**ROBERT WEINBERG:** Good morning, class. Nice to see you again. Hope you had a great weekend. If you didn't, it wasn't because of the weather. So here I am, once again a member of the walking wounded. And we're talking about carbohydrates today, as you may recall, or at least we were at the end of our discussion last time.

And we made the point that these multiple hydroxyl groups on the carbohydrates, on the one hand, determine the identity of various kinds of sugars, just the orientation, the three-dimensional orientation of them, for one thing, and for another that these multiple hydroxyl groups represented the opportunity for forming covalent bonds with other monosaccharides, as is indicated here in these disaccharides, or covalent bonds end to end to create large molecules, which will increasingly be the theme of our discussion today.

i.e. When I talk about large molecules, we just use the phrase generically of macromolecules, since, in principle, these end to end joinings of molecules, which involve the dehydration and the formation of these covalent bonds like right here, can create molecules that are hundreds indeed even thousands of subunits long.

So here, if we're talking about a polymer, we refer to each one of these subunits of the polymer as being a monomer and the aggregate as a whole is being a polymer. And here, we touched upon the fact toward the end of last lecture-- in fact, at the very end-- that one can cross-link these long linear chains of carbohydrates.

Here, we see the fact that glycogen, which is a form of glucose that is stored in our liver largely and to a small extent in the muscles, actually is cross-branched. If one draws on a much smaller scale glycogen molecule, one might draw a picture that looks like this, that it looks almost like a Christmas tree with multiple branches.

And the purpose of this is actually to sequester the glucose, to store the glucose in a metabolically inactive form, until the time comes that the organism needs, once again, the energy that is stored in the glucose, upon which occasion these bonds are rapidly broken down, and the glucose is mobilized and put into the circulation for eventual disposition and use in certain specific tissues.

While it's encumbered in these high molecular weight polymers, the glucose is essentially metabolically inactive. The body doesn't realize it's there. And we can, as a consequence, store large amounts of energy in these glycogen molecules. And it can be stored there indefinitely.

Now, the fact is that this idea end to end polymerization that I just mentioned can be extended to other macromolecules, which also become linked end to end in specific kinds of polymers. And here, we are moving now into the notion of talking about amino acids. And we're talking about proteins.

If we look at an amino acid, what we see is it has an important structure like this. Here's a central carbon and four, in principle, distinct side chains, where r represents some side chain that can be any one of-- as we'll see shortly, any one of 20 distinct identities. But all the amino acids share in common the property that they have this overall structure.

And as you may recall from our discussion of last week, at neutral pH, it wouldn't-- an amino acid of this sort, whatever [? r-ase, ?] wouldn't look like this at all because the amine group would attract an extra proton, causing it to become positively charged. And the carboxyl group would release a proton, causing it to become negatively charged.

And as you might deduce from this, at very low pH, due to the greatly increased concentration of protons, free protons in the solution, the equilibrium would be driven more in favor of reattaching a proton to the carboxyl group, just because there are so many of these protons around.

Conversely, at very high pH, where the hydroxyl ions are in predominance, they obviously tend to scavenge protons, reducing the level of protons to a very low level in the water. And under very high pH conditions, this proton would be released and pulled away by the hydroxyl ions, causing this amine group, once again, to return to its negative charge state.

Now, the fact of the matter is that these amino acids exist in a very specific three-dimensional configuration. And that's illustrated much more nicely here than I could possibly draw on the board, which in any case would be hopeless, which is that you can see in principle that once you have four distinct side groups coming off of a carbon, that there's, in principle, two different ways to create them.

And this is sometimes called chirality. Chiral you see is the form right here. The hands are chiral. If I try as much as I will to superimpose one hand on top of another, it doesn't work because they're mirror images of one another, which are asymmetrical. And as a consequence, we see a similar kind of relationship occurring here, where we see that these two forms of amino acids could, in principle, exist. And they're not interchangeable, unless one breaks one of the bonds and reforms it.

These two forms are called the L and the D. And it out that the L form is the one that's used by virtually all life forms on the planet, i.e. there was an arbitrary choice made some time about 3 billion years ago and more to use one of the two three-dimensional configurations and not to use the other. The other is found in certain rare exceptions. But virtually all life forms on this planet use the L form.

That said, by the way, this begins to indicate some of the arbitrary decisions that were made early during evolution because we could imagine, on another planet, if life were to exist there and if it were to depend on amino acids, that evolutionary system might have chosen the D form. So this is a luck of the draw. This is actually the way things evolved here.

