

**PROFESSOR:** It's going to be a great lecture today. It's about proteins. I love proteins. Don't forget the handout, yeah. OK, so I'm going to briefly wrap up the lecture we were doing on Friday because there were a couple of things that I wanted to make a note of, and then we'll move on to section 2.3 about amino acids, peptides, and proteins.

Now, in the last class, I introduced you to the lipidic molecules, and you can pick them out of a lineup because they are rich in carbon-carbon and carbon-hydrogen bonds. As you can see here in these line-angled drawings, the majority of a lot of these molecules is carbon-carbon or carbon-hydrogen hydrogen. They are molecules that are mostly hydrophobic, so there are some terminologies here. Whoops.

Hydrophobic, which can also be referred to as lipophilic. You either-- you can hate water and love fatty acid or fatty types of materials, so those both terms are synonymous. And some of the lipids are what are known as amphipathic, and they include hydrophobic and hydrophilic components.

There are a couple of tiny terms that I didn't mention explicitly, so I just want to go ahead and do that now. For example, in this phospholipid structure-- and we'll talk about these-- they have long chain fatty acids attached via esters to this glycerol unit, so there's one here and the second one here, and then what's known as the polar head group. In those fatty acids, they could be fully saturated. It means they have no double bonds in the structure, so that term saturated is equivalent to no double bonds, so no carbon-carbon double bonds. Or they could be unsaturated, where there is a double bond within it, so that's one or more double bonds.

And those double bonds take on a particular shape because there's not freedom of rotation around double bonds the same way there is around single bonds. So those single bonds, you can twist them around and twist them around, but the double bond geometry is fixed. And so double bonds, we refer to them as either trans, where the two groups are on opposite sides, leaving the double bond, or we refer to them as cis, where the two groups are on the same side. And we tend to use that cis and trans sort of naming system in a lot of other contexts as well, but you almost always want to remember that trans is as far away as possible, cis is closer than trans.

All right, so I'm just going to take you forward to the phospholipid structure. This is a very important semi-permeable membranes are made up through the non-covalent, supramolecular association of phospholipid monomer units. Here's a monomer unit up here. You see it has an amphipathic structure, with a lot of hydrophobicity but also hydrophilicity, and these molecules assemble into supramolecular structures that form the boundaries of your cells.

Saying they are semi-permeable tells us a little bit about what can go through them. If they were fully permeable, anything could come and go and they wouldn't be much use frankly. It's like leaving the door open the whole time. But because they're semi-permeable, only a few things can come and go without extra help, and other things need active mechanisms to go through. So let's take a look at the boundary here.

So when you see a membrane bilayer, they are-- they're often shown looking like this, where every one of these units is a phospholipid, and there's water on both sides of the phospholipid, because that polar head group is interacting with water on both sides. So down here could be the inside of the cell. Up here could be the outside of the cell. And a lot of cells, especially eukaryotic cells, the ones that make us up, have a lot of endomembranes, membranes within the cells. For example, forming the boundary to the nucleus or to the mitochondria. Yes?

**AUDIENCE:** Why is there a [INAUDIBLE]

**PROFESSOR:** Oh, this guy must be-- so this looks like it's probably a saturated fatty acid. So what do you think this one might be, folks? Unsaturated. And what's the double-bond geometry?

**AUDIENCE:** Cis.

**PROFESSOR:** Cis. Yeah, it is. It's like a-- it looks like a ballerina or something. OK, so we have a lot of concerns, and we'll see later about how things get in and out of cells. But most commonly, things like oxygen or water and other small hydrophobic molecules can pass readily in and out through the semi-permeable barrier, but other things, things that are charged, things that are big, need a different mechanism to get in and out. And we will see later on how proteins provide the opportunities to cargo things into cells or out of cells, even very large entities, and there are certain mechanisms whereby that happens through a semi-permeable membrane, OK?

I want to show you the other feature of membranes. They are self-healing. What this means is if you poke them, you poke a hole in a cellular membrane, You? Basically push apart those non-covalent forces. Once you take the thing away, be it a needle or a very fine glass capillary, they seal right back up to close to close the hole in the cell wall, so that kind of tells us that they're non-covalent forces.

