessarily accompanied by an explanation (the because clause is missing). It's a reliable way of predicting what will happen, but we don't necessarily know why.



Who can you trust?

'News' is not always a reliable synonym for objective, viewpoint-free information. The term 'fake news' pops up in reference both to spurious claims and uncomfortable but factual ones. This is only the latest manifestation of an age-old question: who can you trust? A scientific approach to understanding comes with a very straightforward answer: yourself. On the other hand, you don't have time (or, often, the resources) to reproduce every experiment ever done, so some thoughtful compromises are needed. These can be ranked as concentric circles of trust. But in the end, keep in mind that the final test is yours—you should be an open-minded but skeptical observer, listener, and questioner.

At the center of the hierarchy are things you observe directly and repeatedly, with your own senses and with machines that extend them. Of course, even here, both ill-intentioned presenters and your own limitations can lead you astray. Consider the 'structure' shown below. You know it can't actually exist, because you have an understanding of 3-D structures and have likely seen illusions like it before. But there it is--which is a useful reminder about trusting even what you believe you see or know!

The next circle of 'trust' can be those whose reports have earned your respect—because your efforts at verification have born out their claims, because their demonstrations of knowledge and reasoning have established them as reliable sources. But if we stop here, then the vast store of knowledge about the world will be locked away from us forever. What can we do? Scientific communities seek to fill this void by forming self-checking groups. One of the pillars of scientific knowledge is that any claim can be checked by any participant. For a claim to be judged 'true', it must hold up to repeated challenges of logic as well as multiple tests by multiple participants. This is the principle of reproducibility. Again, the ideal would be for you to perform these tests yourself, but a realistic approach is to consider the views of other practitioners… unless and until you have some reason to doubt. Then it's time to roll up your sleeves and do your own testing.

In the end, a working understanding of the world must be a powerful explanatory framework for how things work and why they are as we observe them to be. If your worldview explains what you observe and predicts outcomes not yet known, then it is provisionally sound. If it fails, or if it grows to be complex and full of arbitrary patches, then it's likely time to attack the weak spots and seek to build something better. Examples of simplification abound—from the Periodic Table (Chapter 3) that reduced the materials of the world from an infinite number down to combinations of a few dozen core components, to Darwin's theory of Evolution, that reduced the dizzying complexity of the living world to a small number of principles driving

diversity. This realization should offer great hope to the aspiring biology student! A focus on principles can render powerful understandings (and the capability to explore in infinite directions) without cataloging an infinitude of facts. Such is the goal of this text.

Lessons learned

> *'Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house.' --Jules Henri Poincaré*

So… why embrace the scientific approach to problems and knowledge? Shortest answer: it works. Do you have a relative saved by a drug or therapy? Do you enjoy… virtually any aspect of modern society? Have you avoided dying of the bubonic plague, are you free of intestinal parasites, do you enjoy eating the fruits and vegetables bigger than thimbles, with soft outsides and small/few seeds? These benefits spring from a scientific approach to gathering and verifying knowledge.

By contrast, consider this story told by (physicist and Nobel Laureate) Richard Feynman, speaking about the distinction between science and pseudoscience (astrology, ESP, etc.):

> *"In the South Seas there is a cargo cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to imitate things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he's the controller—and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land."*

Importantly, a scientific approach to the world is not something limited to a classroom or particular "science-y" task. It's a broadly applicable way of learning and understanding. Science is not about 'authorities'—while we should all appropriately weight the conclusions of those with expertise, their conclusions are subject to our understanding and appraisal. You have a right to understand; you have a right to confirm—in some sense, a duty to do so. Tradition is meaningless, unless it springs from the concept "this has been tested many times, and always we observe expected/predicted results." This text must be held to the same standard--every attempt has been made to provide originating observations as well as the logical framework supporting conclusions. If you find yourself troubled by things that don't fully make sense, strike out on your own! By and large, scientists are amazing people--the amount of time and energy dedicated to science teaching resources on the web is astounding.

While "good natured skepticism" is a general component of scientific thinking, be particularly wary of sloppiness of language. A classical example is the term 'theory'. In everyday usage, it can mean as little as a whimsical conjecture: "I have a theory that there's a gremlin in my house who always hides my phone in the last place I'm going to look." In science, theory is a 'reserved word' that means an explanation of how things are or how things work that makes testable predictions… and whose predictions have been tested, and where outcomes are

always in keeping with predictions. While science is always undergoing examination and is subject to revision based on new information, a theory is the closest thing to a fact you're going to find.

In this light, it's interesting to consider the Pomo language. Where English modifies words based on (among other things) the time of action (past, present, future), in Pomo, suffixes are added to conspicuously convey how 'trustworthy' a given claim is. The four primary suffixes are "-le" indicating the report is hearsay; "-ink'e" meaning detected by the senses other than sight (I.e. Direct, non-visual observation), "-ine" (derived from logical reasoning), and "-ya" conveying direct knowledge. (S. Mclendon "Evidentials in Eastern Pomo with a comparative survey of the category in other Pomoan languages" in Studies in Evidentiality" , A. Aikhenvald and R. M. W. Dixon, Eds. 2003).

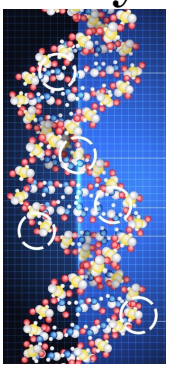
Further explorations

Talk is cheap. No one that does research wants to focus on what is already known; they want to figure things out. PatternMaster is a unique game that challenges you to discern hidden patterns (shades of the five examples previously in the chapter…). You'll find yourself repeatedly generating and testing hypotheses. A variety of patterns are there to be solved, consisting of different levels of ponderability.

PatternMaster—a game about making observations, generating models, testing them… and discovering the underlying rules of its geometrical 'universe'. The link takes you to rationale, instructions, and both web (including iPad) and downloadable Mac/Win versions.

3. Molecules and the molecular world

What you will learn



> *When a journalist asked the great physicist Richard Feynman what single sentence would best encapsulate all science so far if it were to be the sole surviving scrap of all we knew, he replied, "The world is made of atoms." —The Secret of Scent, Luca Turin, p. 28*

Understanding cells requires insights into their machinery… and all the machines are built of atoms. You can master an awful lot of biology by exploring a very small number of atoms—on the order of a half dozen—and a straightforward set of rules about how they are joined and interact.

In this chapter, you'll meet the atoms of life, learn about their key properties, and how atoms combine with each other to enable the interactions at the heart of every machine in every living organism. We'll see how the Periodic Table organizes that atoms in such a way that seeing is understanding. Equally important, we'll look at why the Periodic Table is… periodic and how that lead to its discovery. And we'll take a look at what's weird about water.

The little pieces: atoms

As early as 1829, the hints of patterns in the behavior of the *elements*--the most basic (known) units of the physical world, were starting to emerge. The clues took several forms. There were a number of 'behaviors' that researchers could characterize, including the relative weights of an individual 'unit' of the elements. What does it suggest to you on being told that…

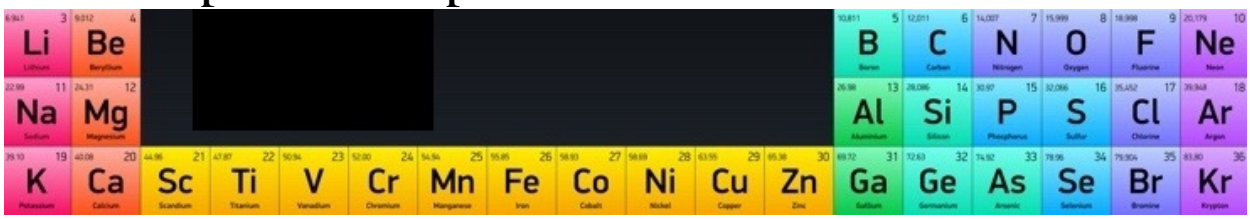
Some elements combined in a 1:1 ratio with oxygen. These were calcium (weight 40), strontium (88) and barium (137).

For other elements, two units were required to 'satisfy' one unit of oxygen… and these also had the intriguing property of exploding when mixed with water. These were lithium (7), sodium (23) and potassium (39)

Others combined with the previous group generated salts (hence their names--halo [salt] gen[give rise to] = halogens). Chlorine (35.5), bromine (80), and iodine (127).

Pause a moment here; challenge yourself to see what German Chemist Johann Dobereiner pointed out in the data. There are patterns here!

Elemental patterns: the periodic table



Did you notice that the central weight is always close to the average of the other two?

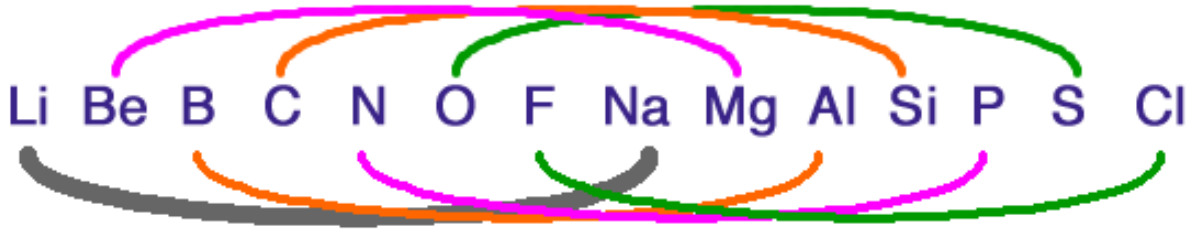
Calcium + barium is 137 + 40 = 177/2 = 88.5; (strontium weight = 88)

Lithium + potassium is 7 + 39 = 46; 46/2 = 23 (sodium weight = 23)

Chlorine + iodine is 35.5 + 127 = 162.5; 162.5/2 = 81.25 (bromine is 80)

This lead to the idea of threesomes whose properties were related and whose weights had a suggestive, but unexplained property.

But wait--there's more! Many more elements were known at the time; enough to fill in another pattern. Check this out--if lists of elements were created, it was discovered that 'related' elements (a la the properties described above) occurred at every 8th position relative to one another, i.e.:



In the image above, the gray line indicates that first element in the list, lithium (Li) shares properties with the eight, sodium (Na). Similarly, the second, beryllium (Be) has behaviors in common with the ninth, magnesium (Mg).

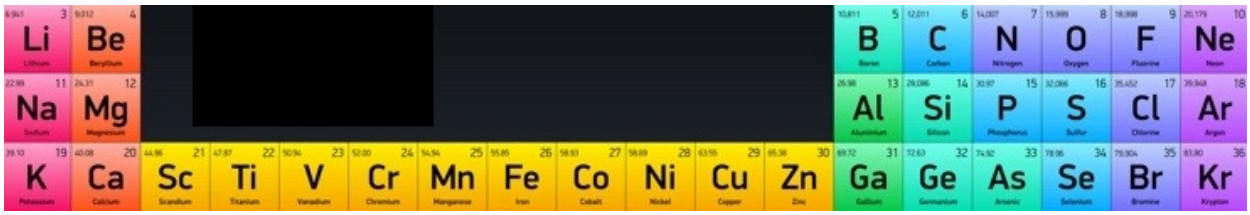
So… the triads were arising because every eight element represented some sort of 'reset' (if you're confused about the missing end-row elements neon and argon (Ne, Ar) the short answer is that neither had been discovered… and the longer answer as to why is coming. Hint: why are the elements in those columns called the noble gases?

The multiple patterning among the elements took its first concrete and useful form in the early periodic tables of Dmitri Mendeleev. But what he *did* with them is what really caught the scientific community on fire. His nascent periodic table contains 'holes'--his confidence in the patterning and relationships lead him to skip places in the table rather than to put elements in

positions where they "didn't belong", i.e. where their properties didn't align with neighbors above and below. But Mendeleev went another step. Using his periodic table as a MODEL describing the world, he made PREDICTIONS about elements *that would be discovered.* He was correct in predicting the properties and discovery of both gallium (Ga) and germanium (Ge).

