Lab 12A • 06/04/12

How can we assemble amino acids together to make what are called peptide chains – polymers that contain multiple units of amino acids linked together.

We're dealing with alpha-amino acids – the simplest one of which we can get by taking acetic acid and simply substituting an amino group at that alpha position. That gives us the molecule glycine. Since there are two hydrogens also at that alpha position, there is no stereocenter at this molecule. It's the one amino acid that is not stereogenic. If we put just one alkyl group on there, that is the amino acid alanine.

Let's draw this in Fischer projection form. The convention is to put the carbonyl group at the top of the Fischer projection, which means as far as where my eyebrow is, it's going to be on the righthand side of the paper, so I can look at that carbonyl. I want to make the Fischer projection in such a way that the dash and the wedge would end up being my side arms – if so, they need to be pointed out towards me, so my eyebrow's going to be on the right, and I'm also going to be looking up at the molecule. If I do so, we end up with the carboxylic acid up top, the amino group down at the bottom. Which side would the methyl group be on? The right, because if I'm looking with my eyebrow to the righthand side looking up at the carbonyl, so that the dash and wedge are pointed down at me. The dash means it's in the back of the plane, which, from my perspective, is on my righthand side; that means yes, the methyl group is going to be on the right.

That's correct, but it's not the standard way to do one of these Fischer projections. The standard way is to have the R group at the bottoms, so that the amino group is to the left or to the right. We've done some practice already in twisting this stereocenter around. I'll rewrite it, keeping the carboxylic acid at the top, putting the methyl group down below, which now makes that amino group on the left – which is what makes this l-alanine. The convention for amino acid Fischer projections is that the acid goes to the top, the R side chain goes at the bottom. When that amino group is on the left, that means it is an l-acid.

Having drawn this Fischer projection, let's back out of the Fischer projection, back to a line structure. If I do it the same way that I've been doing carboxylic acids – putting the carbonyl on the right – then it means the methyl group now is going to be in the plane of the paper. Looking with my eyebrow in the correct position, since on the Fischer projection the amino group is now on my left, that's pointed out towards you, which means it's going to be written with a wedge.

Let me show you a series of structures, and you'll we why I'm writing these in just a moment. These are eight different representations of l-alanine. Notice the top four keep the amino group in the plane of the paper; the bottom four keep the R group in the plane of the paper. The top four represent the carbonyl up to the right, down to the right, up to the left, and down to the left; similarly, the bottom four are up/right, down/right, up/left, bottom/left [joy of grading these types of problems]. You need to know that you might use more than one of these forms depending on what kind of structure you're trying to write.

An alpha-amino acid is substituting at that alpha position. The simplest substituent is to have a hydrogen there, which gives us glycine, which is not stereogenic. Then we have all of the other amino acids that have the possibility of being in its l- or d- form. I've shown you l-alanine; let me just, for comparison's sake, show you one d-alanine, to show you it is the enantiomer. That's an individual amino acid.

You've seen the eight classes of amino acids. The first class is the alkyl group [valine (with a V); leucine (one more carbon between V and attach point)]. There are these abbreviations for amino acids: alanine happens to be abbreviated 'ala'; valine, 'val'; leucine, 'leu' – they're [normally] logical. The one-letter abbreviations happen to be logical in this case as well – 'A' for alanine; 'V' for valine; 'L' for leucine. Glycine, it's 'gly', or 'G'. Unfortunately, there's some amino acids that have the same starting letter, so some amino acids don't just have the one-letter abbreviation you might expect. If a protein contains a few thousand amino acids – let's just say for argument's sake – it's a lot less space to represent them by one-letter abbreviations than writing out the sequence of names over and over and over again; that's why we have the abbreviations.

