Lab 12B • 05/31/12

Alpha-amino acids

The simplest [amino acid] is this compound – to put just an amino group right at that alpha position and having no other carbons or other substituents; this is the molecule glycine. There are different abbreviations that are sometimes used for these amino acids. The amino acids do join together to make different proteins. Specific proteins have very specific functions in the body. Part of the effort [to understand biological systems] is to what is called sequencing of these proteins, which means figure out what is exactly that chain of amino acids. Since these are huge molecules, to more easily represent the molecules, there's three-letter abbreviations for each amino acid and a one-letter abbreviation for each amino acid. For glycine, the three-letter abbreviation is pretty logical – gly – or its one-letter abbreviation is G. Many of the amino acids do have the same [first letter of their names themselves], so [some of the them] have one-letter abbreviations [that do not match the first letter of their names]. This simplest amino acid has no stereocenter to it because the carbon has two hydrogens attached to it. All the other amino acids do have a stereocenter to them, because the structure of all of the remaining amino acids, which are going to be alpha amino acids, have an R group in that position.

Let's draw a Fischer projection for this. I'll set my visualization point like this, where I'm looking up upon the molecule because to carbon, the dash and the wedge, are what's going to be pointed towards me, because I want to be on the horizontal axis of this Fischer projection; In oder to do that, you have to be underneath. Right now, it makes sense to put that amino group in the plane of the board; we're just focused on the fact that it's an amino acid. What I'm going to show you once I redraw this Fischer projection is that we can refer to amino acids with d- and l- configurations, the same way we refer to sugars. There's a standard way that that's done, which instead of putting the amino group in the backbone, it's instead to have the R group in the backbone and then see which side the amino group ends up on. To do so, I'll rotate this around. We see that that NH2 is on the right side of the molecule. When drawn this way – this would be the standard way to draw the amino acids – when that group is on the righthand side, then that is the d form of the amino acids. If we put a methyl group on, then we get alanine, so this would be d-alanine.

Let's go backwards from the Fischer projection to a line structure again. I have this have this habit of writing the carbonyl up for my carboxylic acids. We could write this two different ways, though: if we take this from the Fischer projection, then I would put the methyl group in the plane of the paper. Would I use a wedge or a dash to write the amino group? A dash. Why? Because if I'm looking up – which I need to be if I did choose to put the carbonyl in that direction – if I'm looking up and I'm supposed to have the amino group in the right, based on my perspective, looking up like this, right is away from you, so that means it is going to be written with a dash. If I were to rotate (the bond) around and bring that amino group from being in back of the plane and pointing down to pointing up, then that's going to bring the R group goes this way, which is why it ends up a wedge. If the d form is the d form because the amino group is on the right, then, logically, the l form is when the amino group is on the left, so this is l-alanine. Translating from the Fischer projection back to the line structure, this time, if the R group is in the plane of the board, then we'll have the amino group as a wedge, and if we have the amino group in the plane of the board, then the R group is going to be a dash.

It turns out that the l amino acids are the biologically-active ones. D sugars are biologically active; l amino acids are biologically active. An amino acids is a difunctional molecule, which presents some challenges if you're trying to do synthesis. Before we get into that action of joining the molecules together, let's just talk about peptides, which are a string of amino acids. You can somewhat loosely refer to peptides as amino acid polymers. The reason that we sometime end up going back and forth between representing the structures with the amino group in the plane of the board or not is because it's more convenient to write it in the plane of the board when we're going exactly this, linking ano amino acid to the other. Notice I've used three l amino acids. The one on the left and the one on the right, they're configured exactly the way that I've written the l amino acid previously. Notice that the amino acid in the middle, to keep that zig-zag form, I've written with the carbonyl pointed down; that means it's as if I've flipped the molecule 180° on the paper, which is why that R group is written with wedge instead. If you look, we have three different amino acid units here, so this is what is known as a tripeptide. If you can remember the term 'amino acid', when we list the amino acids, that list is called a sequence. The sequence is a list of the amino acids from the amino end to the acid end. To remember which way to list them, just remember the term itself – amino acid. Instead of end, sometimes terminal or terminus is used, just like in a bus station. A list of amino acids from the amino terminus to the acid terminus.

