COMP 5970/6970-004
Computational Biology: Genomics and Transcriptomics
Lecture notes 18: 3/22/2022

Haynes Heaton

Spring, 2022

Lecture Objectives

- Biology from the big bang to heat death
  - Holding entropy at bay
  - Evolution - what exists either persists or replicates
  - Storing data biologically for evolution
  - Stable and redundant biological data storage and replication
  - Central dogma of molecular biology

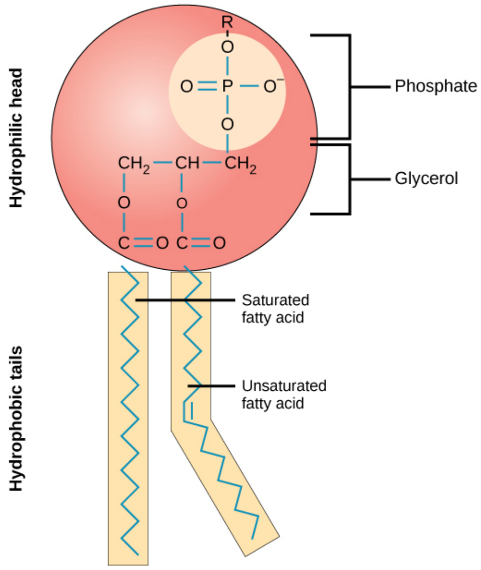
Biology from the big bang to the heat death of the universe.

0.1 The beginning

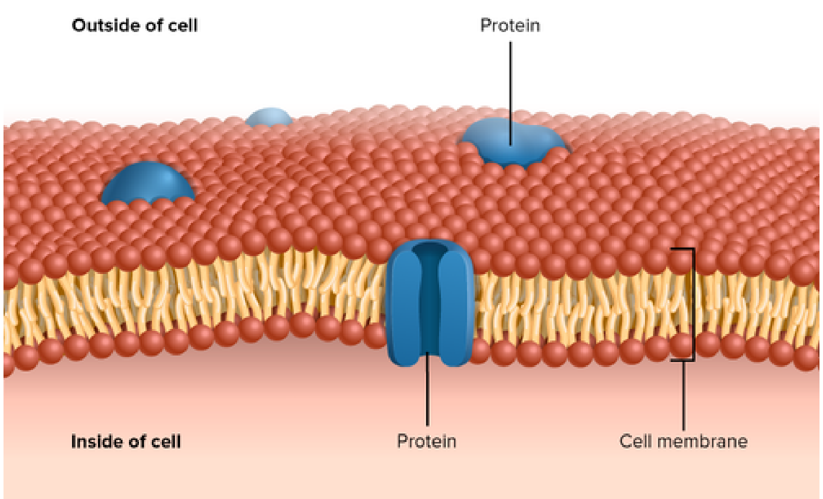
As the theory goes, around 13.8 billion years ago, there was no matter and depending on the definition there was also no space or time. Then out of a singularity, an explosion happened referred to as the big bang. After rapid expansion, the universe cooled enough to form subatomic particles and later atoms, primarily hydrogen. Clouds of hydrogen eventually coalesced into the early stars. Stars produce heat and light through fusion of atoms. Stars like our sun primarily fuse two hydrogen into one helium atom. Much larger and older stars fuse heavier atoms to form heavier elements. After one round of star formation and supernovae, enough heavy element atoms existed to create planets including rocky planets such as earth. Earth was formed some 4.5 billion years ago. Over the following few hundred million years, conditions on earth were inhospitable to life with much volcanic activity and relatively frequent extinction level asteroid impacts from the orbits of the giant planets. After these orbits changed and earth cooled down to form a solid crust and stable oceans, it is theorized that still extreme conditions allowed for chemical interactions to produce organic compounds from inorganic compounds. And some of these compounds were capable of self replication. These could be thought of as the first life on earth. But such molecules in the open are subject to oxidation and interaction with highly reactive chemicals. In combination with lipid membranes, the first cells emerged as much more stable forms of life. While it is not believed that the emergence of life was a singular event, it is believed that all of the organisms alive currently were derived from a single cell, or the **Last Universal Common Ancestor**, or LUCA.