But there's more. The periodic table acknowledges the relationships among elements, but doesn't provide us with insights as to *why things are as they are*. In other words, there were periodic 'laws'--descriptions that were reliable, but not explained. We now understand both the columns (different elements have similar properties) and the rows (every eighth [at the time] element in the early rows has similar properties). The reasons are explained in the next section.

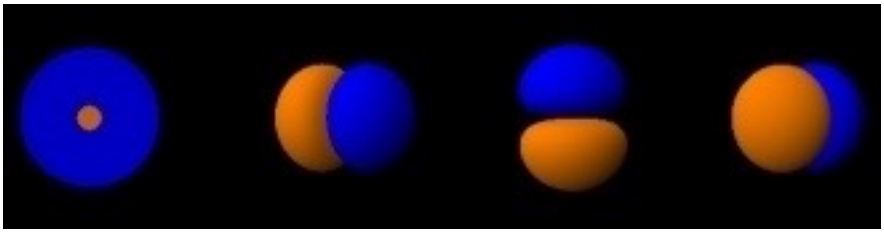
Homes for electrons



There are two clear messages in the patterning of the periodic table: first, that some elements are very similar in physical properties such as reactivity and ratios of combination. Second, that these are not scattered at random, but seem to recur with a… periodicity. When the table was first being discovered (a better term than 'invented', as there is a pattern in nature underlying the *representation* that was created), there was no deeper knowledge about 'elements' other than they appeared to be the indivisible building blocks of matter.

When physicists discovered that there were smaller units that made up elements, the patterns in the Periodic table started to become clear. An element is a particular kind of *atom*; atoms in turn consist of three kinds of building blocks: large particles at their core form the *nucleus*--positively charged *protons* and neutral *neutrons*. *Electrons* form a cloud around the nucleus--far, *far* around the nucleus. If the nucleus were the size of a pea… then an atom (as defined by nucleus + potential locations for its electrons) would be the size of a baseball stadium. Each successive element is defined by having one more proton in its nucleus than the one that came before. And in their 'raw' state, the number of electrons matches the number of protons. While the number of protons defines the 'core identity' of an atom, its electron count generates its *behavior*. So bearing in mind that there's a nucleus with an appropriate proton count, let's focus on electrons for a bit.

While it's generally convenient to imagine the electrons 'orbiting' around the nucleus, there are times when this view will prevent you from thinking well about atoms and molecules. Electrons don't really have 'locations'; there are regions around nuclei where they are more or less likely to be… and the *shapes* of these "likely places"are wonderful!



Potential locations for most of the atoms of biological molecules. The nucleus would be at the center. You see pairs of colors because electrons 'fill' these positions as pairs, with the 'home' for one shown in blue and the other in orange. (Source https://en.wikipedia.org/wiki/

Atomic_orbital)

There are strict rules for how electrons 'fill' these potential locations… and these rules generate the order in the Periodic Table! Let's look at how. *First*, as stated above, two electrons occupy each 'kind' of location (because of older thinking about electrons as 'orbiting' nuclei, these positions are still called 'orbitals'). Secondly, there are several 'types' of orbitals. In the image shown above, the circular/spherical ones are the lowest energy forms (called 's' orbitals), the right three are three different kinds of 'p' orbitals.

What's the point of going through all this? Take a look at some numbers and see if you you can spot the correspondence between patterns. Recall the data that lead up to Mendeleev's work on the periodic table: for several of the elements, properties began to repeat… every 8 elements. Look at the image above… Given that each orbital accounts for two electrons (orange and blue shapes)… how many does that leave you with?

Aha! We're getting warm here--the Periodic Table is indeed about *electrons*--the number of electrons an element 'houses' determine its properties. But wait--we're being sloppy with the numbers. If the *eighth* element is the repeat, then there are only *seven* in each row:

1-2-3-4-5-6-7

8???

But the image above suggests a complete 'set' would be eight, and if you're looking at the Periodic Table, you're seeing that todays rows (the first several) indeed hold eight. What's up?

Now historical context becomes critical. Recall that scientists of Mendeleev's day weren't working from answers at the back of the book; they were sorting through the data they had at the time. Elements were added to the list as they were discovered/purified/characterized. Could an entire column of elements have escaped notice? Absolutely--the *rightmost* elements all have a complete set of electrons… and for reasons we'll look into shortly, this means they're very, very unreactive, and their base forms are all gases. So they obey Mendeleev's rule that elements in the same column behave the same, but their common behaviors were ones that made them unlikely to be discovered! Indeed, their lack of reactivity--being 'too snooty' to interact with other elements--is what gives rise to their common name of "the Noble gases".

You may be asking yourself: "How then do we end up with more than the first row of atoms? Aren't we out of spaces at eight?". Good question. It turns out that there are multiple 'shells' of orbitals that can exist. So each row in the periodic table has a 'full set' of electrons in

orbitals matching the end of the previous row, an as well as a one-electron 'start' on another set. In other words (ignoring neutrons), sodium is everything neon is… plus a proton and one electron. Calcium is all that argon is, plus two protons and two electrons.

Periodic behavior explained: electrons

So… how does electrons-in-orbitals translate into the behavior of the elements? Everything comes down to the quest to fill orbitals--to reach a state where all the outermost electron 'homes' are occupied (minding their *p*'s and *s*'s, if you will).

Before going on, let's look into an omission that may be troubling you. You've likely noticed that we've been working with a 'headless' Periodic Table--the first row, hydrogen (H) and helium (He) haven't been included. What's up? While *almost* all atoms have those 'p' orbitals (capacity 6 electrons), hydrogen and helium are special in that they have only the 's' orbitals, capacity 2 total. So a hydrogen atom 'seeks' only one electron to be complete, and a helium atom is 'full up' with its two.

Now--how orbitals give rise to atomic behavior. Given that there is one 's' orbital in each collection and 3 'p', a 'full house' is… eight! Now the 'familial relationships' of elements in a column begin to make sense. Re-visiting the 'Nobel gases' at the right of the table, Neon, argon and krypton are 'born with' their eight electrons (i.e. in pure, elemental form a neon atom contains 10 protons and 10 electrons, an 'inner shell' of 2 electrons and an outer shell of eight), and so are content as-is. Moving one column in, fluorine (F), chlorine (Cl) and bromine are just one electron short, and all alike in this regard. In the first column, lithium (Li), sodium (Na) and potassium (K) all have a single electron and are seven shy of a complete set… or, as we shall shortly, have one 'loner' in the outermost set, and could 'fall back' to an existing full set if that one would just go away.

Those descriptions aren't just talk. This is *exactly* what is going on. Let's take table salt as an example. You've no doubt heard it referred to as sodium chloride or even 'NaCl'. It indeed consists of a mixture of sodium and chlorine atoms. Why those two? What 'works' about that pairing? Well, the Periodic Table and its insights about orbitals explains all! Remember that chlorine has seven electrons in its outer shell (it's the seventh element in its row). Chlorine needs *just one more* to finish up its outer 'shell'. What to do? Let's take a another look at that first column. Sodium, for example, needs a massive *seven* additional electrons to finish the set it has started… But what if we 'think different'? If it gave away its single outer shell electron, it would 'reveal' the full inner set (I.e. In terms of electrons, it would be just like neon). This is exactly what happens! Sodium 'gives' an electron to chlorine… disrupting the previous balance (protons = electrons) of each. Having given up one of its electrons, the sodium has a +1 charge (being the 11^{th}, element, it begins with 11 protons and electrons; now it's 11 protons, 10 electrons). Chlorine, on the other hand, gained a negatively charged electron, and sports an overall charge of -1 (18 electrons to go with its 17 protons). Now the two atoms 'like' one another on a new basis—the positive charge of the sodium units is attractive to the

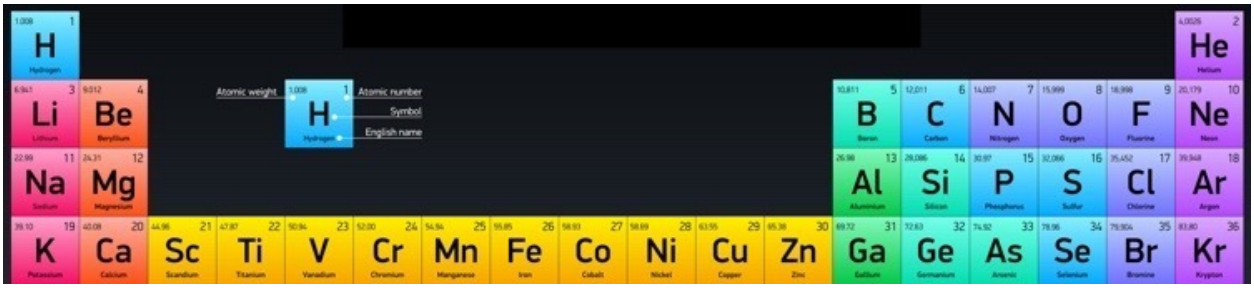
negative of chlorine.

Putting atoms together: molecules

Counting to eight: covalent bonds

Sodium and chlorine represent one extreme of atomic problem solving. Besides the give-and-take relationship these two atoms use to achieve the 'all filled' status of eight electrons in an outer shell, *sharing* can be a thing. At the atomic level, if two atoms share an electron both of them get to count it toward their eight. This is the basis of all *molecules*—assemblies of atoms held together by their electron *sharing*. Let's take a look at how it works.

Recall that hydrogen is special because as a first-row element, it's only concerned with getting to TWO electrons, and is 'born with' one. This makes two hydrogen atoms 'right' for each other--if each contributes its atom to the greater good, then an entity consisting of two hydrogens, and each hydrogen atom is satisfied because it has a part share of a pair of electrons. In a real sense of the phrase, they 'complete one another'. The relationship is relatively stable, and as such, is designated a *covalent bond*. Covalent bonds are (semi) permanent relations between two atoms that arise because of the sharing of a pair of electrons, one contributed by each participant.



This key, combined with the insights from the Periodic Table lets you figure out how many partners most of the biological atoms are seeking. You've already met chlorine, with its position one column short of perfection. This tells you it's lacking only one electron, and you've seen that it can steal one outright from chlorine. As a Periodic Table adept, you're likely already thinking that lithium and potassium could step in for sodium, whereas fluorine and bromine could play the part of chlorine. You'd be right in every case; indeed, this awareness was on of the pieces that lead to the charting of the Periodic Table.

Moving along "Biology's row", how many electrons will each of the following give or take?

__ Lithium (Li)

__ Beryllium (Be)

__ Carbon (C)

__ Nitrogen (N)

__ Oxygen (O)

Let's make further use of these powers and build a water molecule. How many electrons does an oxygen atom seek? How many does a hydrogen have to offer? *What, then, would constitute the simplest complete 'team' of oxygen and water atoms*? This would be a molecule--a collection of electron-sharers, joined each to another: H_2O.

<<<Coming: interactive 'molecule-builder' using Lewis structures and drag-and-drop>>>

While sodium and chlorine solve their problems by a

We've glossed over the striking difference between how sodium and chlorine achieve "eight-ness" vs. The way other atoms do, such as when hydrogen and oxygen cooperate to form water. What's going on here? How can we account for (and learn to *look for*) this difference? There's more to an atoms column in the Periodic Table than just how many electrons it needs to reach an outer set of eight. In a very real sense, how close a given element is to 'naturally' having eight tells us *how badly it 'wants' another*. This 'pull' or 'desire' a given element has for electrons is called *electronegativity*. A good rule of thumb is simply: the *closer* an atom is to achieving a full outer shell, the greater its greed for an additional electron is. Looking at the first eight-element row of the periodic table suggests that fluorine must be pretty intense about an additional electron, whereas carbon would be more 'meh'… and lithium almost couldn't be bothered (matching the behavior of sodium in the previous section--and with a similar outcome: LiCl forms Li+ and Cl-). This is exactly the case; indeed, fluorine is the most electron-grabbing (most *electronegative*) element in the table (elements further down its column are somewhat 'consoled' by the larger numbers of total electrons they have.

