SCIENCE OF GENETICALLY MODIFIED ORGANISMS II + III

- - -

(Prepared by David Ng, Michael Smith Labs, UBC – db@interchange.ubc.ca)

Moving on to the central dogma...

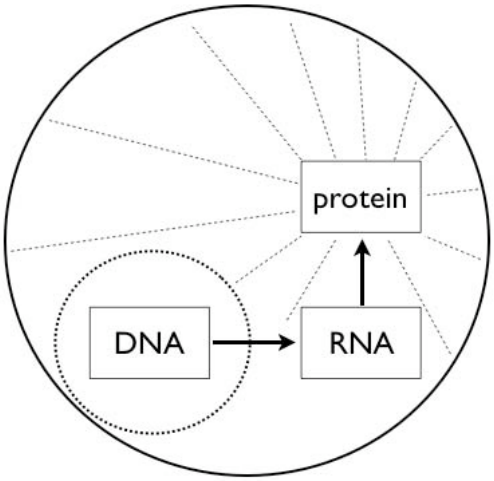
...Which is very much about proteins: in fact Martha Stewart would say, "proteins are a good thing." In life, they are the true movers and shakers of any organism, and are the molecules that actually go through the daily business of living. In short, these are the building blocks that give your various tissues their shape and their function. I bring this up now, because all of this talk about genomes and DNA is only illuminating if you recognize the fact that your genetic code is simply the instructional package for making all of the different types of proteins needed for life.

And things you need for life include: proteins that regulate chemical reactions (for instance, in the conversion of the food we eat into energy); proteins that transport key molecules from one place to another (like the pump implicated in cystic fibrosis); proteins that become the basis of cell structure (like how the architecture of certain tissues is achieved); proteins that facilitate cellular communication (how all the different bits and pieces of your body can work together). In truth, the diversity in protein makeup is responsible for all the diversity in life itself. In other words, bring on the bacon.

In itself, how proteins come about from your DNA code is quite clever. Proteins are built by piecing together molecules that are collectively called *amino acids*. It's a bit analogous to DNA in that if you recall, your DNA code consists of specific combinations of nucleotides. However, whereas our DNA is composed of an alphabet of only four different letters (A, T, C, and G), proteins are built with a much larger alphabet of 20 different letters or 20 different *amino acids.* Again, for any given protein, the determination of which of the 20 amino acids to use and in what order they are to be pieced together is dependant on the nucleotide code itself. This might sound a little confusing but in essence the production of proteins is dependent on dealing with two types of code. More specifically, each combination of *three* nucleotides (often referred to as a *codon*) will signify a particular amino acid. For example, the nucleotide T, followed by a G and another G (or TGG) codes for the amino acid Tryptophan (abbreviated ‘W’), the sequence ATG codes for the amino acid Methionine (abbreviated ‘M’), and so on. The sequence TGGATG would therefore code for two adjacent amino acids, Tryptophan and Methionine. Altogether, there are three letter codons for all 20 different amino acids. In this manner, a long sequence of nucleotides can potentially and theoretically be translated into a long chain of amino acids - i.e. a protein molecule.

To illustrate this two code system, the best example that comes to mind, is the use of Morse code to send messages overseas. In this situation, you essentially have a binary code (dot or dash, two options), which when rearranged into units of three, can translate into one of the 26 letters. For example, *dot dot dot* is the same as the letter 'S', and *dash dash dash* is the same as the letter 'O'. This is a two code system. Your first part being the Morse code element, and the second part

being the formation of words from letters. In our biological example, the first code involves the use of nucleotides to provide information for which amino acids to use, whereas the second code dictates the length and combination of amino acids to form a functionally relevant protein.
In truth, the relationship between proteins and DNA is a little bit more complicated. First, it turns out that the overwhelming majority of the human genome doesn't do anything, and is basically considered to be garbage, junk, filler or if you want to be particularly nasty, crap. This accounts for an astounding 97% of your genome having absolutely no function or significance. This introduces an interesting logistical problem in that humans are using what is essentially a polluted genetic code. In other words, there has to be a system that allows the deciphering of the good stuff from the bad. You don't want to waste your time decoding your junk regions in that it could translate into some random, useless or potentially harmful protein.



eukaryotes!