What kind of linkage is there between the different amino acid units? What functional group is it that we have? An amide, which is why these peptides tend to stick together, in the sense that they do not decompose easily, cause amide bonds are not all that reactive. If we're wanting to synthesize a polypeptide, we need to somehow make an amide linkage. How could we put two or more amino acids together?

First, a logistical issue: is we had two amino acids – l-valine [isopropyl group that looks like a V] and l-leucine [related to valine by putting one more carbon in between the V junction and the main platform of the amino acid.

Side chain [are] the R groups that are on the amino acid itself. [preview of classes of amino acids] What if we wanted to put valine and leucine into the sequence val-leu [alanine - ala - A]. Valine is val or V; leucine is leu or L. If we wanted to make this dipeptide valine-leucine, how could we do it? What are our ways we know to make amides? Can we easily, directly react an amino with an amino acid to make an amide? It is possible: you heat it up enough and you will dehydrate the compound. It'd kinda be like a series of protonate-open-attack-deprotonate, but it does require that heat to remove water. What happens first is the amino would just neutralize the acid. The amino group can react with the acid group. Do we have another way we can make amides?

A carboxylic acid is difficult to make other functional groups from because of the fact that it's an acid, which means if we take anything basic to react with it, it's just going to neutralize instead. Therefore, it would make more sense to make a more reactive group than a carboxylic acid to then react with something else to form a derivative of a carboxylic acid. What two functional groups are more reactive than a carboxylic acid? Acyl halides – how do you make an acyl halide from a carboxylic acid? Thionyl chloride. That's our answer right there: we could take a carboxylic acid, react it with thionyl chloride, turn it into an acyl halide ... and end up with a mess. Why? Because if you take valine, for example, and try to react it with thionyl chloride, in principle, yes, you could get this – but, since you're in a solution that also has the initial carboxylic acid in it, then you'd just make a dipeptide, without ever having the chance to react with leucine. If you tried to put thionyl chloride into a solution of both of these compounds, then you'd end up with a bigger mess, because both the compounds could make the acyl halide, both compounds could react with each other, so you have no control over what kind of addition would occur. Because amino acids are difunctional, you cannot directly link two amino acids together without getting side reactions and getting unwanted combinations. SInce amino acids are difunctional, it's not possible to control the order in which amino acids add to one another – without using a protecting group strategy.

If our problem is that both molecules have an acid and an amino end to them, then why don't we make it where we block the amino end of one of the amino acids, and what if we block the carboxylic acid end of the other amino acid? That's exactly what we could do. First I'm going to describe a method that we could use if all we cared about was joining just two amino acids together. I'm then going to talk about how that's modified a bit so that you could repetitively join amino acids together.

First, we have C protection, how to protect the carbon terminus. What reactions do we we know that could temporarily block the reactivity of a carboxylic acid? Do we know a way to convert a carboxylic acid into something that it would be easy to undo, but while it's in that form it would would not react with the same types of reagents as a carboxylic acid. What if we made a less-reactive carboxylic acid derivatie? What if, for example, we esterified? If I'm going to be joining these two amino acids together, which of the two of them would I want to esterify in that manner? The right one, because leucine, the amino acid terminus, is going to be on the righthand side, if we block that, that guarantees that valine is going to make the connection that we want. We're going to esterify leucine. Standard Fischer esterification; methanol is cheap and easily available, so I'm showing making the methyl ester. Now, that end won't react with, for example, thionyl chloride or other reagents a carboxylic acid would.