The property of electronegativity comes into play when atoms 'get together' to help solve each other's shortage of electrons. The sharing isn't always equal; indeed, it is generally only *truly* equal when two atoms of the same element share--hydrogen H2, oxygen O2, etc. Instead, the amount of time each atom spends with the shared electrons depends on their 'pull' for electrons--their electronegativity. To return to the case of water, oxygen is significantly

MORE electronegative than hydrogen.

In order take make all these electrons easier to keep track of, a wonderful graphic was developed by Gilbert Lewis, where the four sides around an elements symbol (i.e. the letter C for carbon) each represent potential positions for two electrons. Thus Neon would be represented as

More Sharing: Double And Triple Bonds

There's nothing preventing a pair of atoms from contributing more than one electron to their partnership. When each atom contributes a part of electrons, the sharing is referred to as a double bond, and signifies either as = or ::, thus CO2 is O=C=O or O::C::O. That's not the end of it, either—nitrogen gas is two stones of nitrogen joined by a *triple* bond—6 shared electrons, three from each parent atom.

The magic of carbon

You've likely heard that life on earth is 'carbon-based', or that the entire field of organic chemistry is dedicated to the study of molecules containing carbon. What features qualify carbon to be the core building block of life itself?

There are two features of carbon that help make it a good “backbone molecule” for life. Firstly, the fact that it has four electrons in its outer shell means it's long to do a lot of sharing to get a piece of four more electrons to achieve eight. In other words, each carbon atom can partner with four other atoms. Four bonds also allows for complex, branching structures (amino acid side chains), long molecules “decorated” with hydrogen or other atoms (fatty acids), rings with other groups ‘hanging off’ (bases) or collections of similar arrangements ‘packaged’ together (sugars commonly consist of four or more H-C-O-H units). On the other hand, carbon-based molecules can be small—methane (natural gas; CH4) and carbon dioxide (O=C=O) are gases.

--> Interactive: building with carbon

Another winning feature of carbon arises from its middle-of-the-road electronegativity. With this property, carbon atoms are inclined neither to give away electrons during sharing nor to steal them. One consequence of such partnerships is that it takes a moderate amount of input energy to pull the participants apart—the bonds are relatively strong.

Thus, carbon’s qualifications for the core building material of life are the ability to contribute

or complex structures, and the fact that constructions made with carbon cores don't fall apart too readily.

Molecules

Water, the weird one

 <table style="position: absolute; width: 200px; height: 80px; left: 50px; top: 58px;"><tr><td style="text-align: center; color: #000; font-size: 24px; font-family: Arial,sans-serif;">Click to Run</td></tr></table>

Oils: anti-water

Who does what

What you will learn

Pinning down what 'life' is can be challenging. Broadly speaking, most modern definitions include concepts of "doing things" (changing or interacting with energy sources and environment), making copies of oneself, and evolving--giving rise to descendants that are not perfect copies (and that therefore can be improvements or represent abilities useful in new places). In this chapter, you'll come to understand some of the insightful experiments that turned our thinking away from proteins-as-information and crowned DNA as the genetic material. Protein fans should not be dismayed--they reign supreme as both building materials and diverse and precise doers of the cell.

A chicken in every pot

A chicken in every pot: Hershey, Chase and distinguishing DNA from Protein

What are 'instructions' made of

What are 'instructions' made of? Turning nice bacteria into killers

4. Proteins: Life's machines

What you will learn

Almost all of the tasks in your cells are carried out by protein machines—and all tens of thousands of them are assembled using the same 20 pieces. These same units are used by all living things we've investigated. Each protein is assembled as a linear string of these building blocks. As you'll learn later, a single factory does all the assembly, and the strings form themselves into their functional, 3-D shapes! To facilitate assembly, all twenty building blocks have the same 'chassis' bearing a 'side chain' consisting of a small number of atoms—chosen from carbon, hydrogen, nitrogen, oxygen and sulfur.

But a second critical level of 'assembly' must take place. All proteins are *chemically* linear chains of amino acids. But as they are synthesized in water, something magical happens—they *form themselves* into their functional, 3-dimensional shapes. How this happens, and the biological necessity for this magic trick will be laid out for you.

These principles will be made concrete as we look at one of the workhorses of your body—*hemoglobin*, which carries out the paradoxical task of *grabbing* oxygen when it's in your lungs and *letting it go* when in your tissues and muscles. The carrier itself remains the same; how can it behave oppositely in those two locations? And how does a dumb chain of atoms 'know' what to do when? All shall be revealed!

From string to sculpture: protein folding

Can you do it alone

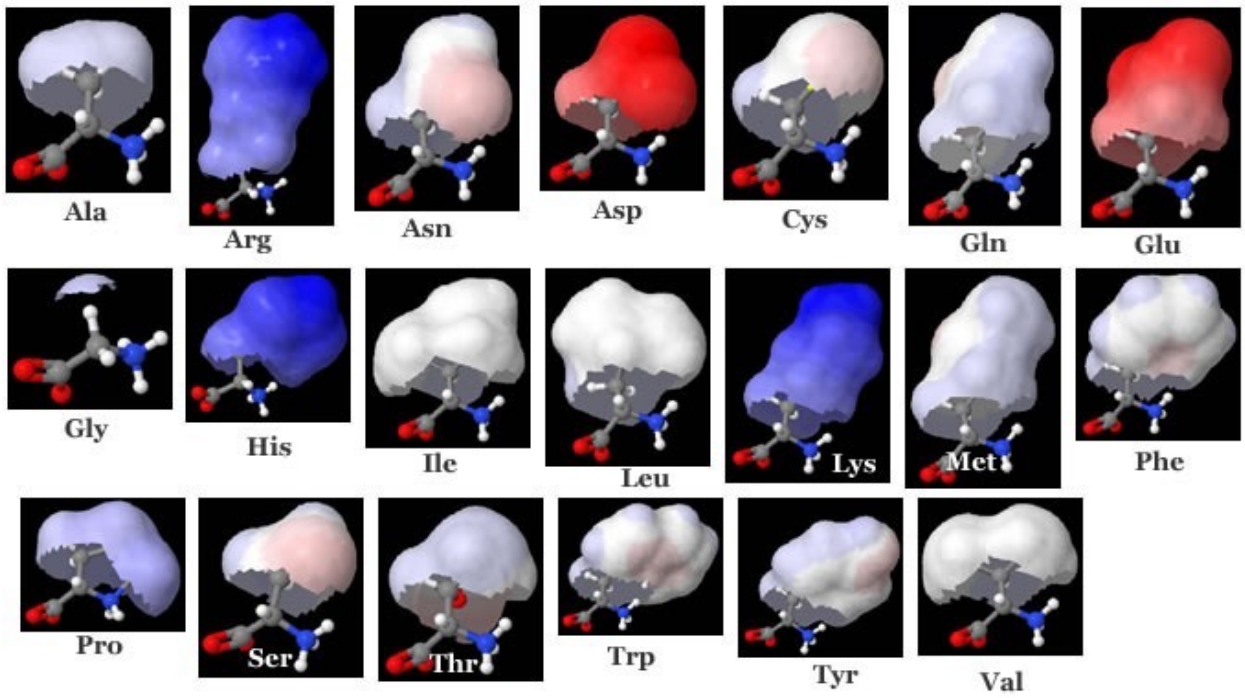
Can you do it alone? 'Scrambled' ribonuclease sorts itself out

Parts list: the amino acids

Meet the amino acids

If you want to build something interesting, you're almost always going to need a variety of pieces. This is particularly true if you want your product to *do* something. In chapter 3, we looked at how combining atoms in different ways can give rise to molecules with different surface properties (neutral, partially charged). Amino acids are simply twenty specific collections of atoms whose shapes, sizes, and properties constitute the construction kit for building virtually all of life's machines. And since everything that happens in your cells is done by one machine or another, understanding the pieces is critical.

The image below shows the twenty amino acids found in every living thing ever characterized on the planet earth. As with the nucleotides, their names aren't always informative, but we must call them something. For now, the three-letter abbreviations are shown; these are often used simply to save space when speaking of many amino acids. A typical cellular machine is a string of as few as 100 amino acids or many thousands. The key, of course, is *which ones* they are. As introduced in the previous chapter, the amino acids are colored in keeping with their key surface property: charge. White indicates an absence of charge (neutral) and various shades of red and blue correspond to amount of negative and positive charge, respectively.



Hydrophobes: hiding from water

Turning corners: Gly and Pro

Charges

Building bridges: Cys

sensing pH: His

5. A machine for doing: ATPase

What you will learn

In Chapter 4, you got acquainted with the amino acids and the general notion that you could 'build anything' with them. Alright then—let's build something! In this chapter, you'll look at an example of the kind of machine that 'harvests' the ubiquitous fuel packets floating around the cell, molecules known as ATP—adenosine triphosphate. We'll start by looking at where this 'energy' is in an ATP molecule, then consider why it doesn't just explode on its own, and then look at how the correct 3D array of amino acids can harvest the in a useful way.

How the cell pays for things: ATP

6. DNA and RNA: Life's Instructions

Overview: Life's instructions

Rather than believe that Watson and Crick made the DNA structure, I would rather stress that the structure made Watson and Crick... what I think is overlooked... is the intrinsic beauty of the DNA double helix. It is the molecule which has style, quite as much as the scientists. —Francis H.C. Crick

DNA (RNA) and information: overview

The wonderful machines introduced in chapters 4 & 5 must come from somewhere; despite their remarkable capabilities, none of them are contain the instructions for making themselves.What do the instructions look like? And how is it that the instructions are capable of being passed to each cell during the process of making you (100 trillion cells) from the single fertilized egg you once were? And lastly (though reserved for chapter 7) how are the abstract instructions interpreted and implemented as construction? The answer to these questions lies in the chemical structure of your genetic material, DNA... and if we dig far enough, we can even begin to think about how life itself began... by looking at the instructions each of us still bears.

What you'll learn

1) The Big Idea is that a chemical consisting of four simple components can 'mean'

something... and a string of these can in and of themselves 'instruct' and attract a partner string.

2) These features make copying and sharing the 'information' (specific sequence of units) trivial... allowing for reproduction as well as sending instructions from the storage domain (nucleus) to the building domain (protein-making machines

3) The instructions are easy to read, and allow us to observe and compare the specifications for any organisms known

4) The instructions are vulnerable to chemical changes... which directly change their meaning

Chemical codes: how molecules 'mean' something

Introduction: Chemical codes and 'meaning'

As recently as the 1800s, even scientists studying life believed there must be something magical separating it from non-life, a 'vital force' that animated living things and so distinguished them from mere matter. By asking the question in this way, they placed the answer out of reach. While right to wonder how a small number of atoms (primarily carbon, hydrogen, nitrogen, oxygen, phosphorous and dwindling amounts of a few others) combine to give rise to a mere handful of properties (varying degrees of positive and negative charges as well as non-charge) could give rise to form and function, the error was not to consider "what can such simple things achieve?". We'll start there.

