Lecture 13

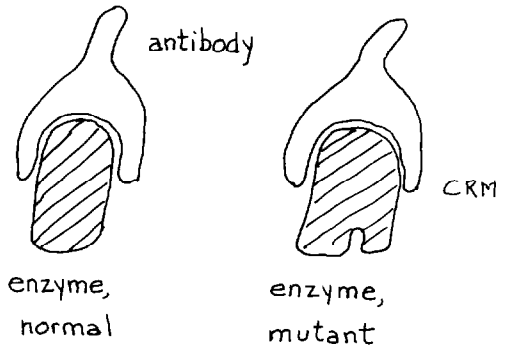
We have seen that the sequence of amino acids in a protein is determined by a series of codons on an mRNA, each codon consisting of three nucleotides. Each codon lines up a tRNA molecule (with its amino acid attached) by base pairing between the codon on the mRNA and the anticodon on the tRNA. Remember that pairing is always between strands of opposite polarity. Each amino acid is specified by one or more different codons, a property we have called degeneracy. But there is also redundancy: for a given codon there is often more than one tRNA. Evolution has provided a safety factor against mutations that would cause disaster if they incapacitated the only tRNA for a given codon. (Keep in mind that tRNA's, like all RNA's, are transcripts of specific DNA sequences.)

Missense mutations

Let us see now what happens when a mutation changes a codon in the gene specifying a sequence of amino acids. If the new codon happens to correspond to a tRNA that inserts the same amino acid, there is no change in the protein. Most commonly, however, the new codon corresponds to a different amino acid. That is, a different amino acid is inserted in the protein. We talk of acceptable missense mutation if the protein with the substituted amino acid is still tolerably functional; that is, the substituted amino acid is similar enough to the one it replaces or at least does not cause too much distortion. Otherwise, the protein fails to fold or if folded is nonfunctional: we call these unacceptable missense mutations.

td mutations

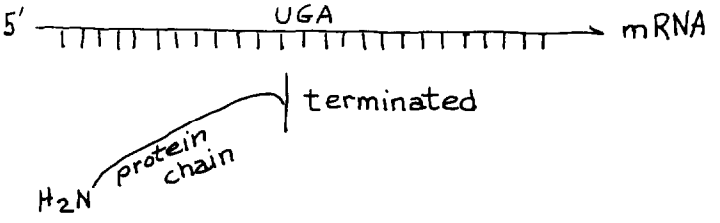
An interesting class of unacceptable missense is temperature-dependent (td) missense: the resulting protein is functional in a permissive temperature range, usually lower than the nonpermissive range. (Can you figure out the reason, from what you know of the bonds that keep together various parts of a protein molecule?) If a totally unacceptable missense mutation happens, the faulty protein may be completely undetectable. Sometimes, however, one can actually find the useless mutant protein: if you have an antiserum that reacts specifically with the corresponding normal protein, you can use it as a reagent to search for the mutant form of the protein.



If you find it and isolate it, you may find out which amino acid has been changed. A mutant protein detected in this way is called a cross-reacting material or CRM (pronounced crème, as in crème de menthe). This technique is very useful in tracing modified proteins in genetics studies.

Termination

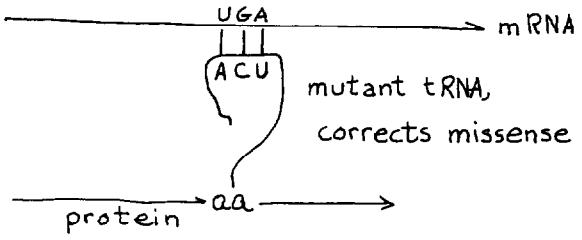
You can readily guess that there is another possibility: a base change in a codon may turn it into a nonsense codon. There are three nonsense codons UGA, UAA, UAG, which correspond to no amino acid. They are the termination signals. When the protein synthesizing machinery encounters one of these codons synthesis stops. There is no tRNA corresponding to these codons, so that no amino acids can be inserted. But the signal is somehow recognized by the ribosome and causes the growing polypeptide chain to fall off the ribosome complex, to detach itself from the tRNA, and to become free.



Suppressors

The story of how the nonsense codons were discovered is quite interesting. Since it involved the study of bacteriophage growth we shall wait a few lectures before returning to it. For the time being let us grasp the main principle. Consider for example a mutation that changes the DNA triplet CTT, which generates mRNA codon GAA, to become triplet ATT, corresponding to codon UAA on the

mRNA. Since UAA is a termination signal, synthesis of that protein stops, the cell produces only a fragment of that protein, and if the protein is a vital enzyme the cell dies--unless...unless the cell (say, a strain of E. coli) happens to have a mutant tRNA, which can more-or-less efficiently put an acceptable amino acid at the place where the UAA codon exists. This phenomenon is called suppression of nonsense. A mistaken tRNA corrects a mistaken codon and saves the cell. [Don't you wish there were more suppression of nonsense altogether?]