Let me explain a little terminology [related to polypeptides]. [A] polypeptide is an amino acid polymer, which is a protein. For each polypeptide, there's what's known as the sequence, which is the order of amino acids found in a polypeptide. If you can remember the [phrase] 'amino acid', then you can remember which way the sequence works, cause you start from the amino end of the molecule and name towards the carboxylic acid end. Imagine that we had the following: valine-alanine-leucine, which could be abbreviated val-ala-leu, or VAL. What is it? This is an example of a tripeptide. Tri peptide means [it] contains three amino acids. Here's what that tripeptide would look like. Notice the backbone, the type of connection we have between the amino acid residues – sometimes that's the term used: this is an amide linkage, relatively unreactive. That's something we should hope for, because that means that amino acids or proteins stick together under rather harsh conditions. There's two of these amide linkages, but in the chain, each unit only has three atoms, so as we go from amino acid to the next to the next, the direction I'm writing the carbonyl flip-flops, just cause I'm trying to maintain this zigzag structure that we default to.

Then, because of that, that means from each amino acid to the next, if I write one with a dash, the next one, the R group's written with a wedge, and then a dash, then back and forth and back and forth. This is why you need to have ability to write amino acids in multiple ways. There is a tripeptide.

How could we synthesize something like this? The first approach I'll take is to say: what if we just wanted to put two amino acids together. Then, we'll talk about a modification that uses an important development known as the Merrified resin, what's known as a polymer support synthesis.

How do you make an amide? We're trying amide bonds to link these amino acids together. We could make an acyl halide. If, for example, if we had acetic acid, we could react it with thionyl chloride, which will get us acetyl chloride. Then, we could react with some kind of amine in order to make an amide. Will this approach work as well with amino acids? I gave us this example of valine-alanine-leucine. Let's say I just tried to put valine and alanine together. Since I want valine's carboxylic acid to be connecte to alanine's amino end, I'm going to make the carboxylic acid of valine the acyl halide; that's the way I want the reaction to occur. We do have two different ends of the polypeptide; sometimes the word 'terminus' or 'terminal' is used to describe the different ends. On the righthand side, we could call this the acid terminal or terminus; sometimes it's called the C-terminal or terminus, because the main functional group there, a carboxylic acid, has carbon as the center of it. At the opposite end, we have the amino terminus or the N-terminus. I want the acid terminus of valine to react with the amino terminus of alanine, so I'll take valine, react it with thionyl chlorine – and I'll end up with a mess, even before I get a chance to react it with alanine. Why? Because the molecule is difunctional. It could react with its own self even before I put the alanine in; that's why I get a huge mess, and that's why you can't do a direct, easy dipeptide synthesis by the same techniques that we used to make a plain, old amide. Amino acids are difunctional; as such, they cannot be easily joined together without the use of protecting groups.

If you wanted to do the following – notice my use of abbreviations [how to correctly approach abbreviations]. We're going to see a totally different reagent that's not just used in amino acid synthesis, but it is very useful here because it does allow joining of amino acids under much gentler conditions. We're going to make two things: for one amino acid, we're going to protect it's nitrogen, we're going to N-protect it. For the other one, we're going to protect the carboxylic acid group, that's C-protection. That leaves only these two functional groups able to react. I'm protecting the N-terminus of one amino acid and the C-terminus of the other; two amino acids can be joined in a specific way.

