

MITOCW | 7.012-2004-L04

Among the issues that some people asked that should be discussed in greater detail should be the structure of proteins.

I'll touch on it very briefly this morning, different kinds of bonding, tertiary and quaternary structure, condensation or dehydration reactions. And, in fact, many of those issues should be addressed in the recitation sections.

That's the ideal place to begin to clarify things which although they were mentioned here may not have been mentioned in the degree of detail that you really need to assimilate them properly.

And I urge you to raise these issues with the recitation section instructors. That's exactly what they're there for.

Having said that I just want to dip back briefly into protein structure, even though we turned our back on it at the end of last time, just to reinforce some things that I realized I should have mentioned perhaps in greater detail. Here for the example are different ways of depicting the three-dimensional structure of the protein. And, by the way, we see that these are beta pleated sheets in the light brown and these are alpha helices.

There are two of them here in green, one going this way, the other going this way, a third one going this way.

And the other blue areas are not structured, i.

., they're not structured in the sense that they are in any way obviously alpha helices or beta pleated sheets.

Here's a space-filling model, a space-filling depiction of a protein. We talked about that last time. Here is a trace of the backbone, of the peptide backbone of the same protein where the side chains are left out, and obviously where one is only plotting the three-dimensional coordinates of each of the backbone atoms, CCN, CCN, CCN. Here is yet another way of plotting exactly the same protein in terms of indicating, as we just said, the structure of these alpha helices in the other regions. That is the secondary structure of this protein. And here's yet a fourth way of plotting, of depicting the same structure of the protein where roughly one is depicting the configuration of the amino acids in terms of a large sausage. Excuse me. If one were to use a space-filling model we'd go up to here. So these are just four ways of looking at the same protein with different degrees of simplification.

Another point that I thought I would like to reinforce and make was the following. We've talked about transmembrane proteins in the past.

That is, proteins which protrude through a membrane from one side to the other. And a point that I realized I'd like to make is that if we look at a transmembrane protein here's one that is starting out in the cytoplasm of a cell. And, by the way, the soluble part of the cytoplasm is sometimes called the cytosol.

Here is the lipid bilayer that we talked about at length and here is the extracellular domain of this same protein.

Now, how is all this organized? Well, the fact of the matter is we discussed the fact that this hydrophobic space in the lipid bilayer is so hydrophobic that it really doesn't like to be in the presence of hydrophilic molecules, including in this case amino acids.

And what we see here is the fact that almost all of the amino acids in this region of the protein, which is called the transmembrane region of the protein because it reaches from one side to the other, are all hydrophobic or neutral amino acids which are reasonably comfortable in the hydrophobic space of the lipid bilayer.

There happens to be two apparent violators of this, glutamine and histidine. You see these two here? I mean glutamic acid and histidine. Glutamic acid and histidine.

One is negatively charged and therefore is highly hydrophilic.

The other is positively charged and is therefore highly hydrophilic.

And on the surface that would seem to violate the rule I just articulated. But the fact is that as it turns out in the particular protein these two charges, these two amino acids are so closely juxtaposed with one another that their positive and negative charges are used to neutralize one another. And as a consequence in effect there is no strong charging or polarity in this area or in this area. The take-home lesson is that somehow proteins manage to insert themselves and to remain stable in the lipid bilayer by virtue of either using only stretches of hydrophobic or nonpolar amino acids or they use tricks like this of neutralizing any charges that happen to be there. Note, by the way, that because there are hydrophilic amino acids down here and there turn out to be hydrophilic amino acid around here, arginine, and here there's a whole bunch of basic amino acids. Note that this keeps the transmembrane protein from getting pulled in one direction or the other because this arginine likes to associate with the negative phosphates on the outside of the phospholipids.

And the same thing is here. And all that means is that this transmembrane protein is firmly anchored in the lipid bilayer, a point we'll talk about later in greater detail when we talk about membrane structure. One other little point I'll mention here in passing, which we'll also get into in greater detail, is that once a protein has been polymerized that polymerization is not the last thing that happens to it once it's polymerized and folded into place because we know that proteins undergo what is called post-translational modifications. And, as we'll talk about in the coming weeks, the process of synthesizing a protein is called translation. And when we talk about post-translational modification what we're talking about is opening our eyes to the possibility that even after the primary amino acid sequence has been polymerized there are chemical alterations that can subsequently be imposed on the amino acid side chains to further modify the protein. One such modification, by example, is a proteolytic degradation. And when I talk about proteolytic degradation, I'm talking about the fact that one can break down a protein.

