BARBARA IMPERIALI: So we are moving along. Lecture 6 is the last of the biochemistry lectures. We're going to be talking about nucleotides and nucleic acids. And you'll understand these terms in a moment. I'll clarify them for you.

But this is a tremendous stepping stone to the next portion of the class. So I show you a few images here. I'm going to reshow you some of these in a moment when we talk about addressing understanding the noncovalent structure of DNA, which is so critical to understanding information storage and information transfer.

But for now, let's just have a quick peek forward. After this section, I'm going to be covering molecular biology, so how to go from DNA to RNA to protein. And then Professor Martin will take over with the basic structures and functions of cells and then genetics.

But for all of this, we're going to need nucleic acids. And I'll explain to you why here. So nucleic acids form fundamental units for information storage, storage. And that is the DNA that is in our nucleus and in our mitochondria, and then information transfer.

And if I get a little bit of time at the end, I have three or four quick slides that you don't have on your handout because it's sort of a floating topic on the use of DNA and DNA-based computing, because it's a nanoscale structure that one can program to do different things. And I think you might enjoy that.

So in this picture of the components and what's known as the central dogma, that is how DNA is converted into messenger RNA, which, through the help of transfer RNA and ribosomal RNA, we get proteins. The key elements on this slide are DNA, messenger RNA, ribosomal RNA, and transfer RNA. And those are all made up of nucleotides being brought together into polymers that are nucleic acids. So obviously, we really need to crack the structures of these and understand how the structure informs function.

Remember, we did that for proteins. We've done that for phospholipids. We thought about it very briefly for carbohydrates. But the thing that I really want to stress to you with the fourth of these macromolecules is looking at how the last component of the biomolecule's structure really informs function. And it's really cool to think about how it's done.

So how is that chemical molecular structure something that we can understand from the

perspective of function? So what we need to do, first of all, is think about what nucleotides are and understand their structure so that we can move forward to understand how they come together to build these macromolecules. They're so pivotal and essential in life for programming the biosynthesis of our proteins.

And now we're understanding more and more about not only that, but also how RNA, not DNA, is involved in a large number of regulatory processes. So it's not just DNA, double-stranded DNA goes to a messenger, and so on. Also, a lot of regulation occurs because of a lot of the other nucleic acids that are within the cell.

So I'm going to go here because I want to describe the composite components of nucleotides so we understand their structure and their properties. So what are nucleotides? And you look at these structures up on the board. They look kind of complicated. So let me deconstruct them for you. It'll make life a lot easier.

So they're two familiar building blocks and one new one. So the familiar building blocks are, first of all, carbohydrates. So the key carbohydrate in nucleic acid is a five-carbon pentose sugar, which looks like this. You can count the carbons, 1, 2, 4, 5, and 5. And you can reassure yourselves everything is there with respect to the carbons by translating this line-angle drawing into a drawing where you put all the hydrogens on and you know where everything is.

There are two types of five-carbon pentoses that are used in the nucleic acid. They are ribose, which is shown here with all OHs on all of those carbons, and two deoxyribose, which is a building block of DNA, whereas ribose is a building block of RNA. What else do I need to tell you?

You'll see this later on. That ribose sugar ends up being connected to what are known as nucleobases. You do not necessarily need to draw those, because you've got them on your handout to put sketches on. So I put them on the board for so I don't have to stand here and draw them for you. And I want to explain certain things.

So the nucleobases in the numbering system-- and I'm going to keep on reiterating this so you'll get familiar with it-- number the carbons 1 through whatever it is, or rather, the atom numbers when you're walking around the ring. So when we talk about the ribose component, they have what's known as a prime numbering system to differentiate it from the numbering system in the riboses. So this would be 1 prime, 2 prime, 3 prime, 4 prime, and 5 prime.

Why is that? This becomes incredibly important when we talk about putting together polymers of DNA and the direction in which DNA is assembled in life, and also, even when we describe 2-deoxyribose, or a ribose, because this would be called 2 prime deoxyribose in the nucleic acid.

