BCHM2071 – Protein Biochemistry

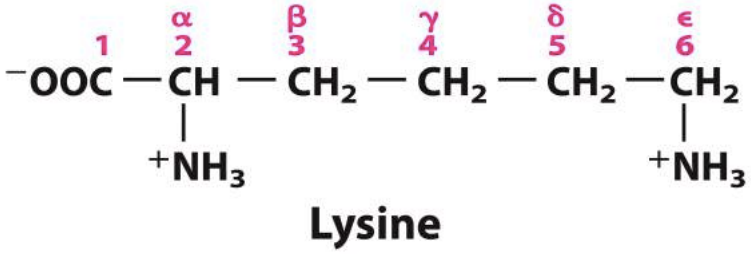
Proteins

Classifications of Proteins

- Functions:
  - Catalysis: Enzymes increase rate of rxn e.g. DNA polymerase in replication
  - Transport other molecules e.g. haemoglobin transporting O2
  - Structure e.g. collagen in connective tissue, keratin in hair
  - Motion of cells e.g. myosin and actin in muscle tissue, actin for cell motility
- Shapes: Fibrous, globular
- Compositions: Simple (only AA residues), conjugated (have prosthetic group i.e. non-AA part)
  - Lipoproteins: Lipids
  - Glycoproteins: Sugar
  - Metalloproteins: Metal

Amino Acids

- Proteins made up of alpha AA (20-23 types) with diff functional groups
- Diff combinations of AA → diff function
- α carbon usually has 4 diff substituents → chiral center w/ L and D isomers, tetrahedral structure
  - Acidic carboxyl COO- group
  - Basic amino NH3+ group
    - Proline does not have NH3+ as in ring structure
  - α H bound to α C
  - R group unique to AA
    - H for glycine → non-chiral
- Naming:



  - Unnumbered 3 p78
    Lehninger Principles of Biochemistry, Sixth Edition
    © 2013 W. H. Freeman and Company
  - Organic: Start from C with higher oxidation number
  - Biochemical: Start from α C and go down R group
- L and D isomers:

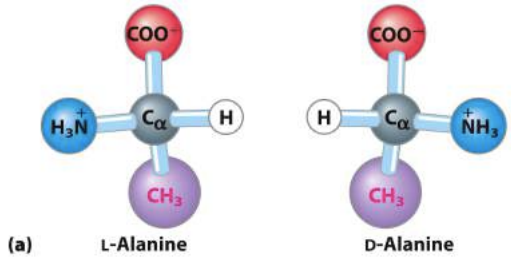
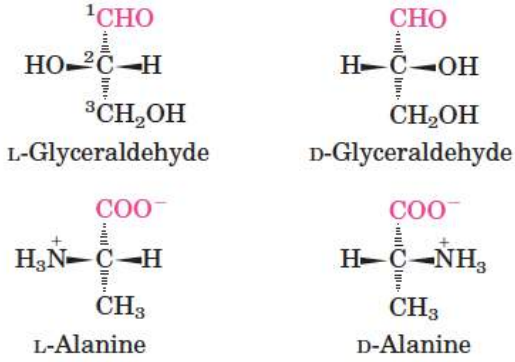
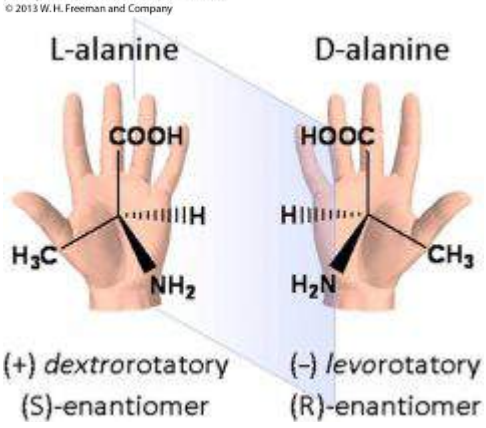
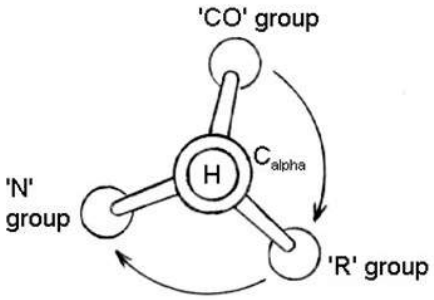