Proteolysis is the breaking down of a protein. And when we talk about degradation we're talking about destroying what has been synthesized.

In the case of many proteins, once they're synthesized there may be a stretch of amino acids at one end or the other that simply clipped off therefore creating a protein which is smaller than the initially synthesized product of protein synthesis, i.e.

the initially synthesized product of translation.

Here we see yet another kind of post-translational modification, because it turns out that in many proteins which protrude into the extracellular space there is yet another kind of covalent modification which is the process of glycosylation in which a series of sugar side chains, carbohydrate side chains is covalently attached to the polypeptide chain usually on serines or threonines using the hydroxyl of the side chain of serines or threonines to attach these oligosaccharide side chains.

We know from our discussion the last time oligosaccharide means an assembly of a small number of monosaccharides.

And each of these blue hexagons represents a monosaccharide which are covalently linked and also modify the extracellular domain of this protein as it protrudes into the extracellular space.

So I'm just opening our eyes to the possibility that in the future we're going to talk about yet other ways in which proteins are modified to further tune-up their structure to make them more suitable, more competent to do the various jobs to which they've been assigned.

Let's therefore return to what we talked about the last time, the fact that the structure of nucleic acids is based on this simple principle. Here, by the way, I'm returning to the notion of this numbering system.

We're talking about a pentose nucleic acid. The fact that there are two hydroxyls here right away tells us that we're looking at a ribose rather than a deoxyribose which, as I said last time, lacks this sugar right there. Note, as we've said repeatedly, that the hydroxyl side chains of carbohydrates offer numerous opportunities for using dehydration reactions, or as they're sometimes called condensation reactions where you remove a water, where you take out a water, dehydration, or we can call them condensation reactions to attach yet other things. And, in fact, in principle there are actually four different hydroxyls that could be used here to do that. There's one here, there's one here, one here and one here. There are four different hydroxyls. The 1, the 2, the 3 and the 5 hydroxyl are, in principle, opportunities for further modification.

In truth the 2-prime hydroxyl is rarely used, as we'll discuss shortly, but the main actors are therefore this hydroxyl

here in which a condensation reaction has created a glycosidic bond.

That is a bond between a sugar and a non-sugar entity.

Glyco refers obviously to sugars like glycogen or glycosylation we've talked about before. Here a bond has been made between a base, and we'll talk about the different bases shortly, and the 1-prime hydroxyl of the ribose.

Over here at the 5-prime hydroxyl yet another condensation reaction.

Sometimes this is called an esterification reaction.

And again esterification refers to these kinds of condensation reactions where an acid and a base react with one another, and once again through a condensation reaction, yield the removal of a water. And let's look at what's happening here, because not only is one phosphate group attached to the 5-prime carbon, to the 5-prime hydroxyl.

In fact, there are three. And they are located, and each of them has a name. The inboard one is called alpha, moving further out is beta, and furthest out is gamma.

And it turns out that this chain of phosphates have very important implications for energy metabolism and for biosynthesis.

Why? I'm glad I asked that question. Because these are all three highly negatively charged. This is negatively charged, this is and this is. And, as you know, negative charges repel one another. And as a consequence, to create a triphosphate linkage like this represents pushing together negative charged moieties, these three phosphates, even though they don't like to be next to one another. And that pushing together, that creation of the triphosphate chain represents an investment of energy. And once the three are pushed together that represents great potential energy much like a spring that has been compressed together and would just love to pop apart. These three phosphates would love to pop apart from one another by virtue of the fact that these negative charges are mutually repelling. But they cannot as long as they're in this triphosphate configuration. But once the triphosphate configuration is broken then the energy released by their leaving one another can then be exploited for yet other purposes.