And what we begin to see now is, if we talk about proteins or-- if we want to be more specific and use the more biochemical term, polypeptide-- we see, once again, we have an end to end n joining system, which is a bit different from that which the monosaccharides employ to create long chains of glycogen or of starch because here we see, once again, a dehydration reaction, where an amine group and a carboxyl group are caused to shed the hydroxyl and the proton, causing the formation of a peptide bond.

And here, we see this important, very important biochemical entity, a peptide bond consisting here of this carbonyl and this nitrogen fused in this specific way. And of course, if you recognize this as being a peptide bond, then you can understand why proteins are sometimes given the term polypeptide.

In some cases, if one has very short stretches of amino acids linked end to end like this, we talk about these being oligopeptides, where oligo- is the general term used in biology to refer to a small number of things rather than a large number of things. And once again, we have here the possibility of extending this infinitely. There are no constraints, in principle, on making this 500, a 1,000, even 2,000 amino acids long, where each of these, once again, is an amino acid.

And once again, I'm being very coy about the identities of R1 and R2, which, as I will indicate very shortly, can be one of 20 distinct alternatives. Here, you see that we're continuing this process of peptide bond formation. And most importantly here is the realization that there is a polarity of elongation here. It doesn't move with equal probability left or right or right to left.

We start at the amino end here. This is the amino end. And this is the carboxyl end-- the amino end and the carboxyl end. And invariably, again because of the way life has evolved on this planet, the new amino acid is added on the carboxyl end. And so when one often talks about proteins, one refers to their N-terminal and to their C-terminal ends, these referring, obviously, to the amino group at one end and the carboxyl end at the other end.

So the polarity is always a directed synthesis adding it on the C-terminal end. In other words, to use a shorthand notation, we think about protein synthesis as going with this polarity, end toward C. Things are growing at the C-terminal end progressively. And each time, one can imagine the addition of an amino acids on the end of it. So again, it can be extended in principle indefinitely.

Keep in mind as well something that's implicit in everything I'm telling you, but I won't always mention it explicitly. And that is virtually every biochemical reaction is reversible. And therefore, if one is able to form a peptide bond, one is able to break it down by biological means as well, i.e. by introducing a water molecule back in and thereby using the process of hydrolysis, which is the breakdown of a bond through the introduction of a water molecule, to destroy the previously created bond.

To use an MIT phrase, the reversibility is intuitively obvious because if you are able to make a-- well, I don't know if it's still used. But it was used in the late Stone Age around here. Anyhow, so any biochemical action must be reversible because if, for example, this polymerization were irreversible, then all the protein that was ever synthesized on the surface of the planet over the last 3.5 billion years would accumulate progressively. And obviously, that doesn't happen.

And therefore, macromolecular synthesis, to the extent it proceeds forward, obviously must go in the other direction as well. And the resulting concentration of a complete protein is known as its steady state. So we might make a protein at one rate and break it down at the same rate. And its steady state concentration represents the compromise between these two, i.e. the concentration of such a protein that we might observe at any one point in time.

Indeed, the term steady state could be expanded to any process in which there's a synthesis and there's a breakdown of something and the equilibrium concentration, which results is, once again, called the steady state of that molecule. Now, let's get down to the nitty gritty, which is obviously something which we can't avoid for very long, which is to say the R's, i.e. the side chains.

Once again, here, we see an arbitrary artifact of very early evolution in the biosphere because there are, in effect, 20 different side chains creating 20 distinct amino acids, which are used in proteins by all organisms on this planet. Again, there are rare exceptions. Certain fungi and certain bacteria are able to make unusual amino acids. But these are the basic building blocks of virtually all life forms on the planet.

99.99% of all the protein that is created is synthesized through the polymerization of these 20 amino acids. And by the way, one of the amino acids, glycine-- over here, you see it right here-- violates this rule of chirality. And you will recall before, I said that because there are four distinct amino acid-- four distinct side chains around the central carbon, sometimes called the alpha carbon, you always have a handedness of amino acids.

This cannot this notion cannot be respected in the case of glycine, seen up here, simply because we don't have four distinct-- here's the central carbon, where I'm pointing with the red. And here, these two hydrogens are equivalent to one another. They're not four distinct chains. There's only three distinct chains here. So glycine violates this rule of chirality, of left and right-handedness.

