

PROFESSOR: OK, well let's move on then, and just talk about the amino acids. Amino acids side chains. And you won't have to memorize these structures. We will give you a chart if you have a problem.

On the other hand, you need to get very familiar with them, so they're old friends even if you can't quite remember how many methylene are in a chain, or something like that. And you will find that they fall into certain categories. And I'm just going to try and give you examples of the major categories.

There are negatively charged side chains. An example would be amino acids known as aspartate, or Asp, in which the side chain which corresponds to the R1 or to the R2 over there, has a methylene group, and then a carboxyl group. But at pH sevenish, which is the pH that you find inside a cell, that carboxyl group would be deprotonated so it would have a negative charge.

The other negatively charged amino acid is glutamate, which also, as you'll see has a carboxyl group.

There are positively charged amino acids. A good one to illustrate this is Lysine, in which there's four methylene groups, and then an amino group at the end. However, again, at pH 7, the kind of pH that you find inside the cell, that amino group is going to get protonated. And so it will have a positive charge on. If you have a Lysine side chain, and Arginine, and in most cases, Histidine, are examples of other amino acids that can have a positively charged group. And why I'm going through all of this, I hope, will become apparent in a few minutes.

Some of the side chains are not positive or negative charged, but rather, they're polar. And we just talked about polar bonds the last time, where you have, the more electronegative an atom is, the more greedy it is for electrons. And if you recall, if you have a carbon carbon bond or a hydrogen hydrogen bond that's nonpolar, and the electrons were distributed equally, the oxygen is greedier for electrons and so there is a little bit of a negative charge there and a little bit of positive charge on the

hydrogen.

Well, that same principle applies to amino acid side chains. Take, for example, the amino acid Serine, which has a methylene group and then a hydroxyl group. Well, here we are. There's an OH bond, so there will be a little bit of a negative charge on oxygen with a positive charge on there.

There's another alcohol called threonine, which also has hydroxyl groups. And you can make amides of both Aspartate and Glutamate, to give Asparagine and Glutamine, and both of these are also polar too. So what I'm hoping you're beginning to get a sense of, you can do an awful lot with the properties of a peptide chain, depending on which amino acids you dangle off the side. And ultimately, that order of amino acids is what's going to be determined by what's in the gene encoding that protein.

Then there are quite a number of amino acids side chains which are hydrophobic. They're sort of fearing water, if you will. The simplest is Alanine, or Ala, which is just a methyl group, or Leucine, is perhaps, a little more obvious because that's got this. And you can see that that's a kind of-- draw it like that. This is, very much, a kind of structure that's not going to want to interact with water.

And then, another example would be Phenylalanine, or Phe, and that one is a methylene group and, then, a benzene ring. So most of you know, have some sense of the properties of benzene, a very, very organic solvent. So here you put a side chain like this, it's very much a residue that doesn't want to interact with water anymore than Benzene wants to interact with water.

And then there are three special cases. One of these is Glycine. In this case, it's just a hydrogen. One of the consequences of that is that since it's just a hydrogen, that's going to be a very, very flexible place, if we have a chain of amino acids and there's a Glycine there, it's going to be very little of way of constraints introduced, either by steric constraints or by interactions.

Another very special one is one called Cystine, Cys. And it's the same idea as

Cysteine, there's an ethylene. But instead of having an OH, it has an SH group. And that may not seem to be a great consequence, the sulfur is a little bit larger. But sulfurs have-- a sulfide group here has a sulfhydryl group here has a very special property, and that is, it can oxidatively dimerize with another sulfhydryl.

So if you have a side chain, and there's a cysteine that has an SH group, and another, either part of the same chain, or part of the different polypeptide chain that also has a cysteine, and they're in an oxidizing environment, and they're also close enough together to interact, they can form a bond like this, which is known as a disulfide bond, and it's the only one of the amino acids that's capable of reaching outside the chain in either hooking to a different part of the chain or to a completely different protein.