Handshaking: information-as-matching

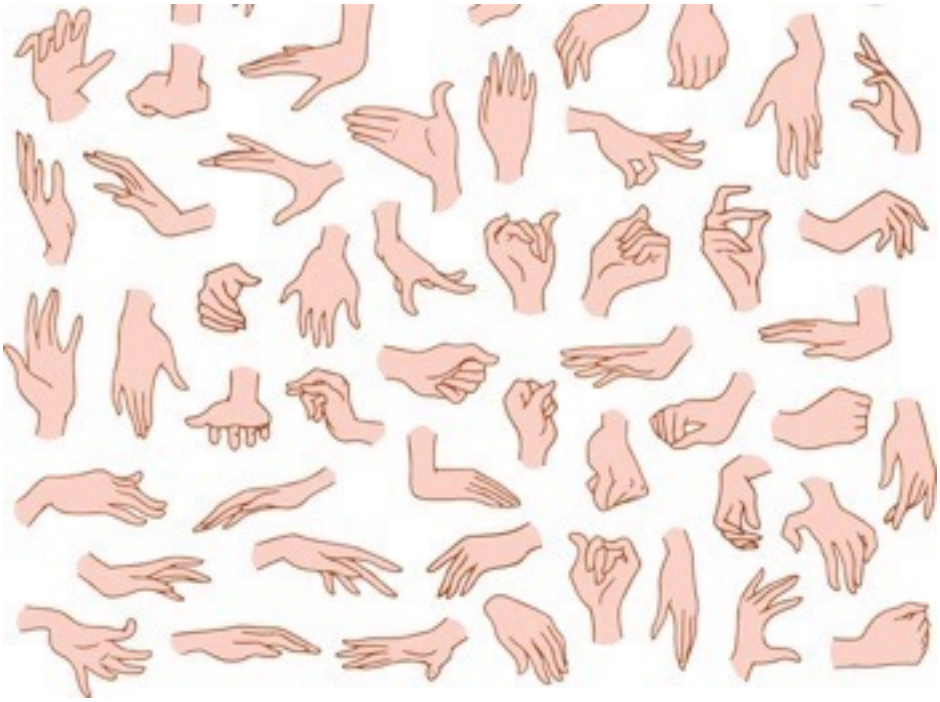
The first challenge is to consider what the properties of a molecule that 'means something' must be. If we are to have 'instructions' for a human body, or for behavior, or even for making a simple protein, then there must be a way of constructing a molecule such that it can 'communicate' something. The first (and most important) communication that we'll consider is "who is my partner?".

Consider a handshake. Imagine yourself walking through, shaking hands with the strangers that you meet.

You've never seen these people before, they come in all shapes and sizes... but are your palms sweaty contemplating the humiliating possibility that you'll fail in this simple courtesy? Likely not. A handshake is a very adaptive, general interaction between two participants. But there's a dark side to that flexibility. Suppose it was revealed that the dark-haired, mustachioed gentleman whose hand you shook midway through the exercise was a millionaire. And he'd give you a million dollars if you could recognize him by going through the handshaking

process again... blindfolded. What are your chances of recognizing him this time?

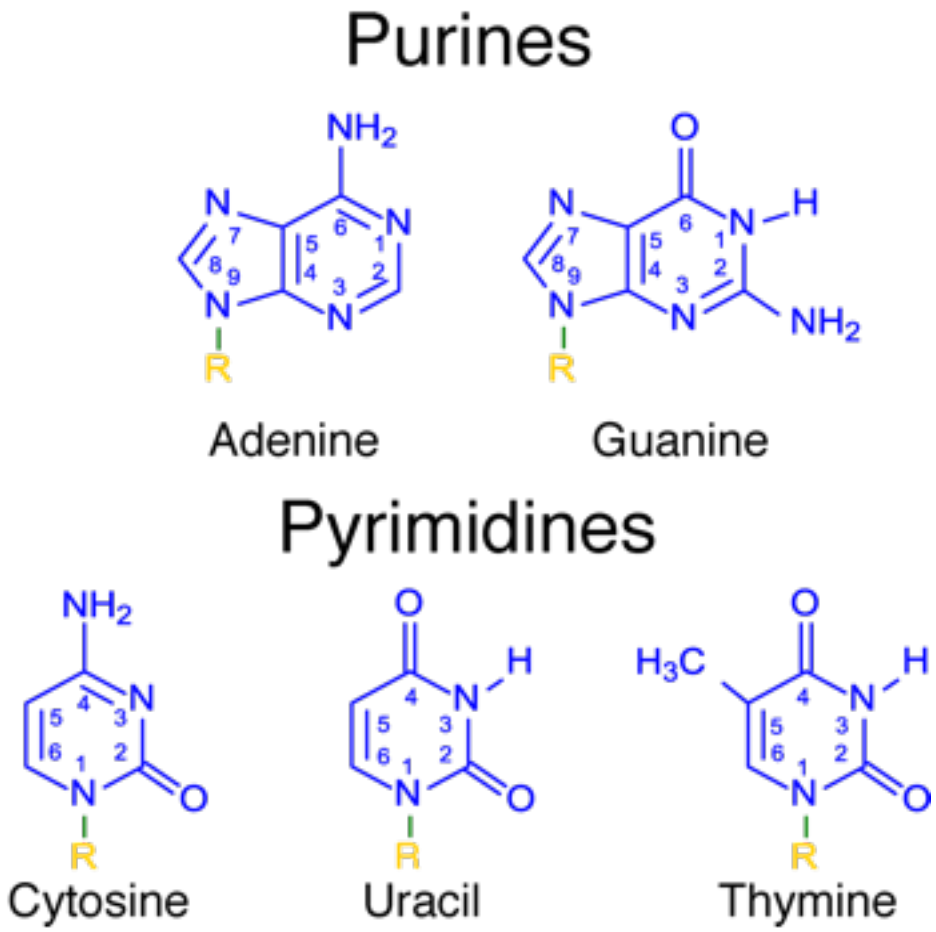
That's the essence of the challenge we face at the molecular level as well: handshakes, being general and adaptable, contain very little information about the partners involved. The process is 'successful' no matter who is involved. Before looking for a molecule with better properties, consider what would be required to improve the handshake. Take a moment and imagine yourself a member of a secret society. You are tasked with developing the society's secret handshake. It can't be a flamboyant affair--no flapping about or squawking like birds. In the quick interaction, you must be able to recognize fellow members from the rest of society--and they must recognize you. If (in real life) you've a friend nearby, even better. Jot down the characteristics of your handshake, both in detail as well as the PRINCIPLES or characteristics that you think would be hallmarks of most secret handshakes.



In all likelihood, in examining your handshake you found it to be structurally rigid, and that it 'interacted' with your shaking partner in some form-fitting way--there was a complementarity to the two structures. You're well on your way to describing the kind of molecule that is going to be the keeper of the genetic information. Now let's add a somewhat surprising piece. Consider Shaquille O'Neal. He strides into the Room of the Secret Handshaking and comes up to you with your carefully structured hand... and he wants to shake it. Given his size and strength, do you think he succeeds? Absolutely. Hand shaken. This little thought experiment reveals the third critical feature of our informational molecule: the interactions between partners are going to be WEAK so that each of the 'real' ones is vitally important, and no part

of the vital interaction is 'drowned out' by some overwhelming component.

Congratulations! You've just generated a very accurate description of the key features of DNA (and its shoddy, short-lived cousin, RNA). The acronym expands to deoxyribonucleic acid, and we'll deconstruct the syllable string as its components become relevant. The key components that match your description of handshaking are the 'bases' (caution: while the bases are basic in the pH sense of the word, that association really isn't relevant here; 'base' has simply become the term we're stuck with for these chemical entities).



The bases (including uracil, from RNA) [source:https://commons.wikimedia.org/wiki/File:Nucleotides.png]

The key features you called for will be the focus of the coming sections.

Rigidity: Notice that each of the bases shown above consists of a carbon 'ring'--because each carbon is linked to two neighbors, there's nowhere for it to go. Further, the double lines between some of the carbons represent double bonds (2 shared pairs of electrons), and their consequence in this particular case is to lock down the structure so it's flat as a pancake and unable to 'wriggle' into any other configuration. In this way, the glowing balls are always presented in exactly the same spatial configuration--exactly as your secret handshake is.

Word alert: Note that there are two classes of base: double-ringed ones (guanine, adenine) and single-ringed ones (cytosine, uracil, thymine). The former are 'purines', the latter 'pyrimidines'.

Useful mnemonic: PUR As Gold: Purines are Ade, Gua.

Base matching is more detailed and is explored in the next section

The basis of molecular interactions

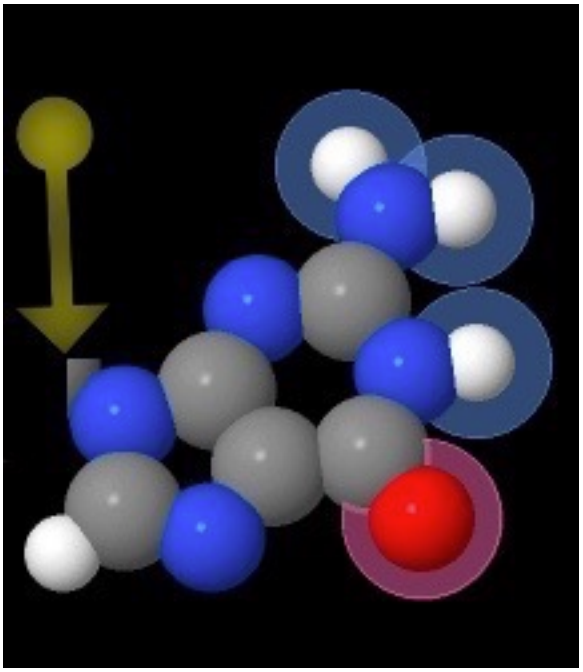
Molecular surfaces: the basis for specific matching

In order to dive in to complementary matching, we need to review some coloring conventions; molecular matching derives from the surface properties of molecules... which arise from the properties of the joined atoms within.

Review: atom, charge colors

One of the most important bits of information 'written' in the patterns of the Periodic Table is the degree of "electron greed" (electronegativity) of different atoms. To understand the surface-matching that goes on between bases in cells, it's vital to think of them not as bunches of colored balls, but as surfaces with properties. For a quick review, let's apply the principles of electronegativity to the 'handshaking' surface of a base:

Interactive: ID donors/acceptors exercise



The base guanine. Arrow indicates the point where it attaches to parts we'll discuss later; for now, notice the charges: partial positive (blue) and negative (red). These are indicated ONLY on the part of the base that engages in handshaking.

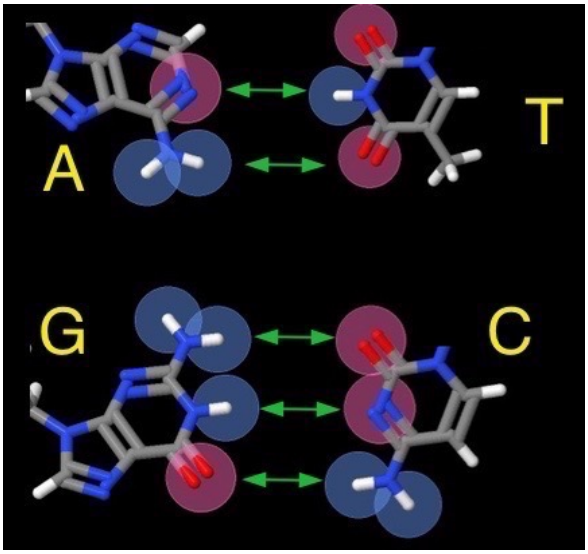
The critical element is the glowing spheres, which you'll recognize from the previous chapter as hydrogen, oxygen and nitrogen atoms. Only the subset that are involved in 'handshaking' between two bases (basepairing) have been indicated in this way. If your chemical analysis skills are engaged, you'll realize that each of the glowing atoms is either partially positively charged or partially negatively charged. This achieves two more of our goals: complementarity (if one partner is (+)(-)(+), the other must be (-)(+)(-)) and weak interactions--these are all of hydrogen bonds, whose partial charges must compete with water to form, and aren't 'worth much' under most biological circumstances. So every one must count--only the 'perfect' partner, with no conflicts and making several good interactions will 'successfully' shake hands.