0.2 Holding entropy at bay

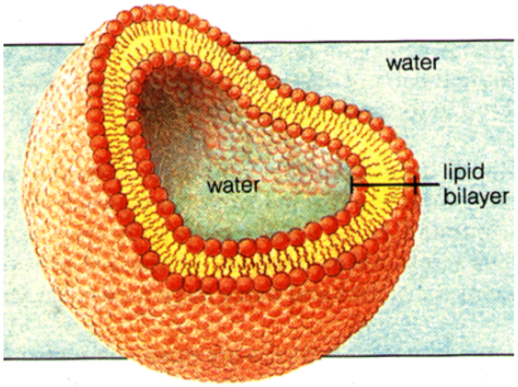
The second law of thermodynamics states that the total entropy of an isolated system can never decrease. However, subsystems that are not isolated can reduce their energy through work. This is exactly what biological organisms achieve. And they do so in an energy efficient manner by selectively separating themselves from the greater system. For instance if you did work to put all of the air molecules in one corner of a room, without some barrier to keep them there, they would very quickly repopulate the entire room. The accomplish this with semi-permeable membranes created with bilipid membranes. These are created from molecules with one polar end and one nonpolar end called phospholipids. A phospholipid has a glycerol molecule with a polar phosphate group on it and two nonpolar hydrocarbon tails.

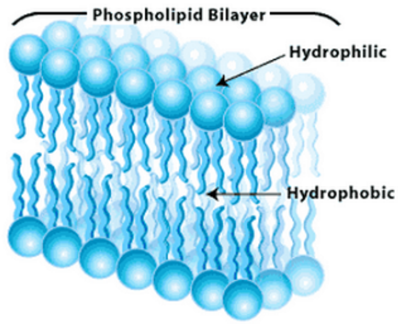


These phospholipids self organize in aqueous (water based) solution into bilipid membranes because their nonpolar tails are hydrophobic and have a lower energy state when with other hydrophobic molecules such as the other hydrocarbon tails of phospholipids while their phosphate group is hydrophilic and have a lower energy state when near other hydrophilic molecules such as water and other phospholipid phosphate groups.



This lipid bilayer is at its lowest energy state when no hydrocarbon tails are exposed to the aqueous solution. The only way to accomplish this is for the lipid bilayer to wrap on itself in 3D space forming a spherical shape. This is known as a micell.





These self organizing molecular barriers turn out to be extremely useful in separating low entropy environments from high entropy environments with minimal energy. As with the metaphor about pushing the air to one corner of the room, this is akin to putting that air in a compressed canister. Once the initial work is done to put it there, little energy is needed to keep it there. For the bilipid membrane, hydrophilic molecules will not cross the barrier often due to the high energy state they must pass through when interacting with the hydrophobic hydrocarbon chains. Proteins embedded in this bilipid membrane may allow for the passage of some but not all molecules or may allow for only one-way passage of certain molecules. In this way, cells separate their low entropy interiors from the high entropy exterior and maintain that state with minimal energy.

0.3 Evolution as observation bias

What exists either persists or replicates. And what replicates must exist long enough to replicate. Biological systems strive for both to different degrees. The cell maintains this low entropy state with its cell membrane and has developed many different mechanisms for self-repair. And more complex multicellular organisms have developed systems to fight off external attacks (the immune system) and heal injuries. In terms of replication, very simple systems might follow chemical signals to make more or less of each component and split into two when possible.

0.4 Brief history of the study of biology

Natural science often advances with new human built technologies. So it is today and so it was in the 1600s with the advance of the microscope. Robert Hooke was a polymath who used early microscopes and telescopes to make various observations including the first known observation of a cell. He used his microscope to view some fungus and saw a honeycomb type structure and he coined the term "cell" to describe each separated entity.

It was long known that traits were passed down from parents to offspring. The rules of this inheritance of traits was not known until an Austrian friar in the mid 1800s by the name of the Right Reverend Gregor Mendel, experimented with pea plants. He found that some traits were dominant and others recessive. An organism would need two copies of a recessive "factor" in order to have that trait but one copy of a dominant "factor" would express itself.

