BIOL591: Introduction to Bioinformatics
Overview of biological information and how it works

Reading in text: Nothing in the book is really pertinent.

Outline:

A. What is bioinformatics?
B. Genes and organisms
C. Membranes and self-assembly of cells
D. Metabolism: monomers, polymers, central metabolism
E. Control points of metabolism
F. Summary

A. What is bioinformatics?

Bioinformatics may be described as the application of informatic tools to biological problems, but what are informatic tools and why do biological problems need them? We will consider two general classes of problems that need help in a big way.

First (Fig. 1A), there is the transformation of information from DNA, through RNA, and into the linear sequence of amino acids that make up protein. We understand well the first two of these steps, but the last, the folding of the amino acid chain into an active three-dimensional structure, remains shrouded in mystery. In large part, the information remains the same in content, though not in form, from beginning to end. This is shown in Fig. 1B:

Note that the rules for transformation of DNA information into RNA information are ridiculously easy to follow: take one strand of DNA, change all the deoxyribo-sugars into ribo-sugars, change thymines (T) to uracils (U), and you have it. The rules to transform RNA information into amino acid sequence information are similarly simple: the rules constitute the genetic code, which take three letters of RNA sequence and translate them into one amino acid. While a human would have no trouble applying these rules to a few hundred letters of DNA,

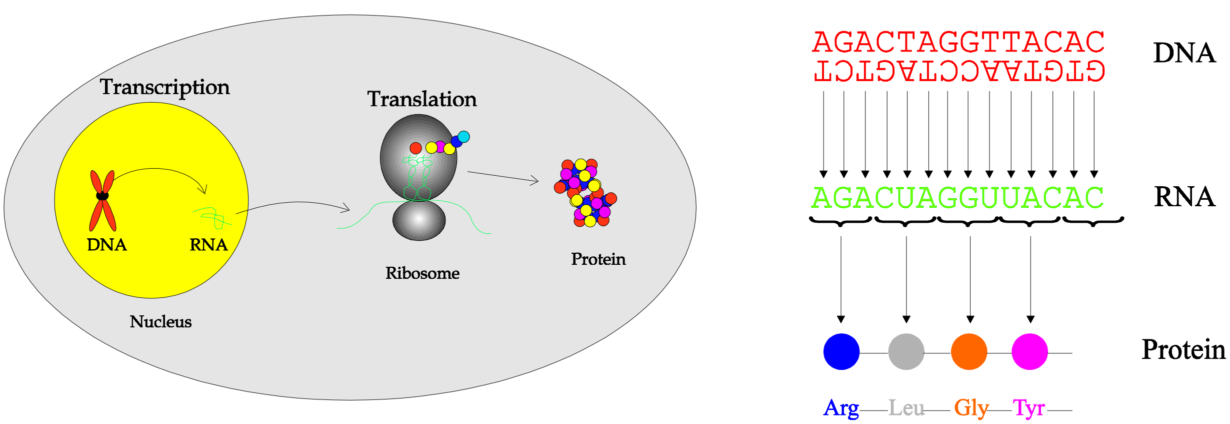

Fig. 1. Flow of information from DNA to protein. Double-stranded DNA is *transcribed* into single-stranded RNA, which is *translated* into the linear sequence of amino acids, which is folded into an active protein.

applying them to even a small genome, e.g. the 4.6 million letters of *E. coli,* would get tiring real fast. But applying mindless rules millions of times is what computers are good at.

A second class of problems are those related to the complexity of real biological systems. Imagine an automated bank teller. People shovel money into it, some of it foreign, some of it local. One machine within the teller converts Canadian dollars into 70 U.S. pennies and a second machine converts Mexican pesos into 10 U.S. pennies. A third machine collects pennies, and when it has accumulated 100 of them, it converts them into a dollar. A fourth machine converts twenty one dollar bills into a single twenty dollar note. A fifth machine monitors the number of twenty dollar bills and if a certain threshold is exceeded, it tweaks the dollar manufacturing machine and slows it down.

I've just described a system with only a handful of component parts, but even so, it would be very difficult to predict how much of each type of money is inside the system at any given moment. You'd need to know the speed at which each machine works, the degree to which the regulatory machine modifies its target, and of course how fast users put in and take out money. Even then, you have a sizable problem to solve. A living cell, however, is comprised of thousands of interrelated processes. Clearly, fast computation creatively used is necessary to get a handle on such a complex system. And just as metabolism within a single cell is complex, so are interactions between cells, interactions between organisms, and interactions between populations and their environments.