And in fact, proteins that tend to get excreted out into the media, either by bacteria or other things, often have a lot of disulfide bonds. Because when you link the peptide chains together like that, it tends to make a very tough protein that's hard to break down and can be very, very robust.

And there is one other special category of, one other special amino acid that's known as Proline. You have the alpha carbon atom, the carboxyl group, and then there's the amino group here. But this carbon is linked by a little ring with three methylenes to that amino acid. Again, this may seem sort of an unnecessary detail or something, but this is the way life evolved on earth.

This is an amino acid, but because of this ring structure, this bond is not able to rotate. So wherever a Proline shows up in the sequence, it puts some structural constraints on the conformational space that that chain is capable of getting itself into.

So when we study protein structure, this is at the heart of how proteins work, we'll spend quite a bit of time in the ensuing lectures talking about the central dogma and the idea that the linear order of the amino acids, in a protein, is determined by the sequence of the DNA and how that's encoded. But at the end, what you end up with is a linear sequence of amino acids, all joined together by peptide bonds.

And there's an incredible number of conformations possible. These things could go all over the place in all kinds of different ways. Yet, only one form, in general, is the biologically active conformation or maybe there's a couple of them and it switches back and forth as part of a machine action, or are part of what it does. But by and large, for every protein there'll be one, or just a couple of conformations. And so understanding proteins, what many, many people are interested in is trying to understand how you can get from that linear sequence and determine the three dimensional structure.

There are techniques, X-ray crystallography and NMR techniques now, which enable us to get the structures, solve the structures of proteins. In fact, there's the structures of tens of thousands of them are in a database called the Protein Database. And we're going to be talking about a little protein viewer that you'll be using that, in fact, once you've used it in your problem set, you could go open the structure of any protein whose structure has ever has been solved, if you want to do it.

But what we haven't yet figured out is a reliable way of saying, here is a protein that consists of a particular chain of amino acids. I'm going to predict its three dimensional shape. So we understand parts of it, but there's parts we don't know. And I'm going to take you through the first part of understanding protein structure.

And before we do that, I want to just talk about the levels of protein structure and the terms that are used to describe these. When people talk about the primary structure of a protein, what they're talking about is the sequence of amino acids, and it's possible I'll abbreviate those as AA, at some point without thinking about it. So just in case I do, that's a fairly commonly used abbreviation for amino acids.

So that simply, Phenylalanine joined to a Proline joined to a Glycine joined to two Cystines joined to something else, but that's not terribly useful in terms of telling what the protein does. Then there's secondary structure. These are regions of local folding and they're driven by, guess what? Hydrogen bonds. And we'll talk about how that goes in just a moment.

Then the term, tertiary structure, is the term used to describe the entirety of the folded protein. If I went in and determined the structure of a protein using x-ray crystallography, this is what I would see. It would be the tertiary structure. And there are other forces that we haven't discussed yet that contribute to that tertiary structure.

And then, a quaternary structure means that there's more than one polypeptide chain. And it could be as simple as an enzyme that's got two subunits and you've got to have them both there in order for it to work, or as I think you're beginning to probably get the sense from my use of the term protein machines, there are structures that have multiple interacting proteins and have complexities that rival some of the mechanical things that we build ourselves.

So the interesting story, a little, bit how the insights into secondary structure were first arrived. Some of you may have heard the term Linus Pauling. He was at Cal Tech, a Nobel Prize winner, very, very influential scientist, in a variety of ways.

The key insight that Linus Pauling had came in the late 1940s. People had been doing X-ray crystallography on minerals and things like that, and the basic idea was you had a crystal of some type, you bounced electrons off, you got a diffraction pattern. Then you could work backwards and figure out the structure that was generating the diffraction pattern.