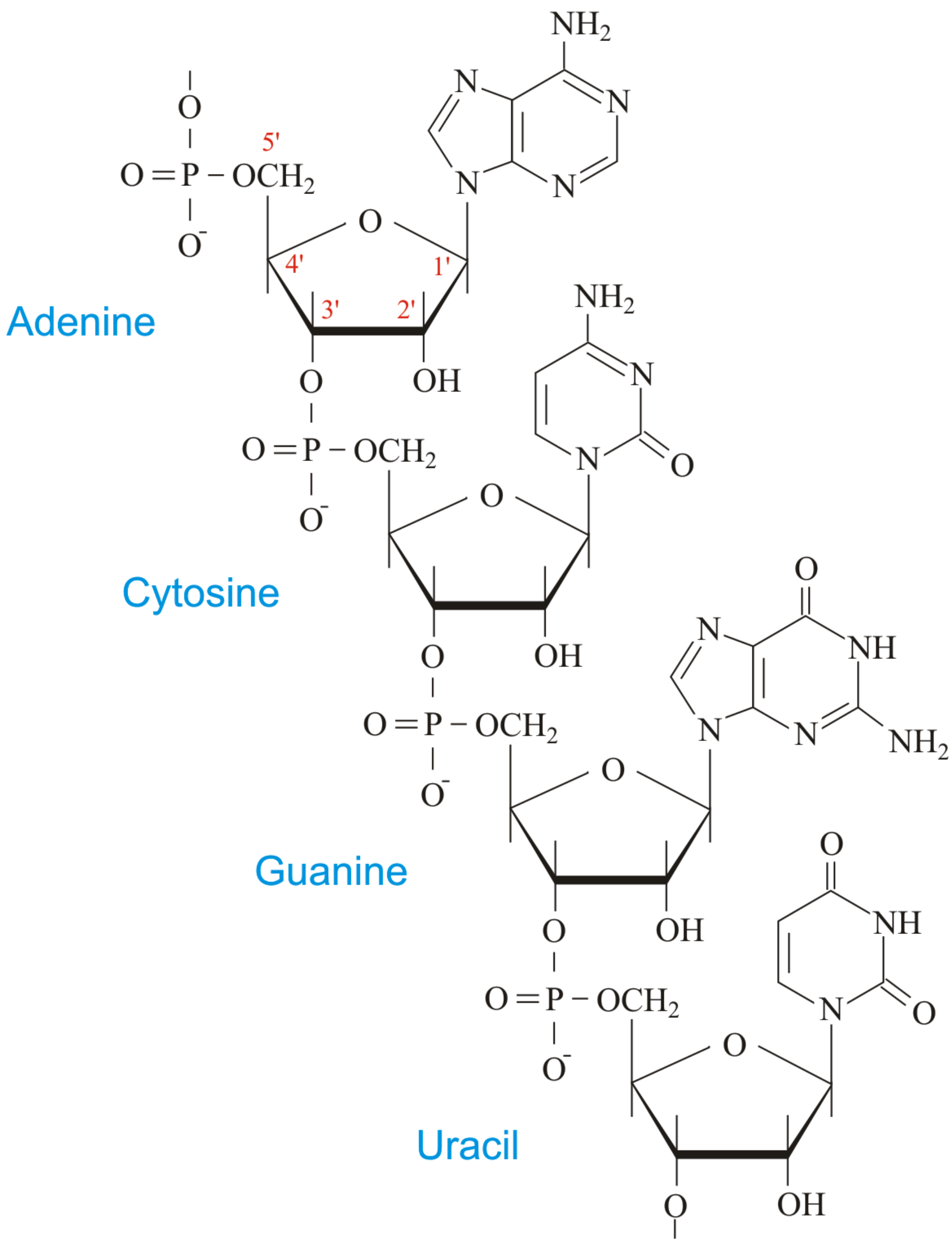
Around the same time in England, Charles Darwin introduced the idea that species evolved in a branching pattern from common ancestors in a bid for survival through natural selection.

It wasn't until the 1950s that the structure of the molecule governing these inherited "factors" was discovered. Rosalind Franklin used x-ray crystalography to inspect DNA and RNA at King's College London. These images were shared with Francis Crick and James Watson in Cambridge which led to the discovery of the double helix structure of DNA. This discovery also pointed at a viable replication strategy due to the redundant information contained in the molecule. For me this sends chills down my spine. I happened to work in the same building two years ago. And the story goes that when the discovery was made, Watson and Crick walked down the the Eagle pub and announced that they had discovered the secret to life. I recommend the scotch eggs at the Eagle pub. The beer, like most British beer, is luke-warm and not very fizzy.

Later we will address the modern history of DNA sequence and sequencing projects such as the Human Genome Project.

0.5 Storing data biologically for evolution

But eventually a blueprint is needed to map out what to do in different environments and for replication to be possible there, the blueprint must also be replicated. In computers, binary numbers are used to store information. In biology, **nucleotides** are used to store information. Initially, it is believed life used **ribonucleic acids**, or RNA to store information. RNA is a polymer of **ribonucleotides**. Each nucleotide consists of a purine (Adenine (A) or Guanine (G)) base or pyrimidine (Cytosine (C) or Uracil (U)) base and a ribose ring. These ribose rings can be bound to one another in series with a phosphate group creating a phosphate bond.