B. Genes and organisms

There is a relationship between the information within genes and the complex nature of whole organisms, just as there is a relationship between the characteristics of subatomic particles and the state of the universe. Actually, most of us, I suspect, are quite comfortable with this idea. After all, doesn't DNA determine practically everything about an organism?

It's true, genes composed of DNA contain the information required to specify a functional organism. But there is no gene that specifies intelligence, none that specifies beauty or even eye color. There is no blueprint within DNA for an organism, a cell, or protein. In fact, genes don't DO anything. They merely contain passive information.

But passive information can determine an organism (Fig. 2), What a cell can do (e.g., how a skin cell differs from an amoeba) and what a cell is doing (e.g., how a firing nerve cell is different from one at rest) depend on the chemical reactions taking place within it. These reactions are controlled by enzymes. Without enzymes to catalyze the reactions, they would not take place to an appreciable extent.

It follows, then, that the enzymes that are present and active in a cell will determine the cell's type and state. If you tell me what enzymes are in a cell, how many, and how active they are, I could, in principle, tell you that the cell is from the liver of a mouse and not from the spleen of a codfish. I could tell you that the cell is embryonic and not terminally differentiated. I could tell you whether the cell is responding to heat stress. (Interestingly, I could not tell you the race of the individual who donated the cell). In principle, I could tell you everything there is to be told about the capabilities of a cell, just by knowing what enzymes are present.

Whatever controls enzymes controls the cell. The function of the gene is to provide the information required to synthesize enzymes. Even though genes play no direct role in the functioning of a cell, they determine what proteins are present and, in the end, what the cell can do

Fig 2: How genes determine organismal function

- **Function of organism** is determined by **function of its cells**
- **Function of cells** determined by **chemical reactions** that take place within them
- **Chemical reactions** occur or not according to presence and **activity of enzymes**
  - Enzymes restricted to certain species (e.g. mammals make lactase to consume milk sugar)
  - Enzymes restricted to certain tissues (e.g. enzymes to make neurotransmitters confined to nerves)\
  - Enzymes restricted to certain developmental states (e.g. enzymes required to make yeast spores found only during sporulation)
  - Enzymes restricted to certain environmental states (e.g. heat induces production of enzymes to repair damage due to heat)
- **Enzymes** are proteins, their **activity** determined by **protein structure**
- **Protein structure** is determined by **amino acid sequence**
- **Amino acid sequence** is determined by **genes**
- Therefore, **genes** determine **organismal function**

SQ1. What at root do enzymes control?

SQ2. How do genes exert control over a cell?

SQ3. Suppose the entire sequence of DNA from a human is determined. What information can we extract about the person?

A. The adult size of the person
B. The proteins the person was capable of making
C. The number of legs the person had
D. The color of the person's hair

C. Membranes and self-assembly of a cell

Those of us who find it difficult to assemble a swing set appreciate the enormity of the task of assembling an entire cell. How can genes, lacking insight or example, direct the construction of the complex structures found within cells? Fortunately, cellular structures for the most part assemble themselves. Membrane biogenesis is a prime example of such a process.

A cell and many of its internal components are defined by their membranes. Membranes serve many functions. One basic function is to keep the innards of a cell or organelle inside, and in this view one might think of cells as soap bubbles in water. But biological membranes are much more than mere walls. Much of what the cell does takes place on or in membranes. For example, most of the ATP used to power the cell is made through structures on membranes.

