

Part I

Conversational Molecular Biology: An Aspirant's Guide

1 A Brief History of the Language

Molecular Biology of the Gene by Watson, et. al., Chapter 3

1.1 Why speaking English is not enough

It is crucial that anyone entering the neurosciences be fluent in the language of molecular biology, and there is no excuse for any of you leaving this program without a comfortable knowledge of the subject. There are two main reasons for this:

1. The tools of molecular biology allow us to formulate and answer questions which must be tackled on the way to constructing a theory of brain function.
2. The history of molecular biology serves as inspiration for investigations like the ones we are undertaking in neuroscience, showing that such studies are not completely hopeless.

This will be a very short introduction and will help answer the questions on the qualifying exam, but I hope to be invited back to shore up the knowledge that you need to have about molecular biology.

1.2 In the Beginning there was a Void

Questions concerning the creation and workings of life and the universe have, for thousands of years, been the domain of religion and mysticism. These have been opened to scientific examination in the last couple centuries, but even today people disbelieve the explanations science gives in the areas where scientists claim the answers are known, and people refuse to believe that science can ever explain phenomena in areas like thought or consciousness,

where science has no firm foundation. We, as human beings, decide and move, create art and religion, inspired by a spirit which we have believed to be separate from the physical, knowable world, and made of different stuff. When the organic chemists synthesized urea in the laboratory from inorganic material it was a profound demonstration that living organisms are made from the same material as non-living organisms, and the same chemistry governs us all. But well into this century, after physicists had overturned Newtonian beliefs about the structure of the physical world, there were still great debates about the substance of life and the mechanism of inheritance. Shrodinger was not the only physicist to turn his attention to biology, when he wrote his monograph, *What is Life?* trying to develop a framework for the understanding of this very ephemeral question. The discoveries in the ensuing 40 years have very convincingly created a vocabulary and a grammar for genetics, and the result is a fully developed language, which took many words from early genetics and pinned them down to clear and simple meanings. And really, the basic vocabulary of molecular biology today is so simple to learn, understand, and use, that we forget that before the 1940s, genetics (at least in terms of a molecular formulation) was composed of a confused and incoherent babel of languages. I'm emphasizing this because neuroscience, which sounds a lot like 1930s and 40s genetics, can look to the development of molecular biology as an example of the language-building process that we will have to begin in order to have some clarity about the question, *What is Thought?*

2 What is Life?

2.1 The Problem

There is an immense diversity of living organisms, each of which, from the lowest to the highest, is able to produce offspring, nearly identical to the parent. In addition, we know that even human beings require only two cells to encapsulate everything necessary to produce another human being. In order to understand genetics, we want to look inside these two cells and find the material that makes a human being.

The first step is to look and find chemically, what does the cell contain? So in the 30s and 40s, using the chemistry then available, what chemicals do

you think they found in the largest quantities? (DNA, RNA, proteins)

2.2 DNA, RNA proteins, ha ha. DNA, RNA, proteins, ha ha.

DNA and RNA are nearly identical. They are both composed of long chains of four *nucleic acids*. The difference is that RNA has an additional alcohol (hydrogen-oxygen) group on each nucleic acid *base*. This alcohol group is a major reason for the instability of RNA molecules. I'll skip ahead in history a bit, to tell you that each nucleic acid base has a particular partner which it is attracted to. The four symbols for DNA bases are A (*adenine*), as in *ATP*, T (*thymidine*), C (*cytosine*), G (*guanine*) as in *G-protein*. Adenine and Thymidine form pairs, as do Guanine and Cytosine. DNA in cells is most often found in long, long, complementary chains, (sometimes millions of nucleic acids long). When you hear about chromosomes, they're nothing more than extremely long chains of DNA, surrounded by various proteins and such, which squash the DNA down to a size that will fit in the cell *nucleus*. And proteins are chains made of *amino acids*, all variations on a theme (amino group, carbon attached to one of 20 side chains, and a carboxylic acid). The appendix to this paper, taken from **Molecular Biology of the Gene**, shows the molecular structures of amino acids and nucleic acids. You'll be surprised at how small and simple these molecules are.

So researchers interested in the molecular basis of life had a suspicion that proteins, DNA and RNA were involved, and they further were learning that chemical reactions were carried out in living cells by special molecules which they called *enzymes* (check it on your vocabulary list). They then made the discovery that these enzymes were proteins, made as chains of connected *amino acids*. Now there was also a traditional genetics idea of a gene, as the basic unit of heredity, and this concept of a gene was somehow related to the traits of an organism. There was evidence gathering that genes are responsible for the synthesis of single enzymes. So, if proteins direct most metabolic processes and synthesis of new chemicals in the cell, what's the genetic material made of? How do you test this idea?