Secondly, the location of your DNA and the location of protein synthesis are different. This of course, makes no sense because how can you translate your DNA into proteins if the two molecules reside in geographically distinct places? Here, we find that your DNA is found within a small physically enclosed area of the cell called a *nucleus*, and proteins are awkwardly made *outside the nucleus*. Although this nucleus could be viewed as simply a mechanism to "house" and protect your genomic DNA, it does create a rather unfortunate conundrum in that the all important DNA code is not accessible to the machinery necessary for its translation into proteins.

In a rather crafty way, biology has managed to solve these problems through the use of a middle-man known as the *messenger RNA* molecule or *mRNA* for short. For the sake of clarity, mRNA is fundamentally similar in structure to DNA having nucleotides. There is a slight difference but it's visually quite minor - it could actually be the basis of a challenging 'spot the difference' comic. However, thinking conceptually, mRNA is comparable to a no-nonsense piece of genetic code that is constructed from only the useful parts of your DNA. This is similar to having study notes for a particular subject where only the crucial parts are highlighted and regurgitated. Consequently, problem one is solved. Here we have a strategy that can weed out the good from the bad and hence no crap.

Additionally, mRNA is special in that it is a string of nucleotides with the ability to move and ultimately leave the nucleus. You have to remember that your genomic DNA living inside the nucleus of a cell is akin to an elephant stuck in the upstairs toilet. It is simply too big to pass

through doors that might otherwise be situated along the walls of the nucleus. mRNA molecules do not need to be so big. They are much more manageable in size because for each molecule, they contain only the sequence of one protein (not all of them), and more importantly they contain only the *necessary* sequence of that one protein (no junk). This means that problem two is also solved, as mRNA acts as a mobile representative of the genetic code that can now get out and come into contact with components required for protein production.

- - -

Confused? Don't worry, it's alright if it seems a little perplexing right now. I know many people who have had nightmares over this stuff. If you do find yourself waking up in the middle of the night screaming nonsense about RNA and elephants, try thinking of the following analogy.

Because you are such a wonderful person, you wish to prepare a nice chocolate cake for your friend, and to do this, you visit the library to look for a good cake recipe. For some unexplained reason, you are also a huge Martha Stewart fan, which is why you decide to look for a cake recipe in one of her many 'Martha Stewart Living' magazines. After searching for several hours, lo and behold, you find a promising recipe in her 'Weddings Issue', but notice that the magazine itself has a sticker on it that says 'for reference only.' This is a bit of a bother because it means that you won't be able to take the magazine out of the library, and hence, into your kitchen where you had plan to spend most of your time being a wonderful person. Furthermore, despite your best efforts, you can't seem to find any semblance of a photocopy machine anywhere, since this is the sort of library this is, and since the analogy wouldn't work otherwise. Begrudgingly, this small nuisance forces you to look for a pen and a piece of paper so that you can manually scribble down the recipe to take home. As you do this, you quietly think to yourself that your friend had better appreciate all of this effort.

No offence to Ms. Stewart, but I find her magazines are always full of extraneous and in my opinion useless information. Do you really need to know the history of the chocolate cake? Do you really need to know about the appropriate cutlery used for serving cake? Do you really need to see and evaluate 15 different colour schemes for acceptable presentation? I don't think so. All you really need to concern yourself with is the ability to make the cake taste good. This is why, when you go to the bother of copying down the recipe, you don't include all of the nonsense - you just copy down what you need. In short, this turns out to be just a few lines of ingredients and directions scrawled neatly on your piece of paper. The crucial point is that you can now freely walk out of the library with the recipe in hand.

Next, of course, is a trip to the local grocery store where you would get all the necessary cake ingredients and maybe indulge yourself with the smutty magazine about child actors gone bad. After which, you would head home and bake a wonderful chocolate cake which is met with such praise, that you are glad you didn't waste your time using table setting number six for the occasion.

A strange story indeed but here is how the analogy works. First, you need to envision the entire library with all of its resources as the genome, and also envision the building itself as the nucleus. The complete recipe found in the magazine actually represents the genomic sequence for one particular protein. As mentioned before, Martha publications tend to have a lot of useless information, some of which is not even directly related to the production of the cake (advertisements and historic footnotes). This is identical in premise to the crap in your genomic material, and the concise notes you scribbled down symbolize the messenger RNA molecule. This, as mentioned before, is twofold important because, (1) it represents the minimal amount of

information needed and, (2) it represents the ability to leave the nucleus (in this case, the library) and the ability to go to places where protein production can take place (in this case, the rest of the world, but more importantly the grocery store and your kitchen). Finally, the cake itself represents the protein. Remember I said that in living systems, it's really the proteins that are the real movers and shakers? They are certainly the most interesting parts of the big biological picture, and wouldn't you say that the cake itself is the most interesting part of this process?