So I'm going to bore you with that numbering system because I'll start to use it very commonly. And it will make a lot of sense as we start to assemble the DNA macromolecule when we talk about the way it's built and drawn and written. The numbering system will be important because we'll constantly refer to 5 prime and 3 prime. That's just a little preview for later.

The next component of the nucleic acid is a phosphate. Phosphorus looks like this. But in nucleic, in the nucleotides, these are joined to other units as phosphoesters. But you want to remember that in phosphorus, you have 1, 2, 3, 4, 5 bonds to phosphorus, and you commonly have a negative charge on one of those oxygens. And in the structure of DNA, you actually have phosphates occurring as phosphodiesters. And you, once again, you will see that when we see the intact structure of DNA.

So what are nucleotides? Nucleotides are a combination of a carbohydrate or a sugar, a phosphate and a nucleobase. That's the third component, the one we're going to learn about now. So the nucleobases look like this. There are two families, two flavors of nucleobase. There is one flavor-- let's get this cleaned up a little bit here-- that has two rings. And it has the shorter name, purine.

And there's a different family or flavor of nucleobases that has one ring, and it has the bigger name. And that, to this day, is the way I remember purines and pyrimidines. Small name, big structure; big name, small structure. If that's helpful to you, go for it. Use it. I haven't patented it or anything.

So in nucleic acids, there are two different purines. They are known as adenine and guanine. You do not need to know these structures. I actually only know my favorite three of the five to draw easily. And the other two, I'm always stumbling around the ring. So don't worry about that. We all get to know the ones we work with every day. For me, it's uracil, it's adenine, and it's cytosine, but not the others.

But what you do need to understand is a little bit about their structures. Because when we start to talk about the noncovalent structure of nucleic acids, principally, the double-stranded helix

of DNA, we need to know where the hydrogen bond donors and acceptors are in these structures. So if you want to indulge me, you can take a look at these structures.

This hydrogen would be a donor. You can see that it's a hydrogen on a nitrogen. This nitrogen is interesting. It has 1, 2, 3 bonds to nitrogen, which means there are a lone pair of electrons also on that ring system. So that would be a hydrogen bond acceptor. And the adenine nucleobase can accept and give a pair of hydrogen bonds.

And you can work that out for all of the others. So in guanine, there is an acceptor, another acceptor, and a donor, and so on. So those rings in the nucleobases are very important because they have places that you can hydrogen bond to.

Now, is everyone feeling comfortable about this? Does anyone want to ask me a question that might help clarify, because it's quite-- yeah, do you have a question?

AUDIENCE: [INAUDIBLE] What does uracil [INAUDIBLE]?

BARBARA IMPERIALI: What does-- sorry?

AUDIENCE: Uracil.

BARBARA IMPERIALI: Uracil. These are all-- sorry. All these nucleobases have fancy names. So, so far, I've shown you the structure of adenine, guanine, cytosine, and thymine. Uracil, which is not drawn on the board, is very similar to thymine, except this methyl group is a hydrogen.

Knowing the names is also complicated. I really care that you understand the hydrogen-bonding patterns; not to draw the whole structures, but to identify hydrogen-bonding patterns; not to remember fancy names, because there's no logic to those names; but really, to remember ribose, deoxyribose, phosphate and phosphodiesters, purines and pyrimidines, just the sizes of them to pick them out. Does that make sense, what I want you to know, and what you can remember if you think it's interesting?

Now, in nature, we use the nucleotide building blocks or the nucleotides in many different ways. It's not just in DNA and RNA. And so here, I'm showing you some really important nucleotides that are found in nature. And I'll give you a little bit of information about their signaling.

So here are the components that you can pick out. There is, in this case, a ribose sugar. In this case, it's phosphate, but it's a phosphate triester. So it's got three phosphates in a row.

And here's a nucleobase, which is a purine. And this is adenosine triphosphate. So it's one of the bases, one of the nucleotides used in energy, energy transfer. In a lot of metabolic processes, we use ATP as a molecule that has energy that can be unlocked for chemical processes.

There's another one of these, which is guanosine triphosphate, where the nucleobase is different. They're both purines, but they have different structures. You can see them there.

