

Chapter 3 Amino Acids, Peptides, Proteins

Problems

1, 2, 4, 5, 8, 9, 10, 13, 14, 15, 16, 18

Also start memorizing AA:

names, structures, abbreviations, 1 letter codes, pKa's of side chains

3.1 Amino acids

Amino acid is monomer

condensed into peptide (small) protein (large)

Both can be degraded back to AA's by hydrolysis

(At neutral pH hydrolysis is slow but faster in acid or base)

Large bulk of proteins broken down to 20 different amino acids

A. Amino acids share Common structural features.

i. C atom center, Amino group on one side carboxylic acid on the other, and an R group that can be H or more complicated will explore in a minute

ii. 20 common AA called **standard AA's** other AA's derived from these AA's usually by post-translational modification (ie AA was changed after it was made into a protein)

iii. All 20 AA's have trivial names

All have a 3 letter code

All have a 1 letter code

I expect you to know codes and structures for all AA's

Naming substituents on AA's is not standard organic chem

base C is called alpha C

then out from there is beta, gamma, delta, epsilon

iv. All AA's except Glycine are chiral

All AA's except Glycine have 4 different constituents around C α .

Why is this important?

Chiral center

non-superimposable mirror images (Figure 3-3)

enantiomers

optically active

(Will rotate plane polarized light and display differential absorption of circularly polarized light)

What did you learn in organic about nomenclature in this system?

a. L-D (oldest) dates back to 1851 Emil Fisher relate structure to L or D glyceraldehyde Figure 3-4

Under this nomenclature AA's are L. **All prepared by living things are L.** Ingest D and will die of starvation.

b. based on how rotate plane polarized light + or -
AA's can be either

c. R or S AA's can be either. If good at this give a few a try,
not used much in Biochem so won't go over

B. Classification by R groups

Figure 3-5

Have looked at common part, now lets look at different part, the R group

What are useful chemical properties

Charge +/-

Acid or base (tied to above base usually +. Acid usually -)

polarity

Table 3-1 Be able to look at an amino acid (or your memorized **Figure 3-5** structure) and place into 1 of 5 groups

non-polar aliphatic

G, A, P, V, L, I, M

non-polar aromatic

F, Y, W

Polar uncharged

S, T, C, N, Q

+ charged (Base)

R, K, H

- charged (acid)

D, E

Note: some of these classifications are not clear cut. Tyrosine is somewhat polar

Also note: should remember roughly the pKa's of sidechains

Backbone COOH is usually about 2 and NH is about 9.6

Special notes on special groups

i. Aromatics

Give protein absorbance in 280 nm range (**Figure 3-6**)

ii. cysteine may occur as free SH, or may be reduced to form cystine (**Figure 3-7**)

iii. His with pKa of about 7, is use often as a catalyst because can

act as either proton donor or acceptor at neutral pH

C. Uncommon Amino Acids also have Important Functions

Other AA's observed in nature

Figure 3-8

- i. 4-hydroxyproline (found in plant cell walls and collagen)
- ii. 5-hydroxylysine (found in collagen)
- iii. 6-N-methyl lysine Found in myosin, a muscle protein
- iv. γ carboxy glutamate (found in blood clotting proteins)

All of above are made by incorporating base AA into protein, then modifying after protein is made (post-translational modification)

Ornithine and citrulline are intermediated in amino acid synthesis, but are not incorporated into proteins

Selenocysteine (looks like cysteine with a selenium instead of a sulfur)
 Synthesized from serine
 Incorporated into a very very few proteins in specific organisms
 A rare bird

D. Amino acids act as acids and bases

Since have common form of NH and COOH
 have both acid and base functions
 called zwitterions when have both + and negative form
 also called amphoteric

Look at titration curve of an AA without a charged R groups

Figure 3-10

COOH pKa ? (2)

NH₃ (9.6)

Is this normal for an organic COOH and NH₃? **Figure 3-11**

NO COOH is lower than expected.

Usually more like 4 (acetic acid is 4.8)

NH₃ is also lower than expected.

methylamine 10.6

Can you explain this?

A lower pK_a means is more or less acidic?

More acidic

More acidic means is gets rid of H more easily

Why can get rid of H more easily?