And here, by the way the side chain, which in all of these cases is depicted as extending off to the right of each amino acid, the side chain is simply an H, simply a proton, a hydrogen atom. In fact, what we see about these amino acids is that the side chains have quite distinct biochemical properties. And that begins to impress us with the notion that proteins and their biochemical attributes can be dictated by the identities of the amino acids that are used to construct them.

We can talk about the notion of nonpolar versus polar amino acids, i.e. amino acids which have poor affinity for water-- they don't have a separation of plus and minus charges. And as a consequence, they're a little bit, or quite a bit hydrophobic. Now, you will say, well, how can they be hydrophobic? Because here, this oxygen is charged and here this amine group is charged. That would make it highly hydrophilic.

But keep in mind, when I'm talking about these amino acids, I'm not talking about them when they're in a single amino acid form. I'm talking about their properties once they have been polymerized into a state like this. And once they're polymerized into a state like this, the NH2 and CO charging that is, the charge here and the charge here become irrelevant because this oxygen. And this amine group are both tied up in covalent bonds.

And this acquisition of a proton and this shedding of a proton over here cannot occur because both of these atoms, O and N, are involved in covalent bonds. So therefore, when we talk about non-polar and polar amino acids, keep in mind, we're focusing on the biochemical properties of the side chain because the central backbone of the polypeptide-- and the central backbone is defined quite clearly here.

Here's the central backbone. And you see it has a quite repeating structure-- N-C-C, N-C-C, N-C-C. This is invariant. What changes and what defines the biochemical attributes of this oligopeptide or a polypeptide are the identities of these side chains, which, again, are plotted on this particular graph-- and you have a different version in your book-- off to the right.

Here, you see we have a proton, a methyl group, a valine, a leucine, an isoleucine. And the differences between this suggest these are all quite aliphatic, quite similar to the propane that we talked about last time or the hexane. That is to say, these are quite hydrophobic side groups. And as such, if there were a polypeptide, we can imagine and you put the polypeptide in water, we can imagine that these amino acids would not like to be directly confronting the water because of the fact that they're hydrophobic.

Methionine is also a bit hydrophobic. I'm equivocating there because the S has a slight degree of hydrophilicity. It has a slight degree of polarity, but not really that much. And these aromatic side chains here, because they have these benzene rings-- and those consequently are called aromatic-- these are quite highly strongly hydrophobic. So they really hate to be in the intimate contact with water.

Here, on the other hand, let's look at these side chains because here we have strongly polar molecules, side chains again. Keep in mind, we're focusing on the side chains. Here, we see serine with a hydroxyl group that can form hydrogen bonds with the water, threonine, which has its own hydroxyl group, asparagine, which has two atoms here-- this carbonyl and the NH2-- both of which can form hydrogen bonds with the water, as can glutamine.

So these are quite hydrophilic. They're not as fanatically hydrophilic as these charged molecules, where the side chains are not just capable of forming hydrogen bonds. In this lower group here, the side chains are capable of undergoing ionization. So they're actually strongly charged.

And here, we see here the carboxyl group in aspartic and glutamic acid has actually discharged its proton, becoming negatively charged. These are acidic amino acids by virtue of the carboxyl groups they have. Basic amino acids here-- arginine, lysine, and histidine-- all acquire a positively charged side chain by virtue of these nitrogens here, which have a strong affinity for pulling away protons, for abstracting protons from the aqueous solvent.

And so we have a whole gradient of hydrophilicity down to hydrophobicity. And here, we have intermediate structures. We also have some very special idiosyncratic kinds of amino acids. Here is tyrosine. And tyrosine is a little bit schizophrenic. Again, it has this highly hydrophobic aromatic group here, the benzene ring, which hates to be in water, and the hydroxyl group, which actually is a friend of water.

So here, we have something where its role is quite equivocal. Here, we have cysteine, as indicated here. And what's interesting about the cysteine group, in this case, is the SH group, the side chain, the SH group, because this group is able to form bonds with yet other groups from other cysteines.

So let's just look at this cysteine here for a moment. You see there's a CH2, and then there's an SH. So let's imagine-- I'm not going to draw all the atoms here. But let's imagine here we have a CH2 group-- I'm not drawing the backbone-- SH over here. And we could imagine another protein chain, another polypeptide chain down here-- again, I'm not drawing the backbone.