Let's see what the reagents we could use to protect are, and then let's look at this alternate reagent for doing the coupling of two amino acids together. First we'll cover N-protection. N-protection involves a molecule that's called di-tert-butyl dicarbonate; it's also called BOC anhydride. Carbonate is a structure where you have two oxygens on a carbonyl. Dicarbonate is where you join to carbonates together, which, if you did it in the way that I have shown here, what kind of functional group do we have in the middle of this dicarbonate? An anhydride. The di-tert-butyl part of the name means that, on each end of this dicarbonate, we have a t-butyl group. Why do we use such a funky molecule? Because whenever we use a protecting group, we need an easy way to remove it. t-Butyl groups can be more easily hydrolyzed than other esters; we're going to take advantage of that later on. We're going to use it based on this reactivity that it is an anhydride. If we take an anhydride and react it with something, in a symmetric anhydride like this, you just take one half of the molecule, get rid of it, and you attach whatever would react at the position where this oxygen is located. We're going to use this to protect the amino end of valine, because I'm going to try to join valine to alanine, in this direction that's shown, I need to protect the N portion of the valine. So I'll throw valine at it. Valine is difunctional still, so which group, the amino or the acid group, is going to react with the anhydride? The amino end, because an acid group wants to get attacked by a base; a base wants to attack an acid, whether we're talking about Brønsted-Lowry or Lewis acids and bases. An anhydride likes to be attacked by alcohols and amines. Is it impossible for an anhydride to react with an acid? In order for an anhydride to react, something has to attack that carbonyl. What's more likely to attack the carbonyl: the oxygen of a carboxylic acid, or the basic nitrogen of an amino group? Basic nitrogen of the amino group, so you're only going to get one product out of this. I protected the amino group. Often this protection is abbreviated, so you can say t-BOC. Now that amino acid is ready to be joined.

Now let's move to C-protection. One of the simplest ways to protect is to make an ester, because we know that we can reversibly produce esters. In this case, we'll take alanine – because that's the one that we do want to protect this carboxylic acid end – do a classic Fischer esterification with methanol and acid. Now that end is protected. [Better esters available] We're now ready to join them. We're going to use a reagent called DCC, which stands for dicyclohexylcarbodiimide. [why name?] This will eventually turn into a carboxylic acid derivative. It's the decomposition of DCC into a really stable carboxylic acid derivative that drives the reaction forward. What's the structure of DCC? Those are cyclohexyl rings, not benzene rings.

I'm going to switch to an easier set of molecules. Imagine that we have some kind of carboxylic acid and some kind of amine that we want to join together. If I were to throw DCC into a solution with one or the other of these two compounds, the carboxylic acid or the amine, which one do you think DCC would actually have a chance to react with? The acid? Why? Because the nitrogen with a lone pair is a base, and we have a carboxylic acid, so they would simply neutralize each other first, which is the first step of this mechanism. Acid-base neutralization. We make a carboxylic acid salt, which can react again with this protonated DCC to push one of the carbon-nitrogen double bonds open.

It might not be immediately obvious, but the DCC, once it's attached like this, effectively acts as a leaving group, with the oxygen. We haven't made an acyl halide, but we've made something that acts like an acyl halide. I won't show a proper mechanism, but I'll point out at this point that, effectively what's going to happen, is that amino group is going to attack the carboxylic acid. This is like we were trying to do aminolysis, which means react a carboxylic acid derivative with an amine in order to make an amide. The reason this can happen is because this is not an ester, really, like it looks like; this thing attached at the ester position is more reactive than just an alkyl group. That's the driving force for this reaction. [not a full mechanism] If I simplify [the mechanism] I get the following. We can see that there is a carbon that's got two oxygens and a nitrogen attached, all at the same time. That is this thing that looks like an ester in the middle of being aminolyzed. Notice that I've written six of these atoms to suggest a ring. [correct as concerted mechanism?] The DCC groups ends up coming off with the oxygen as a leaving group, because it is favorable for the carbonyl to form by pushing over the carbon-nitrogen double bond, which the lone pair on nitrogen, in response, can grab the neighboring hydrogen, which, losing that hydrogen, helps this other side of the molecule also form a carbonyl. It is a pericyclic reaction in this case. We get two things as a result: the amide that we want, plus this decomposed by-product of DCC. This is the reagent DCC, which is used to avoid using harsh reagents like thionyl chloride. If you notice, we didn't need any external acid or base sources; we just needed the carboxylic acid itself that we started with to initiate this process, so this can be executed with fairly mild conditions, and it is a very efficient reaction, which means it has a high percent yield.