So this is a really cool video of someone doing micro-injection into eukaryotic cells. The needle points to the cell, approaches the surface. You can drop something into the cell, and then the cell closes and maintain-- regains its integrity of the barrier, so this is a very cool observation. People do this. They have to not drink too much coffee because it's quite complicated to do a lot of micro-injection, because you can really cause carnage in your cell population if you're not very dexterous with the micro-injection but people can be very good at it.

So I just want to ask a couple of questions before, give you a couple of things to think about before we close up. The lipids. So here's a typical lipid bilayer, where I've highlighted a single lipid. And the colors, those are the head groups, and all in white and gray are the hydrophilic components, and just one of the phospholipids is highlighted, and that would be this molecular structure here. So first of all, what do you think the non-covalent forces at that membrane interface may be? That is, what's going on here at the interface? What are the types of interactions that you might have there?

Give you a minute to think about it, and I want to show you that I'm actually giving you a clue here, because you can see the structure, negative charge, positive charge, but also remember this is a barrier to water, so there are other things going on with the solvent that the membrane is sitting in because there's water surrounding that barrier layer. Anyone want to tell me what the answer is, and why? Yeah, did you-- are you-- yeah.

**AUDIENCE:** Hydrogen bonding.

**PROFESSOR:** Yeah, between what and what?

**AUDIENCE:** Like the oxygen and [INAUDIBLE]

**PROFESSOR:** Right, so water. Water is a good hydrogen bond donor and acceptor, so there will be hydrogen bonding. What about amongst all those lipid head groups, what's the other major force? Yeah?

**AUDIENCE:** Electrostatic force.

**PROFESSOR:**

Between the different charges. So the correct answer here is both of them. Don't think it's just electrostatic, it's both. It's electrostatic amongst the head groups, hydrogen bonding between all that sort of dense bunch of charge, and the water. And then the other question, what type of molecules can get across? I've already answered that question to you.

Salts are going to need ways to get in and out. Small proteins are too big to dissolve in that membrane through passive mechanisms, so we're going to have to figure out how to get proteins in and out of cells. Neurotransmitters, such as this, this is GABA, or gamma aminobutyric acid. It's charged. It just can't get through without a transporter of some kind, and it's actually proteins that end up doing the heavy lifting of the transport processes that we'll see.

OK, so moving along. This section will be about the building blocks of your protein macromolecules, which I want to remind you comprise 50% of all of the macromolecules, so that suggests it's a pretty important class of macromolecules that has a lot of different functions. Now, the amino acid building blocks-- blocks look pretty simple. They're called amino acids because they have an amine, the carboxylic acid, and there's a carbon that is tetrahedral between the carboxylic acid and the amine.

And the simplest of those is when those are both hydrogen, but most of the amino acids are differentiated from that-- this one I've showed you on the board. This amino acid is glycine. Usually, when it's just a lonely amino acid in aqueous solution, it's in a different charged form, just consistent with what we talked about in the last class. And I put it here.

So this is glycine. It's one of the 20 encoded amino acids. That means the amino acids that are made through ribosomal biosynthesis through a code that's provided by the messenger RNA, so they are encoded by messenger RNA. Later on, you'll see all of the beautiful mechanics of those processes.

Now, this table looks pretty complicated, so I'm going to deconstruct it a bit. But what I first of all want to assure you is that these-- you will always get a handout with these structures on them. We are not asking you to remember these structures. You might become familiar with some of them, but you do not have to remember them.

You'll have a table that shows them, but on that table, I won't necessarily give you the information on what their properties are, because those are things that you should be able to

spot by looking at their chemical structures, all right? So that's important. So these are all line-angled drawings, so you see the carbon. The hydrogens aren't shown in there.

The charges are shown for what's called the side chain, because most of the amino acids have a side chain. The amino acids are also chiral, but you'll learn more than you ever wanted to know about chirality in 512, so I won't weigh you down with any of those properties. So there is a side chain that dictates the properties of the amino acids.