Word alert! "Complimentary" means you say nice things about someone or something. "Complementary" means there is something about one element that fits with or is somehow appropriately matched to the other.

Molecular hands shaking: two pair

Complementarity of the four bases; green arrows indicate favorable partial charge pairings;

the two matched pairs found in DNA (Ade + Thy; Gua + Cyt) are shown below:



Important things to notice in the figure:

•The bases are made of four common, familiar atoms (carbon = gray; hydrogen = white; oxygen = red; nitrogen = blue)

•The 'pairing face' of each base is 'decorated' with partial charges, with (+) shown as blue halo and (-) shown as red halo (while this convention is annoying given the atom colorings, it's from the Real World)

•Each base 'pairs up' perfectly with its partner (blue halos across from red halos) (Note: adenine has a 'dead spot' because the hydrogen attached to a carbon is 'too short' to bridge the gap and lacks a charge)

•Two of the bases are larger (A, G; these are the purines) and two smaller (C, T; the pyrimidines)

•Each pair consists of a purine and a pyrimidine (G:::C; A::T [in this text, you'll always see G:::C with 3 colons to remind you of its three hydrogen bonds]). The vital consequence is that while the members of a given pair are different in size, each formed pair is identical in size.

Principle: The units of DNA 'recognize' their partners through rigid, direct, chemically complementary contacts

Word alert: The terms 'purine' and 'pyrimidine' aren't immediately useful, but we'll have need

of making the distinction periodically. One handy mnemonic: *PUre Ad Gold*: the PURines are Ade and Gua.

As you may have learned elsewhere, the full names of the bases are adenine, thymine, guanine, and cytosine. Click the link below for pronunciations.

Audio: Base names

Alas, there is no profit in trying to remember the names by poring over their structures or searching out their deep meanings; adenine is named by virtue of having been isolated from the pancreas (the Greek root for 'gland' is aden-), thymine from the thymus. Cytosine was pulled from the cytosol of some cells, and poor guanine? Originally derived from guano.

The reality of the matching of bases can be hard to capture even with fancy 3D images. The following video demonstrates the principle in a more intuitive way: using shapes that bear magnets on their interacting surfaces. Just as the partial (+) (blue tape) and (-) (red tape) charges on the bases attract, the opposite poles of the magnet do so as well. The key here is that only partners that are complementary in both shape and polarity (mimicking charge) actually succeed in 'handshaking'.

[To be created: a simple video using magnets marked with colored tape/nail polish and displayed in 'faces' of three; demonstrate how perfectly complementary faces 'snap' together readily, even when released 1 or more centimeters away. They don't 'fit' when askew, and when one of the partners is changed to a non-match, the two are disinterested]

Interactive: basepair exploration tool

Both science and learning are fundamentally based on the idea that individual investigation and verification are core values. In order to enable your investigation of claims made about the bases, this eText has incorporated a modeling tool called BasePairer. With it, you can directly examine the structure of the bases, their potential interactions, the changes that give rise to mutations, and even think about life on other worlds. You'll get a chance at each of these in this chapter or beyond. As with all tools, it's first important to take a look at how it works. The short video below will acquaint you with basic capabilities and the meaning of different symbols.

Video: URL https://thinkbio.guru/HowTos/BasePairer_Intro/media/BasePairer_Intro.mp4

Let's go back and pick up an important idea that we glossed over previously. Click the link

below to open the BasePairer tool. It will open showing the purine guanine on the left and the pyrimidine cytosine on the right. Initially, they are shown in identical orientations. To simplify things, the atom colors have been suppressed and the charges are darkened.

Interactive: pairing guanine with cytosine

Questions: Manipulating cytosine to pair with guanine

Compare: nucleotides v. amino acids

Looking back: biomolecules may not be as different as you'd think

Remember back in our protein days, we discussed how the amino acids each contained a unique part joined to a stereotypical base... with the important consequence that each could provide unique functionality, but could be approached and operated on in the same way (generating the peptide bond between them). In order to allow you to see the correspondences, let's review the elements of an amino acid.

Scientists can miss the obvious too

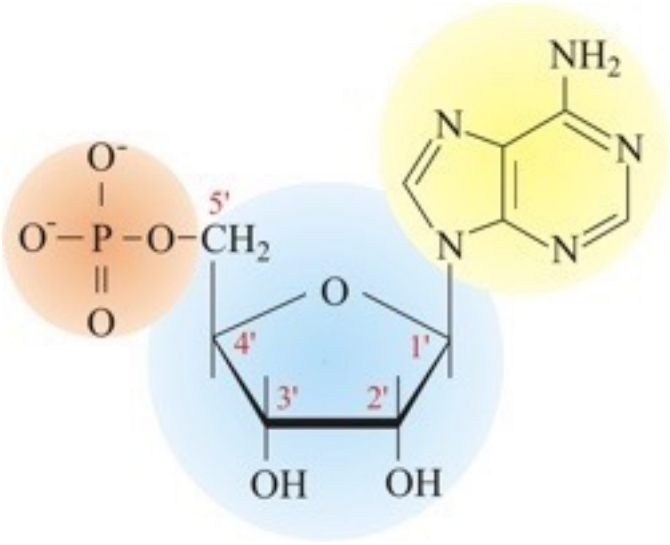
At the beginning of the last century, no one even knew which molecule stored your genetic information. While it's always easy in hindsight to pick the winner in a horserace, it's also instructive to learn from the mistakes of others so one's own thinking is more sound. One of the arguments against DNA as the genetic material was that it did not contain enough 'variety' to have much to say. It was known that DNA consisted of four different 'building blocks' (the bases you've seen, attached via pieces you'll meet below), whereas proteins drew from an 'alphabet' of twenty amino acids. While it's certainly true that the complexity (number of combinations) of a string of amino acids rises faster than that of a string of bases, if you just 'go further' (add more units) you can reach any amount of distinctness (information) you want with either code... or a code of just two letters (Morse code, for example!).

Questions: How many words (3 letters, 4 choices => 64)

The instructions in each of your cells run to ~3,000,000,000 basepairs, strung together into 46 chromosomes--(more concretely: you've got approximately one meter of DNA in every cell of your body!). At this point, it's not really known how much of that potential information actually 'says something'. Current accounting is that your critical genetic information is chunked into about twenty thousand genes, and that perhaps as much as 98.5% of the potential information in your cells has nothing valuable to say. Such judgements, however, are subject to change as we learn more; whole classes of functional molecules are still being discovered. So our current inability to recognize a given sequence as 'important' should not be considered a definitive judgement on its value.

DNA sentences: strings of bases

In any event, if all your information were floating around free, it wouldn't mean anything--much as a book would lose value if you cut out each letter and poured the whole into a shoebox. It's a particular linear sequence that holds the key. The bases are strung together by virtue of each of them being attached to two further elements--a sugar called 'ribose' and a phosphorous bound to several oxygens--a 'phosphate'. The combination of base, sugar and phosphate is termed a nucleotide.



Source: http://www.periodni.com/gallery/nucleotide.png

The numbering is inescapable for some of our later discussions, so let's dispense with it now. The Rules of Organic Chemistry dictate that the biggest 'chunk' of a molecule gets ordinary numbers (the base in this case, though the numbers are not shown). The next group is numbered using the 'prime' symbol (1', 2', 3'...). In this case, the phosphate is attached to the 5' carbon of the sugar. The 'next' nucleotide in the string would be attached to the oxygen which is attached to the 3' carbon. There would be another phosphate, which in turn would be attached to the 5' carbon comparable to the one shown.

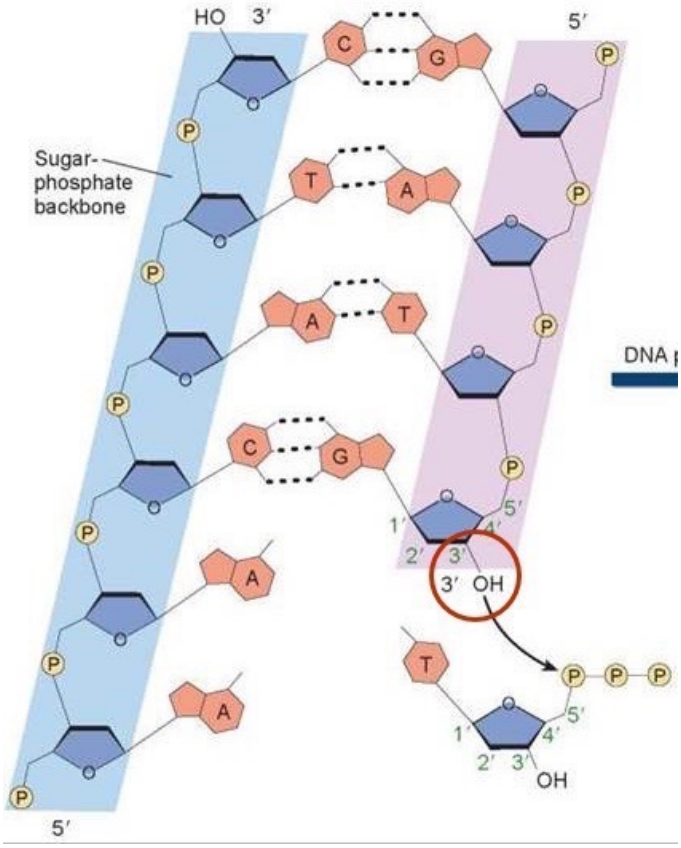
Prime numbers

In the image below, click on the carbon that would be numbered 4'

Connections: Chemistry of joining

The mechanism of attachment of a new nucleotide to an existing chain uses the same

chemistry we examined in the chapter on 'spending' ATP. Up very close, the chemistry is identical--the electrons on a 'hungry' oxygen 'go after' a phosphorous. The differences are in detail. The -OH group, instead of being part of water (i.e., the 'rest of the molecule' is a hydrogen H-OH) is the 3'-OH of the ribose of a nucleotide on the EXISTING molecule, and instead of attacking the outermost phosphate, it's the innermost one that is the target:



Will be replaced shortly; NOT permissioned; source (copied by source?) http://schoolbag.info/biology/living/living.files/image313.jpg

Key features: labeling of positions, shows -OH and phosphates. Needs to be changed: the incoming base should be PAIRED, but not JOINED

At the end of the reaction, instead of the terminal phosphate being 'joined' to water (it 'looks' like it has 'broken off' but in reality it acquired an -OH from water), in this new case the joining is to the 3'-OH that attacked. So a covalent bond is created between the 3'-OH of the existing chain and the 5' position of the newcomer. Since the 3' end of the existing chain is

now longer, we speak the synthesis of DNA as proceeding from 5' to 3'.

Looking: a short DNA sequence

Getting specific: stringing lots of 'letters' together

While the bases are often abbreviated A, T, G , C it's important to keep in mind that the structures you have been looking at show what's really important. Just as each of the four bases presents a unique surface (in terms of charges and positioning), a longer string gets more unique (and more 'informational') with each addition. Let's take a look at a string of nucleotides in preparation to looking at some of the amazing consequence this specificity offers.

Keep in mind at all times that the matching between the bases is shape- and charge-based; it's not like our operation of a code where an intelligence is required to look up a letter, find the correct match according to a table, and write down the new letter. In a very real (though weak!) sense, if an adenine is present, a thymine 'wandering by' will stick to it momentarily via interactions on the basepairing face of each. When a string of nucleotides of a given sequence 'bumps into' its perfect complement, suddenly we're talking about a lot of interactions... and a fair amount of 'sticky' holding them together. A perfectly matched pair of chains, each only 20 nucleotides long, would easily 'stick together' up to a temperature of 50-60°C (you may know that your body temperature is 37° on the centigrade (°C) scale.