What about the amino group? Di-ter-butyldicarbonate. tBOC. Carbonate would be a carbonyl with oxygens on either side of it. Dicarbonate is this kind of structure, where you have two of those linked together. The di-tert-butyl means on either end of it we have a tert-butyl group. If you look at the structure of this, what do we have? It's part ester, but it's part anhydride. What happens when we react this with an amino acid like valine? We're doing this with valine because we want that amino end blocked to ensure only the carboxylic acid end will react. Showing a partial mechanism, if that nitrogen attacks either of the carbonyls, this compound is symmetric so it doesn't matter which of those carbonyls you attack, then you could push open that carbonyl. [saving space by skipping tetrahedral intermediate] We kick out that whole rest of the molecule. It's an anhydride; anhydrides are like acyl halides, they can react with amines to make an amide. The amine would have a hydrogen on it at first, which means when it first makes its new bond, it's going to be positively charged. We also lose a hydrogen, and we end up with this amide, which you can abbreviate. It is an amide, which means it won't react with a carboxylic acid. We've protected the carboxylic acid end of one molecule; we've protected the amino end of another molecule. Now we can put the two compounds together.

There's another issue that can come up, which is that thionyl chloride is a pretty reactive compound. Some of these amino acids have, for example, alcohol groups in them. If we to try to use thionyl chloride to convert just the carboxylic acid portion of that molecule into an acyl halide, there's guarantee we wouldn't also simultaneously convert and alcohol into an alkyl chloride. Thionyl is used to make alcohols alkyl halides instead. There's another reagents very, very commonly used. Dicyclohexylcarbodiimide (DCC). It's structure is this. [why is it called a diimide] This is going to result in eventually forming a type of carboxylic acid derivative. DCC we use without any other helper reagent to link two amino acids together. [simplifying – abbreviating amino acids] I'm going to represent the C-protected leucine that I made in this way, and then I'll represent the BOC-protected valine in this way. Let's see what the mechanism for DCC would look like. DCC is not going to react with the esterified leucine because esters have nothing that would want to react with the nitrogens here, but what about the BOC-protected valine; is there something on that molecule that might want to react with DCC? Is DCC acidic for base? Basic, because it has the nitrogens there. Imines, like amines are basic. We have a carboxylic acid on the BOC-protected valine. Can't that carboxylic acid react with DCC just by doing a simple acid/base reaction? Yes, that's what starts this process off.

First, just acid/base neutralization. We now have a carboxylate. We also have this positively-charged nitrogen next to a carbon. Again, if this was like an imine, then this would look like protonate-open-attack, which is what happens next. [open/attack, resonance] We end up with this. There's now an interesting series of steps that occurs. SInce there's still more acid around, we can imagine that there's an acid around available to have the carboxylic acid portion here protonated. Now we have the other amino acid that we protected at the C end come in; that was the leucine. Recognize that, technically, this protonation and this attack should be separate steps. We started to protonate-open-attack of an amine trying to react with the ester here. The reason that this ends up being favorable is we're about to tautomerize the rest of what's remaining with the DCC. If you look at it, we now have a carbon-nitrogen double bond in the same place where we have an oxygen that's singly-bound to that carbon. This would rather be a carbonyl, because then you would have an amide on both sides. There's would be so much stabilization of that carbonyl because we would have two nitrogens both donating their lone pairs through resonance. Right now, this is not nearly as stable a structure. It's the decomposition of the DCC that is the driving force of this reaction. [cyclic reaction] The other nitrogen from the DCC can grab a proton from this carbinolamine (or the equivalent). That means that the carbon-nitrogen double bond can open up through resonance. To simplify further, I'll just extend that arrow to show it coming from the double bond. That's because the carbonyl wants to form, which means this carbon-oxygen bond breaks, and since the alcohol portion here got deprotonated, that bond has to have somewhere to go. If you look at what we end up with, the DCC has decomposed into being a urea derivative. What we have left over, if I turn if back around, writing it from amino end to the acid end, it's going to look like this. We've now successfully made our dipeptide.

How could we remove the protecting groups? It does turn out that an ester is more reactive than an amide, so the methyl ester, we could hydrolyze. If we hydrolyze, we're at the same time going to be able to take care of that tBOC group. Let me expand out that structure again. If we use a strong acid that wouldn't cause any other side reactions, trifluoroacetic acid is often used for this purpose. Trifluoroacetic acid could react with an ester to turn it back into a carboxylic acid. The ester portion of this is more reactive than the amide portion, which is why the amide portion is left alone. You might recognize this as one of those types of carboxylic acids that's not thermodynamically stable, so spontaneously, this loses carbon dioxide, and so we've deprotected the amino end. The ester end I'm not showing, but that would be just simple hydrolysis. If this ester can be hydrolyzed, the other ester could be hydrolyzed as well to give us our carboxylic acid.