And then finally, the last one I show you here is a nucleotide that has a cyclic phosphate. But it still has a nucleobase, a ribose, and a phosphate. And this is cyclic AMP. And when come back after Professor Martin has talked, we'll talk about the role of cyclic AMP as a second messenger.

So these two molecules, in addition to being building blocks for DNA and RNA, also are forms of energy where you can use ATP or GTP as a form of energy in a lot of metabolic processes. And in fact, though, when we start constructing proteins using the ribosomal system, you'll notice we use GTP as a form of energy, not ATP. It's interesting how nature chooses to do that. Any questions about this?

One tiny wrinkle left to deal with, and that's a little bit more about those building blocks for the nucleic acid, and one more item that it's useful to understand the name of. So here are the five nucleobases, two purines, and three pyrimidines. In DNA, we have AT, G and C, so A, T, G, and c.

So we have different building blocks. Three are common to both polymers. One is different. Uracil and thymine are exchanged when you go from DNA to RNA. The pyrimidines are cytosine, uracil, and thymine.

And in RNA, you have a AU, G and C. So there are reasons for these differences, and I'll nudge into some of those chemical differences in a moment. So the information up there is the same information that I have on this board.

The next thing I need to talk to you is we very commonly use the term, or two terms, nucleoside and the nucleotide. How irritating is that? The nucleoside is just the ribose plus the nucleobase, but no phosphates. As soon as you put on phosphates, they become nucleotides.

So for example, nucleobase, ribose, and in this case, a phosphate on it. And that becomes a nucleotide. No matter how many phosphates they are, it's called a nucleotide.

I'm less concerned that you will remember that nomenclature, more that you know what it's all about, because otherwise, it might become a little bit confusing. So just remember, if you can remember that. But I think I've tried to define the things I would like you to remember-- the building blocks, the numbering system, the phosphodiester linkages, and the nucleobases, as far as understanding where donors and acceptors are for hydrogen bonding.

And there's one thing. So we call that a nucleoside, whereas we call it a nucleotide when it includes the phosphates. And there's one thing that you want to notice, is that the bond from the nucleobase to the ribose is a glycoside bond. It's a bond to a carbohydrate.

So that's why it's called a glycoside bond. There are glycosidases that cleave the bond from the base to the sugar. Those are very important when we have mutations in our DNA, and we want to cut out the sugar to fix it so it doesn't get misread in the biosynthesis of DNA, in the biosynthesis of messenger RNA.

So that bond is important. We may often talk about it, but only when we get to learning about how DNA sequences are corrected if there are mistakes in those sequences. And that will be later on.

So let's start to now look at the polymers. Now, I want to tell you that by the early 1900s, people pretty much knew the structure, the noncovalent structure of DNA. And I'll describe it to you now.

DNA is made up of nucleotides. And this is its basic structure, where you have a phosphodiester backbone linking riboses, and each of those ribosomes is modified with a purine or a pyrimidine. And that is the basic structure of a nucleic acid polymer, only it's very, very, very, very long.

So let's take a look at the components here. Look at the bonds. And maybe on your notes, just highlight the bonds and some of the things I'll talk about.

So first of all, the numbering system here, we always talk about a nucleic acid. And we describe the sequence of the nucleic acid based on from 5 prime to 3 prime. Because the phosphodiester bonds join the 5 prime-- there should be a number there-- and the 3 prime

sites. So the linkage would be here, would be 5 prime and 3 prime joining to the ribose molecules.

So the architecture of that nucleic acid is a polymer that includes a phosphodiester backbone linked by phosphate esters-- that's 1 phosphate ester; that's the other one-- on two of the OHs of the ribose sugar. When this is DNA, there's no OH group on that carbon site. That would be the 2-prime site.

You can see-- you can pick straight out that this is DNA. The sequence is then defined by what the identity of the base is here. So this would be guanine, adenine, thymine on that sequence.