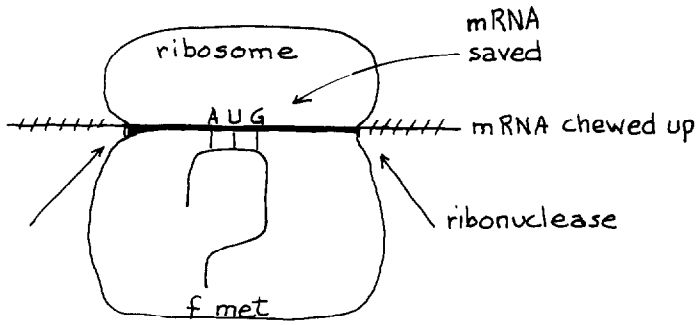
There is another worry, however. If suppression requires a mutated RNA, what happens to the codon that was read by that RNA? The way out is that the suppressor tRNA's must be, not the major tRNA's that are used everywhere to put in amino acids, but some minor, low-grade tRNA's. The cell can afford mistakes made by these tRNA's in the same way that people can afford losing their appendixes but not their hearts.

Notice some dangers of suppression. How could a cell survive if an amino acid were inserted whenever a natural termination signal occurs? Proteins would never end. The safety play is that suppression is usually inefficient, a few percent at most. It causes little trouble to a cell if 1 percent of its proteins are not terminated properly, whereas even 1 percent of an essential protein made at the site of a mutant gene may keep the cell going and growing. Practically all bacteria do have some levels of suppression, which suggests that the mechanism is important in evolution for maintaining and trying out mutated genes.

Each tRNA molecule must be recognized by the enzyme that puts the appropriate amino acid on it and by the codon on mRNA. If a suppressor mutation occurs in the anticodon, the new anticodon may pair either with the codon for another amino

acid or with a termination codon. Remember that each tRNA represents the product of a specific gene.

So much for termination. Let us now see what happens at the other end of the protein. There must be an initiation signal, which puts methionine (or formyl methionine) at the start of the chain. But the code has 61 signals for amino acids and three termination signals; there is nothing left for a start signal! In fact, the start is signalled by one or more special sequences of nucleotides that end with AUG (methionine) but are positioned in a special way on the ribosome. The "start sequence" is still poorly understood. One way to try is to use a pure messenger RNA, for example a viral RNA, in a Nirenberg system that lacks all the amino acids except the starter methionyl-tRNA.

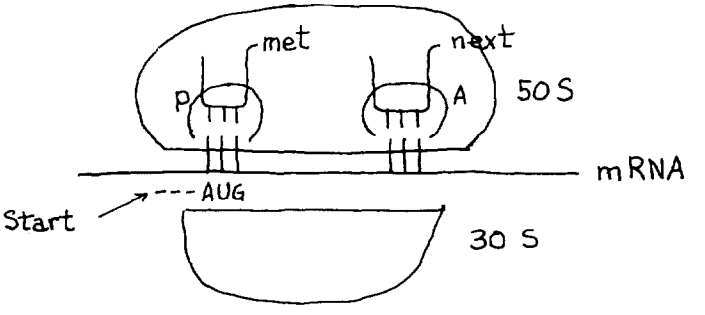


The starting stretch of mRNA becomes trapped in the groove of a ribosome. Then one adds an enzyme that digests away free RNA so that only the part trapped in the ribosome remains and can be analyzed. It seems that there is not one start sequence, but a set of start sequences, all ending with AUG.

The special tRNA for formyl methionine (or for methionine in animal cells) has the anticodon UAC corresponding to codon AUG. Why does it not start a new polypeptide chain every time it finds an AUG codon for methionine? Because it is a special tRNA, which works only at the beginning of the proteins.

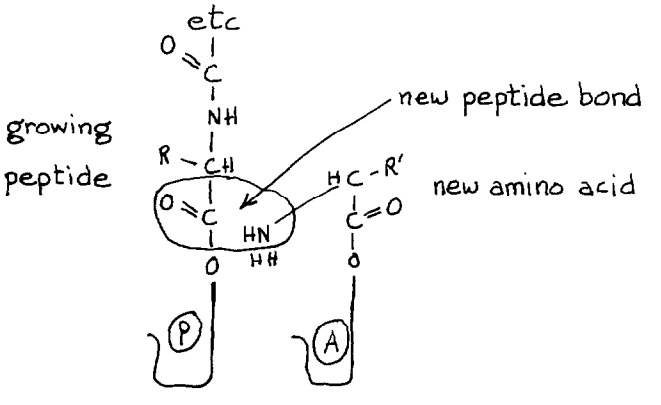
To explain this, let us leave the genetic code for a moment and go back to the biochemistry of protein synthesis. We already know that proteins are made on ribosomes. Each ribosome consists of two subunits, a smaller and a larger one. Each

subunit has one rRNA molecule and 30 to 60 different protein molecules. In a cell free ribosomes separate into two parts. The "starting" tRNA, with



its methionine or formyl methionine, associates with the small subunit and with the start sequence of mRNA. Then a large ribosomal subunit joins in and the complex is ready to go (provided a number of other ingredients are also present; let us forget about them).