And that had, then, been extended to proteins. And it was discovered there were certain proteins that would crystallize and you could bounce electrons off and get a diffraction pattern. And at least a category of these proteins, and analysis of the diffraction pattern suggested it was some kind of helix, and there was a repeating element of about 5.4 angstroms, roughly.

And so, Linus Pauling was very interested in protein structure. And I think it was in late 1948, he was visiting England and he caught the flu, just like some of you have been catching. And he spent a few days reading detective stories and then he got bored. And so he tried to take on this-- think about this problem. While he was lying

in bed.

And he made a simplifying assumption. He said let's just forget about all the side chains. Maybe they don't really matter in terms of this basic property. Maybe it's determined by the backbone of the peptide chain. So he took a strip of paper, started pleating it.

And he was a very good chemist, so he knew about this partial double [? blind ?] character of the peptide bond and the constraints that it put on the structures that the protein could take. And in doing this, he realized that if he folded the thing into a helix, kind of like this, into a right handed helix, that things worked out such that the carboxyl group in the backbone was just beautifully positioned to form hydrogen bond that was on one of the nitrogens.

He called this an alpha helix. There were 3.7 amino acids per turn. And the distance from here to here was 5.4 angstroms. And if we just-- sorry, I meant to put that up earlier, or did I go backwards? Anyway, there are all the amino acids and they're in your book.

Here is just the backbone of an alpha helix. And the orangey yellow colored bonds are the hydrogen bonds. And I hope you can see how the spiral goes. And you can also see, as it goes by, you can look right down the hole down the middle of the helix.

So let's put on some amino acids now. And again, you'll see, as it goes by, you can look right down the helix. But you see how the amino acids stick out onto the side. And if you look, for example, there is a Phenylalanine and a Tyrosine, they're aromatic groups that are very hydrophobic.

And up here there's a Lysine, so that's this side of the helix is charged. That's a glutamate. So there's a couple of charged amino acids on this side of the helix. Up here we've got a water hating part and somehow this is, I think, reminding me that I left something out. Let me just fix that up while I'm at it.

The other hydrophobic amino acids, I forgot to say those are Isoleucine, Valine,

Methionine, Tyrosine, and Tryptophan. Those are in your book. Those are other examples of hydrophobic amino acids. But I think, even in this little example of an alpha helix, you can see, depending on which amino acid was where, along that little region of alpha helix, it would very much influence what that part of the protein was capable of doing.

There's a second region of secondary structure that's very important. It's called a beta sheet. The one I'm showing you is an example of an anti-parallel beta sheet, although you can have parallel beta sheets as well.

But what I've done here is to take one strand of a polypeptide chain and I've written it out this way. And then I've taken a second-- what has happened? Oops. That's interesting. The stool just broke. OK. Fortunately, I noticed.

So what we have here is that the possibility for hydrogen bonding between this hydrogen of amino group and this oxygen, again, so we can get hydrogen bonds formed like this. And this makes what are called a beta sheet structure. And they can build up as well. This next one gives-- you can see how you can put one beta sheet on top of another.

And both of these are two major types of secondary structure and the way an alpha helix is represented is something like this. This would be an alpha helix. And a beta sheet is written as an arrow like that.

And so most proteins tend to have structures that consist of, for example, an alpha helix, some kind of turn, maybe a beta sheet, another turn, another beta sheet. Now maybe a turn, maybe an alpha helix going this way, some combination of regions of secondary structure. And I've got just a couple of examples of that.

Here you can see a domain of a protein with some beta sheets in purple, alpha helix in green. Where that's a piece of a protein coming from what's known as the bracket one gene. Some of you may be aware there's a familial susceptibility to breast cancer that was discovered. It's a complex protein. Part of it, and a very, very important part of it, is this piece known as the BRCT domain. It's the bracket one c

terminal domain, consists of beta sheets alpha helix.

Here's a protein I've already shown you the structure of, but maybe you recognize now, that green fluorescent protein is mostly beta sheets. It's the only beta sheets is going down here. There's a little bit of an alpha helix up there. And there's a bit of one over here?