Now, by convention, if we write out this sequence, the way the sequences are written, are 5 prime to 3 prime direction. So if I look at that, I would be able to name it as an A, G, T sequence, because we always write the sequences 5 prime to 3 prime. We can remember that later on because we actually also build sequences 5 prime to 3 prime.

So there are some conventions in biology and biochemistry. You want to remember that by convention, we write peptides N terminal to C terminal. But we also build them N to C. So that's why the convention is strong, and it's good to remember, because it can get you out of a lot of trouble if you remember those things.

Now, when we are building a DNA polymer, we grow that sequence. You'll see the biochemistry for all of that polymerization in the next class. It's amazingly cool how the entire contents of a cell, the DNA, can be replicated in amazing time frames, but all through growing those chains from 5 prime to 3 prime.

So when we add another building block on, we remove a molecule of water. So that's a condensation reaction. And we form a new phosphodiester bond. So in the biosynthesis of DNA, you keep on adding new nucleotides to the 3 prime end. There's a chemical reason for that.

When we build DNA, we don't just cram the two groups together. We, rather, come in with a triphosphate and use that activated triphosphate as the new building block. And you kick out triphosphate. And you'll see that when we talk about DNA synthesis.

But what I want you to remember here is this is another condensation reaction. We talked about them when making peptides. We talked about them when making carbohydrate

polymers. And now we're seeing, once again, a condensation reaction to make a nucleic acid polymer.

Now, the last term that's kind of worth mentioning is the word nucleic acid. What's that about? I don't see any carboxylic acids. It turns out the polymers of DNA are very acidic because the OH group on those phosphodiester backbones is very acidic. So you give up H plus. And this is in its most stable form as O minus. So when

DNA was first isolated, it was isolated from white blood cells by isolating the nucleus. And it was found that it was a very acidic material packed into the nucleus. That's why it was called nucleic acid, acids in the nucleus. Before people even understood anything about the composition, it garnered that name, nucleic acid. So we talk about polymers of nucleotides, we call them nucleic acids.

Then with respect to writing our sequences, we could write them in this way. So pdGATC. That would be that structure. What do all the little extra Ps and Ds stand for?

The P stands for whether there's a phosphate at this end. The D stands for whether it's a deoxy sugar as a building block. Going all the way to the other end, there's no little p at the other end. So it means that OH is free. Does everyone understand that shorthand writing?

There's another way I could know this was DNA without needing to put deoxy on each of the building blocks. Does anyone know how I know immediately it's a stretch of DNA? Yeah?

AUDIENCE: No uracil?

BARBARA IMPERIALI: Yeah, there's no uracil, and there's thymine instead. So in principle, as long as there's a T in there, you know it's DNA. As long as a U in there, you know it's RNA.

Now, let's talk about the noncovalent structure, because I really feel that that's the most exciting part of this entire endeavor because the covalent structure really doesn't allow us to understand how DNA stores information for building proteins. It doesn't tell us that much about it. It looks like a cool polymer, but we can't really understand the details by not looking at the covalence of the noncovalent structure.

So there was one key piece of information, and it's called Chargaff's data. And this piece of scientific information ran around the scientific community in the early '50s because it seemed incredibly important. And what Chargaff's data was, he collected all kinds of organisms, and

then their nuclei, and then measured-- or their DNA-- and then measured the ratio between the purines and the pyrimidines.

He measured the ratio of the large ones and the small ones of the nucleobases. So how many of these relative to how many of those? And what he found by looking all across organisms from all domains of life is that there was a one to one ratio of purine to pyrimidine.

So that became very interesting, because what it suggested was that in some way, the noncovalent structure of nucleic acids had some correlation between the number of the purines and the number of the pyrimidines. And what you can imagine is it sounds like we're always pairing a small one with a large one by looking at that number. So this is really, really important because it's like the light bulb that went on with respect to understanding the structure of double-stranded DNA.

So despite all kinds of variations, some organisms have a lot more GCs. Some have more ATs. But no matter what, the ratio is always one to one. And this ultimately led to understanding the noncovalent structure of double-stranded DNA because it provided clues to how there could be some way that information was coded, but then could be replicated.