Bringing this back to our dipeptide discussion, we could take the BOC-protected valine, plus the esterified alanine, and join them together by DCC. I'm going to unabbreviate the BOC side, because it's going to react momentarily. What we're going to do next is undo the protections. I'll show you the N-deprotection, at least. This is the acid trifluoroacetic acid. [pKa = 4.76 for acetic acid] How would the fluorines affect the pKa value? Inductive effect, which means it makes it more acidic, so we get a pKa of [0.23], which means a pretty strong acid. The t-butyl group is very hydrolyzeable in acid compared to other esters [because it's E1]. With the trifluoroacetic acid, what happens is we knock off the t-butyl group. The process doesn't stop here, because does this structure look familiar to you at all? This is one of those carboxylic acid forms that's not thermodynamically stable. Any time we have a carboxylic acid where, on the other side of the carbonyl, there's also a heteroatom, those molecules are not isolatable normally. If we had -OH, that would be carbonic acid; this is a carbamic acid derivative that's going to decarboxylate. That leaves us with a now-unprotected amino end, so we unmasked half of this dipeptide. Esters are more reactive than amides, so if you're careful about it, you can hydrolyze the methyl ester and, in theory, not affect the rest of the molecule.

Wanting to put the focus on multiple amino acids being assembled, let's switch to that. To make a tripeptide or a larger molecule, we use what is known as [the] Merrifield resin. Let me explain the practical reasons why we do this. Yes, I showed you a route that could be used to synthesize a dipeptide. Notice that was have a bare amino end. If we had another N-protected amino acid, we could link it to what we've got right here and make a tripeptide, deprotect the amino group, DCC another protected amino acid, and keep doing it over and over and over again. But there becomes a real-life problem, which is you're making bigger and bigger and bigger molecules that become less and less and less soluble in anything, at least easily. There becomes a synthetic problem, a practical problem, not being able to deal with these compounds, once they get too big. That's where this Merrifield resin synthesis comes in, because it reduces the issues of purification, in a very sneaky, cool way.

We need to know what the resin is. [A] resin is kinda like a polymer. This particular one is a derivative of styrene. What is styrene? Benzene ring with a double bond on it. It turns out that styrene is a very reactive molecule. In either a strong acid source, like aluminum trichloride, or if we had a Grignard reagent, which is a strong base source, or if we had light, which causes the formation of radicals – under acid, basic, or radical conditions, you can get styrene to polymerize. What happens is one alkene attaches to the next attaches to the next, so you get these increments of two carbons, and on the second of each of those carbons, we're going to have a phenyl group on it. That is polystyrene. Depending on exactly how you do this, it's going to end up making what looks like little beads, just like the stuff that's in styrofoam cups. What if we did a variation of this? What if we took two different monomers – if polymer means multiple molecules strung togehter, then monomer is the individual molecule that goes into making it up – I'm taking styrene now with a different functionalized styrene. This kind of position that I've just circled, a carbon that is attached to a benzene ring and something attached to that carbon, [is known as] benzyl. Benzyl halides are great substrates for both Sn1 and Sn2 reactions. We've got some kind of way to initiate polymerization. [block copolymers] This is a functionalized polymer bead – a little sphere.

Why is this kind of substrate so useful? I'm going to simplify this with a P for polymer. The very first thing that's abbreviated is a benzene ring; there a benzene ring next to this methyl position, so it's a benzyl group. That helps us understand this: that if we were trying to do a longer peptide synthesis, what I was describing earlier is the approach we would still take. We'd protect the carbon end of one molecule and attach an amino acid to the amino end, and then attach another one to its amino end, and so on and so forth. If we had this big, long amino acid chain, whatever's the end of the chain is the first thing that we put on the polymer. [valine-alanine-leucine] If I want to synthesize that particular tripeptide, I'm going to put leucine onto this polymer support first. How am I going to do that? Use sodium hydroxide to deprotonate the leucine. You make a carboxylate, which is not that basic, but it's basic enough to be able to do an Sn2 reaction on the polymer chain. Why is this so magical?