Figure 3-3
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company







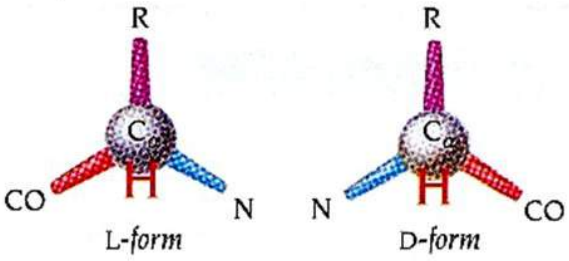


Figure 1.3 The "handedness" of amino acids. Looking down the H–C_α bond from the hydrogen atom, the L-form has CO, R, and N substituents from C_α going in a clockwise direction. There is a mnemonic to remember this; for the L-form the groups read CORN in clockwise direction.

- Named respective to structure of L-glyceraldehyde by aligning C atoms
  - L isomers are S enantiomers (except cysteine)
- L isomers have α amino group on left; D isomers have α amino group on right
  - Can also identify in absolute configuration using CORN rule; anticlockwise = L isomer
- Proteins only contain L isomers

- D isomers → mirror image of original protein, or may not form structure

- Generation of AA:
  - Cells can generate specifically L isomers as enzymes are asymmetric
  - Not all AA can be generated by body (essential AA)
- Classification of AA:
  - Non-polar/aliphatic

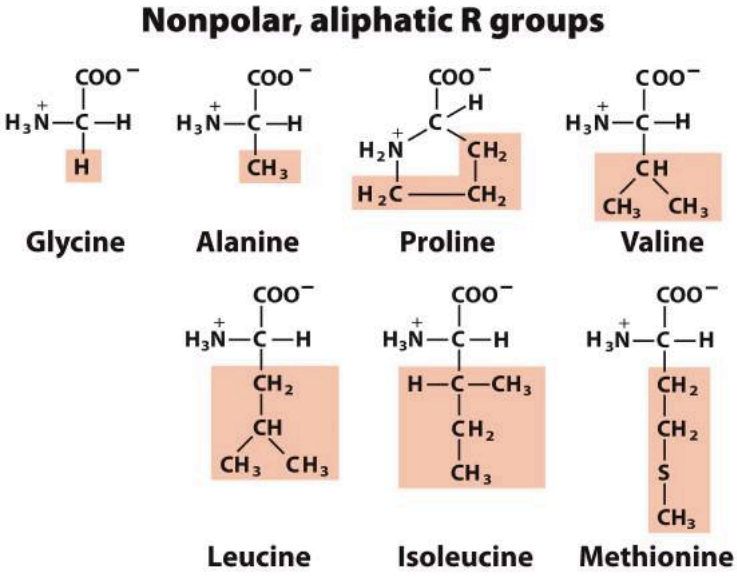


Figure 3-5 part 1
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

    - Hydrophobic interactions – cluster together
    - Hydrophobic side chains buried inside protein if protein is water-soluble e.g. in cytosol
      - But better to expose hydrophobic rather than bury charged
    - Exposed in proteins in lipid membrane
    - Methionine is starting AA for all peptides
  - Aromatic

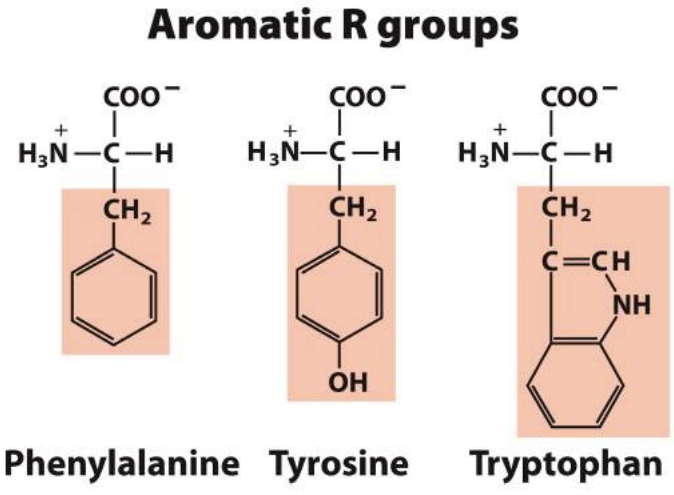
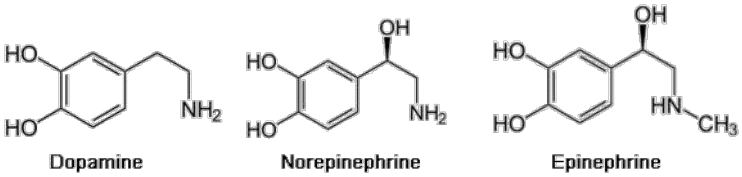