Reviewing the steps: we wanted to make a dipeptide sequence. I pointed out that to join the two [amino acids] together, you have to block one end of each molecule, cause otherwise we can't control how they're assembled. I showed you C protection: a simple way to do it, but not the best way, but our first pass, is to make a methyl ester. We also have to protect the amino end of the other amino acid; that's where this BOC protecting comes from. The reason the BOC protecting group is used is we could protect the amino group my making an amide; that's really unreactive. But the whole point of this is to make an amide. It doesn't make sense to make an amide protecting group if we're trying to make an amide functional group, but how could you deprotect in the end? BOC is s very sneaky reagent, because yes, it's an anhydride, so it's going to add to the amino group in this way, but it's also an ester, so we can hydrolyze the ester portion – that's what we do last – without the amide portion being affected – until, after removing that ester portion, we have a decarboxylation. Once you do have your carbon-protected amino acid, your nitrogen-protected amino acid, you're controlling which amino group and carboxylic acid group react. I've shown you this alternate reagent, DCC, that can be used to join the reagents without using harsh conditions like thionyl chloride, cause this happens under virtually neutral conditions; the carboxylic acid itself provides a proton to start this process off.

Merrifield resin synthesis

What's the name of this molecule? Styrene. If you were to react this with either an extremely basic reagent like a Grignard reagent, or if we react this with light, or it's even possible to react this with acid, then we can make polystyrene – exactly the same type of polystyrene that you think of when we have styrofoam cups. If you crush it, it's like it's formed of these little beads of material. What if we were to purposely synthesize this polystyrene in such a way that we produce these beads of material. Now, instead of just making it from plain styrene, we add a bit of functionalize styrene: p-chloromethylstyrene. Then you make what is called a co-polymer. Polymer means making a huge molecule out of multiple repetitions of a smaller molecule. Co-polymer means to do it with two of them. You don't make it completely this substituted styrene, because what we want is a bead that every few places, every once in a while around that structure, around that bead, we have these functionalized styrenes. I'm highly simplifying the way I'm drawing this structure. Here's the way I'm going to write it: he resin, the bead, with this chlorine on a methyl group sticking off. But if I'm simplifying this way, you've got to keep in mind it's a chlorine on a methyl group on benzene. What do we call these kinds of structures where we have substitution at one carbon away from benzene? [not allyl] Benzyl. It's very reactive. Benzyl chlorides in general are good substrates for Sn2 reactions. [essential amino acids]

Let's say that I want to make the tripeptide VGA, valine-glycine-alanine. Alanine's going to be the carboxylic acid end I don't want to have react. Here's what I'll do. I'll take alanine [leave out lone pairs][arabic script, has lone pairs] Alanine, I'm going to react with sodium hydroxide. If I then take the resin, a carboxylic acid salt is only a weak base, but it's a base, which means, technically, it could be a nucleophile. The benzyl halide is a particularly reactive substrate, so even if normally maybe you didn't have the best base to use right here, this will still work. What happens? Nothing – not until you somehow force the hydrolysis of the ester.

As long as you avoid that, then every reaction you do with this now functionalized bead has a one-step clean-up to it. You dump the beads out of solution, and you wash them with some water and dump them right back into your next reaction flask. This platform-assisted synthesis works in the following way: you start out with the bead, you put your first amino acid on it by deprotonating and doing an Sn2 reaction. You could then dip the bead into your next solution, pull it out, put it into the next one, pull it out, put it into the next one, the next one, the next one,. There's an automated synthesis that works at least twenty or so times for making small amino acid chains. If we were to do this without using this support, it's called, then as we add one more amino acid and keep on adding and adding them, we're making huge molecules, huge molecules that start to be difficult to work with in solution; big enough peptide chains won't easily dissolve. But if we put it on a bead that isn't going to dissolve anyways, at least we have a way then to separate out these compounds from solution and purify them and then continue to the next step, without having do distill or do chromatography or do anything else on it. [nobel prize?]

