Lecture 21A • 06/06/12

Determining the order of amino acids in a polypeptide chain

The sequence of a polypeptide itself is expressed from the amino end or terminal or terminus to the acid end. Let's say we had something like lysine-phenylalanine-cysteine. How could we determine this order? There's a reagent we could use to do this. To determine the types of amino acids we have, we could completely decompose this chain – if this polypeptide is nothing more than a series of amide linkages, undo the amide linkages, which we could by hydrolysis, subjecting this to a strong acid, concentrated solution for a long period of time at elevated temperatures – which means we lose any order the amino acids had been in. That wouldn't give you sequencing, that would just allow you to know the proportion of amino acids in a polypeptide chain. What I want to show you here is how we really could, one amino acid at a time, determine the structure of a polypeptide chain. The reagent we're going to look at is phenylisothiocyanate, which is used in the Edman degradation. [The mechanism is] a series of tautomerizations and ring opening and closings. The Edman degradation uses this reagent, phenylisothiocyanate (PITC). It has this structure [similar to the isocyanates seen in the Curtius and Hofman rearrangements] When we react this with something like water, we're going to end up with a carbamic acid. [Carbamic acids] decompose; that's why at this point one carbon is lost. Remember that whatever compound you start with effectively loses a carbon in the process. Isothiocyanates means substitute the oxygen with a sulfur.

If you look at this, this is kind of like the sulfur equivalent of a ketone fused simultaneously to an imine, which is why we end up with this carboxylic acid-style derivative. What I'm about to react with this is going to attack the carbon in the middle of the molecule. What amine I going to react with this? Let's take a generalized amino acid chain, a polypeptide. I'm going to not include stereochemistry, because, as we will see, it won't matter. I'll show two amino acid units, and I'll simplify the rest of this structure by just saying that the rest of this chain is just beyond this last amide linkage. At the end of the chain, of course, is a carboxylic acid – but, is a carboxylic acid likely to be the thing that attacks this phenylisothiocyanate – other than maybe providing a proton source? No. So what of this molecule is more likely to react with the phenylisothiocyanate? Would this compound more likely be a Lewis acid or a Lewis base? What is a Lewis acid? Something that accepts a lone pair. What's a Lewis base, therefore? Something that donates a lone pair. Sulfur could donate a lone pair to something, it could [be] alkylate[d], for example, if we had an alkyl halide. Do we have anything that would want to be attacked on this polypeptide chain? If we called the phenylisothiocyanate a Lewis base, where would the lone pair attack the polypeptide? What about in reverse? What argument could you make that phenylisothiocyanate is a Lewis acid, which means it's going to want a lone pair? Isn't there a fairly good-sized delta positive right there on that carbon? We have a base – remember that amines are bases. The first step of the reaction here is that the amine attacks that carbon, exactly like water over here attacks the isocyanate. What's produced is this. Here is the phenylisothiocyanate, here's where it shows up in the product. Here is a thiourea. Urea is a carbonyl with nitrogens on either side of it; the compound urea itself would be a carbonyl with just amino groups on either side. This is a thiourea because it's got a sulfur there instead, and it's actually a thiourea derivative since we have substituents attached to the nitrogens.

We don't stop here. Here's where the first tautomerization occurs – the nitrogen not next to the benzene ring could, in theory, make a carbon-nitrogen double bond if you were able to push the sulfur-oxygen double bond open. Why might that happen? Because with this reagent, under acidic conditions, maybe somehow the amide over here could get opened up. This is not a set of mechanism arrows; this is just a set of transformations that are going to occur. That lone pair pushes open the carbon-sulfur double bond, which can attack the carbonyl over here, making a five-membered ring – pushes the carbonyl open. A proton comes from somewhere, and we end up with this next intermediate. Look at what we have here: this that I just circled would be somewhere between an amide and a carboxylic acid. Visualize, for a moment, a carboxylic acid. Imagine that you did one cycle of protonate-open-attack-deprotonate, somehow adding an amine in. In real life, amines and carboxylic acids neutralize each other, but just pretend. If you had a carboxylic acid, it has carbonyl and an -OH group. Push open the carbonyl, now you've to oxygen and oxygen; something else attacked it, that would be a nitrogen. Now, switch to sulfur, cause it acts a lot like oxygen, and that's exactly the situation that we have here. In other words, what we would want to happen is a carbonyl would want to form one way or another, because having these three heteroatoms on that carbon makes it a particularly reactive carbon. That's the justification for what happens next, which is that, effectively, this ring closes.

