

HAZEL SIVE: All right. Let's move on to the second topic of our discussion today, which we will start today and then continue on Friday. And this is a discussion of the proteins. The proteins, probably the most fascinating class of macromolecules, 55% of the dry mass of a cell, and proteins function everywhere. They do almost everything.

Strictly speaking, proteins are not hereditary information, although Professor Jacks will talk with you about a class of proteins called prions, which kind of are hereditary information. So not hereditary info, but they do almost everything else. They form the structure of the cell. They form a major class of catalysts called enzymes that we will talk about on Friday. They function in defense as in the immune system. They allow cells to move, dot, dot, dot. Okay?

We'll talk about proteins on and on and on as major players in the fabric of life. Their monomer is an amino acid that is abbreviated AA or little aa. And it has a particular structure that forms around a central carbon, which is called the alpha carbon. And from this alpha carbon, there are four groups that emanate, something called R, which is the functional interesting group of each amino acid. So R is characteristic of each class of amino acid. We'll talk more about this in a moment.

And then we'll put in a hydrogen and then there is a nitrogen or an amine group, which I'm going to draw here as NH_3^+ . And the other group is a carboxyl group, which I'm going to draw here again as ionized. So R, actually, I'm going to give a special name to. We'll call it a side chain. And this side chain can be a charged group. It can be polar or non-polar.

And let me see what I have as the first thing. Yes. Here are some of the side groups in amino acids, for example, lysine. Here is the alpha carbon, and here's its side group. You can see it ends in an amino group. It's positively charged. Glutamic acid, I keep forgetting glutamic acid. This is a better screen. Glutamic acid ends. It's got a carboxyl group at the end. It's negatively charged. It's an acid.

Here's one, tyrosine with a polar uncharged group. This benzene ring and the

hydroxyl gives some polarity to the molecule and here's leucine with a hydroxyl gives polarity. Leucine is non-polar. It's really a hydrocarbon and cystine has this interesting very reactive sulfhydryl group at the end of its R chain, which is capable of reacting with another sulfhydryl group and forming a disulfide bond.

The polymer, actually I think I could write it here, the polymer of amino acids looks like this, and it's called a peptide if it's less than 20 amino acids and a protein if it's about more than 20 amino acids, but that's kind of loose. You should know the term peptide, polypeptide mean something a bit bigger than a peptide, protein, they're all the same chemical structure -- just refers to differences in the amount of sequence in the polymer.

How does the protein polymer form? I'll draw that for you now. And it forms by making a particular bond called a peptide bond that you should know and be able to recognize. So let's draw amino acid one with carbon. And we're going to put its amino group, its hydrogen, and then its carboxyl group. And we're going to add to it. So this is going to be amino acid one, and we're going to add to it another one with a second side chain.

And here, we'll draw out the hydrogens on the amino group. This oxygen and the hydrogens on the amino group are going to interact, and they are going to undergo a condensation reaction. And actually I realize when I drew for you the nucleotide condensation reaction, that was when we drew the dinucleotide that was forming, I realize I didn't put that there was elimination of a molecule as the dinucleotide was forming. It's a bit complicated, and that was why I didn't do it.

But in this case, this is a kind of classic condensation reaction, which will eliminate water, and you'll end up with something that looks like this. So here's your alpha. I'm going to circle the alpha carbon so we don't get where they are. And then we've got a carbonyl group that's joined to an amine and then the other alpha carbon, R2. So here is amino acid one, amino acid two. And here is a dipeptide, or diamino acid. It doesn't really matter what you call it.

The peptide bond is this guy, and it's a very, very important bond. It holds the

protein chain together, and you need to be able to recognize it. Two other features of this as well, which will be reminiscent of the nucleic acid case, and on one end -- the ends are different. Okay? The ends are different.

On one end, there is a free amino group, and this is called the amino end or the N-terminal of the dipeptide or of the protein. And on the other end, there is a free carboxyl group, and this is termed the carboxyl or the carboxy, or the C-terminal of this dipeptide, or indeed of any protein.