To show you what happens next, we want to put the amino acid onto this amino terminus that's available. That means whatever amino acid next that we need to use has to have N-protection to it, so that only its carboxylic acid reacts. Our next amino acid, if we're trying to make VGA is to have glycine. So, I'll take glycine, react it with BOC anhydride, so we have the protected glycine that has that reactive carboxylic acid group. For shorthand, if I call the support resin that's got the first amino acid, A, and if I call the protected amino acid that we just made, B, A plus B plus DCC equals a dipeptide. At this point, you'd literally just filter out the beads, the reagents go away, you dip this in trifluoroacetic acid. It turns out that as long as you control the conditions carefully, the polymer support will not react with this; turns out that that BOC group ends up being more reactive. That means you can deprotect this amino group, and it's now going to react with the next amino acid. So, we could take valine ... let's jump ahead, let's say that already we've made the protected valine. React those with DCC – now we've got a tripeptide. Trifluoroacetic acid again will deprotect the amino group. [for whatever reason] HF reacts with benzyl ethers. If you look at that protecting group, yes, we have an ester from the perspective of what you see here, but if you look back up at what we have, we've got this benzyl group connected to an oxygen to that positions. From that benzyl group's perspective, it's a benzyl ether. It reacts with HF to give us the released, deprotected tripeptide. That is the Merrified resin synthesize it. To summarize it one more time: you protect the last amino acid in the chain by putting it on the resin first; that way, if the carboxylic aid end you could protect once and never have to keep on protect, deprotect, protect, deprotect.

Once you have that amino acid attached, you then protect your second amino acid, join it together, deprotect, take the next amino acid, join it together, deprotect. You keep assembling until you've got all your amino acids that you want. You do your final deprotect, then you release from the platform, and you've got your peptide chain. [pre-lab]

_____

alpha-amino acids

l-amino acids are biologically active
peptide – amino acid polymers
sequence – list of amino acids in a peptide from the amino terminus to the acid terminus (amino –> acid)
dipeptide val-leu

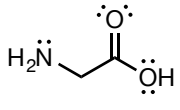
Since amino acid are difunctional, it is not possible to control the order in which amino acids add to one another without using a protecting group strategy.

C-protection
N-protection – di-tert-butyldicarbonate – t-BOC
dicyclohexylcarbodiimide
amino deprotection