Keep in mind, just to reinforce what I said a second ago, the difference between a ribose and a deoxyribose is the presence or the absence of this oxygen. And now let's focus in a little more detail on the bases because the bases are indeed the subject of much of our discussion today. And we have two basic kinds of bases. They're called nitrogenous bases, these bases, because they have nitrogen in them. And if you look at the five bases that are depicted here you'll see that they are not aromatic rings with just carbons in them like a six carbon benzene.

Rather all of them have a substantial fraction of nitrogens actually in the ring, two in the case of these pyrimidines.

And here you see the number actually is four.

In fact, one of these nitrogenous bases indicated here, guanine has actually a fifth one up here as a side chain.

This is outside of the chain, it represents a side group. And if we begin now to make distinctions between the ring itself and the entities that protrude out of the ring, they really represent some of the important distinguishing characteristics.

It's important that we understand that pyrimidines have one ring and these have two rings in them. The purines have a five and a six membered ring fused together, as you can see. The pyrimidines have only a six membered ring. And what's really important in determining their identity is not the basic pyrimidine or purine structure. It's once again the side chains that distinguish these one from the other. Here in the case of cytosine we see that there's a carbonyl here, an oxygen sticking out, and there's an amine over here. We see uracil which happens to be present in RNA but not DNA which has two carbonyls here and here.

Obviously, therefore what distinguishes these two from one another is this oxygen versus this amine.

And here we see the thymine which is present in DNA but not RNA.

And this will become very familiar to you shortly.

This looks just like uracil except for the fact that there's a methyl group sticking out here. Now, very important for our understanding of what's happening here is the fact that this methyl group, although it distinguishes thymine from uracil is itself biologically actually very important.

It's there to be sure and it's a distinguishing mark of T versus U, but the business end of T versus U in terms of encoding information happens here with these two oxygens sticking out. They're the important oxygens, here and here. And therefore from the point of view of information content, as we'll soon see, T and U are essentially equivalent. It may be that one of them happens to be in RNA and the other in DNA, but from the point of view of understanding the coding information they carry it's these two carbonyls here and here which dictate essentially their identity.

We have the same kind of dynamics that operate here in the case of A and G where once again this one has only an amine side chain and this one has a carbonyl and an amine side chain right here.

Now, very important there is a confusing array of names that are associated with all this. I don't know if you can, well, it reads reasonably well. Because once a base, and I just showed you bases which are unattached to the

sugars, once bases are attached to the sugars they change their name slightly. So keep in mind that here, when we talk about these nitrogenous bases, the bases are just free molecules where in each case this lowest nitrogen is the one that participates in the formation of a covalent glycosidic bond with the ribose or the deoxyribose underneath it. And here we can see one indication of how that, you see this N, in all cases via a condensation reaction, forms a covalent bond with a five carbon sugar, once again deoxyribose or ribose. Once the base associates with the sugar, that is the base plus the sugar is called a nucleoside.

So when we talk in polite company about a nucleoside we're not talking about free bases. We're talking about the covalent interaction of a pentose binding to a base. The pentose could be one or the other of these two. And that's what a nucleoside is.

If on top of that we add additionally one or more phosphates then we even modify our language even further because a base attached to a sugar which in turn is attached to a phosphate is called a nucleotide. The nucleotide, the T is there to designate the fact that there's actually, in addition to the base and the sugar there's a phosphate which is attached and extends off the end. And there are slightly different names. For the purposes of this course we won't get into this very arcane nomenclature because it is, to be frank, and you know I always am frank with you, confusing. Here is U.

And when uracil, the base becomes linked to a ribose it changes its name from uracil to uridine. Cytosine changes its name to cytidine when it becomes a nucleoside by a covalent linkage to either ribose or deoxyribose. Thymine becomes thymidine. And the same nomenclature exists, the shift in their names exists in the case of the purines as well, adenine becomes adenosine and so forth. We need to focus mostly on the notion of A, C, T, G and U. Those are the things we need to think about. And why is this nomenclature confusing? Well, here the nucleoside ends with osine, O-S-I-N-E. You see that here? You say that's easy to remember, but look up here. Here the base ends with O-S-I-N-E.