Now, the next thing that became the clue to the structure of double-stranded DNA came from a very talented researcher, Rosalind Franklin, who sadly died way before her time of ovarian cancer, really, in large part because she spent a lot of time near X-ray beams. So that would have caused mutations to her DNA.

And she developed a way to make fibrils of DNA that were ordered enough to collect electron diffraction data. And that diffraction data actually gave a clue to some of the dimensions of the double-stranded DNA structure. And it actually was the clue that told the spacing between the strands of DNA. So it really was a piece of information that you simply couldn't do without.

With Chargaff's data and with this, what was called Photograph 51, it really gave you the clue. And it was really during those years that Watson and Crick were desperately model building to try to understand the noncovalent structure of DNA. And once they had those two pieces of information, they could actually put together hand-built models.

This looks kind of clunky, but I know the room they took this photo in from my years at Caltech. In fact, I can recognize the room. They built not just little tiny molecular models, but big molecular models so they could make measurements to say, the diffraction data told me this

was so many nanometers apart. And they were able to piece together the structure of double-stranded DNA.

But I still haven't shown you how those two strands come together. It's really intriguing, because at that very same time, Linus Pauling, had been-- done very well with the structure of the alpha helix and proteins, also was trying to figure out the structure of DNA. But he came up with a sort of a crazy structure where he thought that it was a triple-stranded structure where the bases actually stuck out, and somehow, this triple-stranded structure coded for replication of DNA.

Now, there's a ton of things that are really awful about this structure. First of all, it's a triple-stranded. But the other terrible thing is there's so many phosphates in the backbone there would have been massive electrostatic repulsion. Those sequences would want to blow themselves apart because you can't cram that much negative all in one place.

But it was really an intriguing sort of sociological phenomena of the time. Pauling was a major pacifist, and he was really, really active in nuclear disarmament. And they said that his mind just wasn't on some of this stuff and that this model came out of him really worrying about other things and not focusing on the DNA structure.

So let's try to explain Chargaff's data by looking at the nucleobases and thinking about how they might come together. So here I show you the structures of the four nucleobases in DNA. Wherever I have an R, you can assume that's part. That's a ribose that is part of the phosphodiester backbone. What we want to understand is, how do the nucleobases come together to form some kind of pair that could be useful to programming their resynthesis?

So I've drawn them all here, but it's not quite intuitive. I need to do a little bit of flipping around to line things up better. And the other thing I need to do is get things at the right angles so you can start seeing how those bases might come together, because Chargaff's data dictates that you have a purine and a pyrimidine, purine pyrimidine. You have pairing between the nucleobases in your double-stranded DNA in a structure that looks more like this.

And in each case, you've paired a purine and a pyrimidine. So what I want you to do is take a look. I've shown you now where donors and acceptors are. You can go back and do this for all the nucleobases. But I'm going to do this for you right now, by showing you the donors and acceptors of hydrogen bonds within those structures, what I've done is I've lined them up beautifully so they look straight at each other, so you can tell that there is a complementarity

between a purine and a pyrimidine that makes very nice hydrogen bonding, which is the noncovalent force that's very important.

Between G and C, I can set up three hydrogen bonds. Between A and T, I can only set up two hydrogen bonds. So the one purine is complementary to one of the pyrimidines. One purine is complementary to one of the other pyrimidines.

And then we can draw those hydrogen bonds in place. That totally explains the measurement from the Franklin data of the distance, the width of the double-stranded helix, because it's identical for both of those base pair options. And that gives you the structure that forms the noncovalent structure of DNA, which is a series of interactions where the solid line is the phosphodiester backbone, but sticking out like steps on a spiral staircase are the bases, where each base is complementary to a specific additional base. So it predicts the Chargaff ratio, and it also predicts the distances.

Now, within all the model building, it became quite clear that the structure, the noncovalent structure of DNA, was afforded by antiparallel strands, where one strand went in one direction, 5 prime to 3 prime, and the other strand went in the opposite direction, 5 prime to 3 prime. When we start replicating DNA, we're going to see that that's pretty convenient. But thermodynamically, it is also the favored orientation.