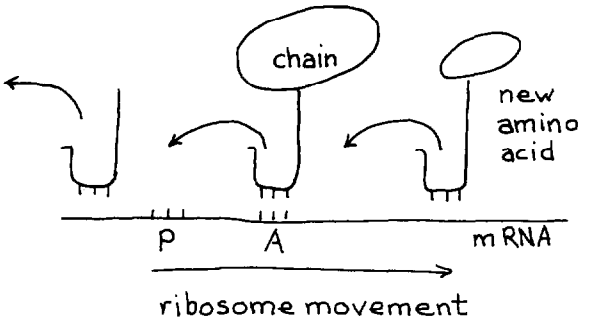
The ribosome has two adjacent sites, A and P, each of which accommodates a tRNA. Usually a tRNA with its amino acid would make contact with its codon at the A site. The "start tRNA" (met or formyl met) is unique because its structure somehow makes it go to the P site of a messenger ribosome complex, where the "start AUG" codon is waiting. In this way, the starting amino acid becomes located at the P site, ready to accept the next amino acid to the A site, according to which codon



is there. The next step is the enzymatic joining of amino acid number two to number one. The α-amino group of the second amino acid attacks the

bond between the first amino acid and the terminal nucleotide of the tRNA. A peptide bond is formed, the first amino acid jumps piggyback onto the second. The polypeptide chain has been started.

Next, the unloaded tRNA on P is shoved off and the loaded one on A moves to P (or rather, the ribosome moves one triplet downstream until its A site is opposite the next triplet). The process repeats itself till it meets a termination signal.



Of course this very complicated machinery requires enzymes and energy. Some of the enzymes are fixed on the ribosome, others are in the cell sap. The energy is provided, partly by the high-group-donor potential of the amino acyl-tRNA bond, partly by GTP; two molecules of GTP are expended for each amino acid added. How the energy is used is still unclear: for example, how does the ribosome move? The whole system resembles a production line: there are still many things to unravel about its moving drives and control switches.

Polysomes

In a production line, a new car does not wait to start until the preceding one is finished. Likewise, as a ribosome slides along a messenger RNA, the start signal becomes again available for another start. Under full steam, there can be a ribosome at work every 200 or 300 nucleotides, each making a polypeptide chain. A messenger with its working ribosomes is called a polysome (see page 108).

An interesting question: does a polypeptide chain begin to form its three-dimensional structure while it is being synthesized or only after it is released? This is important, because the final three-dimensional structure, which after all is the only thing that matters in a protein, may differ depending on whether the folding occurs before the chain is finished or only after it is

completed. If folding occurs continuously, seeking arrangements of minimum free energy for the early portions of the polypeptide, the structure thus established may be different from the state of lowest free energy for the completed polypeptide. For example, a sequence of hydrophobic, uncharged amino acids located in the late portion might tend to be hidden in the center of the protein molecule, but if the earlier part of the chain was already balled up in a tight three-dimensional structure it would be unable to do so. Some small protein molecules, after being completely denatured into loose chains, return spontaneously to the original form. But some of the larger ones may not do so. Because its program is expressed in a time sequence, the protein molecule made in a cell may not be in the configuration that a chemist would call the energetically most sensible form.

Regulation

One more question: are all proteins of a cell made in equivalent amounts? Certainly not. An automobile factory capable of making 17 different car models would be in trouble if it had no way of modulating how many of each model it made, according to demand, cost, and so on. Likewise, a cell whose genes worked all at a constant rate would make too much of some proteins and not enough of some others. A program is useful only if it has both instructions of what to make and instructions of how much to make (that is, if it knows how not to make).

For example, the DNA of *E. coli* is sufficient to code for about 4,000 proteins. If all of them were made all the time in maximum amounts, the rate of protein synthesis would be 4,000 times the rate for any one protein. This is definitely not so. Likewise, a human liver cell or a muscle cell have the same genes, but the liver cells make mostly liver enzymes, and the muscle cells make mostly contractile proteins. Evidently there must be regulation of gene function. Genes are shut off and turned on when necessary, either in response to the external food supply or in response to developmental stimuli during the growth of a complex organism.

At this point it would be useful to understand how gene regulation works--one of the most exciting fields of molecular biology today. But we must make a detour. We cannot go into the regulation

of gene action before we learn more about genes and genetics. The reasons will become obvious if you stop to consider that until now you have learned general properties of all genes--DNA structure, messenger production, protein synthesis, the genetic code, the signals to start and end a protein chain. But the individual genes are all different, in length, in sequence, and in regulation. To understand the experiments that have clarified these properties we must first delve into the gene concept and the organization of genetic materials.