There's a second consequence as well. Given that there are four possible DNA 'letters' (bases) at each position, the likelihood of a creating given sequence randomly becomes a fourth as likely for each letter added. Thus a specific two-nucleotide sequence (such as G->C) occurs only once out of every randomly selected pair. Similarly, a given sequence of 17 nucleotides likely occurs only once throughout your genetic instructions... But wait! There's more! An experimenter could design the complementary sequence (as easy as "A with T, G with C") and it will 'stick' specifically to its partner sequence if the two encounter each other. There are a lot of important and amazing consequences of this matching and we'll take a look at several of them... after we consider the structure of genetic information in your cells.

Principle: A series of bases (joined in the form of nucleotides) constitutes a unique pattern of information as well as a unique chemical surface

As anyone whose computer ever 'ate their homework' knows, any document of value should be backed up somewhere. This safety measure is also embedded right in the structure of your

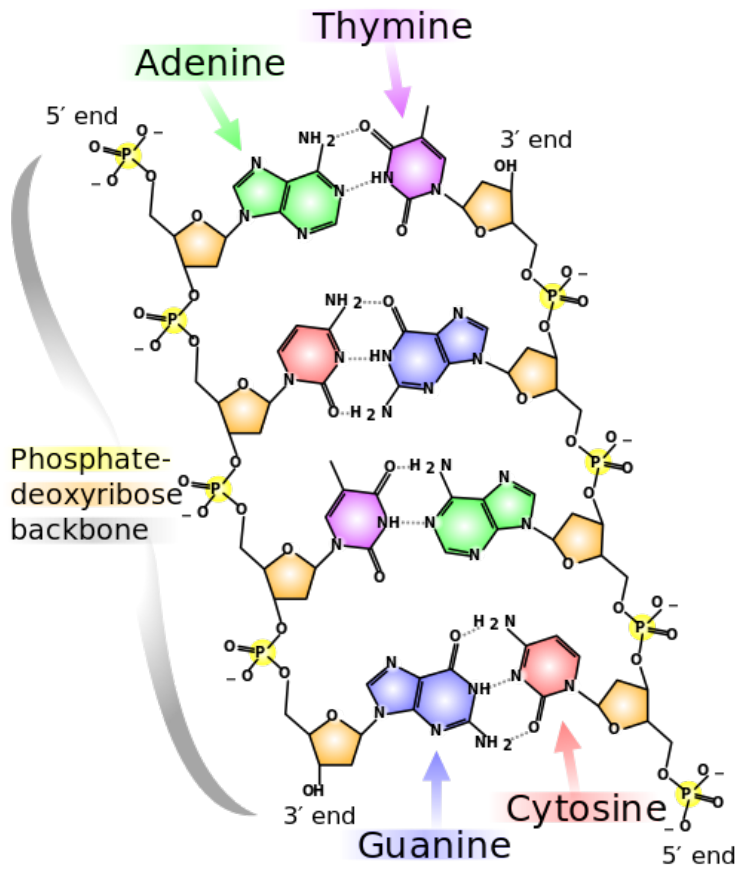
genetic material. The fact of complementary partnering between bases causes two matching DNA strands to 'zip' together to form a long, ladder-like structure... with a twist! In order to optimize interactions between each partnered pair of bases (a basepair), the ladder twists about its centerline, forming a helix... or rather a double helix, with each strand of DNA bases forming one of the twisting sides of the latter.

Pairing faces of a string of bases

Remember earlier when we established that in order to pair guanine and cytosine, one must be 'flipped over' relative to the other? This is where that difference manifests. An arrow indicating the direction of progression from one base to the next on one strand would point in the opposite direction to the corresponding arrow on the other strand. Technically, the two strands are said to be 'antiparallel'. In this textbook, when presenting DNA double sequences we'll present them in a way that reflects their antiparallel nature:

You likely won't see this visual cue on MCATs and GREs, so remember to pause and think "Why was the bottom strand backwards in my textbook?". It's to remind you of this anti-parallel nature.

A chemically correct view is shown below; pay close attention to the sugars.



Source: https://upload.wikimedia.org/wikipedia/commons/thumb/e/e4/DNA_chemical_structure.svg/514px-DNA_chemical_structure.svg.png

If you start on the left strand, the top indication is that the last carbon in the chain was a 5' cabon (on the green adenine). Moving downward, the first part of each new nucleotide encountered is the 5' position of its sugar (which would attach to a previous nucleotide); the backbone 'exits' from its 3' position.. so we're 'goes from' 5' to 3'. On the other strand, tracking through the sugars indicates that the flow is in the opposite direction.

Questions: Predicting the partner

Word alert: why do we say the chain "goes from" 5' to 3' and not from 3' to 5'? When a strand of DNA or RNA is actually being built, an incoming nucleotide is added to the free 3' end of the existing chain--so the chain gets longer on its 3' end, thus 'growing from 5' in the 3' direction'.

Implications of double-stranded storage

When 'at rest', your DNA is double-stranded; both partners are present. This has a number of important ramifications. For one, the chemically somewhat delicate basepairing positions are sequestered in a conversation with each other. But the most overwhelming implication of having complementary strands is that you essentially have two complete copies of all the information: having one, you (or any of many cellular machines--or, in a sense, even water) can find and build the matching partner.

The implication here cannot be overstated: the structure and nature of DNA causes it to be inherently copy-able--virtually self-copying (much more on this later!). Indeed, even one strand of DNA is copy-able in principle; the structure of each base on one strand 'yearns for' its chemical counterpart. All that is needed is a machine capable of realizing this fit has occurred and 'gluing' the new partner to an existing structure. This critical feature (and its implications) was immediately apparent to Watson and Crick upon solving the solution of the structure (they are shown with a very early photo below; Crick is pointing with a slide rule to make the environment more 'science-y').

Click for image source

In one of the papers announcing the structure, the two tossed off a statement that, while true, is so understated as to constitute an ancient humblebrag: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." Since each strand contains everything needed to 'specify' its partner, simply by separating them and allowing new components to come in and match up with the

existing face (and gluing the newcomers together), two complete double strands can be made from one. And that's exactly what happens!

Principle: since the two strands of DNA are chemically complementary, each possesses everything necessary to re-create its partner.

You might also enjoy: if you want a sense of the discovery as well as a 'time capsule' of the questions and thinking of the day, you'll enjoy this helpfully **annotated version of Watson and Crick's seminal paper** on the structure of DNA.

Copying: many tasks; many machines

While it's true that the chemistry of the bases renders them sufficient to the task of finding each other and pairing, stitching the backbone together is hard work, and to get it done on timescales of today's cells requires a lot of sophisticated machinery. While there's stage full of actors on the scene, you can make a lot of sense of the process and players simply by thinking about what tasks need to be done. In this and pretty much any process you come across, nothing ever happens because it would be a 'good idea' or because it 'needs' to get done. There are no wandering intelligences with jack-of-all trades capabilities. As you learned in the investigation of enzymes, most proteins are capable of doing exactly one thing... though they often do that one thing extremely well.

Principle: Every action that takes place in a cell will have a dedicated machine whose structure both recognizes the 'need' for the action and acts to make that action occur rapidly.

Let's consider an ongoing DNA copying (DNA replication) event, the sort of things cells undergo anytime they want to take their single copy of DNA and produce two identical ones from it.

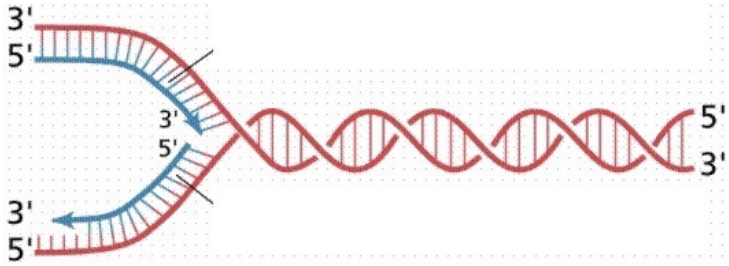
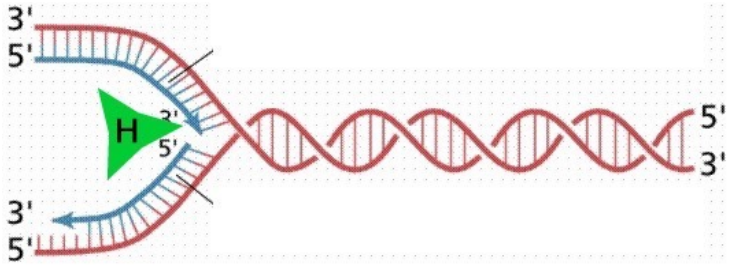


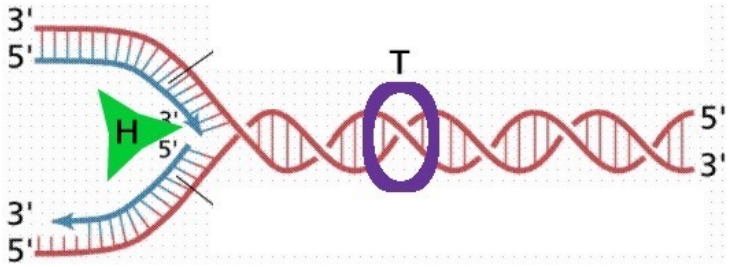
Image not permissioned; will replace with one having these key features: helical representation of un-copied DNA. The two strands should be differently colored and having an actual ribose as well as the 5' and 3' labels reinforces and adds understanding. Blue strands should not yet be present.

As you've come to expect, the favored structure of DNA has the bases nicely paired on the inside, and the two strands wound around each other as shown in the right half of the image above. If we want that to stop being true, somebody is going to have to do the unwinding... There's a protein for that. Since the DNA is in helical form, the "add an *-ase* to indicate an enzyme" rule works well here--the protein that does the unwinding is called "helicase" and would proceed from right to left in the image above, separating the strands. Since it's making the strands do something they're not excited about, work is involved, and so the enzyme must be fueled by ATP.

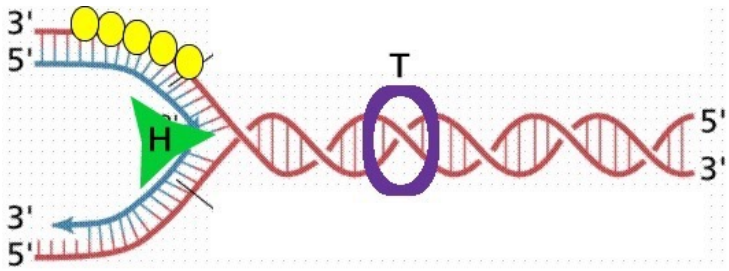


These are crude renderings to suggest final versions. The blue single strands should not be present yet and letters are used instead of labels (here, H = helicase)

Solving the 'wound up' problem creates others. If you take a pair of shoelaces, wrap them around each other a dozen times, then use your finger to 'helicase' them apart by pushing down from the junction toward the shoe, what do you observe? The windings get tighter and tighter as the helicase proceeds, since there is no release for the accumulating twists. Wouldn't it be great if someone was running along at the tight end of your shoelaces, cutting one, unwrapping it around the other, and undoing the cut? Well, your cells have an enzyme for that. Since that operation changes the number of twists in the DNA (aka it's topology) this enzyme goes by the tongue-twisting term topoisomerase.



That solves the problems 'in front' of the unwinding machine helicase... but what about behind? Just because the strands have been forced apart doesn't mean they have to like it. Sure, it'd be great if they sat patiently until their turn... but remember, nothing in the cell happens solely because "a smart molecule would do that". But there is a machine that enforces single strandedness. It's got a great name. It's a protein that binds to single stranded DNA, and it's called... single strand binding protein.