So let's just look at the orientation. Where you would draw one strand of DNA, 5 prime to 3 prime, now I've taken this all down to cartoon level. These are the phosphate diesters, the riboses, the 3 prime end, and the 5 prime end and the bases that come off at the 1 prime carbon. And then when you pair it with another strand, one strand goes in one direction. 5 prime-- whoa, I don't know why this is misbehaving, 5 prime, whoops-- 5 prime to 3 prime. The other strand goes in the other direction, 5 prime to 3 prime.

And when asked this question a few years ago, I couldn't really explain it very well. I just said it had to be because it always has been. But what's really cool is people have been able to solve the crystal structure of a parallel pair of DNA strands. So this is canonical DNA, the beautiful antiparallel structure. And it's very regular, very, very even.

It turns out, though, when you try to pair the two strands in a parallel orientation, they're very uncomfortable, and it's much less stable. So the antiparallel orientation is very important for the thermodynamic stability and the optimum hydrogen bonding interaction of all those bases that are pairing. So it's actually what nature favors because it is more stable. Any questions?

And this, it's on your slides. But you can see just how regular DNA looks so organized, whereas the antiparallel one, the one, the parallel one, really does not afford you good hydrogen bonding interactions at all. So let us now-- so what we've done now is we understand the structure of DNA, the noncovalent and covalent structure of DNA. We understand it's antiparallel.

What we'll do in the next class is show how you can peel apart those antiparallel structures to make unpaired structures. And you can use each of them as the template for the synthesis of a new strand of DNA. So you can get two daughter double strands from a single parent double strand. And that all comes from understanding the structure.

Now, what I want to do is move you just very briefly to the structure of RNA and comparing the DNA and RNA structures, because there are some differences. So let's just work through what the differences are. I have this written down. And the differences are very important for the functional properties.

So DNA, RNA. First of all, obviously, deoxyribose, ribose. And you may go, why, why, why is nature so complicated? Why do I have this extra factoid to remember about RNA versus DNA?

And it's really amazing that the difference between having that hydroxyl on the 2 prime position versus not happening, not having it, makes enormous differences to the stability of the polymer. RNAs breakdown very, very readily. DNAs are stable for the lifetime of a cell, all perfect in the nucleus or at mitochondria. They stay intact.

So there's a stability difference between the two sugars. Because DNA has to be the place where you store your genetic material, it's got to stay good, whereas RNA is the message that you make transiently to program a protein being made, and then you want to get rid of it. So we need the differences in stability that originate from that small feature.

ATGC-- there's the difference-- AUGC in the bases. The most common DNA is double-stranded DNA, whereas RNA forms various structures, so much more irregular structures than the DNA, probably in part because the ribose is substituted differently. So that continuous strand of double-stranded material is not quite so stable in RNA.

We find DNA principally as double-stranded DNA. But the RNA we find as transfer RNA, messenger RNA, ribosomal RNA-- it does go on forever-- short interfering RNA. So various

RNA is used for a lot of purposes, whereas DNA principally stays as the double-stranded DNA. There's a little double-stranded RNA, but it is a precursor to some of these other forms of RNA.

So this slide just summarizes some of that for you, the differences comparing DNA and RNA. And so what we'll see later is how RNA lends itself to these interesting structures where you still have some base pairing, but you have a lot of loops and turns and diversity of structure. And that's really kind of the origin of this RNA world, where RNA structures were not-- could have variety of form that might contribute to different functions beyond just as a message, as a place to store a DNA message. So there are a lot of things that one can understand about DNA by knowing its hydrogen bonding patterns.

So can you guys guess which of these strands would have a complementary strand and be the most stable double-stranded DNA? So this would be one strand. You could draw for each of them it complementary strand. Can you guess the clues to figuring out which would have a most stable organization of the antiparallel double-stranded DNA? What would I be looking for? Yeah?

AUDIENCE: More Gs and Cs [INAUDIBLE]

BARBARA IMPERIALI: So number one, higher GC content because Gs and Cs form three hydrogen bonds. As and Ts only form two. And what's the other determinant, just looking at those structures? Yeah?