Let's think about the fully protonated form H₂A⁺

Has charge on NH₃⁺, wants to get rid of H⁺ from COOH to become negative (COO⁻) and then no net charge on the molecule. Another way to think is that charge-charge

interaction stabilizes the zwitterion so it occurs more easily

Use a different logic on NH_2 . Base is close to COO^- . These electronegative atoms (O's of COO^-) pull electrons toward them. NH_3^+ is more easily deprotonated (more acidic) because these electronegative atoms help to delocalize the electron that remains on the NH_2 as it deprotonates, the NH_3^+ group is more acidic.

E. Titration Curves

Notes don't follow text

Construct a curve like **figure 3-10**

at what pH has a negative charge?

At what pH has a positive charge?

At what pH is neutral called **isoelectric point** pI

formulas from Analytical?

$$(pK_1 + pK_2)/2$$

Now how about a more complicated AA, say Glutamate or His

Figure 3-12

Amino Acids with >2 ionizable groups

What were basic AA's

(Lys, Arg, His)

What were acidic AA

(Asp, Glu)

What do titration curves of these AA's look like

Examples: Glu and His **Figure 3-12**

Glu pKa's 2.19, 4.25, 9.67

His pKa's 1.82, 6.00, 9.17

What are structures at these 2,5,7,9,10 etc

Where is pI ?

What is charge at neutral pH?

Note - His is special because it is only AA in the bunch with pKa close to neutral so depending on situation can be in either ionized or unionized

Before we leave let's look at either Tyr (2.2, 9.11, 10.07) or

Cys (1.96, 10.28, 8.18)

Both have 3 ionizable groups - why not acid or base?

Both acids but extremely weak, so at neutral pH uncharged, so usually don't see much. However when place near another group get very interesting results

3.2 Peptides & proteins

Now start building monomer into polymer

A. Peptides are chains of amino acids

Condense 2 AA's via a peptide bond kicking out water

(like figure 3-13 on board)

The resulting peptide bond has special properties will use in next chapter. Lets see a hint

Is acid or base? (Amide bond - neither)

Polar or nonpolar? (Polar and slightly resonant so even more polar)

Can make hydrogen bonds? (Yes to both sides)

Note equilibrium is toward AA's not peptide

peptides/proteins breaking down

need to input E to make bond

Will see is done biologically next semester, chemicals later this chapter

Actually kinetically slow process, at pH 7 $t_{1/2}$ about 7 years

Acid and base catalyzed so at high or low pH much faster

Peptide kind of a generic term for several AA's hooked together

oligopeptide a 'few' AA's (<30?)

polypeptide 'many' AA's (30-100?)

Proteins 'lots' of AA's (about 100)

In biological systems is always linear. Not branched.

Note: order is important. Who is at which end

AA with free α amino group in main chain is call the amino-terminal or N-Terminal

AA with free carboxylic acid in main chain is called the carboxyl terminal or C-terminal

B. Can distinguish peptides by ionization behavior

Peptides will have ionizable groups at N and C terminus, Plus any ionizable groups from side chains. Thus can do for peptides what did for

amino acids, predict charge at any pH and predict pI .

Can use this to find methods to separate different peptides
 Won't correspond exactly to pKa given in 5-1 because environment changes pKa's. For instance will Terminal COOH and NH₃ still have pKa's of 2 and 9? No - no longer near charged groups so no inductive effect, and will be more 'normal' with 4 and 10

C. Biologically active peptides come in many sizes and compositions

from 2 to 1000's of residues

Nutrasweet (Asp-Phe-methyl ester)
 often have effect at very low concentrations
 many hormones are peptides

Oxytocin (9 AA)
 Bradykinin (9 AA)
 Thyrotropin releasing factor (3)
 Insulin (51 AA 30+21)
 Glucagon(29)
 Corticotropin(39)

Proteins

Cytochrome C about 100 residues
 Titin (from muscle) 27,000 AA - MW 3,000,000

Average MW of a residue roughly 110

Proteins can also be multi-subunit so can have several associated together!

Peptide composition

Each protein has a different # of AA's, different % composition of AA's, different sequence of AA's. All will have different physical properties

Some AA's may be there others may not compare 2 proteins **Table 3-3**

D. Many protein will have other things besides amino acids

may be covalently attached or non-covalently associated
 added group called prosthetic groups

Proteins with groups called conjugated proteins

Lipoproteins -lipids
 glycoproteins -carbohydrate
 phosphoproteins -phosphorus
 hemoprotein -heme
 metalloproteins -metals

3.3 Working with Proteins

How do we know what we know about proteins?

