

[CREAKING]

**BARBARA
IMPERIALI:**

So what we are going to talk about today, and I'm going to do something a little annoying, as usual, I want to take you somewhere here. Yeah, sure, back to carbon, that's fine.

But what I want to show you is the sizes of the molecular players in translation because we are going here now from the smallest carbon atom. And we spent the first four lectures on amino acids, nucleotides, sugars, phospholipids. But we're entering new territory here, and I'm just going to do this in such a way that it comes onto the screens.

What you can see here is, in order to make proteins like hemoglobin and antibodies-- these are made up of amino acids, the really tiny things that are out of view right now-- we need large entities that are made up of either nucleic acids alone or nucleic acids plus proteins. And so the things that will feature today are the transfer RNAs and the ribosome.

And I want you to look at the size of this molecular machine. This is we're getting pretty large now. We're needing to use large machines to make the smaller catalysts that are essential in the cell.

And, just so sort of maybe size doesn't seem so important, but let's just go a little bit further here. And what to me is intriguing is that the size of the ribosome is pretty similar to the size of the rhinovirus, a little smaller than the hepatitis virus, but quite a bit smaller than some of these other viruses.

So the ribosome organelle is a large entity in the cell. And, when you do look at electron micrographs of cells, you can see these dark dots, which represent ribosomes. They're big enough to see, whereas the proteins themselves are not big enough to see. So that's what you're destined for today.

All right, OK, so I've started to place on this board some of the molecular players, the messenger RNA, the transfer RNA, and the ribosomes, which are made up of ribosomal RNA plus protein. And I want to just remind you about the structure of the mature messenger RNA just for a minute because, in the last class, we talked about a lot of manipulations of that pre-messenger RNA, the fresh thing out of transcription.

And now I just want to remind you that the messenger is single-stranded RNA, obviously. It has a 5 prime cap, which has got this funky 5 prime, 5 prime bridge that's resistant to exonucleases. Somewhere in that sequence is a start codon. It's something that says this is the bit I want to translate.

Often, there's a lot of stuff here that you don't translate. It's part of what's known as the ribosome binding site. And there are many features in this part of the sequence that are very important for translation. They contribute to the efficiency of translation.

Generically, we'd call it the ribosome binding site, but there are funny things called Shine-Dalgarno sequences and stuff. Don't worry about any of that. I just want you to appreciate that, the mature transcript, you don't translate the whole thing. A lot of this stuff is structural, functional for other reasons that contribute to the success of translation.

Once you see one of these, the ribosome mows its way through and reads the nucleic acids in the message. So the message is being conveyed over to the new machinery. And then, when you hit one of these three codons, and we'll discuss these properly when we get to them, it's time to stop and finish translating.

The other end the message has a poly-adenine tail. Remember that, once again, is structural to protect the ends of that transcript. Even if some of the hundreds of adenine nucleotides are nibbled off, you don't get into the part of the gene that's critical to be translated.

At a certain stage, though, you might get in. Exonucleases might chew up enough. And they may end up chewing up your transcript, but that probably suggests that the messenger has been around too long, and it's time for it to retire to a better life, OK?

So remember the poly-A tail. And this, once again, plays other functional roles with respect to being recognized as a transcript and being helped to get out of the nucleus. So this was really what we talked about last week.

There's one more feature in here. I'm just going to remind you that the final mature transcript has also been through splicing with removal of introns and the pasting together of exons, which is a wonderful way to diversify transcripts of translation and give you much more proteins than one gene can encode, OK? And we talked about that last time. Great.

So the first thing we have to think about here is how do we go from four bases to a language that includes 20 letters, right? Or it's better really, more precise, for me to call them

nucleotides. It's more precise because the base just represents the ring system that's attached to the ribose. Nucleotide means the whole thing, including a phosphate. So we go from four bases to encoding 20 amino acids.

Now there are a few organisms-- and, in fact, we have one spare one as well, selenocysteine. There are a couple of other amino acids that might be designated as the so-called 21st, 22nd amino acid. They're not found globally in all organisms. We have selenocysteine in just very, very few proteins. So it's something beyond the list of 20 that you saw.