But I'm drawing another SH-- like this. And the fact is, under the conditions of oxidation and reduction that operate, at least in the extracellular space, one can oxidize these two, resulting in the formation of what is known as a disulfide bond.

So here, we have now for the first idea time, the notion that polypeptide chains can be covalently linked to one another through these cross-links, as indicated here. Conversely, if you add a reducing agent, that will add protons back to this and reduce the oxidation state of the sulfurs, once again causing the disulfide bond to fall apart. On principle, these disulfide bonds could be used to link two proteins together.

But more often than not, if you look at the structure of a single protein protein, here's the structure of a single protein. And often, there are intramolecular bonds, disulfide bonds, i.e. bonds from one domain of the protein to another, from one part of the protein to the other. I'll draw them in right here. Here might be a disulfide bond. Here might be a disulfide bond. And I could go on and on. There might be another one over here.

Why do we have these disulfide bonds? Because, as we'll indicate very shortly, the three-dimensional structure of a protein is very specifically determined. A protein can only function when it assumes a certain three-dimensional configuration, when it assumes a certain three-dimensional stereochemical configuration.

When we talk about stereochemistry, we're talking about the three dimensional structures of molecules small and large. And here, we begin to touch on a theme of how these complex polypeptide chains are able to create proteins that have very specific, often very rigid structures in three-dimensional space. Part of this structural rigidity is maintained by these covalent disulfide bonds, which link, tightly link, neighboring regions, or even not so neighboring regions, of a single polypeptide chain, these intramolecular links.

This doesn't preclude there being intermolecular links between two polypeptide chains that are mediated as well by the disulfide bonds. Here's another very peculiar amino acid because what you see here is that the side chain, which is CH2-CH2-CH2, is hydrogen-bonded here to the amine group. It's not swinging out in free space.

I'm sorry. I misspoke. What we see here is CH2-CH2 is covalently bonded to the amine group. You picked that up, right? Whew, I was just testing you. Sure I was. OK. So here, we see a five-membered ring that's created. So here, this thing is not swinging out in free space. It creates a five-membered ring, where the end of the side chain is actually covalently linked to the amino group.

And that also has implications for the structure of proteins because this particular amino acid, whenever it occurs within a polypeptide chain, doesn't have the flexibility of assuming certain configurations that the other ones have whose four side chains are not so encumbered. None of them has total flexibility.

But this one is far more encumbered in the kinds of three dimensional structures that it can assume. And with that in mind, we begin to ask questions about how polypeptide chains assume three-dimensional structure. If we talk about a polypeptide chain, in our minds hopefully-- there's only 28 combinations?

**AUDIENCE:**     [INAUDIBLE].

**ROBERT WEINBERG:**     Oh, am I good or what? OK, anyhow. All right, so look here. And here, you see this is a typical polypeptide chain. Here, we have a three-letter code. In truth, there there's a single-letter code, which was introduced around 1965. So each of the 20 amino acids has its own single-letter code.

And to make a frank and depressing admission, 35 years, 40 years after the single amino acid letter code was instituted, I still haven't learned it. But we can learn these three letter codes, which fortunately are present here. In the single-letter code, L is leucine and A is art-- A is--

**AUDIENCE:**     Alanine.

**ROBERT WEINBERG:**     See, they know it. This is another example of not being able to teach old dogs new tricks. Anyhow, so here, we see the way-- one way by which one might depict an amino acid-- an amino acid chain, a polypeptide chain. And keep in mind this can go on indefinitely.

As we begin to wrestle with the three-dimensional structure of the chain, we begin to realize the following. And that is that after the chain is initially synthesized, it's initially chaotic. And as it extends, it increasingly begins to assume a very specific molecular, three-dimensional molecular configuration, which is indicated down here. So the chaos that operates initially will eventually result in a native configuration over here, which, in many respects, often represents the lowest free energy state.

Since, for the last 40 years, people have been able to have been trying to figure out-- if you knew the amino acid sequence of this primary polypeptide here, if you knew its primary structure-- and when I say primary structure, what I mean is the subsequence of the amino acids. So if you knew the primary structure of the amino acids, you should, in principle, be able to develop a computer algorithm that would predict the three-dimensional configuration, which is shown here in a very schematic way, and which we'll discuss in much greater detail shortly.