Figure 3-5 part 2
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

    - Contain aromatic ring which absorb UV light at 270-280nm → can detect conc. using spectrophotometry
    - Mostly non-polar → hydrophobic interactions

- Tyrosine and tryptophan are more polar because of OH and NH respectively → H bonding
  - Tyrosine is precursor to catecholamine NT e.g. dopamine, noradrenaline, adrenaline
    - Dopamine is NT which controls BP, blood flow; involved in schizophrenia, aggression, reward-seeking behavior, motor function in basal ganglia; deficit → ADD
    - NA and A involved in SNS



    - Hydroxylation of β C in dopamine → noradrenalin
    - Methylation (CH3) of amine (NH2) group in noradrenaline → adrenaline
- Polar, uncharged

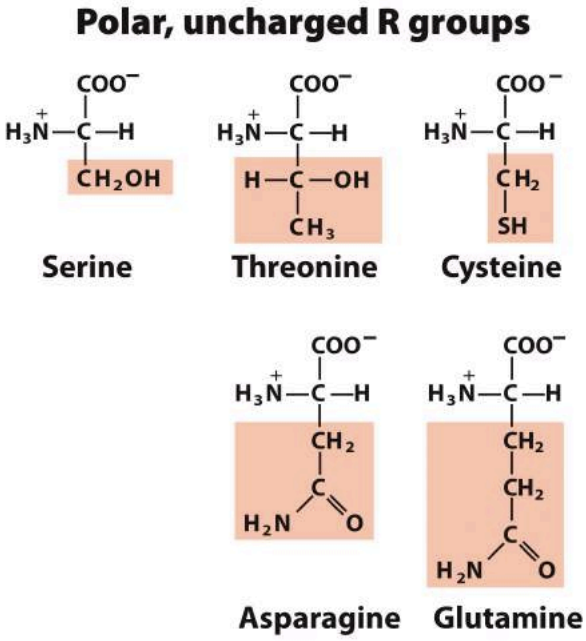


Figure 3-5 part 3
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

  - Hydrophilic: Side chains can form H bonds (H bonded to N, O, F, and S atoms)
  - Cysteine has thiol group → can form disulfide S-S bonds (non-polar, covalent)
- +vely charged: +ve charged group at pH 7 (NH3+/NH2+)