In 1928, Fredrick Griffith, in England, found that a certain strain of bacteria, which had lost its ability to cause pneumonia, could regain that

ability if it was mixed with pneumonia-causing bacteria which had been killed. Oswald Avery, at Rockefeller took this discovery as the basis for his studies of inheritance. He ground up the pneumonia-causing bacteria, separated the DNA, RNA and proteins, and combined each of these fractions with the nonpathogenic bacteria. Can you guess the result? When the DNA was mixed with nonpathogenic bacteria, the bacteria were transformed, and produced the coat necessary to cause pneumonia. So now you know the answer: DNA is the genetic material in bacteria. But the amazing point is that after years of arguments and experiments, it was shown that the same DNA, made of the same four building blocks, *nucleic acids*, is the genetic material in all organisms from bacteria to humans. (If you think this violates the Hershowitz-Toyota law, you are right: and if you ask, "Are there organisms which use RNA or protein as their genetic material", the answer is yes. *retroviruses* and *prions*, a class of proteins like those associated with scrapie, are famous upholders of the H-T law.)

DNA was able to pass traits from one organism to another. But in order to be the genetic material, it had to do two things:

1. Somehow direct the synthesis of proteins.
2. replicate.

The second problem was attacked by Meselson and Stahl at Caltech, who showed, in some ingenious experiments, that DNA has two halves, which split during *replication*, one half going to each daughter DNA. The subject of replication is fascinating, but it doesn't fit into the Central Dogma, which I am building up to.

2.3 How Does DNA Make a Protein?

They had two halves of the solution: DNA transfers genetic characteristics between organisms and proteins are the end product of the genetic material. But there was a third half which couldn't be explained: proteins are often produced outside the nucleus of a cell, in places where there is no DNA. You probably can guess the end: the character which we've been ignoring until now finally finds its role: RNA is copied from DNA, and forms the *template* for production of proteins.

There are a lot of holes in the history I've given so far, but I think you know enough to appreciate the syntax I'm about to teach you. This syntax forms the foundation for the whole language of modern Molecular Biology, and with this syntax you should be able to use some of the basic vocabulary. But as with any language, you really have to practice it, and you can do this by going to talks in Molecular Biology, or join a molecular biology language table.

3 The Central Dogma

DNA RNA Protein

DNA is made of two *strands*. Sections of these strands are separated, and then *transcribed* into RNA. Transcription involves matching each nucleotide of DNA with its complementary RNA nucleotide, and forming a complementary RNA chain from the DNA template (e.g. if the DNA template is 3'-GGCCCTTTTAAAAA-5', the RNA chain will be 5'-CCGGGAAAAUUUUU-3'). The complementary RNA copy of DNA is called messenger RNA (*mRNA*). Transcription is done by proteins called *polymerases*. RNA is much more unstable than DNA, and is also quickly attacked by enzymes called RNases which are everywhere in cells. So RNA must quickly be *translated* into proteins, before it's destroyed. Translation is done on *ribosomes* which match three nucleic acids at a time on the RNA molecule to three nucleic acids on a special kind of RNA, *transfer* or *tRNA*. There are about 20 different tRNA molecules, each of which is attached to an amino acid. The tRNA recognizes the three matching bases on the mRNA, connects its amino acid to a part of the ribosome, and moves on. The next tRNA comes along, recognizes the next three bases on the mRNA and attaches its amino acid to the previous amino acid. And amino acid by amino acid, a protein is created.

This whole process allows us to redefine a fuzzy word from genetics in a clean and satisfying way. That word is the most important word in molecular biology: *gene*. A gene is a piece of DNA which codes for one protein.

4 Amazing DNA (and RNA and protein) tricks

4.1 Restriction enzymes

Molecular biology really became a branch of engineering with the discovery of restriction enzymes. These are proteins that bacteria make, probably as part of a primitive immune system, which cut DNA chains whenever a specific sequence (usually 3-5) of *nucleotide* bases occurs. With hundreds of commercially available restriction enzymes, and other enzymes called *ligases*, which join cut pieces of DNA, virtually any combination of existing DNA chains can be created by cutting and pasting.