And the fact is, after 40 years of trying, one still is unable to do that, i.e. if I were to give the primary amino acid sequence of a polypeptide to the smartest biochemist in the world-- and there are some very smart ones-- he or she could still not tell me what the three-dimensional structure of this protein with total certainly would be. Why? Because there's an almost infinite number of intramolecular interactions that greatly complicate how the protein assumes this structure.

Moreover, if we talk of this as the native state of the protein, we can imagine that there's ways of disrupting that because much of this native state is created by intramolecular hydrogen bonds. Remember, the hydrogen bonds are relatively weak. And if we heat up the temperature, then we can break hydrogen bonds. And therefore, every time we fry an egg, for example, if we want to get down to Earth, we denature.

We break up the three-dimensional, the native three-dimensional structure of the albumin molecules that constitute the egg white. And so when everything turns white, what we've done is to take a native molecule like this, heated it up to temperatures where the intramolecular bonds no longer stabilize-- notably, hydrogen bonds no longer stabilize this three dimensional configuration. And we put it into a denatured state, which might be all the way up here.

And therefore, this acquisition of a native configuration or a native state, native representing the natural state, is also reversible in many molecules simply by heating them up. There are, to be sure, yet other molecules which are different from egg white, from the albumin in egg white, where if you cool them back down, they'll spontaneously reassume their native structure. Many proteins, most, will not do so.

Again, let's go back to this issue of the acquisition of complex three-dimensional structures. And here, we begin to see how some of this structure is acquired and stabilized through these intramolecular hydrogen bonds. And there are many opportunities for these intramolecular hydrogen bonds because here we see one polypeptide chain, here we see another.

And we see that the NH2 group right here-- I'm sorry, the nitrogen group here with the proton side chain and the carbonyl group here with the oxygen are not encumbered. They are, in principle, available to form hydrogen bonds with a polypeptide chain somewhere else.

Now, this other polypeptide chain could, once again, be from another protein, from another polypeptide. But more often than not, we're, once again, dealing with intramolecular cross-links. But in this case, the intramolecular cross-links are not disulfide bonds, which are covalent and hard, stable as a rock in the absence of reducing agents. Here, we're talking about much weaker bonds, hydrogen bonds, which also act between different loops of the protein and serve, once again, to stabilize the three-dimensional structure, the native state of the protein.

And you can see how these opportunities for forming multiple hydrogen bonds can create an enormous degree of stability. And here are some examples of what we now call the secondary structure of the protein. Just a second ago, or several minutes ago to be honest-- and I'm always honest with you, class-- the primary structure is the amino acid sequence. The secondary structure represents configurations like this.

Here is an alpha helix. And here is a beta pleated sheet. And what we see here in this alpha helix is we have a helical structure where the amine group down here, the hydrogen bonds with a residue, which is, I think, 3 and 1/2 residues upstream-- 1, 2-- there's an amine down there-- so with a carbonyl group that's 3 and 1/2 residues upstream of it.

This one, once again, reaches 3 and 1/2 residues upstream. Not all the hydrogen bonds are shown in the background. Only the ones on our side of the helix are shown, on the front side of the helix. But you can imagine that this can perpetuate itself. And each of these carbonyls may associate with a proton from an NH group that's either above or below that particular residue. And this, in turn, can create a helical structure.

By the way, proline doesn't fit well in this. If you add a proline in here, proline is known in the trade as a helix breaker. Why? Because it cannot twist itself around to form an alpha helix. And so if the primary amino acid were to dictate that a proline would be inserted right here, for example, then this helix might exist down below and up above. But it would not be continuous because the presence of a proline is highly disruptive of the formation of an alpha helix.

This means that, in principle, you can make some predictions about the localized structure of a polypeptide by knowing whether or not proline is present, for example. But that still doesn't give you the power to predict the entire three dimensional structure of the finished protein itself.

Now, let's agree that this is the secondary structure of the protein, i.e. the various domains which often form alpha helices within a certain segment of the protein. Or a certain segment of the protein will form beta pleated sheets. And there are several or several other less common kinds of secondary structure.

And here, we deal with tertiary structure. Now, we're getting really interesting. Or maybe you don't like it. But some people say it's really interesting because here are the tertiary structures of some arbitrarily chosen proteins. Here, the tertiary structure of this particular protein-- and the identities of these are not given in our textbook. And I'm sure if we spent two or three weeks, we could find out what they were.