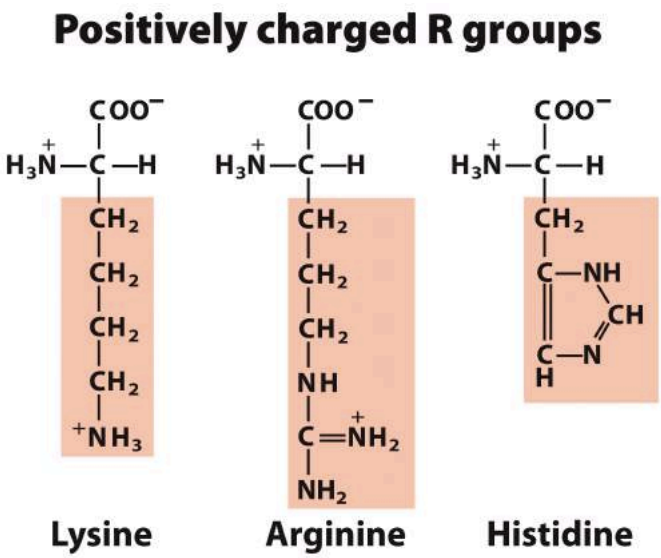


Figure 3-5 part 4
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

- –vely charged: -ve charged COO- group at pH 7

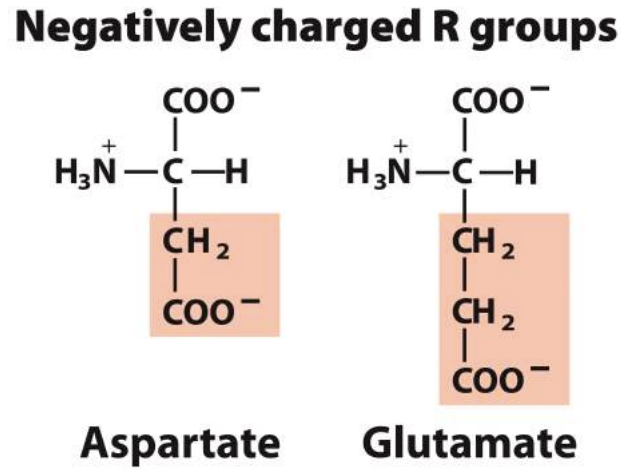
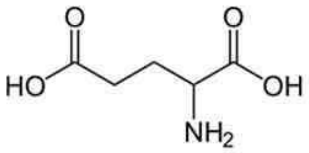
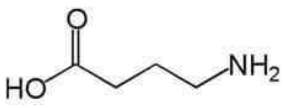


Figure 3-5 part 5
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

- Amidation of R group w/ NH3 → asparagine + glutamine (polar AA w/ slight +ve charge)
- Formation of ketone: Remove α NH3+ group, replace with ketone C=O → α ketoglutarate (from glutamate) + oxaloacetate (from aspartate)
- Decarboxylation: Removal of α COOH group from glutamate → GABA (inhibitory NT)



Glutamic acid, also called glutamate, an excitatory neurotransmitter



Glutamate's metabolic product, gamma-aminobutyric acid (GABA), an inhibitory neurotransmitter.

TABLE 3–1 Properties and Conventions Associated with the Common Amino Acids Found in Proteins

Amino acid	Abbreviation/ symbol	M_r*	pK_a values: pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index†	Occurrence in proteins (%)‡
Nonpolar, aliphatic R groups								
Glycine	Gly G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala A	89	2.34	9.69		6.01	1.8	7.8
Proline	Pro P	115	1.99	10.96		6.48	−1.6	5.2
Valine	Val V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups								
Phenylalanine	Phe F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups								
Serine	Ser S	105	2.21	9.15		5.68	−0.8	6.8
Threonine	Thr T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine¶	Cys C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups								
Lysine	Lys K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups								
Aspartate	Asp D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu E	147	2.19	9.67	4.25	3.22	−3.5	6.3

pKa and pI

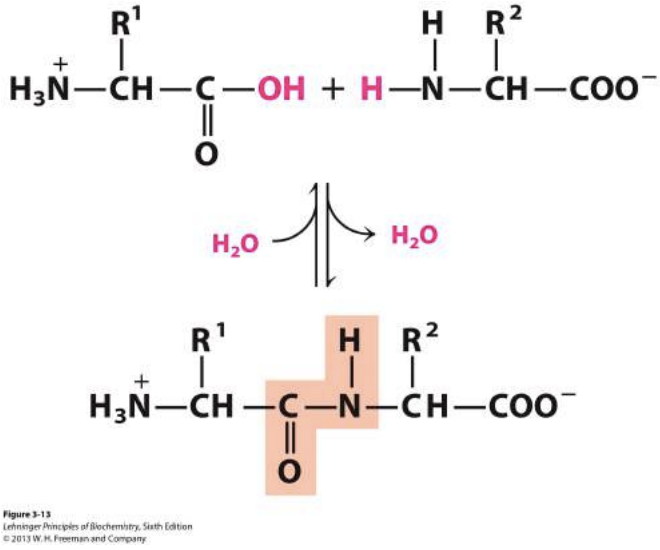
- pKa = pH at which there is equilibrium of protonated and deprotonated forms
- pI = isoelectric point; pH where the AA/protein has neutral charge
    - Avg of pKa's which border the neutral form
    - If pH < pI; more AA are protonated → +ve
    - If pH > pI; more AA deprotonated → -ve
- pKa's of groups (α-carboxyl/α-amino/R groups) depend on surrounding chemical environment
    - Intramolecular in same AA/protein
        - → pKa's of R groups can change once protein formed
    - External e.g. with enzyme environment