4.2 Putting genes into cells

In bacteria, viruses, and even *eukaryotic* (nucleated) cells, many methods exist for putting in new genes, and getting them to *express* specific proteins. But what if I want to know the role of a protein in the immune system, metabolism or behavior of a whole animal?

4.3 Transgenic animals

- Recipe for 1 transgenic animal:
 1. Take one container *embryonic stem cells (ES cells)*
 2. Inject the DNA you want. Let's say the DNA codes for the serotonin receptor.
 3. Find a cell which contains the DNA you injected.
 4. Make sure the cell and all the cell's daughter cells express serotonin receptor.
 5. Place the cell in a growing embryo, and hope one of its daughter cells becomes a reproductive cell (egg or sperm)
 6. Breed the *chimeric* animal with another animal.
 7. Find the offspring that produces the serotonin receptor in all of its cells.

A knockout mouse is produced as above, but the injected gene replaces the animal's normal gene, and since the animal has two copies of the normal gene (maternal and paternal), the *heterozygous* animals—with one gene knocked out—must be cross-bred to give *homozygous* offspring.

4.3.1 Don't count your knockouts before they hatch

Nature is an skilled and inventive foil of experiments. Often, a single gene changed in an animal does not change any behavior or trait of the animal, because other systems compensate. Or, the gene changes too much, and the animal dies before the experiment can give any results.

More specific genetic manipulations are becoming available, allowing expression of genes only in certain tissues and at certain times. These are still very labor intensive, and it's not usually a good idea to volunteer to create these transgenic animals.

Other neat tricks can be found in the vocabulary list.

Part II

Vocabulary List

5 Essentials

DNA Deoxyribonucleic acid, the fundamental genetic material, composed of nucleic acids, and encoding the precise sequence of amino acids in proteins

cDNA complementary DNA. A DNA copy of mRNA. Often used to create libraries.

RNA Ribonucleic acid. Like DNA, RNA is a polymer composed of nucleic acids. RNA is more unstable (short-lived) than DNA, and is used directly in the process of protein synthesis.

mRNA Messenger RNA. A complementary copy, of DNA (where every A in DNA becomes U, T becomes A, G-C, C-G). Encodes the linear sequence of amino acids in a protein.

tRNA transfer RNA. Each tRNA is attached to an amino acid. When the tRNA "anti-codon" matches the three nucleotide mRNA code, tRNA transfers its bound amino acid to a growing protein chain.

library a mixture of molecules, such as RNA, peptides or antibodies, which can be screened to find a single molecule, or class of molecule with the desired function.

promoter A DNA sequence which directs transcription from a downstream gene.

enhancer A DNA sequence which, when present with a promoter, increases the transcription of a gene.

DNA binding protein: zinc finger, etc. A protein which attaches to DNA, often modulating the transcription or replication of the DNA to which it is attached. Many of these proteins have characteristic motifs, such as "zinc fingers" or "helix-loop-helix".

codon A three nucleotide sequence in DNA or RNA which uniquely encodes a particular amino acid

mutation A change in the nucleotide sequence of DNA.

missense and nonsense mutations mutations in the coding region of DNA (i.e. in a gene). Missense mutations lead to a change in amino acid sequence, nonsense mutations lead to a halt in transcription and therefore a shortened protein.

transcriptional regulation Control of the production of proteins at the RNA level: control of the rate of transcribing DNA to RNA. This is accomplished by a variety of sequences within the DNA (promoters, enhancers, repressors, etc.), and their associated binding proteins. Transcriptional regulation is the major control point for the production of different proteins in different celltypes.

translational regulation Control of the production of proteins by changing the rate of translation of RNA to protein. Many translational regulators function by modifying or interacting with ribosomal components.

protein A polymer of amino acids. Proteins form the primary structural and functional elements of a cell. Almost all cellular processes, from cell-cell signaling, metabolism to chemical synthesis are carried out by proteins.

alpha helix A helical formation of amino acids common to many proteins.

beta sheet A planar formation of amino acids, found in many proteins.

enzyme A biological molecule which catalyzes reactions. Most enzymes are proteins.

ase a suffix denoting that the molecule is an enzyme (e.g. kinase, transcriptase, protease...)

amino acid A basic building block of proteins. See attached sheet for the basic structure of amino acids. Proteins are composed of chains (sometimes hundreds of amino acids long), which contain 20 different types of amino acids.

nucleic acids The basic building blocks of DNA and RNA. There are 4 nucleic acids: Adenine, Thymidine, Guanine, Cytosine. RNA contains Uracil instead of Thymidine.