There are other organisms, for example, in archaea, these guys who hang out in bubbling hot pots in Yellowstone, for example, that have another amino acid known as pyrrolysine. I'm going to mention that a little bit later on, but the ones you really need to think about are the ones we're encoding in the global genetic code. This is the ones that are common to everybody, all right?

So, obviously, when you look at the language of bases, one base-- if the language translated directly one base to one amino acid, we could only encode for amino acids. So we know it's not one base, one amino acid. And I know, at this stage, you know that it's three bases, but let's just go through the math or the original questions that were sort of circulating. Like how do we go from this language to the other language?

If two bases encoded each amino acid, we could only encode eight amino acids. That's not enough still. 16 amino-- wait a minute. 16 amino acids, sorry, I can never-- anyway, that's 4 to the 2.

It's pretty good if I can't get 4 to the 2 up at the blackboard. This would be 4 to the 1 power and then-- so what that came down to realizing there were not enough. That wouldn't be a sufficient language to encode the 20 amino acids.

So it's finally deduced that three bases encoded each amino acid. That would give us 64 possible words in the language that needs to be translated. That's a lot more than we need. We only need 20 for the encoded amino acids, so 64 possibilities.

But what else do we need in the language? We need a few more things anyway. Yeah, up there.

AUDIENCE: Oh, I don't [INAUDIBLE]

BARBARA Oh, you weren't. Up there.

IMPERIALI:

AUDIENCE: You need to know when to start and stop.

BARBARA So, exactly, so here this isn't necessarily one of the ones that uniformly codes amino acid
IMPERIALI: through the sequence. We need a precise way to say start. And, in fact, we need a way to say stop. And there are multiple three letter words that say stop.

So it turns out that the genetic code, which forms the basis of this entire concept, has some features to it where it does have some degeneracy. But we'll go through the degeneracy, and we'll take a look at the genetic code because it will tell us exactly how the three letter-- the words made of three bases encode everything we need for translation, all right?

So let's just go back and take a look at this. So we've looked at the messenger. I've told you a little bit about the tRNAs and the proteins. But then we just sort of give you a little bit of the back history.

And, once the structure of double-stranded DNA was deduced, really, everyone was moving on to trying to understand how that converted to the translation to proteins. And there were a lot of workers deeply involved in this. Crick and Brenner realized it was three bases to code for one amino acid, but Khorana, Nirenberg, and Holley, Khorana who was part of our faculty for many, many years, actually defined that genetic code and got all of the details.

Brenner and Crick started to have the ideas, but, really, the definition by doing a process known as cell-free translation where they could very carefully add components to understand how the code, the genetic code, was formulated where they put in specific messenger RNAs and amino acids and transfer RNAs and actually made proteins from that. So that's the work that Khorana and others did. And that was awarded them a Nobel Prize for that work.

And then, later on, things started to get-- you know, these are decades of work I want to point out to you. The ribosomes were discovered. That was a decade later, the sort of details of the structure, but not the structure itself. And it was really exciting in the 2000s when Ramakrishnan, Steitz, and Yonath solved the structure of the prokaryotic ribosome. So each of these things has taken a decade to happen, but they are fundamental, major, important things that we can act on and move forward to understand more.

All right, so let's move to the transfer RNAs. And I've flashed up this slide a couple of times, but I actually now have the movie of the structure of a transfer RNA. So the transfer RNA is a linear segment of RNA, but it's folded up the way ribonucleic acids are with three, four stretches of double strand coming together and loops in between them.

And the one end, this is the 3 prime end. I always try to draw it on the left because, otherwise, things get confusing. It's much easier to sort of see where everything ends up if you put the long arm on the 3 prime side to the left of your picture.

And what I think is cool, when you look at the structure of RNA, we think of messenger RNA as being a sort of rather floppy entity, but it actually really likes to form short segments of double helix. It just doesn't do well with the really long double-stranded structure the same way that DNA does. But this folded up structure is very important. And it was on the observation of these folded structures that the ribosome hypothesis was formulated.

But the two things that you want to remember about the ribosome are that the 3 prime hydroxyl group of the last ribose within this ribosomal sequence is where the amino acid that's going to be loaded into your protein is attached. So that's one point. And then there's another landmark on this structure.