And so this nomenclature which was cobbled together in the early 20th century will bedevil us and generations of biology students to come. Oh well, that's life. Now, one of the things we're interested in and which I talked about briefly last time was the whole notion of polymerization, i.e., how we actually polymerize a chain. Let's look at this illustration which I think is more useful. Recall the fact that I emphasized with great seriousness the fact that nucleic acid synthesis always occurs in a certain polarity.

It goes in a certain direction. You cannot add nucleotides on one end or the other end willy-nilly. You can only add them onto the 3-prime end. And keep in mind that the reason why this is defined as the 5-prime end is that this is, the last hydroxyl sticking out at this end comes out of the 5-prime carbon right here, the 5-prime hydroxyl. And conversely at this end we're adding another base at the 3-prime hydroxyl, at this end, which creates the 3-prime

end of the DNA or the RNA.

In fact, the polymerization always occurs between the 5-prime end of a deoxyribonucleotide indicated here where the bases remain anonymous and the 3-prime hydroxyl. That's the way it always happens.

And here we begin to appreciate the role of the high energy phosphate linkage. Because this high energy triphosphate linkage, which is synthesized elsewhere in the cell like a coiled spring and which contains a lot of potential energy by virtue of this mutual negative repulsion of the phosphate groups, this energy is used to form the bond here between the phosphate in this condensation reaction and the 3-prime hydroxyl.

So that requires an investment of energy. And the resulting linkage which is formed is sometimes called a phosphodiester linkage.

Why phosphodiester? Well, obviously it's phospho.

And there actually are two esterifications that are occurring here. If we look at one of these phosphodiester bonds we see that an ester linkage has been made with this hydroxyl and an ester linkage has been made with this hydroxyl. And for that reason it's called a phosphodiester linkage. Therefore we come to realize that polymerization of nucleic acids doesn't take place spontaneously.

It requires the investment of a high-energy molecule, the investment of the energy that it carries. And when this linkage is formed the diphosphate here, the beta and the gamma phosphates float off into interstellar space. It's only the alpha phosphate that is retained to form the resulting diphosphate, a phosphodiester linkage. And this process can be repeated literally thousands and millions of times. An average human's chromosomes contains on the order of tens, fifty, a hundred mega-bases of DNA.

A mega-base is a million bases or a million nucleotides.

So there you can understand that there's no limit to the extent of elongation of these various kinds of molecules. Now, note by the way yet another feature of this which is that the distinguishing feature between DNA and RNA, the most important distinguishing feature is this 2-prime hydroxyl.

And here we're talking about DNA, but we could almost in the same breath be talking about the way that RNA gets polymerized.

Why? Because this 2-prime hydroxyl or this 2-prime hydrogen in this case is out of the line of fire. The business action is happening right along here. Look where the business action is in terms of the backbone. The 2-prime hydroxyl is off to the side. And whether it's oxygen or just whether it's OH, that is in ribose, a hydroxyl group or just

a hydrogen, as is indicated here in the case of deoxyribose, is irrelevant to the polymerization. And therefore we can guess or intuit, and just because we guessed doesn't mean it's wrong, often it's right, it doesn't really make much difference whether we look at DNA or RNA. Here's a polymerization scheme of RNA and it's absolutely identical to that of DNA.

In this case it's ribonucleotide triphosphates that are used for the polymerization reaction. Now here I just uttered the phrase ribonucleoside triphosphates. Why did I say that? Well, ultimately only the good Lord knows why I said that.

But let's look at this phrase. I said ribonucleoside triphosphate rather than ribonucleotide triphosphate because the fact that I added this on the end makes the T there unnecessary.

The T is there to indicate the phosphate being attached to the ribose or the deoxyribose. But if I'm adding this phrase over here, triphosphate, that obviates, that makes unnecessary my saying ribonucleotide triphosphate. If I'm looking at UTP or ATP, I would say I'm a ribonucleotide if I don't mention the triphosphate. But the moment this comes from my lips then we'll say ribonucleoside indicating that a ribonucleoside, that is a base and a sugar are then attached to one or more phosphate linkages. Now, the ultimate basis of the biological revolution comes from the realization that these different bases have complementarity to one another. That is they like to be together with one another. And if we look at this and we think about the DNA double helix we come to realize that these bases have affinities for one another. And the general affinity is one purine likes to be facing opposite one pyrimidine.