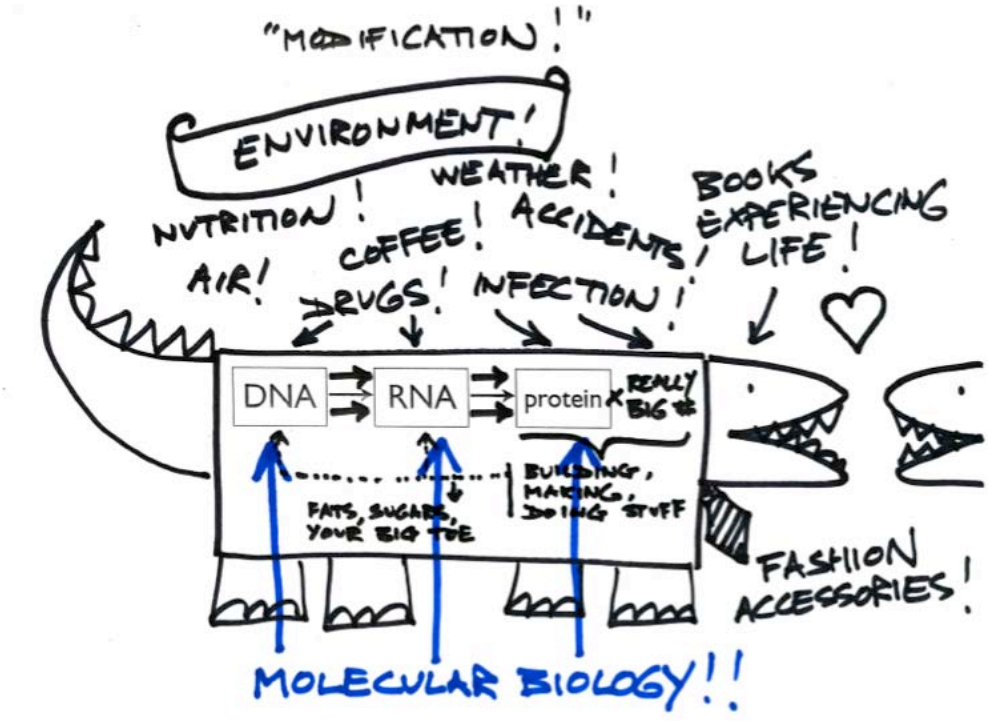
- - -

And... just to add an additional degree of complexity, prokaryotes do it differently (no nucleus, DNA tends to be overwhelming composed of useful bits) – Very Rachel Ray...

- - -

MODIFICATION? WHAT SAY YOU?

Let me count the ways to modify



- - -

So how exactly so you modify organisms using molecular biology?

Well, central to this question is the notion of a "vector." In genetic talk, this refers to "*an agent that transfers genetic material from one cell to another*."

From a more specific viewpoint, a vector needs to provide a few pertinent functions. Here's an abbreviated list to get into the feel of things.

DELIVERY: gotta have a way of getting the DNA inside... Techniques can range from the rudimentary (damage cell membrane and hope that DNA can fall in), to the sophisticated (use a virus or other infectious agent to transfer your DNA inside).

UPKEEP: Naked DNA (no matter how important it is to your research) does not upkeep itself. Therefore, to do this, you either need your "**DNA of interest**" to replicate independently of integration into host genome. i.e. the vector allows the use existing replication machinery of host organism. OR another way to get around this is for the vector to contain information that can drive direct integration into the host genome.

THEY'RE ENGINEERED FOR FUNCTIONAL REASONS: all vectors generally have defined places where you can insert your DNA of interest. This is usually at a place governed by (i) convenience (i.e. easy to put something in this area), or (ii) geography (you put your DNA "here" because it is next to some element – say a genetic element that allows this or that to happen on your DNA of interest." i.e the promoter example we went over.

IT HAS SOMETHING THAT LETS YOU KNOW IF THE VECTOR IS IN THE ORGANISM:
All vectors will generally contain a **selectable marker** (notable exception is where GM crops are involved). This is usually a gene that confers resistance to some sort of drug. Commons ones used in research field include antibiotics such as **ampicillin** (common for e.coli),.

- - -

Some examples of GMOs...

