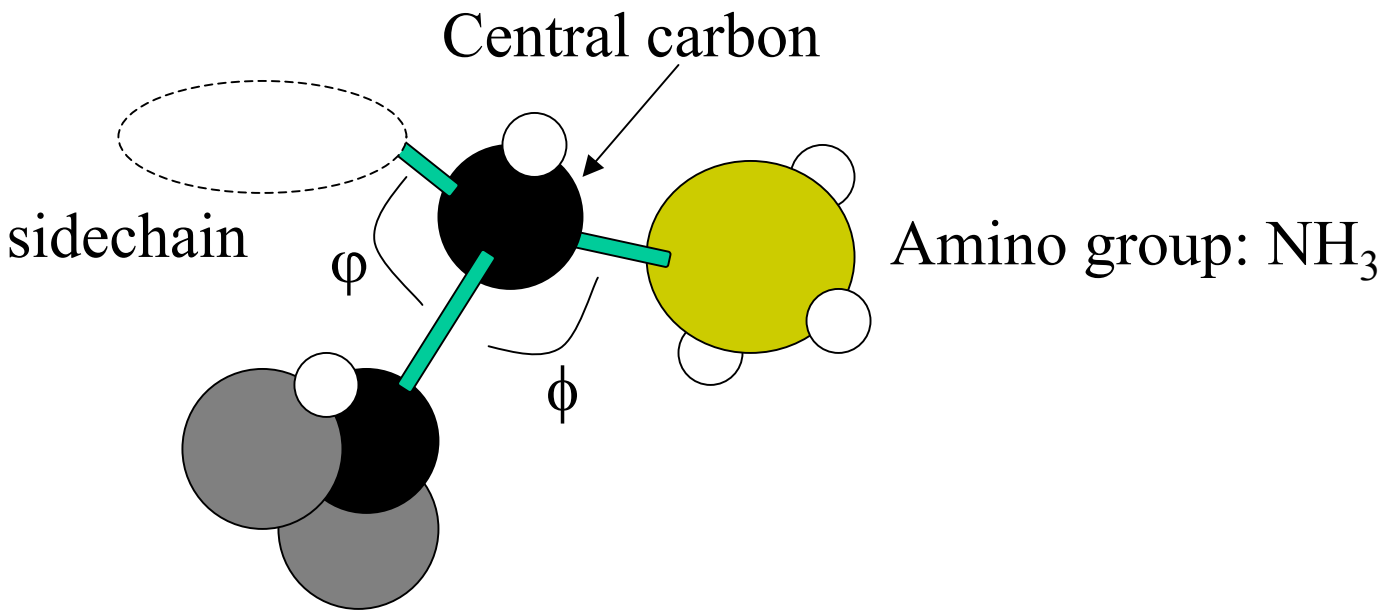
A Database of Proteins

- Proteins are long chains of Amino Acids (AA). There are 20 types of AA that serve as building blocks for proteins.
- Each AA has a specific chemical structure. The length of a protein chain can range from 50 to 1000-3000 AA (200 on the average).
- One of the interesting properties of proteins is the unique folding.

From Haim Wolfson's lecture on Molecular Biology, Jan-2001 and Lawrence Hunter's "Molecular Biology for Computer Scientists"

Basic Chemical Structure of an Amino Acid



Amino Acid Composition of Proteins

- The AA composition of a protein will usually uniquely determine (on specific environment conditions) the 3D structure of the protein (e.g., two proteins with the same AA sequence will have the same 3D structure in natural conditions).
- Researches of 3D structure of proteins have shown that when a folded protein is artificially stretched to a chain, it folds back to it's original 3D structure.

Proteins Functions

- Proteins are known to have many important functions in the cell, such as enzymatic activity, storage and transport of material, signal transduction, antibodies and more. All proteins whose structure is known are stored in the Protein DataBank (PDB) which contains about 10,000 proteins [http://www.rcsb.org/pdb/index.html].

Levels of Structure

A protein has multiple levels of structure:

- Primary structure - Chain of Amino Acids (1 dimensional)
- Secondary structure - Chains of structural elements, most important of which are α-helices and β-sheets. β-sheets can be parallel or antiparallel.

Levels of Structure (II)

- Tertiary and Quaternary structure - 3D structure, of a single AA chain or several chains, respectively.
  - Tertiary: some structures, more packed are 3-10 helices while some small structures that link others are β-turns.
  - Quaternary: some proteins assemble with other molecules; others bind to copies of themselves (dimers); others require prostehtic groups (e.g., heme or chlorophyl).

Chemical Properties of Individual Amino Acids for Protein Folding

- *Glycine*
  - it is the simplest. Its sidechain is a single hydrogen atom. It is nonpolar and does not ionize. It is very flexible.
- *Alanine*
  - Is also small and simple. Its sidechain is just a methyl group. It is very common.
- Glycine and Alanine are *Aliphatic* (are straight chains) like *Valine*, *Leucine* and *Isoleucine*. Long aliphatic sidechains are hydrophobic (this determines how the chain will fold-up into an active protein).

- *Phenylalanine, tyrosine, tryptophan*:
  - Are quite large. These sidechains are aromatic. Phenylalanine and tryptophan are also hydrophobic, while tyrosine, like *serine* and *threonine* contain hydroxyl groups (and thus are more reactive).
- Arginine, lysine:
  - Ionize in basic environment
- *Histidine, gluatmic acid, aspartic acid*:
  - Ionize in acidic environment
- *Cysteine, methionine:*
  - Have hydrophobic chains that contain a sulphur atom which makes them reactive.
  - Methionine is the starting amino acid in all eucaryotic proteins.
- *Histidine:*
  - It is rare but it often appears in the active site of an anzyme.