One pyrimidine opposite one purine. And if we have two pyrimidines facing one another they're not close enough to one another to kiss.

And if we have two purines they're too close to one another, they're bumping into one another, they take up too much space. And therefore the optimal configuration is one purine and one pyrimidine.

And you can see these two pairings here in the case of what happens with DNA. In fact, the realization of this diagram right here is what triggered the discovery of DNA in 1953.

This diagram right here is what triggered the biological revolution.

And though it's been depicted in many, many ways it's worthwhile dwelling on it because this is perhaps the most important diagram that we'll address all semester. Although this doesn't mean we have to spend all semester assimilating it. It's not so complicated.

It's relatively straightforward. And let's look at its features.

Let's dwell on them momentarily because this is a microscopic snapshot of what DNA is composed of. You all know it's a double helix and therefore there are two strands of DNA in a double helix.

And one of the interesting things about the double helix, although we're not showing it yet, we're just showing a little section of a double helix, is the polarity of the two chains that constitute the double helix. Let's look at that polarity.

This one is running in one direction and this one, the opposite one, the complementary one is running in the other direction. And therefore we talk about the double helix as being anti-parallel. Well, I guess I should have a bandage on the other finger to convince you but you get the idea.

They're running in opposite directions.

They're not both pointed the same. And the other thing to indicate is, to repeat what I said just seconds ago, that there's a complementarity between the purines and the pyrimidines. So we use the word complementary with great frequency, with great promiscuity in biology.

Complementarity refers to the fact that A and T here or A and U because I said U and T are functionally equivalent, they like to be opposite one another. There's a purine and a pyrimidine.

And the converse is the case with C and G, they like to be opposite one another. Now, there is specificity here.

You might say any purine can pair up with any pyrimidine, but it's not the case. For instance, A doesn't like to be opposite C and T doesn't like to be opposite G. So one of the things we have to memorize this semester, and it's not many and it's not hard, is that A and T are opposite one another, or A and U, and G and C are opposite one another. That's one of the essential concepts in molecular biology. There are now a thousand things you need to learn, but if you don't understand that then ultimately sooner or later you'll find yourself in a swamp, literally or figuratively. Now, let's look at the different between these two. One of the interesting things is, to state the obvious, the way they're associating with one another, hand in glove, is via hydrogen bonds. That's not any covalent interaction, which means they're reversible. We talked about that.

Which means that if we were to take a solution of double stranded DNA and boil it we would break those hydrogen bonds.

Remember they only have 8 kilocalories per mole and boiling water has far higher energetic content. And consequently if we heat up a DNA double helix and we break those double bonds of DNA that hold the two strands together, the two strands come apart, the DNA ends up being denatured, that is the two strands are separated one from the other. In fact, if there ever were a covalent cross-link between the two strands that's really bad news

for a cell carrying such a DNA double helix. A covalently cross-link from one strand to the other DNA double helix represents often a sign that a cell should go off and die because it has a very hard time dealing with that by virtue of the fact, as we will soon learn or as you already know, the cell has, with some frequency, to pull apart these two strands. And therefore this association must be tight enough so that it's stable at body temperature but not so tight that it cannot be pulled apart when certain biological conditions call for it. You see that in fact here there are three hydrogen bonds and here there are only two hydrogen bonds. That also has its implications. It turns out to be the case that the disposition of this hydrogen and this oxygen here, they're far enough apart that for all practical purposes they don't really make very good hydrogen bonds. And therefore we think of this as having two and this having three. And if you were to try to put C opposite A or G opposite T you'd see that they cannot form hydrogen bonds well with one another. Instead they kind of bump into one another, and therefore are not complementary to one another at all.