And that's what's called-- one of the loops has a special name. It's called the anticodon loop. It comprises three nucleotides that are complementary to the nucleotides in your messenger sequence. So this really is a decoder because, at one end, it's carrying an amino acid, but it's carrying the amino acid that corresponds to the code that's in the messenger via that anticodon loop.

So it's a large structure, but don't mistake it for being something that's just sort of amino acid and anticodon. A lot goes on with the rest of the structure. It's a very important structure in the mechanisms of protein translation and synthesis.

OK, so here I've got to sum that up with a couple more of the ways that you would see the transfer RNAs. You might see it in this globular form. And I pointed out the anticodon loop. The place where the amino acid gets linked is also called the acceptor stem. And, up here, I show you that linkage. And you should-- yes?

AUDIENCE:

I was just going to ask, I see how the anticodons are specialized. How does the 3 prime end of the tRNA know which protein is bound?

BARBARA IMPERIALI: Yeah, and, in a moment, not quite yet, I will show you structures of tRNAs bound to their synthetases, which are the enzymes that load them. So there is specificity throughout that whole thing. It's not just bystander stuffing. It's really involved. And it's a great question, and I hope you'll get an answer that's reasonably satisfied from the structure perspective.

And so, if you look up here at the acceptor stem, the amino acid is joined by an ester bond to the 3 prime end of the transfer RNA, but, hopefully, you can see in here. There's the carboxyl. There's the amine. And CHR designates the amino acid where R would be the side chain of your amino acid. So that's what that looks like at that end of the transfer RNA.

And, coming down to the anticodon loop, you're going to read the messenger 5 prime to 3 prime. And the anticodon loop, when you draw this in this configuration, actually shows you that the anticodon loop is antiparallel to the codon loop to make that good hydrogen-bonding network. So that's why I like to be consistent in the way I render this structure. So the anticodon loop of the RNA complements that triplet codon in the messenger RNA, all right? Yes?

AUDIENCE: [INAUDIBLE]. What's between the G and T?

BARBARA IMPERIALI: What's between the G and T? Hold on. I'm going to-- what's between the G and the--

AUDIENCE: On the loop there.

BARBARA IMPERIALI: Over here?

AUDIENCE: No, above that on the left.

BARBARA IMPERIALI: On the left. Oh, this guy? So cool you ask that. So this guy is what's known as a funny base. It's actually-- that's the symbol for it that you've picked out. And it's a base known as pseudouridine.

And it turns out that these unusual bases show up in RNA sequences. Pseudouridine has an interesting structure where the bond to the ribose ring is not a carbon-nitrogen bond, but, rather, a carbon-carbon bond. And it's a bit more stable.

So, in RNA, there's some of these other unusual bases. And pseudouridine is the most

common of the unusual ones. It can still hydrogen bond, but it tends to show up in these sort of different loops and turn-type places, OK? Thanks for noticing that. Yeah?

AUDIENCE: So do you get the G in parentheses on the yellow part [INAUDIBLE]?

BARBARA IMPERIALI: The G, which is in parentheses, designates that even those sort of bulges between the real loops can vary in length. So it could be more than one. So that addresses-- comes back to that other question. But that stuff in between has variables. It has variable bulges and variable shapes associated with the synthetase enzymes that I'll introduce in a minute that it recognizes, all great questions.

OK, all right, cool. OK, so [HUMMING] we've got all of that. So now let's move ourselves so we've looked at-- we know the messenger well. We're starting to understand the tRNAs. What we need to move on to now is sort of the most important part of the game, which is really taking a look at the genetic code.

So this table is but one rendition of the genetic code. Sometimes, you'll see it in different shapes and sizes. In a second, I'll show you one of the other renditions. But it is the absolute-- the sort of Rosetta Stone for translating messenger RNA to amino acid sequence using codons.

So this sort of-- whoa. Getting a little-- I love translation. I'm sorry. I'm getting a little bit excited about translation. OK, so there are a few features of this genetic code. Number one, you won't have to remember it. It would be foolish for us to think you could, but there are characteristics about the genetic code that are very important.

But, first of all, let me sort of calibrate you. This would be the first of the three letters in the codon. And, by the way, the genetic code gives you the identities of what are known as the codons, which is how we designate the triplet of nucleotides. So you'll hear a lot about codons and anticodons of course.