1. Instead of producing pharmaceutical products in the milk of transgenic animals, there have been reports to produce pharmaceutical products in the semen of transgenic pig.

2. A company called "Genetic Savings and Clone" provides gene banking and genetic services for clients. This also includes the cloning of your favourite pet.

3. There exists a Nude mouse which is a mouse strain that has no hair and also no immune system. These mice have been used for the production of human ears.

4. The song you are listening to is the title track of a Billboard number one CD. The song and CD are specifically written about and dedicated to the first human clone.

(Which one is false?)

- - -

LET'S MAKE A GMO!

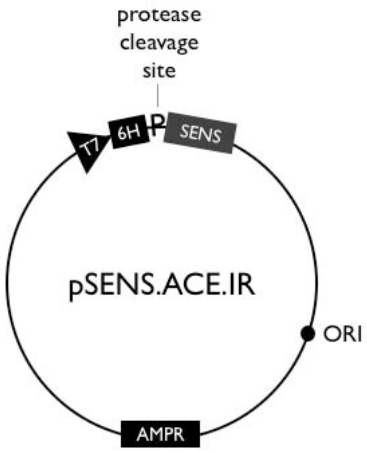
What would we like: we want to make as much SENS protein as possible!!!

Think about the host system, you choose to use. Dr. Sens vs e.coli. (high vs low maintenance, yield potential)
Choose a vector system that works in something like *e. coli*. (pUC18) with "ori" or *origin of replication* (upkeep). e. coli is also one of the easiest organism to introduce DNA into.

If we want to make protein, we need RNA being produced for the SENS DNA. Need an RNA promoter. Let's use one that is very strong (T7 RNA promoter) – this one is so strong that when inside an e.coli, the organism pretty much shuts down and just focuses on the vector you've intoduced. On a good day, you can tweak it so that this vector drives the expression of a protein such that 50% of its cell mass is your thing of interest (good yield!)

O.K. This vector is great at making buckets of RNA for the SENS protein, which in turn can be read to produce the protein. woo hoo!

Wait a minute, if I plan on using this protein for commercial, possible therapeutic use, the regulations on such a thing are pretty severe. One of the logistics I need to ensure is that any SENS protein I use, market or produce has to have very high standards of quality control. Part of this is ensuring that it is very pure.



Hmmm... even with a great 50% cell mass yield, that still infers that there 50% other "stuff." Plus, purify proteins is a real pain.

Maybe, I can tweak my vector a bit more to make my protein easier to purify. INTRODUCING FUSION PROTEIN TECHNOLOGY. 6 histidines as an example. Histidine is an amino acid, but 6 in a row turns out to be loving nickel ions. This means, that technically, if we can engineer our SENS protein to also have 6 histidines at the front of its sequence, we'll be able to grant it the ability to have good binding with nickel ions.

Why would we do this? Well, if you prepare beads coated with nickel ions, you now have an effective way to purify your SENS protein of interest (i.e. only the 6HIS-SENS fusion protein will stick to the beads – by virtue of the nickel ion – and voila, you have easy purification).

Cool.

Except that now the protein we've produced isn't exactly the SENS protein, it's a 6HIS-SENS protein. Hmmm... that might be or might not be a problem, but regardless, somebody is likely going to have an issue with that (i.e. that 6HIS tag you've put on the front there – how do you know it's not messing things up?)

This is a fair argument. So... let's tweak the vector again – let's add a protease cleavage site right after the 6 histidine tag, and before where the SENS protein would be.

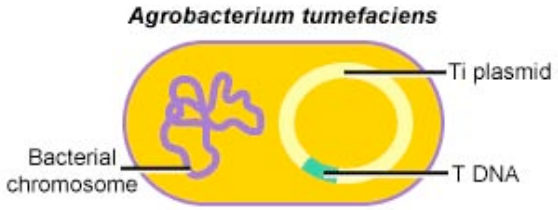
That means, that we can still utilize the 6HIS tag for purification reasons, but we now also have an "out" to get rid of it (i.e add a protease that will cleave the SENS protein away from the 6HIS tag.

TA DA!

- - -

In light of an emphasis on Genetically modified crops, let's spend some time looking at the act of delivering DNA into plant cells.