One tiny detail, the amino acids that are encoded in our proteins are all what are known as alpha amino acids. There are other amino acids. GABA, that I showed you on the previous slide, is not an alpha amino acid. Actually it's, a gamma amino acid. These are called amino acids because the amine group is at the alpha position relative to the carboxyl. Don't need to know a lot more about that with respect to that.

So let's take a look at this set of amino acids, and what you see is amino side chains with rather different properties. I've amassed-- here's glycine at the very top. All amino acids have a three-letter code or a one-letter code. I particularly enjoy using one letter codes and spelling out people's names in peptides and things like that. I'll let you do that in the privacy of your own room. It's kind of amusing to see if your name actually spells out a peptide. Some of us-- if I get a little stopped stuck with Barbara because there are no B amino acid one letters with a B.

The next most abundant type of amino acid have hydrophobic side chains. What that means is they have a lot of CHs, but not a lot else, right? So take a look at them. Alanine has a methyl group, for example, where I've shown the R, that would be alanine.

And they get increasingly big. They're quite large. Some of them have quite extended size chains. Other ones have side chains with rings with double bonds in them. Those are what we would designate in organic chemistry as aromatic. They show-- they are still hydrophobic, but they show different properties to this other set of amino acids.

Some of these amino acids may actually have polar groups in them, but their major feature is that they're hydrophobic. But in an amino acid, such as tyrosine, you could not only have hydrophobic interactions with that ring system, but also hydrogen bonding with the OH on the tyrosine, so some of the amino acids can do a few different things.

The next set of amino acids are those that are polar and charged, and I've shown you the

most common state of all of those amino acids, but you already know that the amine of lysine is likely to be charged. This guanidinium group of arginine, take my word for it, it's charged. It's a bit more complicated to draw. Histidine is also one of those that's annoying to draw, but the negatively-charged side chains with a carboxylate are both negatively charged, and that's something you would remember from the previous class hopefully. And then finally, there are amino acids with polar uncharged side chains, such as those shown here.

Now, this doesn't look like a very exciting set of building blocks. How can life run on things made of 20 relatively simple building blocks with functional groups? And it's that the building blocks are not functional themselves. It is the polymers that are made up of amino acids, and I'll always call them AAs because it's easier for me. The polymers of amino acids are heteropolymers. That means they're made up of a bunch of different monomer units when they're called heteropolymers.

And the other important thing about these polymers is that they are of defined sequence. What is the sequence? It's the order in which the amino acids appear. So I'm writing that down, order. And all the functions of proteins are dictated by the order of the amino acids, so let's take a look at the sidebar here.

So once again, remember a couple of things that we will always give you this table to think about. Ooh, come back. There are a couple of outliers I just want to mention quickly. So I talked to you about glycine, the simplest amino acid with no elaborate side chain. Proline is a little odd because its side chain is kind of in a cyclic structure, and towards the end of the class, I'll talk to you about collagen, whose structure is totally dependent on the involvement of proline in the sequence of the amino acids that make up collagen.

And then the last sort of unusual amino acid is cysteine. It has a thiol, and the one clever thing about cysteine-- I'm just going to put a bit of a peptide here. One cysteine, and then I'm going to put a second cysteine, and these are going to be deemed in a peptidic structure. What cysteine can do is it can exist either with the thiol side chain, SH, or it can be at a different oxidation state where the two sulfurs are joined to each other.

So for the most part, your linear arrangement of amino acids that dictates sequence is solely held by-- together by the covalent bonds and the peptide backbone that we'll talk about in a minute. But occasionally, folded structures, if two cysteines are close to each other and the environment is oxidizing, they will form a cross-link. But they're not what drives folding. They

kind of fall into place later on, but that just sort of sets cysteine apart a little bit for its properties, all right?

OK, so coming down the side here. Amino acids are assembled in a unique linear polymer of defined order, and we designate that defined sequence the primary sequence. And proteins can be 1,000 amino acids, 1,500, 100 amino acids. They can be various lengths where they, you know, we would generally consider the smallest protein to be about 400 amino acids, and you might go up to thousands of amino acids. I'm going to write 2,000 or more here.