But anyhow, here is a protein, a three-dimensional structure of a protein, which is composed of four alpha helices, which go up, another alpha helix, alpha helix, alpha helix, alpha helix. They're depicted here, fortunately, in four different colors. And so we see that when we talk about tertiary structure, we're talking about how the alpha helices are disposed with respect to one another.

The primary structure is the amino acid sequence is not shown here. The secondary structure represents these individual alpha helices. And the tertiary structure represents how these alpha helices are arranged vis a vis one another. Here is a protein which is structured much differently. Here, this is formed of many beta pleated sheets. We saw them in the last figure, in the last overhead. You see it has a quite different overall three dimensional structure.

This could be the beginning of an alpha helix down here, although that's quite equivocal. And here, we see yet another point. And that is, as we said before, the tertiary structure independent of these alpha and beta-- alpha helices and beta pleated sheets may be stabilized by these covalent interstrand cross-linnks formed by the cysteines.

And in the end, if we put all that together, then we come to the realization that the three dimensional structure of a protein, as determined by the art of X-ray-- there we go. I'm not actually dyslexic. I actually have a cousin who I won't mention whose son was so dyslexic that when he came to stairways, he didn't know whether to put his foot up or down. Now, that's difficulty. This is-- this is not so bad-- OK, anyhow-- because I solved it within less than two minutes time.

All right. So here, we see this is what the three-dimensional structure of a protein looks like. This is called a space-filling model because here one draws in, as determined by X-ray crystallography, if we could see what a protein looks like, what it actually must look like, where each of the atoms, including these side chains, is actually depicted.

Before, when we used these far more schematic descriptions like here, we were just talking about the overall structure of the backbone. We weren't really indicating where the side chains were and what space they would fill up. And if we give them the chance, if we put in all of the other atoms, the side chains, and we create a space-filling model where the actual atoms are shown, this is what the protein would look like.

And the fact of the matter is that virtually all proteins have very specific structures. It's not as if they can shift from one structure to another. Once they leave their normal native structure, they will lose their ability to do what their normal jobs are. And this particular overhead happens to bring in yet another theme that we're going to focus on increasingly, which is, what do proteins do in cells? I'm glad I asked that question. One of the things they do is they act as catalysts, i.e. as enzymes.

The fact is, as we'll discuss later, virtually all biochemical reactions require an enzyme catalyst in order to propel them forward. That is to say, if there's a biochemical reaction to occur, almost always, it will not occur spontaneously, the same way that a hydroxyl ion and a hydrogen will join together spontaneously in water.

Almost all biochemical reactions require the mediation of an enzyme, which is a biological catalyst, in order to encourage this to happen. And almost all catalysts in our cells are proteins. So if you have 4,326 distinct biochemical reactions occurring in the cell, that means that there's probably almost as many distinct enzymes, each one of which is assigned to mediate one or another of those distinct biochemical reactions.

And here, we see the fact-- this is an enzyme which happens to be called hexokinase. Recall that the -ase suffix at the end dictates that this is already an enzyme rather than a carbohydrate. And this attaches, in fact, a phosphate group onto glucose. And what happens is that the glucose, which is the substrate which is acted upon by the catalyst, is pulled into this site in the protein, which is highly specialized to mediate the enzymatic reaction.

Almost all of the business of this complex enzyme is carried out right here. And somehow, a lot of the other amino acids that are located at a distance are doing other things, like regulating the activity of the enzyme. But the actual business end of the enzyme is present in what is called a catalytic cleft, an active site of this enzyme, in which the substrates are pulled in and are manipulated and changed chemically by the actions of this particular enzyme.

Now, in saying that virtually all catalysts, but not all, are proteins, I also mean to say that proteins have a second major function in the body. The first major function is to act as enzymes and catalysts. The second major function is to create biochemical structures, i.e. structures of different cytoskeleton proteins, such as I showed you two lectures ago.

And so we're going to come repeatedly to situations where complex structural entities in the cell are composed of different structural proteins. Again, this is just a prelude to talking about these in greater detail. These two major functions-- enzymatic catalysis on the one hand and creating structure on the other.

And so now, we get to really four hierarchical levels of protein structure. The primary structure is the amino acid sequence. And if we dwell for a second on this amino acid sequence, let's realize that any single amino acid can follow any other amino acid. So what that means is that if glycine is the first amino acid, as it happens to be here, serine is only one of 20 different possible second amino acids.