AUDIENCE: [INAUDIBLE]

BARBARA IMPERIALI: Yeah, you are doing-- no. It's actually even more silly. It's more simple than that.

AUDIENCE: Length?

BARBARA IMPERIALI: Length. So all you do is you go along and say, I can make three hydrogen bonds, two, three, two, two, three, two, two, two. So you truly just count hydrogen bonds in its partner sequence, and you can guess which is going to be the more stable because it has the most hydrogen bonds. So we might ask you that. Which one will come apart?

Now, the intriguing thing about DNA is you can peel it. You can heat it, and it'll come apart. But it doesn't denature the way proteins do. If you just cool it down, it comes back together. So another feature of DNA is that you can heat, denature, and then reanneal exactly how it was in

the first place. It doesn't denature to something that's not very useful.

And now the question, can you draw the complementary strand? I always find, of this top strand here, which of these is the complementary strand? Frankly, the best way to do it is sketch out the complementary stand. You can see it kind of upside down because it's really hard to draw things 5 prime to 3 prime when you're also trying to figure out base pairing.

So draw it upside down. Make sure you know the 5 prime and the 3 prime end. And then you can guess the right answer for these types of questions about complementary strands.

Now, one last question, the stability of double-stranded DNA. I've made a whole big deal about hydrogen bonding. That's what holds it together. What other forces could be at play in double-stranded DNA that might contribute to its stability? Any thoughts? What else?

Well, it certainly doesn't look like it's charged, because the predominant charge is negative. There's not an-- you've probably got metal ions there, kind of neutralizing that charge. What would be the other force, and how would I describe it? It's a tricky one.

So we've got these bases, and they're pretty hydrophobic. They're planes. They have electron density on both sides. So it turns out there is some stability gained between the packing of the steps of DNA between each base pair with the next, with the next. So there are hydrophobic forces.

And researchers at Scripps have actually proved this paradigm by making extra DNA bases that don't have hydrogen bonding partnerships, but just provide the stuff that's the flat hydrophobic entity with the right size that can slip into DNA sequences and make stable [INAUDIBLE], make stable not really base pairs anymore, but just be stable in that polymeric structure. Are people understanding and following that?

So finally, when we look at the structure of DNA, there are some trenches where things can bind to, proteins can bind, and we talk about the major groove and the minor groove. But I will talk about those later on when we talk about transcription factors.

Now, I just want to, in really triple-fast time, and I'll put this on the website, there's tremendous interest in using the building blocks of DNA for information storage in computing. So if you look up DNA-based computing on Wikipedia, you'll learn a whole lot about it. Because what's so exciting about it is it's an organized nanoscale material that can be programmed to base pair and form certain structures.

So in the sort of range of different sizes, there's been a lot of interest in DNA as a material for information storage, not for your genetic material, but for plain old information storage. So people have learned how to build structures of DNA where they can construct these sort of cruciform structures by base pairing. They can make the arms of these structures a little bit extended. So you could start joining those things together to make very defined three-dimensional entities.

They went kind of nuts doing this sort of stuff because you can build sort of tetrahedra and other sort of shapes and sizes, all by strands that base pair, that are about 10 base pairs long, that are stable, and only complement certain other base pairs. So you could literally build up-- they often called it DNA origami because you can build up macroscopic structures just by the assembly of strands of DNA that will ultimately fold to form the best complementary DNA to form the structures.

And it's also been found-- as I said, they went completely nuts-- smiley faces and stars and stripes and so on. But the most valuable thing you can-- as I said, you can read more about this-- is to use DNA as logic gates to define and, or, or not, so the sort of three options, and actually use them to program certain puzzles where the DNA will spit out the answer to a particular puzzle through a logic diagram.

So those of you who are interested in computing and these kinds of logic puzzles may want to read a little bit more, because DNA is such a reliable noncovalent structure, where those base pairs are incredibly reliable, that you can start envisioning not just building double-stranded DNA, but building all kinds of architectures or programming things with the sequence of DNA. And that's it for today. And that's the end of the biochemistry section.