When the proteins are smaller, they are not capable of adopting too much ordered structure, and we mostly call them peptides. Peptides are sort of shorter sequences, so peptide sequences. So this would be a protein, and peptides, probably two to 39 amino acids, but these breakpoints are a little bit more vague. So the primary sequence will define the structure of a protein, and we're going to start to talk about the hierarchical structure of proteins as put in place, and that's the primary sequence,

And that primary sequence is kind of a cool thing because it's very specific. It defines-- it's got encoded into its structure, the three-dimensional fold of the protein, OK? All the information for the folded, compact, globular structure that's functional is encoded in that primary sequence. It's a cryptic code. We may not be able to tell by looking at it what it really looks like, but all the information is there in order to program the folding into a globular structure.

So the primary sequence determines the fold, and it's the fold of the protein that mandates its function. It's not the sequence of the protein. The sequence defines the fold. The fold, the three-dimensional form, defines the function, OK? So that's very important.

And I think it's absolutely amazing that with a relatively limited set of building blocks, we can define so many different functions of all the proteins in our body that may be structural, they may be catalysts, they may be things that transfer information from the outside to the inside of cell. All of that is programmed with this rather limited set of building blocks, OK?

Now, let's now talk about peptides because one gets a little frustrated looking at single amino acids. They don't tell us so much about the peptidic structure, so I'm going to draw two amino acids, and then I'm going to tell you one important thing. So let's put R1, and I'm going to draw another amino acid, and I'm putting it in a particular orientation. R2, because that designates that these might be different amino acids. For example, if R1 is H, there's an implied hydrogen here, that would be glycine. If R2 is a methyl group, there's an implied hydrogen there, that

would be alanine, all right?

When nature bonds all these amino acids together, it carries out a condensation reaction to form a peptide bond between these two components of the amino acid, the amine and the carboxylic acid. And now I'm going to draw you the first of the dipeptides that you'll meet. And there are so many things to tell you about these structures, it sort of drives me crazy thinking about, oh, I must remember to tell them that or I've got to remember to tell them that, because the structures are cool. R1, R2.

OK, so this is a dipeptide, two amino acids, and there are some characteristics I want you to remember. When we write out peptides, we always write them N to C. So in that peptide, this would be the carboxyl terminus, and this would be the amino terminus. If you don't always remember to write things in this order, and you tell your friend, oh, go and get this peptide made, and you put it down in the wrong order, they'll make the wrong peptide. So you always-- there is basically an agreement amongst everyone that we always write from left to right, the sequence of peptides.

The next important thing about this structure, as you look at it, there are several bonds joining the polymeric structure. Many of these bonds show free rotations. You can twist them around, there's nothing stopping that conversion. All of these show freedom of rotation.

But the amide, or peptide bond, is unique in that there's restricted rotation about that bond. So it's as if you've got a linear polymer, but every third bond has kind of stuck in a particular orientation, which starts to define a lot of details about protein tertiary structure. It's not complete spaghetti. It's like spaghetti with little bits that haven't been cooked. They're stiffer than the rest of the sequence.

And the other really important thing about the peptide structure is that embedded within that structure, there is the amide or peptide functional group where, remember, this can be a hydrogen bond acceptor, and this can be a hydrogen bond donor. Once you know that, the next few slides will make a lot of sense as we talk about higher-order structure of proteins. So let's just take a look at that with a slightly longer peptide.

By convention, if I'm going to draw a peptide that's methionine isoleucine threonine-- you can look up that names-- those names on the chart-- that would be the MIT peptide. These are the three amino acids. I'm going to condense them into a tripeptide.



When I condense three amino acids, I spit out two molecules of water, and I put in place two amide or peptide bonds. If I go down this backbone, every third bond is going to be fixed, fairly fixed. There's not freedom of rotation around it, and every third bond is going to have the capacity to be involved in hydrogen bonding interactions, as I've suggested here, all right? What else is there here?