Post-translational Modification of Proteins

- Not incorporated by ribosomes (except selenocysteine and pyrrolysine)
- E.g. GABA, adrenaline, serotonin act as NT and hormones
    - Selenocysteine: Replaces S in cysteine w/ Se → higher electronegativity → greater H bonding
- Phosphorylation: Occurs at OH group in residue

    - Important in regulation and signaling
- Glycation: Adding sugars e.g. glucose, fructose
    - Acts as tag for internally produced proteins rather than infection

Protein Synthesis

- Central dogma: DNA → mRNA → peptide
- Transcription: DNA → mRNA
- Translation: tRNA reads mRNA and adds AA coded by each codon (3 bases)
    - AUG (start codon) which codes for methionine → UAA, UAG, UGA (stop codon)
    - Usually more than one codon codes for each AA
    - Condensation reactions of AA → peptide bond
    - 

Naming Proteins

- Starts from amino end (N terminal) → carboxyl end (C terminal)
- Peptides = ≤ 50 AA, < 10 kDa
- Proteins = > 50 AA, > 10 kDa

Peptide Sequencing

1. Separate chains:
    a. Break weak subunit interactions (quaternary structure) using: extreme pH, 6M guanidine HCl, high salt conc.

b. Cleave disulfide covalent bonds:

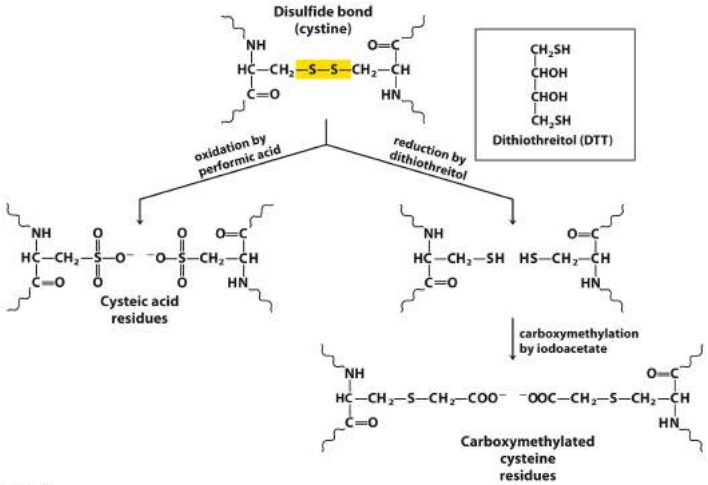


Figure 3-28
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

i. Performic acid oxidation → cysteic acid residues (SO3- ends) which don't react
ii. Reduction with dithiothreitol (DTT) → active thiol groups (-SH) which can react
1. Need to use iodoacetate → carboxymethylated (CH3-COO- ends) cysteine residues

2. Find AA composition of peptide using:
a. Acid hydrolysis: 6M HCl, 12-36h, 100-110C
b. Separate AA via cation exchange chromatography
i. Increase pH gradually → AA's become –ve and eluted
ii. Elution = washing out adsorbed ions w/ pump
iii. React with ninhydrin → fluorescence for spectrophotometry (measuring absorbance of wavelength dependent on concentration)
iv. Obtain elution profile:

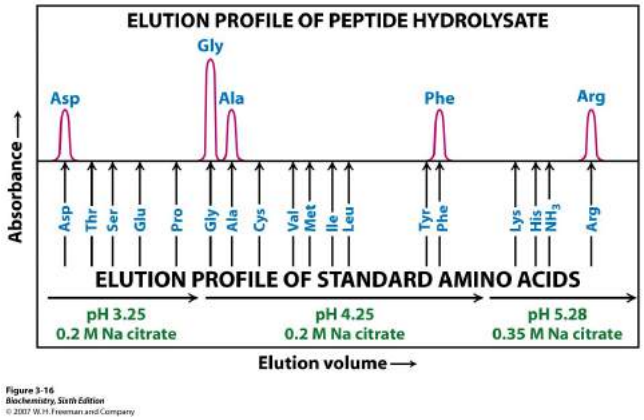
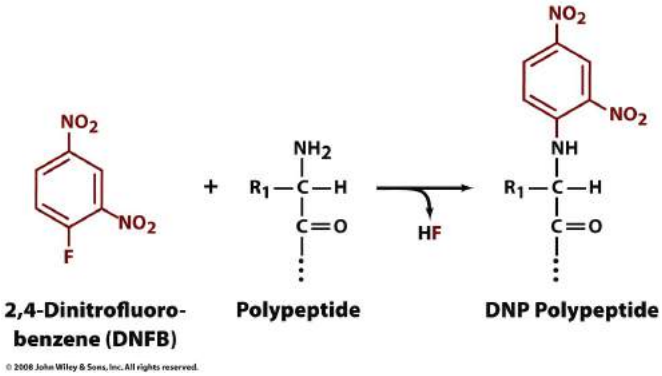


Figure 3-16
Biochemistry, Sixth Edition
© 2007 W.H. Freeman and Company

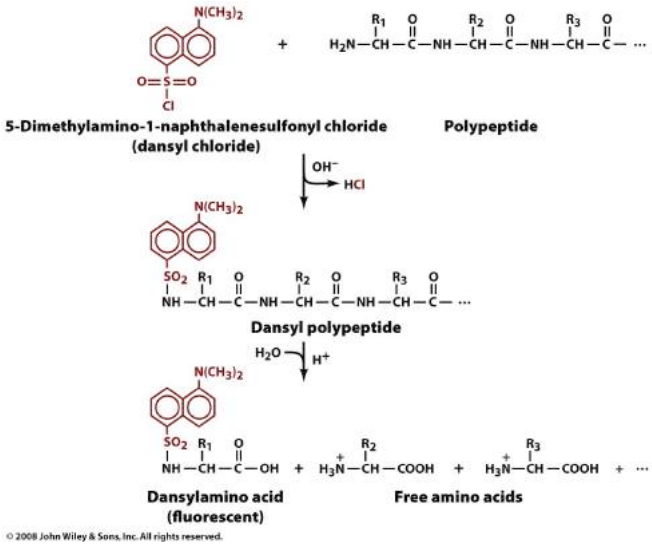
v. Determine peptide composition by analyzing peaks

3. Identify N and C terminal residues
a. N-terminal analysis
i. Dinitrofluorobenzene (DNFB): Sanger reagent



1. Reacts with polypeptide at N terminal → DNP polypeptide
2. Hydrolyse with HCl to obtain DNP-AA (bright yellow) + degrades rest of polypeptide into free AA
3. Run on TLC (thin layer chromatography) using standards to identify N-terminal AA

ii. Dansyl chloride



1. Reacts with primary amines – α-amine and NH2 on lysine → confusable
2. Hydrolyse dansyl polypeptide with HCl → dansylamino acid (fluorescent) + rest of peptide degraded
3. Run on TLC using standards

iii. *DNFB and dansyl chloride cannot sequence beyond N-terminal residue as rest of polypeptide degraded

iv. **Helps to identify number of distinct polypeptides e.g. insulin has two distinct polypeptide chains

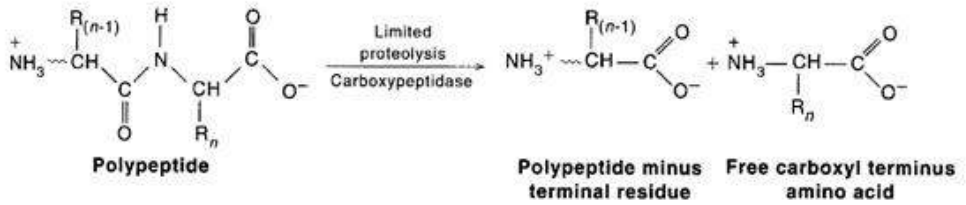
b. C-terminal analysis

i. Hydrazine reaction

1. React with hydrazine in mild acid → reacts at C=O of each peptide bond → hydrazides + free C-terminal AA w/o hydrazine
2. Analyse with TLC w/ standards