Merrifield resin synthesis
tripeptide VGA

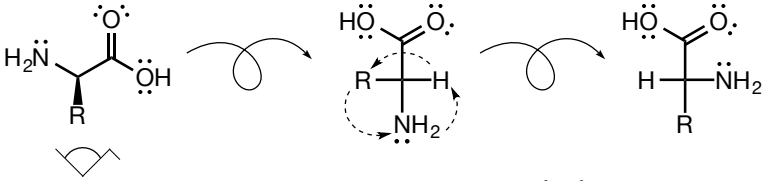
_____

05/31/12 lab • 1

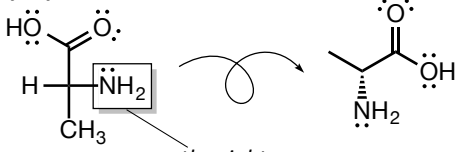


glycine
gly
G

05/31/12 lab • 2



05/31/12 lab • 2

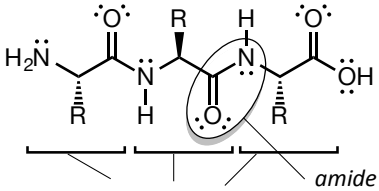


D-alanine
ala
A

on the right

or

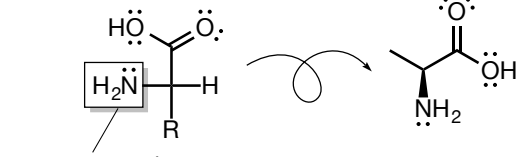
05/31/12 lab • 4



3 amino acids –> tripeptide

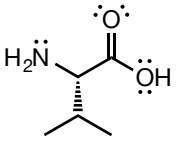
amide – relatively unreactive

05/31/12 lab • 3



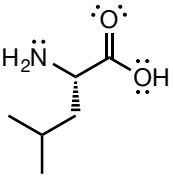