Isolate and characterize in the lab

a bit different than regular chemistry instead of synthesizing a new compound and isolating it from a few reactants or products, are trying to isolate a single protein out of an organism that contains thousands of proteins How do you do this?

A. Separation and purification

Step 1 break open cell in some way

lyse cell- grind cell- sonicate- etc

resulting solution called a **crude extract**

If possible, isolate a particular organelle of cell that contains POI (protein of interest)

See discussion back in chapter 1 on how this is done

Then begin to separate different proteins from each other

i. And oldest method - selective precipitation

usually adding salts like $(\text{NH}_4)_2\text{SO}_4$

First try to get other proteins to ppt

Filter and get rid of other proteins

Now get yours to ppt

Save filtrate, and dissolve in water and back in business

This method usually very crude, but can get protein from 100' of L of cells pptd down to a few mls of protein paste so is useful

Now what?

dissolve back in water.

Now have protein dissolved in water, but what is wrong? Have salt still in water as well. How to get rid of salt?

Dialysis

Place in membrane with pores lets salts out keeps proteins in.

May use dialysis often in purification procedure whenever want to get rid of small ions

ii. Now need **chromatography** What is chromatography?

Method where a mixture is applied to some medium, and as the mixture moves through the medium, different components in the

mixture are retarded differently so materials begin to separate from each other.

Any examples of chromatography from class??
(Bring in a demo??)

Lots of chromatographic methods, biochemistry uses 3 major methods (Figure 3-17)

Ion exchange

Medium is a solid polymer polymer contains charged species

Ions of opposite charge in medium are attracted to solid and stick with different affinity

Least attracted come right through, most attracted may be permanently bound

Your's comes off in between?

If didn't come off

Change ionic strength

Change pH

Size exclusion chromatography

Polymer has pores of different sizes

Small guys penetrate hole, takes longer to get through

Large guys stay outside and get through faster

Nifty technique. Plot log MW vs elution can get MW!

Affinity

Covalently attach something to polymer that protein likes to grab

All other proteins was through

Your protein is stuck

Add excess of ligand

Flushes off your protein, and is now pure

Any one method will not purify completely, so typically will use at least two or more of the above methods

Once pure, then go on and characterize the protein

B. Separation and Characterization via electrophoresis

One technique that is used often in Biochem is electrophoresis

Electrophoresis refers to the movement of charged species in an electric field

-nice thing about this technique can be used both for characterization and separation at the same time so don't need pure sample before use for characterization

-bad thing is that isn't a great technique for bulk separation, too hard to do gram quantities.

How does it work

Place material in electric field

If NET + will move to -

If NET - will move to +

If no net charge? Won't move

How get so no net charge? Hold that thought

What will make move faster?

Larger potential (increase V)

More charge on molecule

What will make move slower

Large floppy molecule more drag

Putting something in way, gel matrix

Who has used in Cell biology or genetic, what was done, how did it work

Now lets try on proteins

If $\text{pH} < \text{pI}$, what is charge on protein? (+)

If $\text{pH} > \text{pI}$, charge is -, move to +

So can see will separate proteins based on how close or far pH is from pI

What else? Size/structure

So pretty complex for an characterization tool

See figure 3-18b for electrophoresis of proteins as fractionated

Can simplify with a couple special sub techniques

SDS gel electrophoresis

Add detergent

Sodium dodecyl sulfate

$\text{CH}_3(\text{CH}_2)_{11}\text{SO}_4^-\text{Na}^+$

Hydrophobic binds to protein

About 1 molecule SDS/2 AA

Protein gets denatured so any structure blown away

Gets uniform coating of negative charge, so migrates in simple manner. In fact migration relative to log MW

Figure 3-19 for SDS Gel

So get measure of MW and purity in 1 easy shot!