Cause when you're done, pour the beads into a cup, filter out the solution, wash off the beads, throw them back in the reaction, do the next one – because your product is attache to these beads, anything else that didn't get attached is just by-product, so you wash it away, so you're guaranteed to have this one amino acid sitting on that polymer bead. Now when you throw that bead at the next solution, then you add one more molecule on; pull the beads out of solution, you've got your product automatically. You can do this over and over and over again, something like 25 time at least if your technique is really good or if you make the whole process automated.

How do we finish this off? Same type of approach we used up above. Let's say we already have an array of BOC-protected amino acids. I've now got leucine on here. The next one I would want to put on is alanine, so I'm going to take BOC-protected alanine, which has its carboxylic acid end free, I'm going to take that with the leucine, [which] has its free amino group ready to react, and I'm going to join them by DCC. Trifluoroacetic acid will hydrolyze the t-butyl ester, which will then release a derivatized carbamic acid, which decarboxylate[s], which means we end up with the amino group free again. [abbreviation approach] Now we can take the next amino acid – which is the first amino acid – which still must be BOC-protected [Bach, Offenbach] Now I've got three amino acid on that polymer chain. If I wanted to add yet another amino acid, I could use trifluoroacetic acid again. At this point, I would have the tripeptide, but stuck on the polymer bead. [mechanism of benzyl ester cleavage] [You can cleave benzyl esters] with HF. Once you've cleaved it, then we've got our released tripeptide. That's the Merrifield resin synthesis.

[lab directions]

_____

Convention for amino acid Fischer projections; acid @ top; R side chain @ bottom

polypeptide ≡ amino acid polymer ≡ protein
sequence – order (and kind) of amino acids in a polypeptide: amino –> acid
example tripeptide (contains 3 amino acids): valine-alanine-leucine (val-ala-leu) (VAL)

Amino acids are difunctional. As such, they cannot be easily joined together without the use of protecting groups.

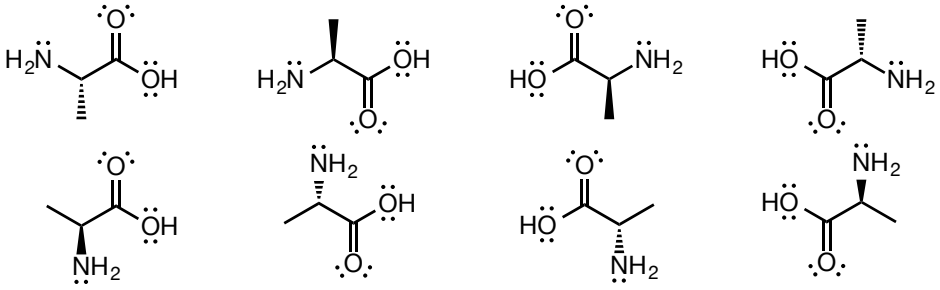
N-protection: di-tert-butyldicarbonate (BOC anhydride)

DCC – dicyclohexylcarbodiimide

_____

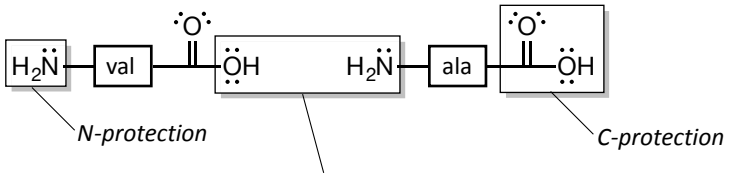
Structures *(remaining structures identical to lab 12B)*

06/04/12 lab • 1



L-alanine

06/04/12 lab • 2



By protecting the N-terminus of one amino acid @ the C-terminus of the other, two amino acids can be joined in a specific way.