Here's an example of a protein that's mostly alpha helix. What's this one do? This is a protein we'll talk about when we talk about DNA replication. It's involved in recognizing mismatches in DNA, for example, the G improperly got paired with the T during DNA replication. There's a system comes along and repairs those mismatches gives you another several thousandfold increase in fidelity, and if you mutate it in that kind of protein in a human, you have a familial susceptibility to colon cancer.

So it doesn't matter what their function is, when you get down to regions of secondary structure, you'll see these recurring things -- alpha helices, beta sheets. And if you understand their properties, you begin to understand some of the basic structure of forces that are giving the proteins their properties.

That's an enzyme called chymotrypsin. What it does, it's an enzyme that catalyzes the cleavage of peptide bonds in other proteins. But there it is. Got a lot of alpha helices, beta sheets, turns. You can go on and on.

I just said, one more up there. That's the Ras protein. That's an oncogene. Mutate that in a particular way, you have a susceptibility to cancer. But it doesn't matter, when you get down to the protein structure, most proteins have beta sheets, alpha helices.

OK. Go back to that one in a second. So we have to understand the rest of the structure of proteins. We have to be able to talk about the other forces that are important for making a protein.

And the third force is pretty simple. That's an ionic bond, and it's just this simple, that if you had a peptide chain that had, for example, an Aspartate with a negatively

charged amino acid on it, and we had, say, a Lysine, four Methylenes, and the NH₃ plus that was attached somewhere else on that polypeptide chain, then we can get an ionic bond, because of the attraction between the negative charge on here and a positive charge on that. So that is one of the things that then a force that can influence the structure of proteins.

The next one is a harder one to understand. It's known as the van der Waals interaction. And here's basically what's going on is that a non-polar bond can have a transient polarity. Sorry about this. And it can induce polarity in a nearby non-polar bond, and that can then give an attraction.

These things need to be very close together, about 0.2 to 0.4 nanometers apart, the two non-polar bonds, in order for this to happen. Does anybody remember the length of the covalent bond, the 0.15 to 0.2 nanometers, so within one or two covalent bonds. They have to be that close. Their strength is about one third, one quarter to one third, to that of the hydrogen bond.

And if you remember, the hydrogen bond was about one twentieth of the force of the strength of the hydrogen bond. But nevertheless, you can have a lot of them because, if you have an extended surface of a protein that's very close together, you can get a lot of these van der Waal interactions. And I'd always found this a somewhat esoteric kind of force.

But in fact, we're familiar with these because that's how a lizard manages to go up a surface. It uses van der Waals interactions. And as I'll show you in a minute, the trick is it's got little hairs on the bottom of its feet that have about a billion split ends and they're so tiny they're able to make van der Waals interactions with the surface. In a minute, I think there's a shot from underneath.

I got these movies from Robert Full at Berkeley, who's worked on these. You could see the lizard kind of peeling its foot off. And here they've made a little robot that can work by van der Waals forces and it will climb up the wall kind of like a lizard.

And here's what's going on at the molecular level. These are the toe pads on a

lizard like this. We're going to be just zooming in now. And you'll see they're covered with hairs, and you keep zooming in more, there's more hairs. And we keep zooming in more, get down to a single hair, there is a 30,000 fold magnification. There's 115,000 magnification. And in the end, a gecko, such as you've got here, has a billion 0.2 micron tips.

And just to compare it to a human hair, over on the edge, then you can see what the gecko hair is like. It's a very, very fine hair and it's able to use van der Waal interactions to stick to the surface.

Bob actually made a Band-Aid by collecting this little hairs out of the thing. And he made a little joke of putting it in a Band-Aid box. But this is interesting because it isn't affected by water. You can peel it off. You can put it back down. And he thinks there were commercial possibilities for using van der Waals interaction.

So, OK, I think we have one more force to go, but I think we will call it a day right here.