When I write the MIT peptide, I write M first, I second, T third. If I wrote TIM, it would be a completely different chemical structure with different chemical properties, so the directionality is important to understand, and there you have it. So now you can go home and practice your name in amino acids and draw them out.

If you draw them out fairly sort of sharply, then you'll never get confused about what end's what and where the substitutes are, but it's important to remember as you're making a dipeptide-- oops, I forget this doesn't work. As you're condensing a dipeptide, when you're putting these R groups on, one goes up, one goes down, but these are nuances of the structure that may be lit for-- good for a later discussion.

So here is now a longer linear peptide, and the suggestion of a globular structure that might be found if that peptide was folded up. And the primary sequence here defines the globular structure, and the process whereby you go from the extended primary sequence to the folded structure is called protein folding. And physical chemists and physicists and computational chemists have for years tried to understand how we could predict the folded structure from the primary sequence. It's not simple because what you're doing is you're solving a massive energy diagram, where as you fold a structure up, you're trying to maximize all those non-covalent forces for maximum thermodynamic stability, right?

It's kind of a three-dimensional puzzle where you're trying to have as many hydrogen bonds, electrostatic interactions, and so on, as you can possibly make. So when computational chemists try to fold proteins, they're basically solving a three-dimensional puzzle where they are maximizing interactions. And there are a lot of ab initio and molecular dynamics programs that are now starting to be able to fold proteins into fairly reliable structures, but they don't always get them right because they haven't gotten all the clues yet.

And also while they may be able to do ab initio or computational folding with small structures, the headache gets way bigger the larger the structures get. So the predictors aren't very good at predicting big structures, they're getting better at predicting small structures. And so just to

reinforce to you, the primary sequence is established by covalent bonds, the peptide bonds, but the globular tertiary structure is based on non-covalent covalent interactions, OK?

Now, I want to ask you this. I love cartoons with science in them, but you know, 10%, 20% of the time, they make mistakes, and I felt this one was particularly pertinent. So a bunch of guys lugging around in a lab and says, well, we finished the genome map, now we just have to figure out how to fold it. What is wrong with that cartoon? What fold? Yeah?

**AUDIENCE:** You want to [INAUDIBLE].

**PROFESSOR:** Yeah.

**AUDIENCE:** [INAUDIBLE]

**PROFESSOR:** Yeah, the genome doesn't fold. It's double helical, duplex DNA or something. You're actually folding proteins, so the cartoon is not quite right, but it's sort of kind of cute. All right, now, when we talk about the non-covalent forces that hold proteins together, I just want you to remember from last time this set of non-covalent forces, because if you understand them and recognize them, you'll understand how they may occur in folded protein structures.

All right, so here's a peptide sequence. Here's a puzzle for you. You can go back and figure out what the one-letter code spells there. Just take out your table with all the amino acids. It's appended to the back of your P-set, and you'll be able to see what that very large peptide spells. All right, I don't want you working it out while you're here. You've got to listen to me for the time being.

OK, so the first order, we get it, there's a primary sequence. The next thing to think about is what's known as secondary structure. It's a higher order than just the primary sequence, and it's established by non-covalent bonds, and it's called secondary-- oof, my writing's horrid today. Secondary structure. And those are interactions that are put in place exclusively by interactions between the peptide bonds of what's known as the peptide backbone.

So if I look at the structure, these are the side chains. The peptide backbone is this continuous linear sequence. That's what we would call the peptide backbone, and the secondary structure is put in place by hydrogen bonding between components of the peptide backbone. So for example, a hydrogen bonds, such as that, or a different hydrogen bonding interaction, such as that. Between the atoms that have lone pairs of electrons and the other atoms-- heavy atoms that hold a hydrogen that's quite acidic.

And there are a couple of major forms of secondary structure. What I'm showing you here is what's known as the alpha helix. First deduced by Pauling, in fact, through model building, he said, proteins could form these ordered structures, and an alpha helix is an ordered structure exclusively made up from the hydrogen-bonding interactions of the peptide backbone.