ii. Carboxypeptidase: Cleaves C-terminal AA residue

1. Analyse w/ TLC w/ standards



4. Sequence polypeptide using Edman degradation

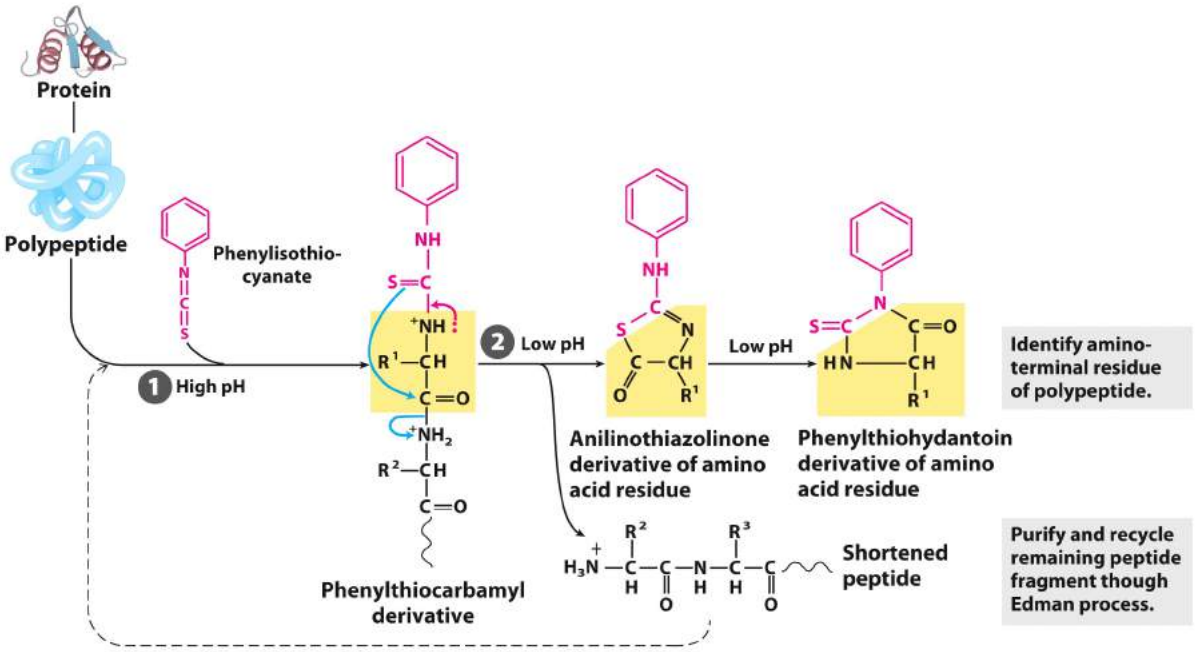


Figure 3-27
Lehninger Principles of Biochemistry, Sixth Edition

   a. Phenylisothiocyanate at high pH reacts at N terminal
   b. Low pH → rearrangement and degradation → derivative of N terminal residue + shortened polypeptide
   c. Isolate derivative of AA residue, run TLC to identify
   d. Recycle remaining peptide sequence through Edman's process
   e. *But cannot keep repeating with long polypeptides as inaccurate yield → noisy results

Limitations of Sequencing

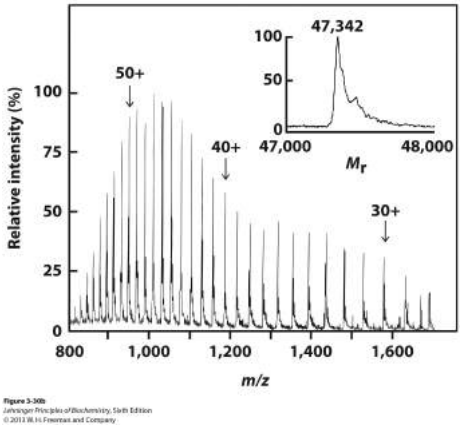
- Polypeptides with 40-100 AA cannot be directly sequenced due to inaccurate yield
- → Need to use enzymes to cleave peptide at particular sites, then sequence shorter polypeptides
- Can determine order by cleaving with different enzymes at diff sites and observing overlaps
    - Can determine location of disulfide bonds by cleaving and sequencing without having first broken disulfide bonds (longer peptide rather than two shorter upon cleavage)