There's another corollary that we can deduce from this diagram, and that is the following. If it's always true that A equal C and G equal T -- A equals T and G equals C. By the way, this is an interesting story. This is the Chargaff Rule. Because about a year or so before Watson and Crick figured out the structure of the double helix there was a guy named Erwin Chargaff in New York at Columbia University who one day figured out that if you looked at a whole bunch of nucleic acids, different DNAs from different cell types -- And in certain cell types what he found was that G was equal to, for example G equals 20% of the bases. Therefore, obviously we know C must equal also 20% because there always has to be a C opposite a G in the double helix, right? G and C always have to be equal. And Chargaff discovered that, in fact, A in such DNA always was 30% and T was also 30%. Well, these together make up 100% which is, we're not in higher math yet, but A and T were always the same. If you looked at another type of DNA he might find that G equals 23% and C also equals 23%. And in this same DNA then A would equal 27%, I guess, and T also equals 27%.

And I hope that adds up to 100%. So he looked at a whole bunch of DNAs and they always tracked one another, A always tracked T, G always tracked C. And then in 1953 up comes these two guys from Cambridge, England, Watson and Crick whom Chargaff regarded as upstarts, as smart-asses who thought they knew all the answers. And Watson and Crick said, gee, this Chargaff rule really is very interesting because it suggests something about the structure of DNA. These cannot just be coincidences.

There's something profoundly important they said, correctly, in the fact that there was always an equivalence between A and T and between G and C. And that represented one of the conceptual cornerstones of their elucidating the structure of the double helix. And so Chargaff who died last year or the year before last, at an advanced age, was for the next fifty years a very bitter man, because he was this far away from figuring out this far. Not this far, but this far away from figuring out, making the most important discovery in biology in the 20th

century. He had the information right there.

And if he thought a little bit about information theory and thought a little bit about the way information content is encoded he could have already predicted, not the detailed structure of the double helix, but at least the way in which it encodes information.

Because, to state the obvious, and as many of you know already, if one looks at the structure of a double helix one can, in principle, depict it in a two or a three-dimensional cartoon.

Here's the way one can think of it. This is the way we've been talking about it over the last couple of minutes. It's a two-dimensional double helix. And from the point of view of information encoding, it doesn't really matter whether we draw it this way or that way. It happens that the double helix is turned around like that, it's twisted around. It's very difficult for biological molecules to be totally flat for an extended period. And the helix is, in fact, something that is frequently resorted to. Witness the alpha helix in the protein. So these are turned around. It turns out that each of these constitutes a base pair, and each of these base pairs is, in fact, 3.4 angstroms apart. 3.4 angstroms thick.

So you have ten of them, the DNA helix advances 3.4 angstroms every ten turns. And ten turns is roughly, oh, I'm sorry. Ten base pairs is roughly one turn of the alpha helix.

So if you go here and you count up ten, we should start again at the same orientation. Another ten is another turn.

Another ten is another turn. In fact, I'm just recalling that I was once a TA in 7. 1 in 1965. And there was a physics professor who became a biologist who always talked about these double helices. And he always talked about the measurements of different DNA molecules. Now, you may know that the term angstrom is named after a Danish person named Angstrom.

That's why it got its name. So whenever this professor, whom I never corrected, God forbid, ever talked about something that was ten angstroms long, he called these ten angstra.

Now, as you know, when you go in a Latin verb from singular to plural it's "um" to "a", right? So he pretended this was a Latin word. What's a good word?

Sorry? What's a common Latin word we use? Sorry?

Millennium. Yeah, millennium, millennia.

So he went from angstrom to anstra. And it went on for a whole year. I never said anything but I knew better. OK, anyhow.

Here you see the genius of Watson and Crick. And, by the way, Angstrom was a Dane, as I said, and not a Roman soldier.

So here we see. OK. So here is the genius of their discovery. And the elegance of it is not how complicated it is.

The elegance of it is how simple it is, because information we see is encoded in two strands. The information is redundant because if we know the sequence of one strand we can obviously predict the sequence in the other strand because it's a complementary sequence. If we always realize that A is opposite T and G is opposite C we can know directly that a sequence in one strand, which may be A, C, T, G, G, C and the other strand moving in the other anti-parallel direction the sequence is like this.

I don't need to know the sequence of the other strand.

I can predict it by using these rules of complementary sequence structure. And that, in turn, obviously has important implications. If we look at the three-dimensional structure, this is more of what's called a space-filling model.

This is the way the x-ray crystallographer would actually depict it. We talked about space-filling models before.