Aspartic acid is only one of 20 different third amino acids, as the third residue. We often call these different residues-- the first residue, the second residue, third residue, fourth residue, and forth-- and so forth. And keep in mind, if we think about the combinatorial implications of that, the first amino acid residue can have 20 different ones. The second can have 20 different identities. The third can have 20 different identities.

That means, if we make a tripeptide-- a tripeptide has three amino acids in it. That means we can make 400 dipeptides, 400 distinct dipeptides. And we can make 8,000 distinct tripeptides. Now, if you imagine that the average amino acid-- the average protein in a cell is, let's say, 150 amino acid residues long. That means that, in principle, one could make 20 to the 150th power distinct amino acid sequences because of these absence of any constraints of which amino acid will follow which other amino acids.

In other words, if the average polypeptide has this many residues, this is the number of distinct 150 amino acid residue-long proteins that one could, in principle, synthesize. I'm not saying all of them have ever been synthesized since the formation of life on this planet. Indeed, since some amino acid chains are 400, 500, 600, even 2,000 amino acid residues long-- I think the one that is affected in muscular dystrophy is more than 2,000 amino acids, dystrophin. Does anybody know here? It's big.

Anyhow, imagine the number of possibilities. So combinatorially, life can make almost whatever types of amino acids it would like by dictating the sequence of amino acids. Now, let's just go and look here again. There's a secondary structure. The tertiary structure is the way in which the different alpha helices here or beta pleated sheets are disposed three dimensionally with respect to one another.

And the quaternary structure represents how different single polypeptides are associated one with the other. So for example, hemoglobin is a tetramer. Hemoglobin doesn't exist as a monomeric protein in solution. It exists as a tetramer. And there's two kinds of globin chains. There's an alpha kind and a beta kind.

And if we look in a very rough and schematic way at the way that a hemoglobin tetramer is arranged, there are two alpha polypeptide chains and two beta polypeptide chains. They're not covalently attached to one another. They're associated with one another via hydrogen bonds and hydrophobic interactions.

And this is the actual native configuration of globin, two alpha and two beta chains. It doesn't exist as a single amino acid in solution. It exists as a tetramer. And indeed, most-- or I shouldn't say most-- but very many proteins exist in these configurations, where the tertiary structure represents four different amino acid chains.

Each of these has an N and C-terminal. Each of these is chemically distinct. These four could probably be taken apart from one another simply by raising the temperature. And they associate like this. And in the absence of this association, if you just had one of these alphas and one of these betas, it wouldn't function well at all. In fact, it might be totally dysfunctional.

One other thing that may be implicit to you, but I haven't said, but that is very important to realize, is the following. Let's imagine that this is the three-dimensional structure of a protein, as it may well be. Let's now think about hydrophobic and hydrophilic amino acids.

The hydrophobic amino acids hate to be present in water. And therefore, they are, we can imagine, in this case correctly, tucked away inside the interstices of the protein far away from the surface. They don't have any contact with water. Conversely, the highly charged hydrophilic amino acids are actually sticking out at the surface.

And this begins to yield yet another insight into how the three-dimensional stereochemistry of proteins is maintained and dictated because the hydrophobic amino acids, through hydrophobic interactions, stabilize the inner core of the protein that is well-shielded from the aqueous solvent. The hydrophilic amino acids are on the outside. They like to be in intimate contact with the water.

So we already now have talked about a number of distinct different interactions that are responsible for creating the three-dimensional stereochemistry of the protein. First of all, there are the disulfide bonds, which create chain-to-chain covalent interactions. There are the hydrogen bonds in which different side chains can-- different chains can interact with one another. And there are these hydrophobic and hydrophilic interactions. And there are some relatively inconsequential van der Waals interactions, which are really not worth discussing, although some people get really excited about them. But we won't.

So here, we now begin to see that we have really interesting polypeptides, unlike the boring polypeptides that are ultimately the way one must judge carbohydrates. Some people think carbohydrate chemistry is really interesting. But entre nous, it really isn't that interesting because you just have the same monomer in 100 or 500 stretches.

Here, a protein is much more interesting because of this enormous variability in amino acid sequence and the consequent ability to create all kinds of chemical reactivities and structures because of these 20 different amino acids.

If we were to imagine life on another planet and we imagine that there were, let's say, amino acid-like molecules that were part of life, maybe that other life wouldn't have exactly the same 20 amino acids as we do. It almost certainly would be a water-based, the way we are. But it would also rely on hydrogen bonds, and hydrophilicity, and hydrophobicity interactions in order to dictate the three-dimensional structure.