So the way you read this table is you read the first letter. So all of these begin with U. All of these begin with C and so on. Then you read the second letter, and there's those designations.

So, if I'm going here, here, it's going to be starting UU. And then, within each block, there are the four alternatives for the other four bases. And then the third one, basically, just designates those. So you can read, for each amino acid, what three-letter codon would correspond to it.

Some people quite like-- oops. I got rid of it too fast. Some people quite like this other rendition where the first of the three letters is in the center. So it's either G, U, A, or C. Then the second one comes out. It's in the brown circle.

And then the third one is the third letter in the codon. And then that tells you what amino acid that corresponds to. And we'll generally tend to just stick with this one table and stick with it. So you might as well get used to that particular table.

Now there's a couple of characteristics in the table. The first thing to notice is that, within that table, there is a codon to start. And it's AUG. And the one thing that's sort of sent to drive you crazy is that AUG codon equals start, but it also equals methionine. And, in bacteria, it equals a modified version of methionine.

So, if the ribosome is reading the messenger RNA, it will look for the first one of these and start reading. But, once you find another one of these further into the sequence, it will put in methionine. Methionine is fairly rare. There may only be one or two more in the protein. And I hate the illogic of that, but, nevertheless, it is the case.

In some organisms, you start with different amino acids, but the most common start is the codon for methionine is the start. So what that means is that every protein you translate has a methionine at one of the termini. Which terminus would it be at?

AUDIENCE: [INAUDIBLE]

BARBARA IMPERIALI: Yeah, OK, so that's the details with methionine and the start codon. Then there are several stop codons. And I've shown you the three of them there. The stop codons tend to be used variously. Some are more predominant in some organisms than other. And some of them respond differently when the process of stopping occurs, which we'll talk about in a second.

But the next important thing to notice about the genetic code is that it's what's called degenerate. Now, when you call something degenerate, it seems like sort of a really nasty thing to call them. And it doesn't mean, oh, it's so bad. It's just a degenerate.

It just means it is specific, but it's not-- wait a minute. I've got the right wording here because I always get this sort of-- degenerate, ah, it's not ambiguous. It's degenerate. That's the best thing. But it's not ambiguous.

What this means is there are all, for several amino acids, more than one codon that specify it. And let's take a look at some of the examples of that degeneracy. It's at its most extreme with residues such as leucine where there are six codons that specify leucine. Alanine has four different codons. They all specify alanine. Lysine has two codons.

And it's generally appreciated that the residues that have more codons tend to be the more common in your messenger RNAs, in your final protein sequence, because you'll find, in your protein, a lot of leucines. So we need a few more codons to stack the deck in favor of putting in more leucines.

But so it's ambiguous-- it's degenerate. But, what I mean by it's not ambiguous, you know what it's going to code for.

The other place where degenerate codons can become important is that there is species specificity. So I'm going to write that in here and explain what I mean. And what that means is that some organisms might prefer two or three of the degenerate codons, and others may prefer a couple of the others.

And what that means, once in the laboratory, is it's really annoying if we want to express a protein in a really convenient bacterial system that we've taken from a mammalian system. Our codon mix may not be the same. And so there are companies now that actually will fix the codons in a gene for you to make them compatible with a different organism for expression.

So it has huge practical implications to be quite honest. And it can be very annoying in the laboratory. OK, and then so codon usage varies amongst organisms.

All right, so now, one last thing, so we've talked about the genetic code. It's the code that's going to be embedded within the messenger RNA. The last thing I want to do is, basically, explain to you one more time, when you're looking to read what your amino acids that get put in may be, you're going to look at the codon. And it will tell you exactly the amino acid.

A long time ago, I used to be confused because I thought I should be looking at the anticodon. And I was trying to translate everything, and it was a real mess. But it's the genetic code in that box that I just showed you is written down for maximum clarity and ease of use. So, whenever you see a particular three-letter code on the messenger, you will then be able to know what amino acid it would code.

So this is kind of interesting. It just reinforces to you that the codon and the anticodon are

antiparallel. And so what I want to do is, basically, in this diagram, you would be reading from the 5 prime to the 3 prime end, as the transfer RNAs attach to the messenger RNA. And that would give you a codon here that would be AUC.