One of the things we appreciate is the fact that the phosphates are on the outside and these bases are in the inside. And because these bases are able also to stack with one another via hydrophobic interactions importantly the bases are protected. The face where they interact is protected from the outside world. What do I mean by that? Well, let's go back to this figure right here. You see the interaction faces between A and T or C and G they're not on the outside of the helix.

They're hidden in the middle. And that's important because it means that these interactions between A and C and G and T, you can see it up here as well, are biochemically protected from any accidents that might happen on the outside.

They're sheltered from that. And that's important because the information content in DNA must be held very stable, very constant. If it isn't then we have real trouble like cancer.

And therefore whenever a cell divides and copies its DNA, its three billion base pairs of DNA, whenever that happens the number of mistakes that are made is only three or four or five out three billion.

A stunningly low rate. And this DNA can sit around.

I told you about Neanderthal DNA that can sit around for 30, 00 years and it's chemically relatively stable.

In part, a testimonial to the fact that this base pairing, the face where the two bases interact across one another, this is shielded from the outside world because it's tucked into the middle, these interaction faces here. This is the inside of the helix.

Here the sugar phosphate groups are on the outside.

In fact, when Watson and Crick were struggling with the structure of the double helix they were in a horse race with a man named Linus Pauling who was really the inventor, the discoverer of the hydrogen bond pretty much who actually got two Nobel Prizes in his lifetime who ended his life believing that if you took enough vitamin C grams of it every day you would never get sick. I don't know what he died of, but probably like Dr. Atkins he probably died of an illness he was trying to ward off. Or he might have died of kidney failure from all the vitamin C he was putting into his body.

Who knows? Anyhow, I digress. The fact is that Pauling thought that, in fact, DNA was constituted of a triple helix, with three strands, and that the bases were facing outward. Well, of course, now we can snicker, now we can laugh, but at the time nobody had any idea.

Now we realize it's only a double helix and the bases are facing inward. And, of course, because Pauling worked with that preconception, he was never able to figure what was actually going on, even though Watson and Crick thought that he had the answer and was about to scoop them. Implicit in what I've just said is the notion that the structure of DNA, which we'll talk about later, allows it to be copied, i.e., now we're referring in passing, and we'll get into this in greater detail later, to the whole process of replication. Because if we have genetic material and we've created in a certain sequence we must be able to make more copies of it. Keep in mind that each one of us, as I mentioned to you some lectures ago, we start out with a fertilized egg with one human genome, and through our lifetimes we produce how many cells? Anybody remember?

I did mention it, right? Is there one soul who remembers it?

Remember the whole story of Sodom and Gomorrah where the Lord says if there's one soul, one righteous soul in the city I will spare the city. And of course there wasn't so he wiped them all out. 30 trillion? Well, sorry. What do we do for him? Something nice. [APPLAUSE] Excellent. OK. You'll remain anonymous, though. You won't be on that video. OK. Ten to the sixteenth cell divisions in a human lifetime. And on every one of those occasions the double helix is copied. I'm telling you that only to give you the most dramatic demonstration of the fact that if you have one set of DNA molecules you need to be able to copy it, you need to be able to replicate it. And that replicative ability is inherent in the double helix as Watson and Crick immediately said and as they noted at the end of their paper when -- I think the last sentence says it has not escaped our attention that this structure, i.e.,

the structure of the double helix, allows for copying, allows for replication. Because if you pull the two strands apart, recall we said earlier that in certain biological situations you need to do that, if the two strands are pulled apart not by putting them in boiling water but by enzymes whose dedicated function it is to separate the two strands.

Then when that happens one can begin to create two new daughter double helices by simply adding on new bases and thereby replicating the DNA. And how that happens is, of course, as you know, IO "Intuitively Obvious". OK. Uh-oh, we're in a dyslexic moment. Now, the fact is I emphasized with great vigor and conviction -- And remember, class, when somebody is convinced of something more often than not they're just wrong in a loud voice. But I nevertheless emphasized with great conviction that T and U are, from an information standpoint, functionally equivalent. They're replaceable, interchangeable. And therefore if we want we can make an RNA copy of a DNA molecule by realizing that if this were DNA we could make an RNA that was complementary to a DNA strand realizing that when the RNA molecule was being polymerized, instead of using T one would use U. All the other three bases are functionally equivalent. And so we could, in principle, and indeed it happens transiently, we could make a DNA-RNA hybrid helix where a DNA molecule is wrapped around an RNA molecule because the two molecules are functionally equivalent. The only difference between the two strands would be, well, there are two differences.