Here's the big important point: at this point in the mechanism, the remainder of the polypeptide chain is ejected. That means that, whatever next I'm going to show, has the R group from that first amino acid that's now been separated from the rest of the chain. That means that we are sequencing, because we're able to pull just the first amino acid away. We're not quite done, because we need to get to the point that we have what we can isolate and analyze from this. Next stopping-off point is this structure, which is a five-membered ring that contains an imine, a carbonyl, and in the ring itself a sulfur. This is known as a thiazolinone. The azol part of the name means five-membered ring with a nitrogen, the thia means sulfure, one means ketone, so thiazolinone means five-membered ring with a nitrogen and a sulfur and a carbonyl in it. And, it has this R group. We're not done, because this is going to open again, the ring, and it's going to close. When it opens, we're going to make effectively a carboxylic acid, so it's like hydrolyzing this thioester. Let's just see what that looks like, then I'll show you what's going to happen next. As part of the opening of this, we're going to have water come in an hydrolyze this.

The carbon-sulfur bond is going to come back over and push this open, so, effectively, in other words, we're doing another tautomerization. The first tautomerization is a response of ring closure; the second tautomerization is in response to ring-opening. Notice we're back to a thiourea when we do this. Now, the last bit: we make an amide, where the nitrogen that is closer to the benzene ring finally gets involved and makes a five-membered ring by combining with that carbonyl. This is known as a phenylthiohydantoin. This can now be isolated. Each one of these is going to be different, depending on the R group, which is going to depend on the amino acid removed. You identify which amino acid had been here, you get the first amino acid in the polypeptide chain. You do the reaction again, pull the next amino group off, do it again and again and again, identify each one of the amino groups, and you have now sequenced the polypeptide chain. [guessing that] Having a thiourea here and a thioimide is thermodynamically more stable than having a thioester with this functionalized imine. The thiazolinone is the super-complicated equivalent of an enol when looking at the hydantoin.

Phenylisothiocyanate (PITC) can be used to pull one amino acid residue off of a polypeptide chain – one and only one – from the amino terminus of that polypeptide chain. The derivative formed, the PTH, the phenylthiohydantoin, can be identified ... the identity of that amino acid can be determined from the PTH derivative that results. The process can then be repeated to fully sequence a polypeptide chain – at least in theory, because if you have a reaction that's 99% successful, then 99% to the 25th power is only 77.8%, so if you had something that was 99% successful and effective and didn't produce byproducts, apply it 25 times, you're down to the point where, really, overall, it's only 78% successful, which means you keep doing it and doing it and you can no longer be confident enough that what you're sequencing is correct. Over time, the number of times this can be used for has increased, but there are also new techniques of sequencing that have been invented. There's a couple of specific amino acids that can be targeted by other agents.

Synthesis of amino acids

I'll start with one that's a variation of the malonic ester synthesis. This uses diethyl alpha-bromomalonate as the starting material. Can this compound undergo Sn2 reactions? What kind of alkyl halide is this? Secondary. Can secondary alkyl halides undergo Sn2? Yes. [ACS exam] Can this undergo E2? Why not? No hydrogen at the right position; you have two carbonyls there, you need a hydrogen one position away from where the leaving group is in order to do elimination. [Carbenes] What would happen if I took that and I took phthalimide; that's the compound used in the Gabriel amino acid synthesis. What if we took the anion of it, neutralized it with potassium hydroxide? Is that a particularly acidic proton that would need to be removed? Is this phthalimide salt a really strong base? Compared to hydroxide, is it strong? No, because that nitrogen's in resonance with two carbonyls [what is pKa]. The whole point of all of this is to explain that yes, this is a perfectly reasonable reaction – that the phthalimide salt will do an Sn2 reaction on the alpha-bromomalonate – which still has a hydrogen at the alpha position. If we were to make a logical choice of bases – for example, if we were to use sodium ethoxide to match the fact that we have an ethyl ester, and if we introduced [afterwards] an alkyl halide, we could then end up with this, which looks like a combination of things: it looks kinda like the Gabriel amine synthesis, and it looks kinda like the carboxylic acid synthesis from the malonic ester synthesis. That's because this is an amino acid, so we're doing something that would give us the amino group at the same time we're doing something that's going to give us that acid group. At this point, we an hydrolyze, and it's going to hydrolyze a whole bunch of stuff. It'll hydrolyze the phthalimide and release this amino group; that's why we get that alpha-amino. It's alpha-amino because we stuck it on in the first place by combining it to this malonic ester. When we hydrolyze the malonic ester, we're going to get a beta-carbonyl-containing acid. That decarboxylates, so what we're left with is our customized amino acid – the only drawback being the fact that we're going to end up with a mix of enantiomers. R has to be primary.

_____

Sequencing –> Determining the order of amino acids in a polypeptide

Edman degradation

*At this point in the mechanism, the remainder of the polypeptide chain is ejected.

PITC is able to cleave only the first amino acid form the amino terminus of a polypeptide. The identity of the amino acid can be determined from the PTH derivative that results. The process can be repeated to determine the sequence of a peptide chain.

Synthesis of amino acids

_____

Structures *(remaining structures identical to lecture 21B)*

06/06/12 lec • 1