on the left L-alanine

05/31/12 lab • 5



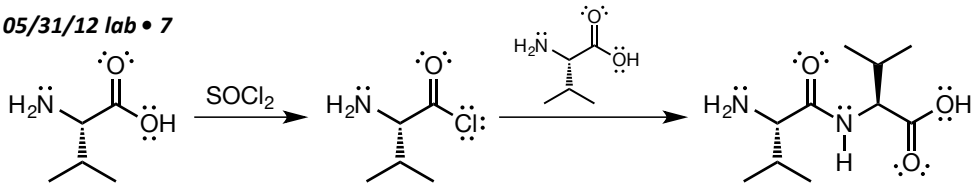
L-valine
val
V

05/31/12 lab • 6

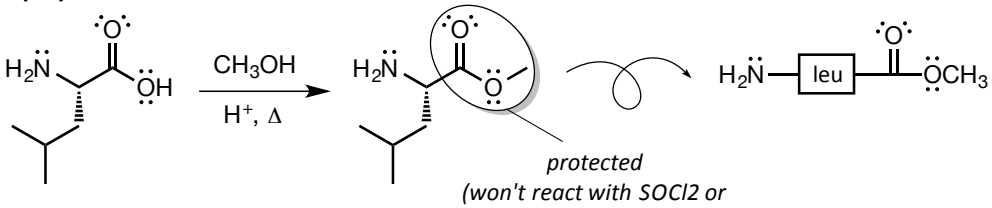


L-leucine
leu
L

05/31/12 lab • 7

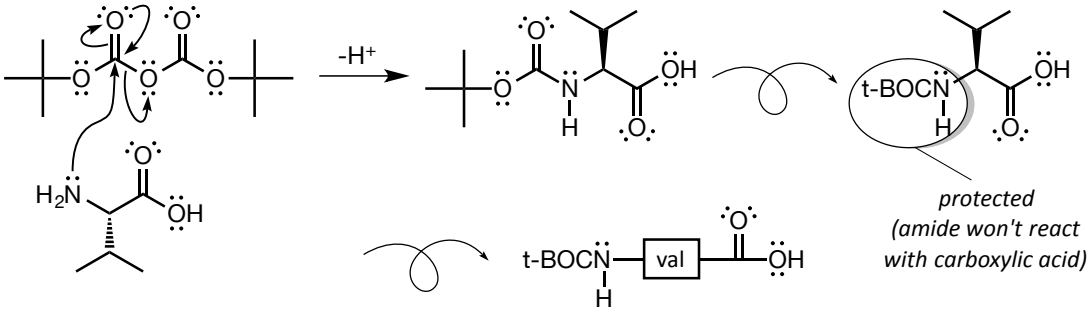


05/31/12 lab • 8



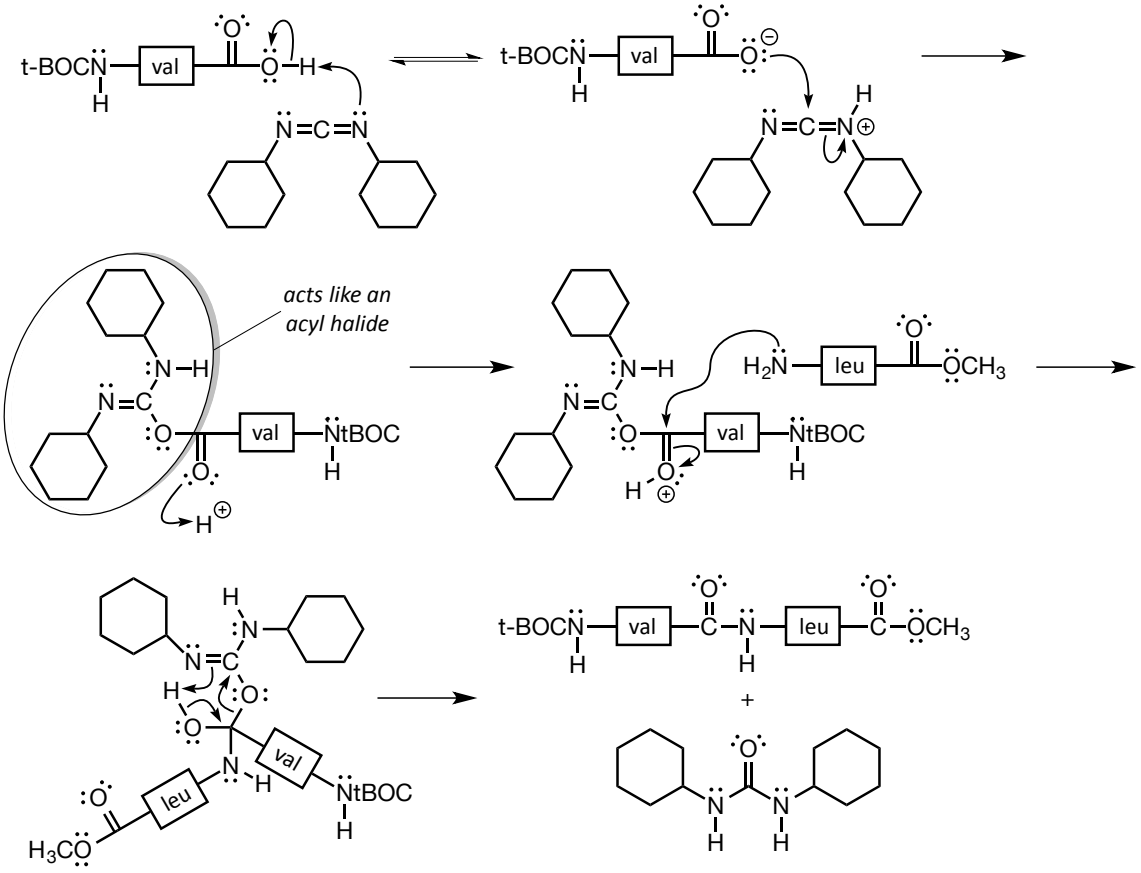
protected (won't react with SOCl2 or other reagents a carboxylic acid would)

05/31/12 lab • 9

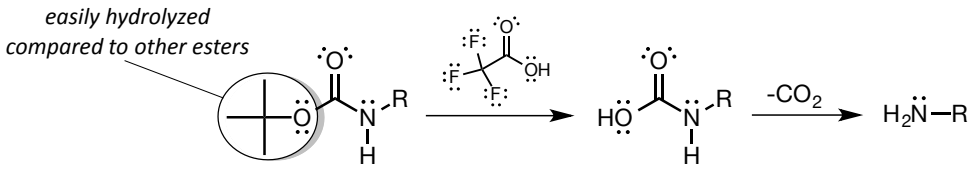


protected (amide won't react with carboxylic acid)

