

CMB311 Fall 2017

Lecture 4 Notes

Protein Structure I: amino acids and peptides

In this lecture we discussed the structure and properties of amino acids, which are the building blocks for proteins.

Questions:

1. What are amino acids, and how do their physical and chemical properties determine their behavior?
2. How might the physico-chemical properties of amino acids dictate protein 3D structure?
3. How are amino acids linked together to form proteins?

Lecture Outline

- I. Amino acid structures, nomenclature, chemical and physical properties
- II. Interactions among amino acids in proteins
- III. Structure of the peptide bond and polypeptide conformations

Background

Just as a reminder, proteins support cell structure, control gene expression, transport materials into and out of the cell, and serve as catalysts (enzymes) that carry out metabolism, among other things. Proteins vary tremendously in molecular weight, conformation, dynamics, etc. Keep in mind that a protein's three-dimensional structure is dictated by the physico-chemical properties of its amino acids and their sequence in the protein. *This is a key concept!*

I. Amino acid structures

General structures of amino acids

Definition: An amino acid (more formally known as an α -amino acid) is a carboxylic acid with an amino group attached to the α -carbon. Remember the common nomenclature for carboxylic acids, in which the first carbon after the COOH (carboxyl) group is the α -carbon. In IUPAC nomenclature, the carbon in COOH is number 1, and the next carbon (the α -carbon) is number 2. In the common nomenclature, subsequent carbons are designated β , γ , δ , ϵ , and so on. Remember, the α -amino group refers to its position relative to the carboxyl group. There is no such thing as an α -carboxyl group in amino acids. Just so you know.

Anyway, the α -carbon is a chiral center due to the attached 'R' group. The amino acid is asymmetric, forming bonds with four different functional groups (there is one exception to this rule as we shall see). There are therefore two stereoisomers, L and D, that are not mirror images of one another. They cannot be superimposed, just as your hands cannot be superimposed. Biology uses exclusively the L stereoisomer for protein synthesis. As a standard, the absolute configuration of amino acids is compared to that of the compound L-glyceraldehyde.

There are 20 different amino acids found in proteins (not including other odd balls that appear occasionally, and do not concern us here). These are distinguished by the structures of their R groups. Each amino acid is given a three-letter abbreviation and a one-letter abbreviation. Both of these are important to know.

Amino acids can be grouped based on their chemical properties. Traditionally, these are:

- Non-polar
 - o Glycine, Alanine, Valine, Leucine, Isoleucine, Methionine, Tryptophan, Phenylalanine and Proline
- Polar
 - o Serine, Threonine, Cysteine, Tyrosine, Asparagine, Glutamine
- Acidic
 - o Aspartic Acid, Glutamic Acid
- Basic
 - o Lysine, Arginine, Histidine

Sometimes you will also see a variation on this classification, with these as a group:

- Aromatic
 - o Tryptophan, Phenylalanine, Tyrosine, Histidine

A note of caution regarding Tyrosine. Some textbooks classify Tyrosine among hydrophobic amino acids while others lump it with the polar ones. Still others sidestep the whole issue and create a separate group of aromatic amino acids. Partly this is because hydrophobicity is not a simple binary characteristic. There is not even a single agreed upon scale of measurement of amino acid hydrophobicity. Tyrosine is certainly more water-soluble than Phenylalanine, a close structural analog. Ultimately, what really matters for us is that Tyrosine is found both in the interior and on the surface of proteins. Hydrogen bonding via the -OH group can compensate for the hydrophobicity of the aromatic ring, and the -OH group does not keep it out of protein interiors.

Also regarding the aromatic amino acids, they absorb UV light, which facilitates measurement of protein concentration by spectrophotometry. Tryptophan absorbs the best, at 280 nm.

It may seem tedious at first to have to memorize the structures of all 20 amino acids, but a number are very similar in structure. For instance,

- Val, Leu, Ile are all extremely similar and have similar properties.
- Tyr is Phe with a hydroxyl group on the aromatic ring.
- Asp and Glu differ by a single CH₂ group in the hydrocarbon chain
- Asn and Gln differ by a single CH₂ group in the hydrocarbon chain
- Asn and Gln differ from Asp and Glu, respectively, by the addition of an amino group

Three amino acids are somewhat different from the others.