So this is going to sound reminiscent of the case as in nucleic acids because like nucleic acids, proteins, because they have different ends, have a linear order. So we'll write it again, different ends and a linear order. And again, it's this linear order that can only be read in one direction or that leads to information that is not symmetric or not randomly oriented that gives the protein its particular properties.

So let's write this out again, formally. On one end, there's the amino end. You can also write this as NH_2 . It's done kind of casually. You can write it as N. Doesn't really matter to us.

And then here's your polymer of amino acids. And here is your carboxy end. You can write COO. I always write COOH. You can also write C. All of those are kind of given as being equivalent and this is your amino end and your carboxy end.

That's terminology, but you can tell from where these free chemical groups are which amino acid was added first and which was added last. And the amino acid nearest the end terminal is added first. The one nearest the free carboxy is the most recently added, always added last.

Now, we went through a bit of a calculation for nucleic acids where there are four bases, and I pointed out to you that you could get a lot of combinatorics out of four bases if you had a polymer that was long enough. For proteins, the situation is actually dauntingly more complex. There are 20 different amino acids. So for a peptide, that's three amino acids, you would get 20 to the third combinations or 8,000.

Proteins can be greater than 1,000 amino acids in length. Proteins or peptides for little ones can range from two to greater than 1,000 amino acids. And so the number of combinations of information that you can get in proteins is enormous, and that means that most of protein space has not been explored by life and can be explored in the laboratory. And that's very exciting in thinking about the various functions that one can find that have not yet been found.

Now, the thing that is different about nucleic acids and proteins apart from their fundamental chemical structure, is that nucleic acids, as we will discuss, are used as a linear more or less, they are read as a linear string. You start somewhere, and you read along the nucleic acid, and you get your information out, and that gives you some kind of a next step.

Proteins, although they have a linear order, are not used in this kind of straight linear way. They are read once they have folded up into three-dimensional structures, and the folding of proteins into these 3D structures is intrinsic and essential for their function. So protein folding -- Actually, let me before we do protein folding, I realize I've been remiss in giving you these slides. Here is something I drew for you on amino acid polymerization from your book, and here we are where we should be.

Protein folding is required for function. The linear order of amino acids in the first peptide chain that is synthesized is called the primary structure. So primary structure refers to the linear order of amino acids. And these amino acids, of course, are held together by covalent bonds as in peptide bonds. But this primary structure, while absolutely essential for protein function, is not the functional protein. It folds up, this primary structure, this linear chain, folds up and folds again and folds again.

The way I think of it is, if you took a piece of string or an old-fashioned telephone cord, and you folded it up, you start rolling it up into a roll, it will roll up, and it'll fold back on itself, and fold back on itself, and fold back on itself. And that's kind of the deal with proteins. So we have to consider something called secondary structure

where the linear chain of amino acids folds back on itself.

So the linear amino acid chain folds up, and it does so in two ways. It forms something called an alpha helix or it forms something called a beta sheet or a beta pleated sheet, and these are characteristic folding structure or folding patterns that are governed by hydrogen bonds. But that secondary structure of the protein is not the functional protein. Those folded hydrogen-bonded alpha helices or beta sheets fold on themselves again to form the tertiary structure of the protein.

So this is more folding of the alpha helices or the beta sheets. And this tertiary structure can involve any number of different kinds of bonds. It can involve covalent bonds like sulfhydryl disulfide bonds, hydrogen bonds, hydrophobic bonds, and all of these things give a tertiary structure. All of them use one linear polypeptide or protein chain.

But then there's a fourth folding that is required for the function of many proteins. This is the quaternary structure, and this refers to, so again, this is of the primary and secondary chain. The quaternary structure is an association between two different proteins in a non-covalent way, usually. So this is association between two different protein chains and two different protein chains that are in some kind of tertiary structure. And usually this association is non-covalent.

The way protein structure is governed and proteins fold is mysterious, and very complicated, and not well understood. And there we go. Look at this interesting stuff you can see on my screen. Here we go. Here, not from your book because I think it's better from a different book, is a picture of an alpha helix, a beta sheet, tertiary structures of folded polypeptide chains. And here is a quaternary structure.

And as we'll talk about on Friday, it's the quaternary structures that are these functional units called enzymes, and we'll stop there.