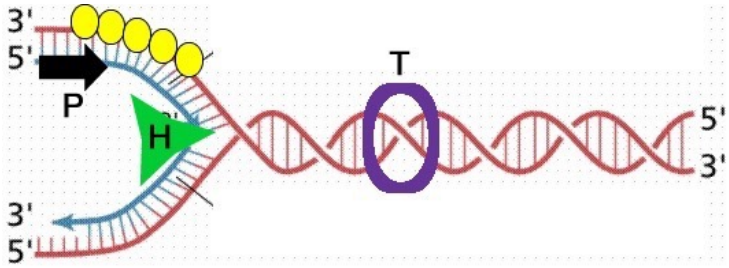


Now we're ready to start thinking about the copying machine itself... but it's not ready to

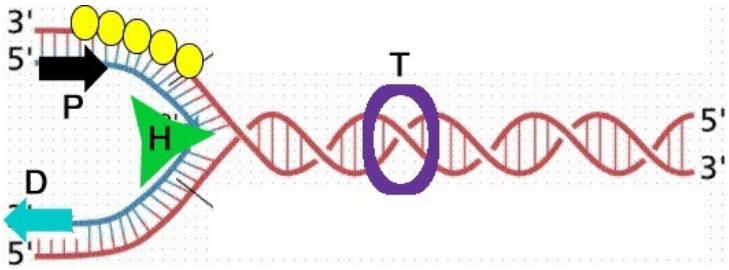
begin. For reasons that are fascinating [Could link out to a discussion, or have this be optional instructor-specified "in-game purchase". I have not found correct answer on the web to link to; it has to do with what is necessary for DNA Pol to proofread--a 'weak' grip on recently added nucleotides, and thus too weak a grip to 'hold' a first nucleotide to start synthesis with], but beyond the scope of the current discussion, DNA polymerase can't just start plunking nucleotides down onto a bare single strand. DNA polymerase can add to a started strand and continue it, but can't "start from scratch". So if a new strand is going to be initiated, there has to be a machine for that; if it's going to build a 'polymer', it's going to be a "polymerase" (actually pronunciation puts the emphasis on the 2nd syllable).

Soundcloud link

Indeed, it's relatively inaccurate RNA polymerase, but since it's one that is only involved in DNA replication, it gets its own name. If you're familiar with the expression 'priming the pump', you'll know that 'priming' refers to special steps involved in starting a process. The RNA polymerase that creates the initial strand that will be extended in DNA replication... is called primase.



At last, we're ready to meet the star of the show, the machine that actually attaches nucleotides to the growing chain of the new strand. As a machine that makes polymers of DNA, it is rightly called DNA polymerase. For our purposes, the key fact is that it can recognize when a newcomer nucleotide establishes a good basepairing fit with the existing strand (known as the template, since it is the 'mold' that provides instructions as to 'what goes here'). Once DNA polymerases detects that a complementary base has arrived and established itself in proper position, it 'cements' the newcomer in place by forming a covalent bond.



Do you have it? **Interactive self-check**

You might also enjoy: Bruce Alberts' career in DNA replication (alongside one in education): the logic, the experimentation, the path

DNA detective (Incomplete)

Instances from the real world (**this section not written; these are resources**)

http://www.nydailynews.com/news/justice-story/snowball-dna-star-witness-murder-trial-article-1.1436214

Principle: DNA sequences represent both potentially unique 'signatures' and 'historical records' of ancestry.

http://www.wired.com/wiredscience/2010/10/king-louis-xvi-blood/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+wired%2Findex+%28Wired%3A+Index+3+%28Top+Stories+2%29%29

> *Sick of taxes, a lack of rights and living in poverty, French revolutionists condemned Louis XVI to the guillotine on the morning of January 21, 1793. After a short but defiant speech and a menacing drum roll, one of the last kings of France lost his head as a crowd rushed the scaffold to dip handkerchiefs into his blood as mementos. Or so the story goes. Lending new life to the demise of Louis XVI, scientists performed a battery of DNA tests on dried blood inside a decorative gunpowder gourd that purportedly contained one such handkerchief. The results, described Oct. 12 in the journal Forensic Science International: Genetics, show the blood belongs to a blue-eyed male from that time period: a possible dead-ringer for the executed king. "The next step is find a descendant either of the king or his mother," said Davide Pettener, a population geneticist at the University of Bologna in Italy who helped with the analysis. "Otherwise we'll have to try to get a sample of the dried heart of Louis XVI's son." The son was Louis-Charles, known as the Dauphin (heir to the French throne) or Louis XVII, and he died from illness or poisoning at age 10 more than two years after his father was executed. His heart is kept in a crystal vase in the Cathedral Basilica of St. Denis on the outskirts of Paris. "It's going to be very difficult to obtain permission from the French authorities, but we may try," Pettener said. An anonymous Italian family who've owned the gourd since at least 1900, possibly the late 1800s, approached one of Pettener's colleagues to do the genetic analysis. Before the family obtained the gourd, it allegedly was a gift to Napoleon Bonaparte, who became First Consul of France in 1799 and Emperor in 1804. "It's a very strange story," Pettener said. "We thought it was a joke at first because we work on population genetics. But we realized it's very important from a historic point of view." The gourd, presently valued at about 500,000 euro ($700,000), is emblazoned with key figures of the French Revolution and bears an inscription that reads, as translated from French into English by the researchers, "Maximilien Bourdaloue on January 21st, dipped his handkerchief in the blood of the king after his beheading." There was no handkerchief in the gourd when the scientific team received it, but there was plenty of dried blood inside to scrape out five small samples. Two laboratories performed three kinds of DNA analysis:*

One probed the Y chromosome (inherited from the father), another scrutinized the HERC2 gene (associated with blue eyes) and the last examined the DNA in mitochondria (the powerhouses of cells, which are inherited from the mother). The tests showed the blood belonged to a blue-eyed man with a rare genetic makeup and not to an animal, nor to anyone in the laboratories, nor the gourd-owning family nor or any one of tens of thousands of people in genetic databases. Pettener added that the blood is also "quite old," making a forgery more unlikely. "A match on the Y chromosome of the Dauphin will immediately authenticate the blood as belonging to the king Louis XVI," Carles Lalueza-Fox, a biologist at Universitat Pompeu Fabra in Barcelona and lead scientist of the analysis team, wrote in an e-mail to Wired.com. "In any case, even with this information, we have historical evidence that this gourd could in fact contain the blood of the king." Images: Paolo Garagnani/Davide Pettener/Elsevier Citation: Lalueza-Fox, C. et al. "Genetic analysis of the presumptive blood from Louis XVI, king of France." Forensic Science International: Genetics, Oct. 12, 2010.

Patterns: Crystals, Chargaff, Chemists, and Crick-Watson

History teaches importance of knowing a good Chemist: Chargaff's rules and the tautomers

The path to the structure of DNA is a fascinating one, full of mis-steps and controversy. One of the most interesting is the failure of the scientific community to achieve insights from "Chargaff's rules". Erwin Chargaff investigated the relative quantity of the four bases in the DNA from a variety of organisms. He quickly discerned (and published) the fact that there were

Some of Chargaff's data on relative nucleotide abundance in genomes

Organism	%A	%G	%C	%T	A/T	G/C	% (G+C)	% (A+T)
Human	29	21	20	30	.98	1.04	49	51
E. coli	25	26	26	24	1.05	1.01	52	48
Wheat	27	23	23	27	1.01	1.00	45.5	54.5
Octopus	33	18	18	32	1.05	1.00	35	65
Chicken	28	22	22	28	.99	1.02	44	56

Data is only useful if you look closely at it and think carefully about what it means. Consider the data for wheat DNA (above). what value did Chargaff measure for each of the base quantities?

Rather striking, isn't it? That's a comparison of the nucleotide distributions within a given organism. Now let's compare between two organisms

How do octopi and E. coli compare?

Looking specifically at the percent of cytosine in E. coli (the famous bacterium living in your gut) and octopus, how different are the percentages in the two organisms (express the difference as a positive whole number)

If you overlook the 'noise' in the data introduced by inaccuracies in the techniques of the day, a clear *pattern* emerges: the relative amounts of A and T are the same within any organism, as is the amount of G and C. On the other hand, there's no clear correspondence between the overall amount of (G+C) vs. (A+T).

Check your understanding: **Chargaff goes to Mars**

Why wasn't Chargaff's data a flaming arrow pointing toward a DNA structure with the basepairing that you learned of above? There are several reasons, and they're quite instructive. First, in the structure of alpha helices and beta sheets the variable part (amino acid side chain or R-group) 'points' toward the outside of the structure. Reasoning by analogy would lead you to expect that the bases would be on the outside and the 'backbone' on the inside (Linus Pauling even proposed such a structure... though it was implausible in that the negatively charged backbones were all sandwiched closely together). For Watson and Crick, there was an even more concrete impediment. When they first sat down to model DNA structure, the chemical forms they had for several of the bases were incorrect in a critical way: they had incorrect dispositions of a double bond in their basepairing face. These alternate forms, now known to be rare, are called 'tautomers'. Because of the 'rule of eight' (outer shell filling), moving the position of the double bond results in one atom not 'needing' a hydrogen and another atom acquiring one. Let's take a look at the consequences for two of the pairs:

First, **compare guanine to its tautomer**

Now... **with which base does the tautomer of guanine fit best**?

Tautomers: misleading alternate forms of the bases
Using the BasePairer representation of the tautomeric form of guanine, which pyrimidine (recall: PURe As Gold) forms a good fit (given that the 'legit' A::T pairing makes only 2 hydrogen bonds, consider "at least two hydrogen bonds" plus "no positions have conflicting charges" as 'good fit')

A cytosine B thymine C neither D both

The upshot of having the wrong structures? Even when it occurred to Watson and Crick to try to model-build by pairing the H-bonding faces of the bases, theirs didn't 'fit' because the structures they were working with were incorrect. All was set right when a chemist down the hall, Jerry Donohue, told Watson that the correct forms shifted some of the charges about... and magically, adenine now 'fit' perfectly with thymine, and guanine paired up with cytosine. And thus the modern era of molecular biology was born.

Chargaff himself never hit it off with Watson and Crick, and perhaps being left off the Nobel Prize for the structure of DNA (which illuminated his work, but for the above reasons was less directly reflective of it than might have been) was the final straw. In an all-too-human moment, Chargaff groused about the recognition Watson and Crick received for the structure saying "That in our day such pygmies throw such giant shadows only shows how late in the day it has become." (Erwin Chargaff in H.F. Judson's The Eighth Day of Creation). To his

credit, he packaged his discontent in a compelling analogy!

Chargaff also walked away with a conclusion valuable to all of us:

> *"When I began to realize how unique were the regularities we had discovered, I tried, of course, to understand what it all meant, but did not get very far. I attempted to build molecular models of the nucleotides… [but] I ran out of atoms and even more of patience… Thus I missed the opportunity of being enshrined in the various halls of fame of the science museums."*
>
> *--ERWIN CHARGAFF, Heraclitean Fire, 1975*

[Base matching in the real world]

CRISPR-Cas9: miracle in the making

https://www.youtube.com/watch?v=k99bMtg4zRk

DNA Detective1: Crimes in the real world

DNA Detective2: Who is related to whom?