ATP Adenosine triphosphate. A nucleoside which is used throughout the cell as an energy source. ATP is also added to a growing chains of RNA. deoxyATP, or dATP, is added to DNA chains.

cAMP Derived from ATP, a molecule central to second-messenger pathways.

G-protein A protein, often found in cell membranes, which converts GTP into GMP, as the beginning of a second-messenger signal sequence.

transcription Creation of a complementary strand of RNA or DNA: usually an RNA copy of DNA.

polymerase An enzyme which transcribes RNA or DNA.

RNA polymerase A polymerase which produces an RNA copy from a DNA or RNA template.

DNA polymerase A polymerase which produces a DNA copy from a DNA or RNA template.

reverse transcriptase A polymerase which produces a DNA copy from RNA.

translation Production of proteins from an mRNA template.

ribosome A large complex composed of many protein and RNA components, which translates proteins from RNA.

restriction enzyme An enzyme which recognizes 4 to 5 nucleotide restriction sites in double-stranded DNA, and cuts the DNA at specific nucleotides associated with the restriction site. Restriction sites are usually palindromic, e.g.: GATC CTAG

ligase, ligation An enzyme that joins two cut strands of nucleic acids. Usually used to fuse two sections of DNA cut by restriction enzymes.

6 Techniques

DNA sequencing Determination of the linear sequence of nucleic acids in a section of DNA

restriction digest An enzymatic reaction for cutting DNA in specific locations.

electrophoresis Chromatography technique, used to separate DNA or proteins by charge/mass. Larger DNA pieces migrate slower than small pieces. DNA can be easily seen by adding Ethidium Bromide (a fluorescent, DNA binding molecule) to the electrophoresis gel, and placing the gel over a UV light.

electrophoresis gel A polymer (either agarose or acrylamide), that looks like a flat, rectangular block of jello. DNA or proteins are injected into the gel, and an electric field is placed across it. Molecules move through the gel, and are separated by charge, size and/or molecular weight.

PCR A reaction to produce large quantities of DNA from small quantities (as few as 1 molecule). PCR works like a pyramid scheme, with each round of reaction producing two copies of the original molecule.

rtPCR A variation of PCR, using reverse transcriptase to copy RNA into DNA, in order to make a template for production of large quantities of DNA copies of the original RNA molecule(s).

chromatography Any of a large variety of techniques to separate a mixture of molecules into its components. Usually this involves forcing the molecules through a separation matrix, in which some molecules move slower than others.

HPLC High Performance Liquid Chromatography. This technique can be used to separate and measure the quantities of molecules from cell extracts, tissue samples, etc. A "column", packed with a gel matrix is used, and the mixture flows through the column.

gas chromatography Similar to liquid chromatography, but high temperatures are used to volatilize the sample into the gas phase before pushing it through the chromatography column.

gene transfer Any of a variety of methods to place new DNA or RNA into cells, and get the cells to produce a desired protein. Some of the methods include viruses, Ca⁺⁺Phosphate, electroporation or direct injection of DNA. [gene expression] placing a gene in a cell so that the cell produces a desired protein. As in "Can you express that channel in xenopus oocytes?"

oocyte An egg cell. Xenopus (frog) oocytes have few native ion channels, and so they are a favorite for expressing ion channel genes to study their physiology.

transgenic Introduction of a gene into whole animals in such a way that the gene is inherited by future generations. This is a labor intensive process. Currently transgenic technology is predominantly used with Drosophila and mice.

hybridization Detection of biomolecules, by using a fluorescent or radioactive probe which attaches to the DNA, RNA or protein of interest.

Southern, Western and Northern blot hybridization Southern blots are used to detect DNA in a sample or tissue with a known DNA probe. Northern blots detect RNA, also using a known DNA probe, and Western blots use antibodies to detect proteins.

antibody A protein produced by the immune system which very specifically binds to small regions of molecules. Typically antibodies can be made for any large molecule (e.g. protein), or small molecule attached to a larger molecule.

ELISA Enzyme-Linked ImmunoSomething Assay. An enzymatic system for detecting the concentration of a protein in a solution. Typically, an antibody is titrated and added to small wells containing the protein of interest. The more antibody binds to the well, the more protein was in the well. The more antibody binds to the well, the larger the fluorescent/colorimetric signal. This method is quick, and is used a lot in clinical screenings for diseases: e.g. the HIV test, pregnancy tests, and many many more.