And you can look at this helical structure. It's a continuous strand of peptide, but there are hydrogen bonds between COs and NHs all the way through the backbone, such that this strand of peptide can fold up into a cylindrical, helical structure, where all those R groups, the side chains of the amino acids, are on the perimeter of that helix. So this secondary structure is an important one because it's very prevalent in a lot of proteins.

The next secondary structure is also held together by hydrogen bonding, and it's interactions between stretched out strands of peptides that may not be close to each other in the primary sequence, but they align in the folded structure. And so for example, what I've shown you here is what's known as a-- this guy is then to say this is an anti-parallel beta sheet. And across that sheet, there are continuous opportunities for hydrogen bonding interaction. If the strands run in opposite directions, it's anti-parallel. If they're in the same direction, it's parallel.

These two secondary structure elements make up a lot of the sort of basics of how proteins start to fold. They're key non-covalent forces, and there are also other smaller motifs. One is called a beta turn, where the peptide sequence may go through a chain reversal, so the sequence would look like this. I'm going to just draw it, and I'll talk to you in a moment about ribbon diagrams. And this piece here would be the turn, whereas that would be the interactions enforced by the sheet.

These are the ordered elements of secondary structure. You don't have to be able to figure them out, but you have to be able to pick them out in order to understand the structure, OK? So even those simple elements still it's hard to make big enough structures to have functions.

So as I mentioned in a continuation of the theme, the protein folding is hierarchical, you can start to put together elements of secondary structure to make things that are a little larger. Helix, turn, helix. Helix with a different kind of turn, maybe put in place by a metal ion or something, or a strand, turn, strand, or now something that's a composite of these two major types of secondary structure, the helix and the turn.

And these really start to be proteins that might be big enough to be able to do something, but

they're all exclusively held together by non-covalent forces between the amides or peptide bonds in the backbone of the protein, OK? Not very exciting just yet. Now, one other little clue that people will-- you might see and you might be confused, people sometimes, when they're drawing sort of a quick picture of a protein, they might draw a helix, but instead of really showing it in detail, they might show it as a cylinder, so you might need to pick that out of a structure.

And then I want to call your attention to that, that in all those motifs, when you join one helix to another, you might need to turn a strand to another strand you need to turn, and so on. OK, so this is like taking your very extended stored of polymer, knowing there are different kinks in it, because of the backbone bonds, but folding it up in a structure that maximizes the opportunity for another order of structure, which we'll talk about now.

All right, so we've seen primary. Secondary is just with backbone. And things start to get much more interesting when we get to tertiary structure, because tertiary structure is enabled by all these other interactions, electrostatic, hydrogen bonding, hydrophobic forces, that can be put in place due to the side chains of amino acids interacting with each other or with the backbone structures. So I'm going to walk you through this, so you can sort of get a sense of how these three-dimensional puzzles work on a very small scale.

So look here, that's a very small motif. And what I'm going to call your attention to is when you fold up these motifs, when the secondary structure is in place, a lot of the side chains are near each other, and they can engage in long-distance contacts. And so for example, I'm going to show you interactions between side chains, between side chains and the peptide backbone, or side chains and water.

But what I want to do is take a look at this and see, can you put any of those potential interactions on the drawing that's on your handout? It's pretty obvious where there's an electrostatic interaction, right? Boop. OK, between plus-- get those out of the way, those are the easy ones. And then interactions between hydrophobic groups, where they want to amass that lipophilic structure, so it's not exposed as much to water, so they cluster, so those are easy.

And then you can start thinking about what are all of hydrogen bonds you could draw. Here I've shown one between side chains, between side chains and backbone, between side chains and water, and those may all contribute to the ultimate thermodynamic stability. Make sure

you get your hydrogen bonds right. Remember, two donors don't interact with each other into acceptors, don't-- so this might describe the folding possibilities of that small motif.

Now what I want to show you-- I'm going to-- let me-- is an ab-initio simulation of a folding process. So let me just get that a little bigger on the screen. So this is computing. GB1 is a very small protein that folds reversibly under appropriate conditions, and what I'm going to do is forward you through this video. This is a simulation. This is all computation. It's not looking at anything by spectroscopy or in solution or anything like that.