If you'd written this the wrong way round, it would look like CUA. So you really want to be reading 5 prime to 3 prime in the codon. And then you can go to your favorite genetic code map and say this thing is read 5 prime to 3 prime. And the amino acid that gets put in is isoleucine, OK? So you'll need to be able to do that quite readily.

All right, next portion, those monster-- loading the amino acid. OK, so, as I said, this is a many building block, many parts thing. So we've got the transfer RNAs. We know where we load them onto the amino acids. We know where we load the amino acids onto the transfer RNA.

But we do not know how that is done. So I want to show you briefly how that occurs. So let me just start by drawing transfer RNAs the way I usually draw them in a very sort of cartoon form.

And, to attach an amino acid to the 3 prime end of the transfer RNA, you have an amino acid residue-- we're just going to go R here-- carboxylic acid, amine. Actually, I'm going to draw them in their appropriate charged states. And what I need to be able to do is faithfully fix this amino acid to the 3 prime OH of the transfer RNA.

And what we do is we need adenosine triphosphate to activate this chemistry. And then the OH at the 3 prime end reacts with the amino acid. So I'm just going to draw that. I'm leaving out steps because, otherwise, it's too many, and you'll be cross with me.

So you attach through an ester to the amino acid from the 3 prime end of these transfer RNA. That's what's done. The ATP makes this chemistry feasible, but there's one more player here. And that's the enzyme that brings them all together, which is known as an aminoacyl-tRNA synthetase, meaning it's an enzyme that makes an aminoacyl-tRNA. And it synthetases it. So that's how its name gets subtracted.

So, with reference to an earlier question, what I'm showing you here are different synthetases for different amino acids that show you that there's a recognition not just for the amino acid that's being loaded, but, rather, for the entire transfer RNA. So some of these look quite different. The isoleucine one interacts in one way. The valine one is a little different. And the glutamine one is different again.

So they vary in the way they interact with the transfer RNAs. So the transfer RNAs are specific for the amino acid that is loaded onto them, but also for the synthetase that does the loading. That's how you get the specificity. Does that address your question from earlier? Hello? OK. That makes sense, right?

So they fall into a lot of different families, but they're quite varied when you look at specific interactions between the synthetase and the amino acid. And, on those synthetases, there will also be specificity for the amino acid side chain of the amino acid that gets loaded.

Now let's look at the ribosome components because they're the last big monsters. Ribosomes have a small and a large-- that's the large and a small subunit. They are made up, as I mentioned over here, of RNA and protein, so small and large subunits. And I've shown them there in two different colors.

The prokaryotic ribosomes are pretty different from the eukaryotic ones. There's a higher proportion of RNA in the prokaryotic ones than the eukaryotic ones, which is kind of interesting. And these complexes are so big, and they're made up of so many protein and RNA strands, that we don't so much measure them by the number of those components, but, rather, by what's known as their sedimentation coefficient, which gives us a sense of the weight of the module or the complex.

So the small subunit-- oh, he's back [INAUDIBLE]. The small subunit would have a 30S sedimentation coefficient. The S stands for Svedberg units. It's how fast it sediments in an ultra centrifuge. And the large subunit would have a 50S sedimentation coefficient. And those correspond to a certain number of daltons.

If you see that S term after a number, that's what it's talking about. It's talking about the sedimentation coefficient, which gives us a sense of the size of the complex.

And, when you start to bring all the pieces together now, what we can see on this slide is the messenger, the transfer RNAs, and the ribosomal all to scale in such a way that it can really explain it. So what I show you here is the small and large subunit. In orange-- well, that's kind of a burnt orange-- is a sneaky little bit of the messenger. In yellow are the transfer RNAs.

And there's one more unit on here that I won't describe too much. It's a protein factor that helps all the processes occur. Generally, it's thought to help the loaded tRNA come to the ribosome, get it in place, and then go away. So it's some of these extra helper proteins that

are involved.

OK, so let's build a protein because I know it's the moment we've all been waiting for, and we're going to walk through how those pieces come together. It's a very clunky animation, but I'm very proud of myself because I did it myself. So I'm just going to show you how these things happen, as you assemble a chunk of polypeptide chain from a messenger.