- Glycine: The R group is simply a hydrogen atom. This means that the α -carbon of Glycine is not asymmetric, so there is no L or D stereoisomerism.
- Proline: Technically, Proline is not an *amino* acid but an *imino* acid. This is because the R group forms a cyclical structure via a C-N bond with the amino group (imino group). This structure constrains the backbone configuration of polypeptides.
- Cysteine: The -SH (sulfhydryl) group of the R group of Cys often forms disulfide bonds with another Cys residue. This is an important structural feature of some proteins, for instance in ribonuclease A (RNase A), which contains four disulfide bonds.

Aromatic amino acids such as Phe, Tyr, and Trp all absorb light in the UV range. Phe absorbs the weakest, and Trp absorbs the strongest. The concentration of a protein in solution can be detected by its absorbance at 280 nm.

Recall the basis for the hydrophobic effect we discussed in the previous lecture. The hydrophobic effect is important because:

- It is the major driving force behind protein folding.
- Hydrophobic amino acids are strongly favored to occur in the interior of proteins.
- Changes in the degree of packing of hydrophobic amino acid side chains inside proteins can greatly affect protein thermal stability.

Study tips:

- Draw the structures on index cards with a sharpie
- Group them based on their chemical properties
- Group them based on their structural similarities
- Note that the two types of groups will not necessarily give you the same answers
- Make a Venn diagram of the two groups
- These will become more familiar as we talk about protein structure

A note on the single letter nomenclature. This is important to learn, even though some of the designations are confusing or seem to make no sense. The single letter nomenclature is used in all bioinformatics work, e.g. protein sequence alignments. You can see why this is if you write the amino acid sequence of a protein. For instance:

Met-Ala-Arg-Arg-Tyr-Leu-Val-Val-Ala-Asn-Gln-Lys-Val-Val-Ile-Cys-Pro-His

MARRYLVVANQKVVICPH

Or, for an actual protein sequence...

```
/db_xref="GeneID:953383"  
/translation="MSLP I R N V A I I A H V D H G K T T L V D A L L K Q S G I F R E G E D V P V C V M D  
S N D L E R E R G I T I L S K N T A V R Y Q D T L I N I V D T P G H A D F G G E V E R V L G M V D G C V L I V D A N  
E G P M P Q T R F V L K K A L E K G L R P L V V V N K I D R P R A D P N T A V D K V F D L F V E L G A D D D Q C D F  
T T L F A S G L G G F A K E S L D D D S E D M K P L F E A I L H H V P P P A G D P N K P L Q L Q V T T L D Y S D Y L  
G R I I I G R I H N G T V K A G Q Q A A L V K E D G S I A K G K V S K L L G F E G L N R I E L P E A S A G Y I V A I  
A G F A D A N I G E T L T C P D E P Q A L P L I K V D E P T L Q M T F S V N D S P F A G Q E G K F V T S R Q I R D R  
L N R E L E T N V A L R V E D G E S A E Q F L V S G R G E L H L G I L I E T M R R E G Y E F Q V A O P Q V I Y R E V  
N G Q P C E P V E Y L V L D V P E A A V G A C I E R L G Q R R G E M Q D M Q T S V N G R T Q L E F V I P A R G L L G  
F R G D F I R I T R G E G I M N H S F L E Y R P M S G D L E T R Y N G V M V A F E E G V A T F Y A M K N A E D R G V  
F F I T P G T K V Y K G M I I G E H N R P Q D I E L N V C K T K Q L T N H R S A T G D E L V Q L Q A P E D M N L E R  
A L E Y I G P D E L V E I T P E S I R L R K V A R K K L V K R "
```

The single letter code is much more compact and suitable for computing purposes.

Most of the amino acids are represented by the first letter of their name:

Ala = A	Met = M
Cys = C	Pro = P
Gly = G	Ser = S
His = H	Thr = T
Ile = I	Val = V
Leu = L	