And what I'm going to do is I'm going to forward you through the structure. This is multi-scale modeling. It's got a lot of details in how it's done, but the starting point is a very denatured protein, all stretched out, right? And what I'm going to do is just show you for a few seconds, you know, this thing's like trying to find its thermodynamic minimum, and it's actually failing pretty badly.

And it does that for about 30-- 60 seconds of the simulations, so I made a point to myself to take you to about minute one, where things start to get fairly interesting. And you're saying, well, what's interesting about that? You see that nascent helix, in the background, the red and the blue, is starting to form strands that are a little bit aligned, and it's trying to find as many connections as possible to satisfy a stable structure.

At a certain point in the simulation, five of the hydrophobic groups are in a little pea. They're in a little hydrophobic cluster, and that's a breakpoint in the folding process, because that gets everything glued together better, so that the rest of it now can start to really find its final place in the folded structure. These early structures are known as molten globules. A lot of the interactions are not yet in place, but the hydrophobic cluster is critical.

But then after that, it's almost as if you're sliding downhill to get all the remaining interactions in place to fold the protein, OK? So protein folding is a puzzle that can be solved computationally by maximizing thermodynamic interactions. So it's sigma this, sum of this, sum of this, sum of that. That's going to get difficult the larger the protein gets, but for small proteins, those simulations really start to make sense, OK?

All right, so let's just move on here. Lost-- ah, good. What did you think of the simulation? It's kind of cool, right? So you can find the link in the sidebar. So just pop these back on now, and that's the folded structure.

All right, so with many proteins, they're much more complex than that. So for example, here's cyclin A. It's involved in cell cycle, and you can see its alpha helix structure dominantly, very clearly, all those beautiful alpha helices. Next to it is the green fluorescent protein, which is a cylindrical structure made up of anti-parallel beta sheets. What's really cool is when you sort of rotate it, you can see all those sheets, but then it does this little sort of curtsy to the audience, and you can look down into the barrel.

And then in some cases, proteins may be a mixture of a secondary structure elements. Here it's a little hard to tell. This is triose phosphate isomerase, but if you look down it, you can see the helices, and there's also a group of beta strands that are held together. So in that protein, it's a mixture of alpha helix and beta sheet.

Now, I'm not going to tell you much about pulling up Protein Data Bank files right now because I want to cover the next topic. And then when we have a few minutes later on, I'll show you. But wherever I show you a structure, I'm trying to show you the Protein Data Bank code, and in the web site, you can see there is a free download of PyMOL, which is the program I used to create all these structures and movies, so you can really look at things.

And believe me, it took me about three years to learn how to use it properly. It'll probably take you about a week or maybe a couple of days. So if I can learn it, you can certainly learn it.

Now, there is one final element of protein structure that people get kind of hung up on, and it's what's called quaternary structure. It's like, aren't we done yet? So in addition to all of these, let's say I have a folded motif, and there's its structure. That would be have primary, secondary, between the strands or the helix, and tertiary structure, right?

But in some cases, proteins hold up to quaternary structure, where it's multiple of these units joined together-- hoo, I could have picked a simpler fold, but that will get you the general gist of it-- all right, where these are actually associated by non-covalent forces. So there's more than one polypeptide chain. In fact, here would be four peptide chains coming together in a higher-order structure that's made up of four of those units.

The prototypic example of this is the protein that carries oxygen around in your blood, which is hemoglobin, and it has four primary sequences that have come together in a tetrameric quaternary structure. Hemoglobin is kind of interesting, because it's made up of two alpha and two beta subunits. If All these subunits were identical, they would be called homooligomers, all the same pieces. If they are different, they are called heterooligomers. We'll see a little bit

more about this when I talk about hemoglobin in the next class, because the features of the quaternary structure are very, very important for the proper transport of oxygen, and single mutations can really mess things up, and you'll see more about that in the next class.