In the absence of this very specific three-dimensional structure, I will tell you that this enzyme could not function. And if you were to take this enzyme, if it were a typical enzyme, and you were to heat it up briefly, even often slightly above normal body temperature, it might denature, i.e. it might lose its three dimensional structure irreversibly.

And once it were denatured-- once it was denatured, this process of denaturation, it might not be able to spontaneously assume that pre-existing three dimensional configuration and therefore would forever be inactive. That means to say, very explicitly, that even though the amino acids that are creating that active catalytic site remain there, their highly specific three-dimensional disposition is critical for the continued actions of this enzyme. And once their three-dimensional dispositions are shifted around through the process of interactions, then we have trouble because the enzyme can no longer do its assigned task.

We're going to go now to an even higher order of complexity, in one sense. We're going to go to the royalty of the macromolecules, which are the nucleic acids. Of course, protein chemists would take great umbrage at the very notion that there are things better than proteins.

But the fact of the matter is, entre nous, there are-- I can't show you that overhead because it's from the other textbook, which is copyright. And we're being filmed. How many people have had the backs of their heads immortalized on these videos? Did you call home and ask anybody to identify you?

**AUDIENCE:**     [INAUDIBLE]

**ROBERT WEINBERG:**     I don't know. But soon, you'll be-- each of us has the limelight for 15 minutes in a lifetime, right? So you'll have your 15 minutes. OK. Here are some nucleic acids. And let's look at these nucleic acids and the way they're put together. Keep in mind, to anticipate what we're going to say next time, once again, we want to make end to end aggregates. We want to polymerize molecules.

And in this case, we want to do so, once again, through a dehydration reaction. And moreover, just to look at the building blocks of nucleic acids, we start again, in this case, with two pentoses. Recall that they have four carbons-- 1, 2, 3, 4. No, did I say four? You know I meant five-- 1, 2, 3, 4, 5. So whenever I say four from now, on I mean--

**AUDIENCE:**     Five.

**ROBERT WEINBERG:** Or whenever I say four, I may also mean four. OK. 1, 2, 3, 4, 5. And let's look at the two basic kinds of pentose molecules that are present in nucleic acid because they define the essential difference between DNA and RNA.

Here's a regular old rather familiar kind of pentose with five carbons. And here's an unfamiliar kind of pentose, which we call deoxyribose. Why? Because if you look really carefully, you'll see that the hydroxyl group here, which should be present in any self-respecting pentose, is missing and is replaced simply by a hydrogen group, i.e. it's lost its oxygen, whence cometh the word deoxyribose, and ultimately clearly the word deoxyribose nucleic acid.

And one of the attributes, one of the virtues of carbohydrates, as we discussed last time, were these numerous hydroxyl groups, which represent opportunities for all kinds of dehydration reactions, which can enable one to build much more complex molecules. And here, we see the structure of, for example, a deoxyribonucleotide whose detailed structure we'll get into next time.

But just let's look at how these hydroxyl groups have been used. The hydroxyl group, in this case in DNA-- and by the way, notice that the structure I've shown here-- there's a side chain attached here and a side chain attached here. And neither of those depends on whether or not there's a hydrogen or a hydroxyl right here.

And look what's happened here. Here, we have this hydroxyl, over here, to which a base has been attached covalently, once again by a dehydration reaction. And here, we have a situation where actually three phosphate groups have been attached to the hydroxyl group in this direction. And this represents the basic building blocks of nucleic acids. Now, one of the things that's going to be really important and that you're going to have to memorize-- I told you weren't going to have to memorize anything. But you didn't believe me, did you?

**AUDIENCE:** No.

**ROBERT WEINBERG:** Good. OK. One of the things you're going to have to memorize is the numbering system here. This is number 1, 2, 3, 4, 5. Or to be totally frank-- and you know I'm always that-- 1-prime, 2-prime, 3-prime, 4-prime, 5-prime. And that numbering system, it turns out, is going to be very important for our subsequent discussions.

Notice here that, for example, it's here at the 2-prime position that this deoxyribose is lacking the oxygen that is present normally in RNA. And with all this in mind, we will wait in great suspense until Wednesday when we actually talk about how this is exploited to make highly complex polymers. Have a good two days. See you Wednesday at 10:00.