Isoelectric focusing

Remember earlier thought

If $pH < pI$ +

$pH > pI$ -

$pH = pI$ doesn't move

Make gel with pH gradient using mixture of organic acids and bases

Figure 3-20

Apply to one end of gel let move until it stops. Now at pH measure pH of gel and have pI. Now you know why we calculated pI's earlier, it a value can find experimentally

Can combine both in one Figure 3-21

This is a cutting edge technique in new field of 'proteomics' Studying changes that occur in proteins can compare host of proteins form one person to next to see how are different, and apply to clinical situation

C. Quantification of protein [Not in text](#)

In all this actually missed an important step

How do we know the protein is the one we want?

Need an ASSAY

Some chemical procedure that identifies out POI

Usually not only identifies, but is proportional to amount

With proteins that are enzymes, usually some measure of enzymatic activity

In coming up with assay need several things

Proper (optimum) pH

Proper reactants/products

Does need any prosthetic groups or cofactors

Is there an optimum temperature

Etc

Put together a set of conditions that can watch either reactant disappear, or product appear and that rate is proportional to amount of enzyme

May not know all of the above for optimum, best assay until after has been purified! -

But can't begin purification until have this assay down

Once have assay can now look at POI as purification occurs

Table 3-5

Note difference between *activity* and *specific activity*

Specific activity is activity/wt of protein. Should be at max when finished

How do you know when purification is complete?

Only 1 protein in SDS gel

Specific activity is at max

3.4 Covalent structure of Proteins

As will see in next chapter that there are four hierarchies of structure in proteins

1° structure - The connection between amino acids - Covalent Structure - Sequence and disulfides

2° structure - local structure of amino acids and nearest neighbors

3° structure - overall 3D folding

4° structure - how peptides associate together in larger polypeptide structure

This section deals with 1, will talk more about 2, 3, & 4 in next chapter (1° structure)

Now we have a protein purified now we want to know its chemical structure, or its sequence.

Let's start with a discussion of the mechanism of how this is done, and then we can talk more about the biochemical implications of the sequences

(book does reverse order so notes do not follow text)

A. Determination of Molecular weight

One of the first things we want to know is the molecular weight. Because that will tell us roughly how many AA's we have. How do you get MW?

Have seen SDS gel and gel permeation chromatography

give you MW +/- 5% so not terribly accurate

Can be further off if funky protein

New technology way is via mass spec

[Box 3-2 in text](#)

How does mass spec work

Put material in gas phase (usually in vacuum)

Put charge on material (various ways - let's not worry about)

Now can use charge to accelerate material in electric field

Once moving can detect mass by observing material

How long does it take to travel

- in a magnetic field how quickly does it turn

From these measurements can get Mass/charge ratio

And from there get mass

Methods established for organic molecules for 20 years
 In past 10 years has been extended to large molecules like proteins

B. Amino Acid Composition

Next step is usually to get overall AA composition. Used to compare proteins with each other and can be used as a check when you have the final sequence to make sure it fits with known data. (Used to be part of 3-25 but dropped)

- i. Hydrolyze protein in either strong acid or base
 (Boil in 6M HCL, 12, 24, 48hrs
 This knocks proteins into AA
 Separate AAs by chromatography and quantitate
 All has been mechanized onto machines called amino acid analyzers
 Need a few mg of protein
 Some acids destroyed in base, others in acids, so usually need both acid and base hydrolysis
 can double check results against MW to be sure you have everything

Note - this only tells you overall % composition, NOT sequence. For that you need to dig harder

C. Sequence Determination

- i. (a start) **Figure 3-25a**
 Take protein and modify with FDNB 2,4-dinitrofluorobenzene also called Sanger reagent after inventor of process **show structure on board**. This chemical spontaneously reacts with all free amines on proteins. Where would you find free amines? N terminus and lys. Do AA analysis again. Will find modified lysines because they will run a bit differently, but not a problem, will find 1 AA in the whole bunch that runs differently because now has 2,4-DNB group attached, and that is your N-terminus so have identified AA 1

- ii. (continued) **Figure 3-25b**
 Let's try to get more than one residue
 the chemical phenylisothiocyanate is another reagent that reacts with free amines

This is called an Edman Degradation after Pehr Edman, inventor
 With this reagent, if you hit the protein with 6M HCL for just a minute (not the boiling for 24 hrs required for hydrolysis)- the reagent forms a phenylthiohydantoin derivative that pulls off the first AA and leaves the rest of the protein unchanged
 run extract on HPLC, identify AA
 hit protein another time to get next AA