And so here's the messenger. It's being read 5 prime to 3 prime. What happens, first of all, is that the small subunit kind of floats along, looking for the place that would be the ribosome binding site that I mentioned here, and then sliding its way to position the start codon in the right place to start the synthesis.

Once that happens, the methionine that's on its tRNA, the start one, gets into place. And, at the same time, the large subunit, completing the ribosome complex, comes together. So you now have large and small ribosomal complexes stuck onto that messenger, ready to carry on.

And, in each case, you're translocating through the messenger RNA. And, in each step, you're bringing in a tRNA that's loaded with an amino acid where the anticodon of the amino acid-- and I've got those little letters shown on the bottom here-- is complementary to the codon that's within the messenger RNA.

So we can start building this protein. AUG is that start codon. Methionine is the first amino acid. And it's always at the N-terminus. Then there's another amino acid comes in. The codon was UUU, and that corresponds to phenylalanine.

And then the next thing that happens is there's a movement such that a new bond gets-- a new amide bond is formed between the methionine and the phenylalanine. And that new bond is an amide bond. It's not any of the others.

So you're literally intercepting this complex on the transfer RNA with the amine of a new amino acid. That's how that comes together. I'm not going to worry you too much with the chemical details, but a lot of them have been illuminated by having the structure of the ribosome. And it shows you, in fact, that it's not proteins that are catalyzing that reaction. It's nucleic acids.

So now I'm just going to move you on through the synthesis, in each case, building that polypeptide chain and then moving, translocating. This guy leaves. The ribosome moves along. This was a good Saturday afternoon's work. We've got to use up all those tRNAs. It doesn't get interesting until you get to the end.

So we're using them, but now we come and hit a stop codon. We have UAG. So what happens here? The whole process slows down because you can either load what's known as a suppressor tRNA-- that's the one, the magenta one, that went up there-- or a protein release factor can come in and bind. But, in either case, there's no new amino acid to come in, and translation finishes and releases the protein in its complete form.

Now the reason I want to differentiate between the release factor and an RNA that has the complementary stop codon is because the RNAs that are known as the suppressor RNAs have now been completely hijacked to make an enhanced genetic code where we can load lots of different amino acids using the suppressor RNAs. And, if anyone would like to chat to me about this, I'd love to because I think it's a fascinating field, but it's a bit beyond the scope of.

OK, so this entire process took the following where you've got small and large subunits, all the elongation factors and initiation and release. And then, in each case, the energy is actually not provided by ATP. It's provided by GTP. But where ATP is important is in loading the amino acids onto the transfer RNAs.

This occurs at about a rate of 20 amino acids per second, meaning you're reading about 60 bases per second, which is pretty consistent with the rate of transcription, not the rate of replication. That's far faster.

OK, sometimes, when you're making a lot of a protein, you will see that ribosomes line up on the messenger RNA. And you'll have many proteins being made at once and at different stages in the game. And this is a lovely electron micrograph that actually shows this process in action. And, when you have a lot of ribosomes on one messenger, we call them polysomes. They have that name.

OK, good, all right, so what I want to do now-- so does everyone feel good about how you translate a messenger into a protein and the various moving parts? You would always know the amino acid structures, have the genetic code. Just be familiar with reading it and making sure you could pick out which amino acid might be incorporated in response to which particular codon.

So, when proteins are made on the ribosome, they have a bit of a choice. They can get made and fold beautifully into active proteins. Those proteins could be modified. They could go to

different places in the cell.

Occasionally, proteins misfold. Maybe the rate of synthesis is too fast, or the environment isn't right. So there will need to be mechanisms whereby proteins get degraded if they're not folded properly, but that's the story for another day. It's not always perfect, but what is known now is that, as proteins are emerging from the ribosome, they're starting to fold almost immediately from that N-terminus to, ultimately, attain their compact shape.

The last thing I want to talk to you about today is what happens-- what are the types of errors we get in translation. Now translation, like transcription and regulation, has some editing mechanisms to fix errors, but, occasionally, there are errors in the DNA that put different amino acids. And the editing at that stage, by the way, to fix errors is editing when you've loaded the wrong amino acid onto a tRNA.