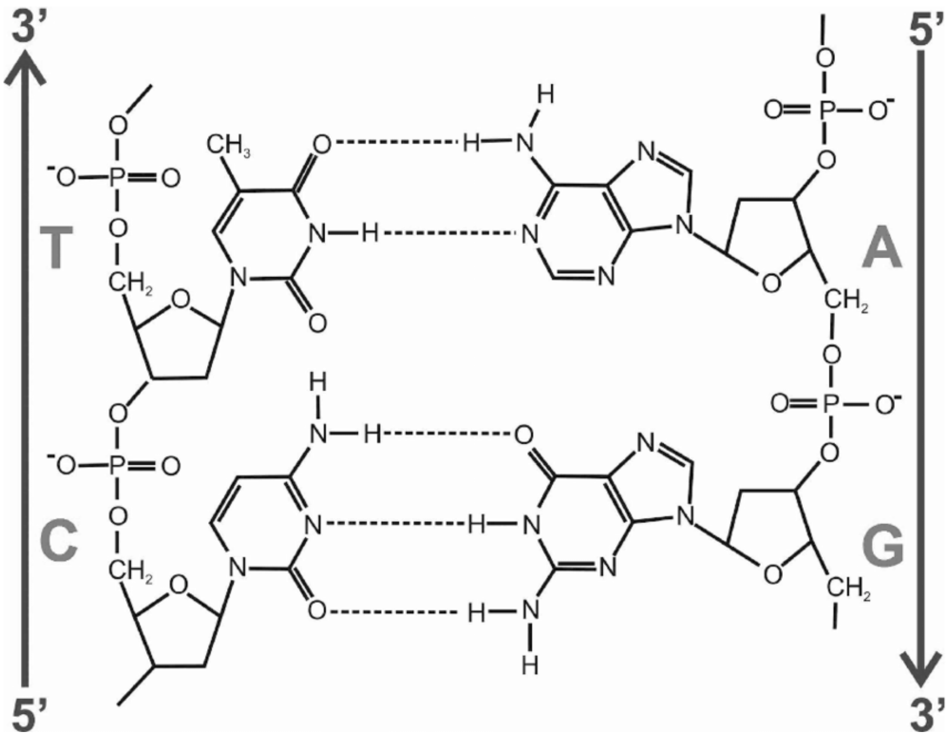


The carbon atoms in the ribose are numbered 1 to 5 as shown above with the 3rd and 5th carbon binding to either phosphate bond. With this we can describe a direction of the RNA sequence as either 5′ to 3′ or 3′ to 5′. This will be important later.

RNA is a very special molecule as it can act as information storage as well as fold into complex 3D shapes and carry out specific biochemical tasks similar to enzymes. These are called ribozymes. RNA is capable of enzymatically replicating other RNA. This is why it is believed that RNA was the first molecule used to transmit genetic information and that before DNA, life consisted of an RNA world.

0.6 Stable and redundant biological data storage and replication

RNA can form a double stranded molecule with its reverse complement molecule, but due to the hydroxyl group on the 2′ carbon of the ribose, the double stranded molecule is not stable and the strands will dissociate at moderate temperatures. Evolution eventually found that removing this hydroxyl group could produce a much more stable molecule, **deoxyribonucleic acid, or DNA**.



DNA is very similar to RNA, but the ribose is missing a hydroxyl group on the 2′ carbon giving less steric inhibition to the formation of the double stranded structure. And DNA has Thymine in place of Uracil which is simply a methylated version of Uracil. As you can see, Each purine can create hydrogen bonds with one pyrimidine. Adenines bind with Thymines and Guanines bind with Cytosines. Of note, Guanine-Cytosine bonds contain three hydrogen bonds whereas Adenine-Thymine bonds contain two. This makes the double strand structure stronger the more GC pairs in the region.

The angles of these bonds and other minor interactions create a helical structure which also strengthens the overall chemical stability. Because of the double stranded nature of this helical structure, it is termed a **double helix**.

0.7 The central dogma of molecular biology

The central dogma of molecular biology states that in general, the flow of information in most organisms goes from DNA to RNA to protein. And proteins carry out the various functions of the cell. This idea is broken in a thousand ways, but is in general true.