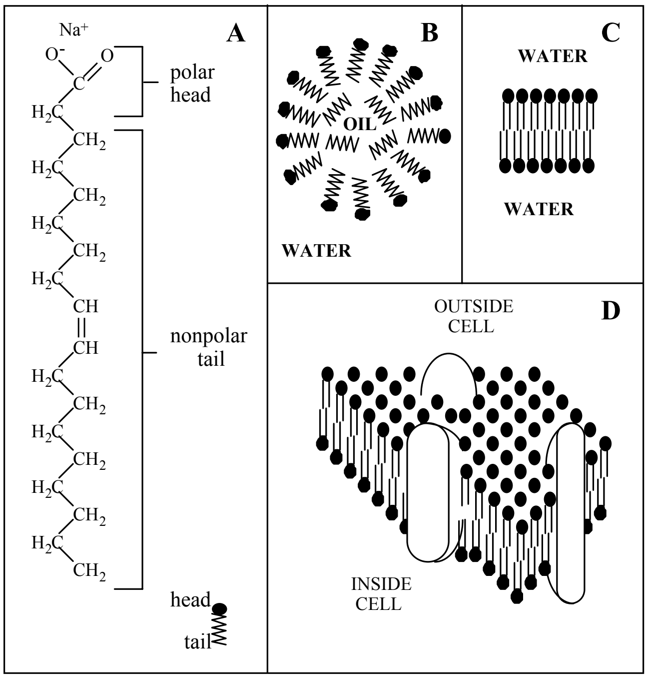

Fig 3. Amphipathic lipids and membranes.

A. Chemical representation and cartoon of sodium oleate, a fatty acid.

B. Oil droplet composed of fatty acids surrounding oil.

C. Phospholipid membrane composed of two layers of phospholipids, hydrophobic tail to hydrophobic tail, with the hydrophilic head groups exposed to water.

D. Biological membrane, composed of proteins swimming in a matrix of phospholipids.

Just as life to a great extent is the play of membranes, so are membranes largely the play of lipids. Lipids (literally, fats) are oily substances that can't dissolve in water. There are a wide variety of them in cells -- fat and cholesterol are two -- but the ones that form the structure of membranes are of a special class: *amphipathic* lipids. These are lipids that are *hydrophobic* (water hating) on one end of the molecule and *hydrophilic* (water loving) on the other. Pure hydrophobic molecules, like oil, separate from water, while purely hydrophilic molecules, like salt, freely dissolve. A class of amphipathic lipids that we see every day (or should!) is soap, one of which is oleic acid (Figure 3A). This fatty acid has a long, 18 carbon hydrophobic chain and an acidic, hydrophilic head group. Oleic acid is also a common component of phospholipids, the primary material of biological membranes. Note that fatty acids and derived amphipathic lipids are often portrayed in cartoon fashion as hydrophilic balloons over a hydrophobic string.

Soaps work precisely because they are amphipathic. The hydrophobic tails surround oily dirt, permitting the dirt to be dispersed in water (Figure 3B). When you rinse off the soap, the dirt goes with it. No one has to teach soap how to form oil droplets or bubbles: because of the interaction between water molecules, bubbles form themselves. The geometry of tiny bubbles is determined by the geometry of the soap. The size of the hydrophilic headgroup (the balloon) of soap determines how curved the bubble can be.

The amphipathic lipids of membranes differ from soaps in part because of their geometry. Fatty acids like oleic acid have bulky head groups relative to their thin tails, so if you put them side by side, you get a curved surface, just as assembling wedged-shaped pieces would form a circular pie. However, most of the lipids of membranes are *phospholipids*, containing two hydrophobic chains along with the polar headgroup (Figure 3C). These lipids are more or less cylindrical, that is, the headgroup has about the same area as the chains, so when you put them side-by-side you get not a curved surface but a plane. To prevent exposure of the hydrophobic chains to water, the plane is of two layers, a bilayer. The bilayer forms because of the amphipathic properties and geometry of its components. There is no need for some molecular machine to insert each component in place.

A cell membrane, then, is a vast bilayer composed in large part of amphipathic lipids (Figure 3D). The bilayer serves to exclude hydrophilic molecules from passing the membrane -- they can't get past the hydrophobic zone. It is thus possible to maintain a concentrated salt solution on one side of the membrane and a very dilute solution on the other, and this is precisely what the cell does.

The membrane is not pure lipid, however. If it were, the cell would be in virtual isolation from the outside world and would starve to death. It contains also protein that sit on the surface or span the membrane. These proteins serve a multitude of purposes. Many are involved in facilitating transport of hydrophilic molecules that the cell wants to be able to traverse the membrane (food, for example). Some proteins are involved in anchoring cell components to the membrane. Proteins find their way to the membranes in large part due to the same hydrophobic properties that govern the structure of the lipid bilayer.

SQ4. What are membranes? What are their functions?
SQ5. What is an example of a household item that is hydrophobic? Amphipathic?
SQ6. Identify the symbols representing protein in Figure 3D.