Carbon-based storage: dangers and outcomes

Things fall apart

Diamonds are forever... but your genetic material must be synthesized by biological machines, and the resulting product, while also carbon-based, is more fragile. As the Talking Heads put it

> *"Things fall apart... it's scientific." (Remain in Light, Talking Heads)*

So far we've looked and the glories of the molecule at the center of life... but that bit about the tautomers should've set your antennae tingling. The folks who put forth the incorrect 'tautomeric' structures of the bases weren't wholly wrong--sometimes the bases DO visit those unfortunate shapes. It's literally a fact of (carbon-based) life that nothing will last forever. Things that carbon-based enzymes can put together, energetic water molecules (and nastier components inside cells) can change or take apart. And when the meaning of a molecule is its shape and 'feel', a change to those is literally like changing a letter in a sentence.

Changing faces: deamination and the meaning of the bases (adenine=>hypoxanthine; incomplete)

You met a tautomer above; let's take a look at one of the other common changes to befall a base. Without going into detail, the chemistry is not unlike what we looked at when popping the terminal phosphate off of ATP. Once again, the key ingredient is plain old water; it's a reaction that is wholly inevitable when water is combined with the structure of several of the bases. And it's going on in your cells right now--something like 10,000 times per cell per day. Rather ominous, isn't it?

Word alert: unfortunately, while the base that is the result of deamination of adenine is called hypoxanthine... but when attached to sugar and phosphate, we call it inosine. Sorry about that! This will become important when we discuss 'wobble' during the synthesis of proteins, and a reminder will appear there as well! For the rest of this chapter, we'll focus on the base and speak of hypoxanthine only.

RNA: cheap knock-off, or legacy code?

RNA: cheap knock-off, or legacy version nucleic acid?

We've been discussing DNA as if it were the only game in town. In the nucleus, that's true--it is the undisputed storage material for all organisms known to mankind (we'll leave viruses out of the tally here). But it has a very famous sibling, RNA that performs many critical functions in the cell--not only is it the molecule charged with shuttling short-lived copies of the genetic information from the storage banks out to the protein-making machinery (next chapter), it is a key constituent of the protein-making machinery itself. RNA has also been 'rediscovered' recently in that it plays diverse roles in controlling which genes are making products and which are not. So how did DNA end up with the cushy storage job, while RNA got stuck with the blue collar tasks and short lifespan?

Let's start with the conclusion, as it will provide context for the arguments that lead up to it. As we'll discuss much more deeply in a later chapter, life today presents us with a fascinating chicken-and-egg problem. If proteins take the action role in the cell, and DNA is the storage material, how could life have bootstrapped its way up from single molecules? The universally accepted answer among biologists is that the third party, RNA (or a now lost molecule like it), exhibits sufficient of the characteristics of BOTH to have been the initial jack-of-all trades, but that over the millennia, DNA evolved as an improvement in the storage category, whereas proteins outstrip RNA in the machine-making department.

Let's take a long hard look at RNA as a molecule as a first step in seeing what's so grand about DNA. There are only two chemical differences: the sugar in RNA has an -OH in one position where DNA has only a hydrogen, and one of the DNA bases (thymine) possesses a small decoration (a methyl group, -CH3) in an out-of-the-way position that is lacking in the RNA equivalent (called uracil... yep, first isolated from urine).

Who's afraid of an -OH?

Let's begin with a review...

Image source: https://en.wikipedia.org/wiki/Nucleoside#/media/File:Guanosin.svg

Questions: That old-time numbering: Which carbon is which?

Total 1 questions

Response Self-Answer

Word alert: The names of DNA and RNA are very similar, and we're now looking at the root of the difference. RNA is "ribonucleic acid", in honor of the fact that it is acidic (the phosphate has lost it's hydrogen), and is found in the nucleus (though as you'll see, many different kinds of RNA molecules are playing critical roles in the cytoplasm as well). DNA is 'deoxyribonucleic acid', the oxygen that has been removed being specifically the one at the 2' position.

As described above, new nucleotides are added because the 3' OH attacks the innermost phosphate of the threesome that 'grows out of' the 5' position. But let's take a very close look at the location (and most importantly, the neighbor) of the resulting bond:

Consider the relationship of the 2' -OH indicated by the arrow... if you think about it, the relationship of that -OH to the phosphate next door is absolutely identical to the attacks on phosphates we've looked at before (popping off the last phosphate of ATP, and more recently, blinding one nucleotide to the next). And the consequences of that similarity are rather dire. For the entire life of an RNA chain, every single 2' -OH is looking greedily at the phosphate that its neighbor at the 3' position has. Periodically, one of them manages to pull off the chemical attack. Just as we saw with ATP, formation of a new bond requires a bond be 'let go' so that all atoms involved continue to adhere to the 'rule of eight'... and the bond let go is often the one to the 5' position of the following nucleotide. In other words... the chain is severed. And the rest of the RNA drifts away. Not so good.

A simple solution

The evolutionary 'discovery' of deoxyribose, a sugar with the -OH replaced by a simple -H solved the problem neatly and completely. However, without going into gory detail, it's worth considering that the cell can't just take an eraser and make the oxygen go away. The chemistry involved is expensive and involves some rather energetic (and dangerous) players. And of course, the enzymes to carry out the reactions had to come into existence. Nonetheless, this evolutionary 'discovery' clearly paid off and became part of the 'line of survivors' that lead to modern day cells, processes, and creatures--like us.

An uglier story: what's wrong with uracil?

One of the differences between DNA and RNA is that for one of the 'letters', DNA uses the base thymine, but RNA uses the base uracil. Take a look at this image to answer these questions

The added methyl appears to have no significant consequences for basepairing.

Word alert: calling one base 'thymine' and one 'uracil' implies that they're really different, but as you've seen, they're totally not. Indeed, a legitimate and more information-rich name for thymine is "5-methyl-uracil'. It's a better way to think about thymine since it reminds you how very similar they really are.

Word alert II: On the other hand, you might think that 'thiamine', which sounds virtually identical to 'thymine', would be very similar. No such luck. Thiamine is a very different chemical, and also goes by the name vitamin B1.

If the methyl (-CH3) group that distinguishes thymine from uracil is just hanging out in irrelevant land, why is DNA's switch to thymine touted as such a critical evolutionary discovery? To get at the truth, we must dig deeper. Our quest begins with a comparison of cytosine and uracil. Use this image to answer the following question:

Comparing cytosine and uracil

Click on the pairing positions on cytosine (left) that differ from the corresponding positions on uracil (right). Click only on the haloed areas

While it might be ideal if cytosine differed from uracil at three positions, being the opposite of

one another at two positions seems pretty good... isn't it? Absolutely. Until that cytosine suffers a bad day in water. Our ubiquitous friend can interact with cytosine in a reaction similar to the 'popping off' of a phosphate in the ATPase reaction. The amino group (-NH2) is replaced by the oxygen of water. The effect? After a little bit of double bond switcheroo (required so that all atoms involved continue obeying the outer shell Rule of Eight) the cytosine has changed into uracil. Not resembles uracil; not has some features in common with uracil--the chemical changes create a de facto uracil where only moments ago, a perfectly satisfactory cytosine stood.

For whom the bell tolls: the traitorous pairing of a deaminated cytosine

While other deaminations wreak havoc among the bases (compare adenine to hypoxanthine above; it's also in BasePairer's 'Deaminations' menu), in an RNA-based code, cytosine => uracil would be particularly devastating. The reason? RNA employs bases adenine, guanine, cytosine, uracil. Like when adenine deaminates to hypoxanthine, cytosine deamination to uracil, causes it to 'mean' something completely different (uracil, like thymine, is a partner for adenine but cytosine is the partner for guanine). But there's a huge difference. Whereas the appearance of hypoxanthine in an RNA- or DNA-based code is clearly call for alarm ("what's that doing here?") the change from cytosine to uracil is 'silent', in that uracil is a legitimate member of the 'RNA club'. So while you can build a 'policing molecule' that scans DNA for hypoxanthine and initiates a replacement (indeed, you have just such an enzyme... and many others with similar policing functions), you must not remove uracils from RNA; it would be equivalent to declaring the letter 'e' unwelcome in this textbook!

So now we come to the heart of the mystery. If the problem is that cytosine becomes uracil, why not solve it by eliminating the use of cytosine? Great idea on paper! You're a better designer than the process of evolution. But that's because while you can jump directly to a 'best' solution; the evolutionary process is constrained: 1) it must work with what it has, and 2) all the intermediates must be alive and leave offspring. Even if there's a great solution 'down the road' somewhere, the genetic material can't declare itself 'closed for the next 1000 generations while we work on the cytosine problem.' Thymine represents a clumsy, after-the-fact, inelegant, expensive solution to 'the uracil problem.' But before making too much fun of it, bear in mind that every single living organism on the planet employs this solution!

Let's break down the pieces of the fix. Recall that the overriding problem is that using an RNA 'alphabet', a deaminated cytosine (i.e., a uracil) is a wolf in sheep's clothing: it's wrong, but there's nothing wrong with it--it's one of the legitimate letters of RNA sentences (guanine, adenine, cytosine, uracil). If you upgrade the alphabet to adenine-guanine-cytosine-thymine, everything changes. Now when a deamination event converts cytosine to uracil, there's a glaring sore thumb standing out in the room--"Who let a uracil in here? Excise it

immediately!" As with hypoxanthine, you do have a policing enzyme for exactly that purpose (though of course, it operates only on DNA!).

Remember that the only difference between uracil and thymine is the 'decoration' methyl (-CH3) group at a non-basepairing position. So it's a label--a marker that contains the information: "I am not a failed cytosine; I'm the legitimate partner of adenine!". An ugly solution to be sure... and one that involves a lot of hidden costs. In order for the solution to arise, the (expensive) capability of adding a methyl group to uracil had to arise first. Ever since the solution appeared, your cells (and those of every other organism) have had to 'pay' (in energy and resources) to place a methyl group on every single adenine-partner that was headed into DNA. In order for the 'solution' to solve anything, the machinery to police and remove the uracils had to arise. And in refining the process, there are even more costs--since your cells synthesize DNA components by taking RNA components and 'fixing' the ribose (to 2'-deoxyribose), your cells are constantly pointlessly creating deoxy-uracil. So there exists an enzyme that 'pops off' the phosphates of these, rendering them unable to be added to growing DNA chains (they will pair with adenine, after all! Poor DNA polymerase will cheerfully put them in, because "does this pair" is all DNA polymerase understands!). All in all, an ugly, expensive solution... but it appears that the increase in accuracy is worth the cost, as everyone you see is doing it!

Summary of principles

Principle: The units of DNA 'recognize' their partners through rigid, direct, chemically complementary contacts

Principle: A series of bases (joined in the form of nucleotides) constitutes a unique pattern of information as well as a unique chemical surface

Principle: since the two strands of DNA are chemically complementary, each possesses everything necessary to re-create its partner.

Principle: DNA sequences represent both potentially unique 'signatures' and 'historical records' of ancestry.

Principle: Every action that takes place in a cell will have a dedicated machine whose structure both recognizes the 'need' for the action and acts to make that action occur rapidly.

End chapter review

Interactive chapter review

Chapter summary paragraphs: Demonstration of Summary Tool
VocabuWary: getting the words out of the way

Landing spot simulating references to other chapters
Periodic table
Protein structure
Translation
ATPase
Enzyme
Origin of life & RNA world
pH, cells and proteins

Resources

DNA replication

Resource: https://medivizor.com/blog/2016/05/25/understanding-oncology-2016-dna/

http://bioscriptionblog.com/2017/03/29/crispr-primer-varieties/

Genetic engineering: good info on golden rice pathway

http://www.learner.org/courses/biology/support/13_gmo.pdf

7. Machines from instructions: Translation

Resources

Tutorials movie: Ile_Val_Synthetase.mp4

16.Sharing instructions

What you will learn

We've looked at what the information held in the cell's genetic material 'says' and instructs. One of the key features of life

ToInclude

Modeling Mendel's data onto possibilities

Given 3:1 ratio, what can your conclude?

Tools of the Trade: molecular biology