One, in the RNA strand we'd have a U instead of a T.

And, two, in the RNA strand all the sugars would be ribose rather than deoxyribose. Right on. OK. Good. So this structure, the simplicity of the structure gives one enormous power in encoding all kinds of information and replicating it.

What it means, as we'll discuss also in great detail later, is that if we have a certain sequence of bases in the double helix of DNA an RNA molecule could be made to copy one of the two strands to make a complementary copy.

And that RNA molecule could then leave the DNA double helix having lifted one of the sequences from it and then move to another part of the cell where it might do something interesting. And therefore to extract information out of the double helix doesn't necessarily mean to destroy it. If one can copy one of the two double strands in a complementary form as an RNA molecule that may enable the information that is encoded in the DNA to be copied without destroying the double helix itself.

Again, that process, which we'll also talk about later, is called the process of transcription.

And so in the course of this morning I have uttered the three words which represent the cannon, the basic

fundamentals of molecular biology. What are the three words? Replication, transcription and translation.

Transcription means when you make an RNA copy of a strand of the DNA double helix. Let's just add a couple more footnotes to what I've been saying just so we are on firm ground for subsequent discussions. It turns out that often in RNA molecules they can form intramolecular double helices.

There's no reason why you cannot make a double helix out of RNA as you can make out of DNA. And therefore you see often in many kinds of RNA molecules they will hydrogen bond to themselves using these complementary sequences. And this is called a hairpin, by the way for obvious reasons. And so many RNA molecules, most of them in fact have these intramolecular hydrogen bonded double helices with confers on them very specific structure.

One other aspect of the two versus three hydrogen bonds is the following. If a double helix has many Gs and Cs then it's going to have more hydrogen bonds holding it together than if it has few Gs and Cs. So let's look at the Chargaff example. Chargaff who lived for fifty years stewing in his own bile in bitterness because he couldn't figure this out, which is exactly what happened by the way.

And so here this has a higher G plus C content, the one on the right than this one. This is 23% or 46% G plus C. This is 40% G plus C.

If it's 46% G plus C that means there are more hydrogen bonds holding the two strands together. And it turns out that if you want to denature a double helix that has high G plus C content you need to put in more energy, you need to heat the double helix up to a higher temperature. It's more difficult to pull the strands apart. One other side comment on what I wanted to say is the following. The presence or the absence of this hydroxyl here in RNA has an important consequence for the stability of RNA and DNA. Let's look at what happens to an RNA chain when a hydroxyl ion, which happens to be floating around at a low concentration, happens to attack this phosphodiester bond. What happens is that this phosphodiester bond will tend to cyclize. It's forming this five membered ring. And ultimately that will resolve and break causing a cleavage of the RNA chain. This phosphodiester bond now forming a cyclic structure here as an intermediate representing the precursor to the ultimately cleaved chain. That means that if you take RNA molecules and you put them in alkali they will fall apart very quickly for this very reason. What happens to DNA molecules when you put them in alkali? Nothing. They're alkali resistant because there isn't a hydroxyl there to form this five membered ring.

And therefore alkali cannot cleave apart the DNA or the DNA phosphodiester bond. If we imagine that OH groups, that hydroxyls, are present at a certain, albeit a certain concentration, albeit a low concentration in neutral water we can see that even at neutral pH with a certain frequency RNA molecules will slowly hydrolyze.

They'll certainly be slowly broken down by the hydroxyl ions.

DNA molecules, however, will not. And that represents yet another important biochemical reason why DNA is chemically stable and why it can carry information over years, decades or tens of thousands of years, because the phosphodiester linkage in DNA rather than RNA is very stable chemically and can hold these adjacent nucleotides together, one to the other. See you on Friday morning.