D. Metabolism: monomers, polymers, central metabolism

Complex soap bubbles within even more complex soap bubbles -- that is a simple view of what a cell looks like frozen in time. In real life, however, a cell is a high-powered organic chemistry factory, busily converting materials from its environment to its own uses. Take us, for example. We eat a meal, parts of wheat, cow, and tomato (spaghetti with marinara sauce), chew well, and in the end we've converted it into our own protein, our own carbohydrate, etc. We eat thousands of different kinds of protein -- it really doesn't matter too much which they are -- but in the end they're turned into our own.

How do we do it? The trick is to take the *polymers* we eat -- carbohydrate, protein, lipid, and nucleic acids and break them down into their component *monomers*. While the polymers are different from organism to organism, we all use the same monomers -- simple sugars to make carbohydrate, amino acids to make protein, and so forth (Table 1). So, a better representation of how we eat is seen in Figure 4. We eat a huge variety of proteins and other polymers, break them down to amino acids and other monomers, interconvert them through central metabolic pathways (although humans are partially defective in this regard) to create the proportion of monomers

Table 1. Polymers and their monomers

MONOMER	ACTIVATED MONOMER	POLYMER
simple sugar (e.g., glucose)	nucleotide diphosphate sugar (e.g., UDP-glucose)	polysaccharide (e.g., starch, glycogen)
amino acid	amino acyl tRNA	protein
acetyl CoA	acetyl CoA & malonyl CoA	fatty acid
nucleotide	nucleoside triphosphate	nucleic acid (e.g., DNA, RNA)

suitable to our needs, then synthesize a huge variety of polymers. It is like cars converging from hundreds of surface streets to a few entrance points to the turnpike, leaving at a few exit points to go to hundreds of other surface streets in another city. Very logical, but something needs to control the flow of traffic.

SQ7 . Pick out from the list below those compounds that are polymers. Which are identical whether isolated from snails or whales?

A. phenylalanine	D. sugar	G. nucleoside triphosphates
B. protein	E. acetyl CoA	H. hydrophilic compounds
C. starch	F. actin/myosin	

E. Control points of metabolism

What controls traffic is enzymes, proteins that act as catalysts in biochemical reactions. There is a limitless number of chemical reactions a compound can undergo, but at low temperatures, most biochemical reactions occur very slowly unless enzymes facilitate them. Enzymes provide the roads. At high temperature, a compound can jump the curb, plow through a few houses, and go wherever it likes, but with little control. At high temperatures we burn up. If you make the roads, if you control the enzymes, then you control the traffic. You can determine what compounds are made or not. And that means that you control the cell.

How the *enzymes* are controlled will be a topic taken up later in the semester.

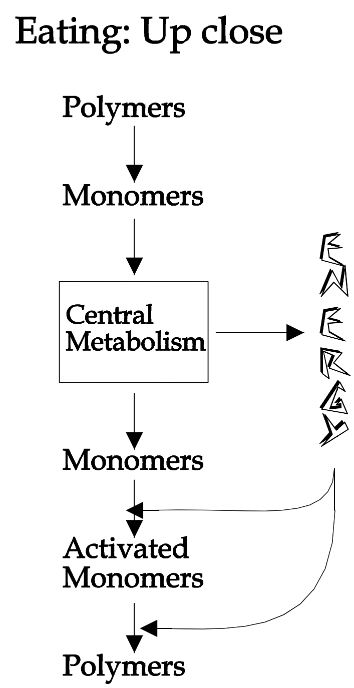

Figure 4. The strategy of metabolism. Polymers are broken down to small molecules, which are interconverted through central metabolism. From these small molecules, the polymers required by the organism are synthesized.

F. Summary

1. What is bioinformatics?
   - Analysis of molecules carrying biological information
   - Analysis of complex biological systems
2. Genes and organisms
   - DNA controls cell function through determining its protein
3. Membranes and self-assembly of cells
   - Cells are delimited by membranes
   - The amphipathic nature of membrane components lead to self-assembly
   - Self-assembly is a common theme in the manufacture of biological structures
4. Metabolism: monomers, polymers, central metabolism
   - Polymers from the outside are broken down to monomers.
   - Monomers are built up into the organism's own polymers.
   - Monomers can be interconverted.
5. Control points of metabolism
   - Reactions are made possible through the catalysis by enzymes
   - Controlling enzyme activity controls metabolism