So just wrap that little bit up, proteins are condensation polymers of amino acids. Each protein sequence is defined by covalent bonding. Native proteins. Most of them that are not have quite quaternary structure are folded through secondary and tertiary interactions, these things that we already talked about, and folding is defined by how to maximize all those non-covalent forces to get the maximum thermodynamic stability with the maximum number of interactions. And subunits may also come together through quaternary structure.

OK, so I'm going to talk to you about several proteins throughout the course, but for now, I want to focus you in on a structural protein that provides mechanical support for tissues. In the next class, we'll talk about transporters and enzymes, and as we move on to signaling, things like receptors and membrane proteins and so on. So the protein I'm going to describe to you is collagen.

It is the most abundant protein in the human body. It plays enormous roles. It's not an enzyme, it's not a catalyst, it's not a transporter. It is one of those structural proteins, where the structure of collagen has evolved to provide a mechanical stability to lots of essential components of complex organisms.

And there are many different types of collagens that are found in different parts of the body. For example, bone, tendon, cartilage, and so on. They are all collage and structures, but they have subtle differences, maybe some have different, slightly different, mechanical properties to adapt to the functions that they perform, OK?

And what I'm going to show you is that a single amino acid change in the primary sequence of collagen can destabilize the structure, so it is no longer viable. And the disease type I'm going to talk to you about is a set of diseases known as collagenopathies, and the particular one is called osteogenesis imperfecta. Osteo always refers to bone because collage and plays a critical role in the structure of bone. Bone isn't just bone, it's collagen involved in it.

And it's also this disease is called brittle bone syndrome. And here's the X-ray of a baby born with brittle bones syndrome, and you'll see that the long bones in the upper arm are all irregular because the bones are brittle, and they'll break even in utero. A lot of babies with this defect can't even be born through the birth canal because it would crush the bones, and many

of them don't survive very long at all.

Some survive with different kinds of cases, but their lives are greatly impacted, and they could just sort of hit a table and the bones would break, all right? There are those sort of serious situations where parents are actually accused of abuse to the child, but the child actually had brittle bone syndrome, and it was just through helping them put their clothes on or taking them upstairs, the bones got broken very readily. So osteogenesis imperfecta really describes a collection of these defects.

Now the collagen tertiary structure is shown here. It's actually made up of a type of helix. It's not an alpha helix. It's a polyproline helix, where the individual subunits in that tertiary in the structure are fairly long and extended, and I show you three strands in this polymeric structure, a yellow, a red, and a green. And these rolled together into a three helix bundle that has a fibrillous structure, and then all these structures come together to make the macromolecular structure that is collagen.

It's not just one of those fibrils. It's bundles of those fibrils in a very organized pattern where you could even see that patterning in electron microscopy. And there are many genetic defects of collagen, and what's so important to think about is if you have a defect in one strand that defect will propagate through every single strand. If this is one strand made up of three polypeptide chains, it propagates all the way through the structure.

And I believe I have little time to just show you, here's the collagen structure. I'm just showing you how it's extended. Those are three independent strands, and there's a set of magenta residues in the middle, which come from a defect in the sequence where a glycine has been changed to an alanine.

So I'm going to show you this movie because it shows you right at the center of the structure, there are residues painted in pink. And what I'm going to do is show you close up of that segment. If you look at those cells they're all nicely organized, except where that defect is, and that defect is caused by the change of a hydrogen to a methyl group on three residues that come together, and that bulges out that fibrillous structure and makes it not as compact and beautiful as it should be in the version that's got the glycine there.

So if you look at it, you can even see that helix gets bulged out and it's not as well-aligned as the rest of the structure. And then that defect gets propagated into all the fibrils and results in



the weakening of the bones. Either the collagen fails to form properly, or the collagen, when it forms, it has much less mechanical stability.

So I think that's a good place to stop and I'll pick up next time with hemoglobin. Oh, one last little thing, a couple of things for you to do. There's a great link on the website to the Protein Data Bank to see how enzymes work. And if you have a little time, it would be awesome if you could just take a quick flick through those parts of the text. These slides are posted with these reading assignments, and they're posted in color if you want to look at them again.