Current Methods of Sequencing

- Using DNA sequence

- Mass spectrometry e.g. MALDI MS, ESI MS to identify mass of protein to determine sequence
    - Ionization of proteins → gas phase
    - Desorbed into vacuum with electric/magnetic field
    - Analysis of path to determine m/z ratio
    - Analyse output and peaks to determine mass of proteins → protein + sequence
    - *Can also identify post-translational modifications via variances in mass found

    - MALDI MS: Ionization with laser
    - ESI MS: Electrospray ionization – Solution passed through ionizing needle



- Mass spectrometry w/ tandem MS/ MS MS:
    - Cleaving polypeptides into smaller fragments e.g. via trypsin
    - Sequencing each smaller sequence

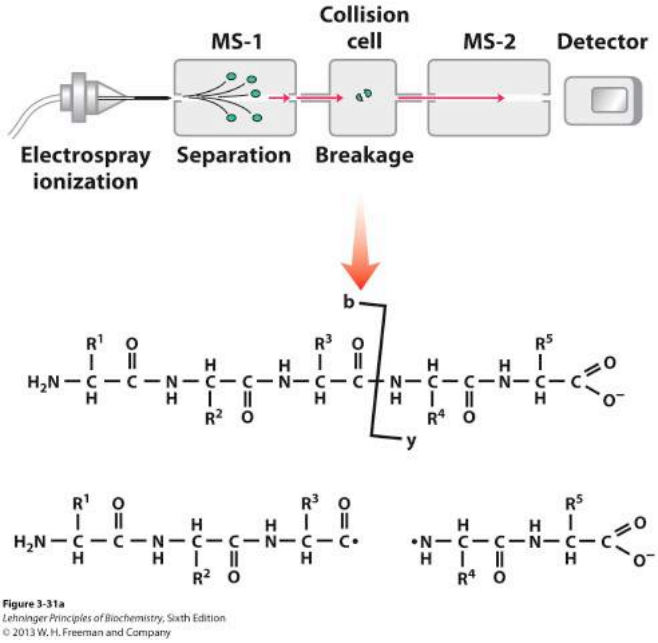


Figure 3-31a
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

1. Electrospray ionization → ionization of peptide molecules
2. MS-1: Separation of peptides, one peptide sequence passes through
3. Collision cell: Fragmentation of peptides (many of same) w/ collision inert gas
    a. Fragmentation at one place @ each peptide → two fragments from each peptide, one with charge

        i. Usually at peptide bond
    b. → Charged fragments of diff length from each peptide (b series w/ charge on N terminus, y series w/ charge on C terminus)
4. MS-2: Measures m/z ratio of charged fragments → many sets of peaks from breaks at diff sites
    a. Peaks differ by one AA → can determine AA by mass → Can determine sequence
        i. Although leucine + isoleucine have same mass → ambiguous

- *No longer use Edman degradation much even w/ sequenator machines as takes a long time

Benefits of Analysing Protein Sequences

- Can infer function from AA sequence
    - Families: Similar AA sequences may have similar function and/or structure
        - Although very diff sequences may have similar function due to similarity at crucial site
    - Special sequences are signals for location, modification, etc
- Compare homologous protein sequences across diff species → evolution
    - May have **deletions/insertions** of sequences
    - Differences reflect **evolutionary divergences**, different functions or uses at different situations
    - Mutations due to natural selection → specialization
    - Generate **evolutionary trees**
- Analysing variants (polymorphic) and **which residues most crucial** (invariant) → reveals structure and function e.g. cysteine and threonine are substitutable, just need OH group
- Analysing which residues are most commonly found shows which are most important for function e.g. lots of +ve R groups → associate w DNA
    - Changes in core regions → changing function
    - Changes in outside regions → change in what other proteins are interacted with