Agrobacterium genetically transforms its host by transfering a well-defined DNA segment from its tumor-inducing (Ti) plasmid to the host cell genome. In nature, the transferred DNA (T-DNA) carries a set of oncogenes [cancer inducing genes], the expression of which in plants cells, leads to neoplastic [uncontrolled, undifferentiated] growth of the transformed [cancerous] tissue and the production of opines, amino acid derivatives used almost exclusively by the bacteria as a nitrogen source [tricks the host cell to make good stuff for itself]. Recombinant *Agrobacterium* strains, in which the native T-DNA has been replaced with genes of interest are the most efficient vehicles used today for the introduction of foreign genes into plants and for the production of transgenic plant species. **(Tzfira and Citovsky, Current Opinion in Biotechnology 2006, 17:p147)**



Some examples:

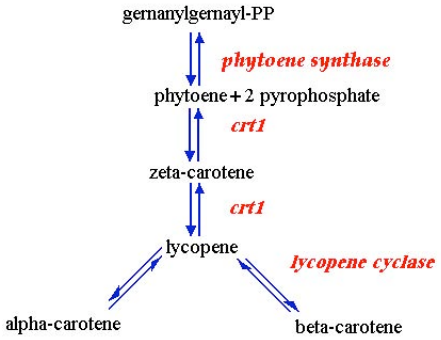
ONE: INSECT PROTECTION
Bt (*Bacillus thuringiensis*) plants (corn).
This is a soil dwelling bacterium, which at a stage of their life cycle can form crystals of a protein insecticidal endotoxins collectively known as Cry toxins. These toxins have a variety of specific targets against different types of insects (lepidopterans – or moths and butterflies), when toxins reach the gut of said insect. Primary target in industrial use is against the European Corn Borer – useful because the Borer spends most of its larval life (when it is most susceptible) inside the corn stem. Therefore conventional Bt sprays are much less effective, as they have to rely on specific moments when the borer is outside.

TWO: HERBICIDE TOLERANCE
Roundup Ready
Roundup is broad spectrum herbicide (from Monsanto), with glyphosate as the active ingredient. Very widely used. Functionally, glyphosate acts by inhibiting the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS). This is an important enzyme necessary for the shikimate pathway necessary for the production of certain ring containing amino acids. i.e. it's necessary for plant survival, but this also happens to be a plant specific pathway.
- Essentially, there exist variants of the EPSPS that are resistant to roundup, and consequently it the code for these variants that are incorporated into Roundup Ready plants.

SOME JARGON to throw at you from the Health Canada website (not necessary that you totally follow it, but some of it should sound more familiar to you now).
"Constitutive expression of the CP4 EPSPS encoding gene was controlled by inclusion of sequences from the rice actin promoter (P-ract1) or cauliflower mosaic virus promoter (e35S) and the 3'-polyadenylation signal of the nopaline synthase gene (NOS 3') from Agrobacterium tumefaciens. The second cassette also contains the maize hsp70 heat shock protein to stabilize the level of gene transcription. The plasmid vector, PV-ZMGT32L, utilized for transformation of line 603, did not contain antibiotic marker genes or a bacterial origin of replication. Segregation and Southern blot analysis across multiple generations was used to demonstrate the stable inheritance of the novel trait."

THREE: GOLDEN RICE
A GM rice produced to synthesize beta-carotene, a Vitamin A precursor, in the edible parts of the rice plant. It does this by introducing the genetic information required for expression of certain enzymes necessary to allow endosperms to form beta carotene.

Here is the metabolic pathway to make beta carotene. The rice plant can totally do this, except that ironically, it can't do this in the endosperm (the white part that most people eat).



T-DNA contains *psy* (phytocene synthase) and *crt*, under a endosperm specific promoter. (*psy* is from the daffodil and *crt1* is from Erwinia uredovora, a soil bacterium). This is because these are the two enzymes missing from the rice's endosperm tissue, so reconstitution will allow expression of beta carotene (orangy colour AND a precursor to Vitamin A production).
WHY vitamin A? Second most prevalent nutrition deficiency (after Iron) in the world, particularly affecting developing countries. Symptoms include blindness.

- - -

SPIN: "*Feed the hungry of the world*" vs "*It's killing butterflies*" (both baloney)

A few quick words to segue into Allen's section...

Health: allergenicity, toxic effects.

Environment: Spread (contamination – Mexico Oxaca Corn), resistance (Roundup), unintended targets (Monarch butterfly)

Economic: loss of niche, terminator technology, seed culture. labeling, corporate control

- - -

You too can win a Nobel Prize!

A few prep words for the lab which will be coming up in a few weeks.