But what happens when the DNA message, the DNA starting point, is wrong, which means the messenger is wrong, which then means we get a different translated proteins? So I'm just going to give you a couple of terms here.

When we have here-- and, on all these slides, I'm going to show you the double-stranded DNA. I'm going to show you the messenger that gets made and the protein that comes forward. And they're all lined up so you can follow them really nicely, always writing 5 prime to 3 prime, except when we have double-stranded because we have to put the bottom strand in a different order.

First of all, this entire chunk would be called the reading frame. It's the portion of DNA that's going to be read and transcribed into the messenger RNA.

And, when you look at the two strands, you're going to have to figure out which one is the template strand. And you would figure that out by knowing that you read 3 prime to 5 prime, but that you transcribe 5 prime to 3 prime, right? So you could recognize which of the two strands you're going to make the messenger out of.

Once you get the messenger, you come straight to the genetic code because you can use the genetic code to translate each of these. Does that make sense? So this would be a situation where you have a wild-type enzyme. Everything is transcribed and translated properly.

Occasionally, though, there are errors that will introduce defects into the ultimate protein. The

first type of error is a nonsense mutation, which might be leaving out a base pair, inserting one, substituting one. And this would, ultimately, cause an error in the DNA that then causes an error in the messenger RNA.

So let's say we delete by mistake, or we're missing the C-G base pair. Then the messenger RNA is now different. And a lot of the base pairs-- the bases slide up to fill the gap. So we have what's known as a frame-shift mutation. We've shifted things.

And what we've done by doing that is introduce wrong codons. So the first codon was read properly. The next one was read properly, even though there was a mistake, but then, all of a sudden, we get in this mess where one of the later codons carries a mistake and becomes a stop codon. OK, does everybody see that? So that would be called a frame-shift mutation that introduces a nonsense mutation and puts a stop codon into your messenger RNA, OK?

The next types of mistakes are ones that are what are called silent mutations. It doesn't matter, for example, if you made an error in the DNA, which ended up with an error in the messenger RNA, if you still code the same amino acid, right?

I go to the genetic code. I'm like, oh my goodness, I've got a mutation. Oh, it's fine. CCC codes for proline, but CCA also codes for proline. So that's called a silent mutation, OK?

Then the last ones are the ones where we start to encounter errors in DNA that result in errors in proteins that may cause genetically inherited diseases. So let's take a look at that.

Here it's a missense mutation where we've got an error in the DNA, which has resulted in an error in the messenger RNA. So, instead of valine-- instead of leucine, we've put in valine. That's not so bad because it's quite what we would call conservative. They're sort of similar amino acids. They have similar personalities. So this is a missense mutation, but it doesn't cause any dramatic changes, probably, in the DNA.

And then the last one I want to show you is a missense mutation that is non-conservative and causes a serious defect. And this takes you back to the beginning of the class where we've put a mistake in the sequence where we've changed a glycine to an arginine. And that's a big change.

And I just want to remind you of the situation in hemoglobin when we had a missense mutation, and we incorporated a valine instead of a glutamic acid, just through one change in the DNA, which made one change in the messenger, which put a drastic change in the protein

that caused sickle cell anemia. So missense mutations are where you put in the wrong amino acids. And those are the ones where you end up, in a lot of cases, with inherited diseases.

Nonsense mutations are not so bad because you probably just truncate the protein. You don't - nonsense mutations aren't so bad because you end up with a truncated protein, which would be degraded. The missense mutations are the more serious ones because you end up with a full length protein that might have a mistake in it. And then that would affect the function.

Am I being clear enough to everyone? Yeah? Good. OK, I am going to tell you that I'm handing over the baton to my colleague, Professor Martin. He'll take over on Monday.

Mouse is pretty happy. He's pretty excited about genetics. And these will be the lectures that will occur. What? You haven't seen him, have you? He's keen on genetics, yeah.

AUDIENCE: Great.

BARBARA IMPERIALI: OK, and that's it. Don't forget my office hours on Monday if you need them. I think this field is fascinating. Once you get used to the mechanics of it, it's really cool to think of how you go from DNA to RNA to folded proteins.