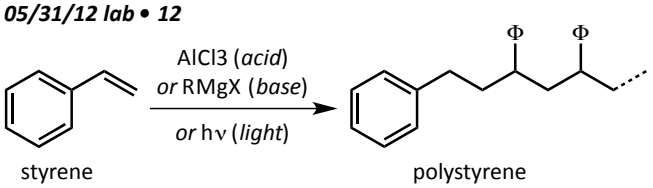
05/31/12 lab • 10



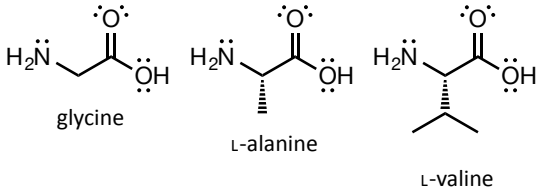
05/31/12 lab • 11



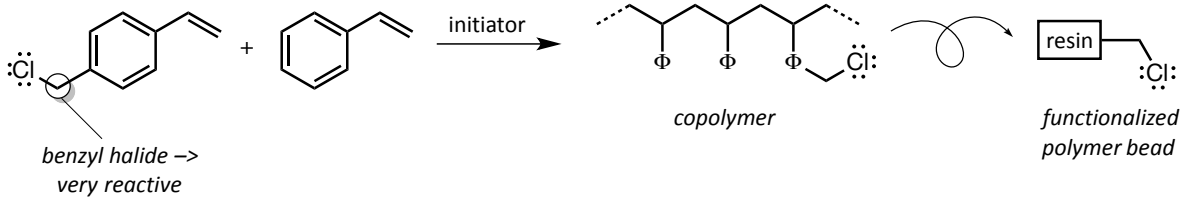
05/31/12 lab • 12

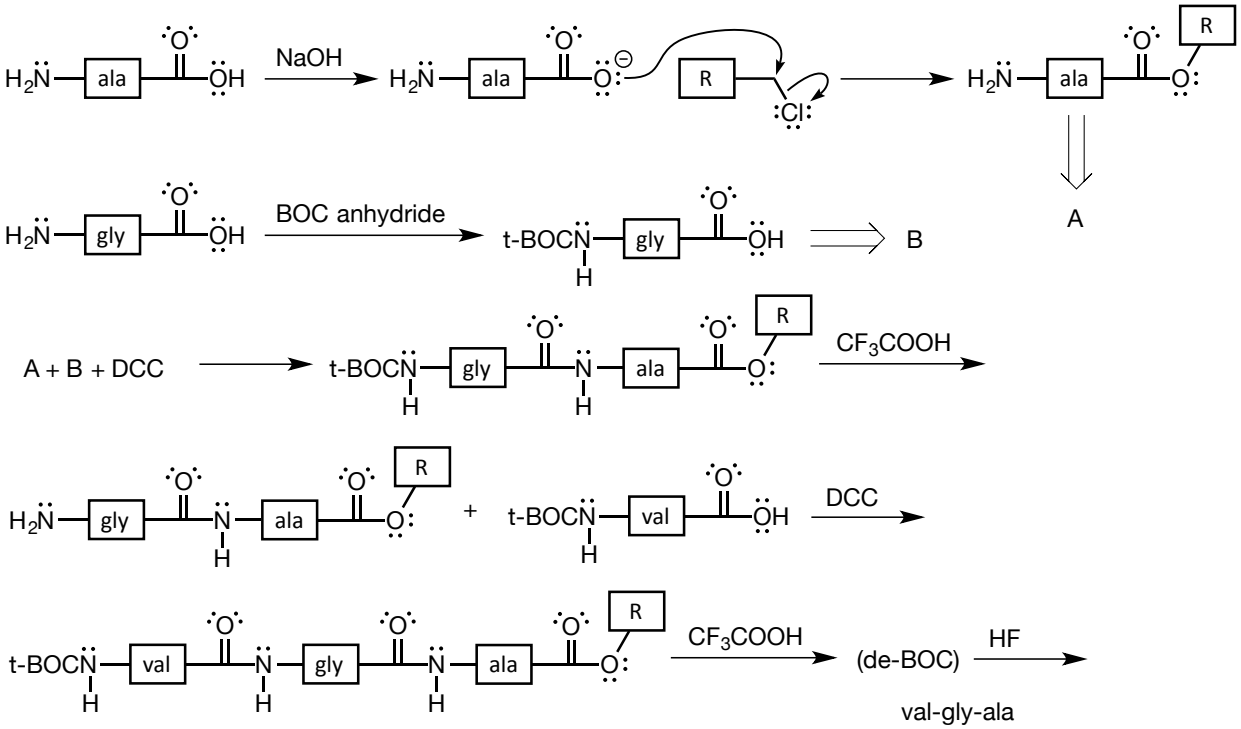


05/31/12 lab • 14



05/31/12 lab • 13




NaOH
BOC anhydride
A
B
A + B + DCC
CF3COOH
DCC
CF3COOH
HF
(de-BOC)
val-gly-ala