Again has been automated into a machine called a protein sequenator
Use a few mg of protein and read off sequence

Not quite that simple

Reaction not 100%

Say only 97%

Lets say sequence is gly-pro-lys

Reaction 1 get 97% of gly off end

Reaction 2 get 3% of gly that missed first time, and $.97(.97)$ or

$.94\%$ fo pro off

Reaction 3 get 6% of pro and now only $.94 \times .97$ or 91% of lys

Can see getting less and less of correct AA and more and more of other AA's

Usually can get to between 20 and 50 before gets so bad that can't follow any further. So can't do entire protein this way

Note machine works fast, get this in about 24 hours

D. Getting the complete sequence

So need to cut protein into smaller pieces

i. Remember that protein MAY contain a few disulfide

Need to take care of these first

Not removed by other procedure and may interfere

Can remove disulfides by 2 methods (figure 3-26)

Oxidation with performic acid

Reduction with dithiothreitol

Will air oxidize back together so need

Additional acetylation step to prevent re-oxidation

ii. Now can cut into peptides

Use with chemical methods (CNBr)

Or enzymes to cut into smaller pieces See table 3-7 for specificity

Get peptides

Sequence peptide

iii. Note: Which one goes first, second, third??

Need second cleavage method (or more)

So get overlaps and put overlaps together

Have linear sequence

Still have to go back to disulfides

Cleave it again

Isolate peptides

If get peptides tied together and have 2 N terminals then are tied by

disulfides. Put with other information to get complete sequence

Summarized Figure 3-27

E. Current techniques

I have outlined classical approach done biochemistry for past 20 years
What are new modern edge approaches?

i. DNA sequencing

Some of you have work with Dr. Sarver, know can sequence DNA
Isolate DNA that makes a protein, and sequence that instead
Works well but need DNA

Potential problems - post-translational modification
Where are disulfides?

Genome - entire DNA sequence of an organism
Proteome - corresponding set of proteins

ii. Mass Spec

Saw above how can use mass spec to get a MW

In machine called MS/MS can be even fancier

Use enzymes to make peptides, but don't bother separating
Make ion out of peptide

Use one mass spec to separate out the peptides by MW
Isolate ions of 1 MW

Inside machine make ions collide with other molecules like
He or Ar

Get fragments - typically knock off 1,2 3, etc AA's

Use second MS to get MW of all fragments

From difference from one fragment to next can tell what MW
of missing AA was, and can then get sequence

One small problem

ILE and LEU exactly same MW

Overall procedure actually more complicated, I have glossed
over details

F. Small Peptides and Proteins can be Chemically Synthesized

As chemist ultimate test is synthesize

have done analysis, think you know 1^o sequence. Best proof you are right
is to synthesize from scratch. If what you make has identical properties to
what is found in nature, then you had it right. How do you synthesize a
protein with a MW in the 1000's?

Do standard chemical synthesis as in chem lab, get about 4 residues.
Need something new and different

Bruce Merrifield invented way to link AA to insoluble resin and cycle reactions to make efficient, simple and mechanized. Got Nobel prize. What I will show you is not his original chemistry, but a modern evolution using his principles

Figure 3-29

Works well. Synthesizers cost 10-50K

One problem, reaction not 100%

So slowly errors accumulate

See table 3-8

Practical limit about 100 AA if everything is optimal, but usually lower

Need to do solution chemistry to link peptides together if you want a large protein

Compare with biology

On synthesizer can prepare 100AA protein in a few days and it is about 80% pure

E coli, does in 5 seconds, 100% correct

G. Amino Acid Sequences Provide Important Biochemical Information

Now have protein 1° structure. What good is it?

Can provide insight into

3-D structure/function

Cellular location

Evolution

Structures of 1000's of proteins available on web

sequences many more sequence available

Search sequence for similarities and relationships

BIOINFORMATICS

Structure/function

Have been trying to predict structure from sequence for many years- still can't do it

But if part of a family, (usually 25% sequence homology)

Can look at family's structure and function

And make a good guess

Sometime not entire protein, but a 'domain' will work

Cellular location

Will see later- certain sequences determine

Cellular location

Sites for chemical modification

Lifetime in cell

Evolution

Look at how sequence changes as a function of evolution

Can trace family trees and phylogeny