BIOC4004 - Industrial Biochemistry
Lectures 4 & 5 - Fri Jan 16, 04

Topics for the Day:

- Methods in Recombinant DNA technology (conclusion)
  - The Polymerase Chain Reaction (PCR)
    - Novel applications of PCR
  - DNA sequencing
    - manual vs automated
- Protein Structure
  - amino acids and their properties
  - primary, secondary, tertiary, quaternary
  - general concepts
- Protein engineering
  - what we can learn from mother nature
  - mutagenesis
  - “synthetic genes”

Detection of specific macromolecules by complementarity:

1) DNA-DNA. A single-stranded DNA (ssDNA) probe molecule can form a double-stranded, base-paired hybrid with a ssDNA target if the probe sequence is the reverse complement of the target sequence.

2) DNA-RNA. A single-stranded DNA (ssDNA) probe molecule can form a double-stranded, base-paired hybrid with an RNA (RNA is usually a single-strand) target if the probe sequence is the reverse complement of the target sequence.

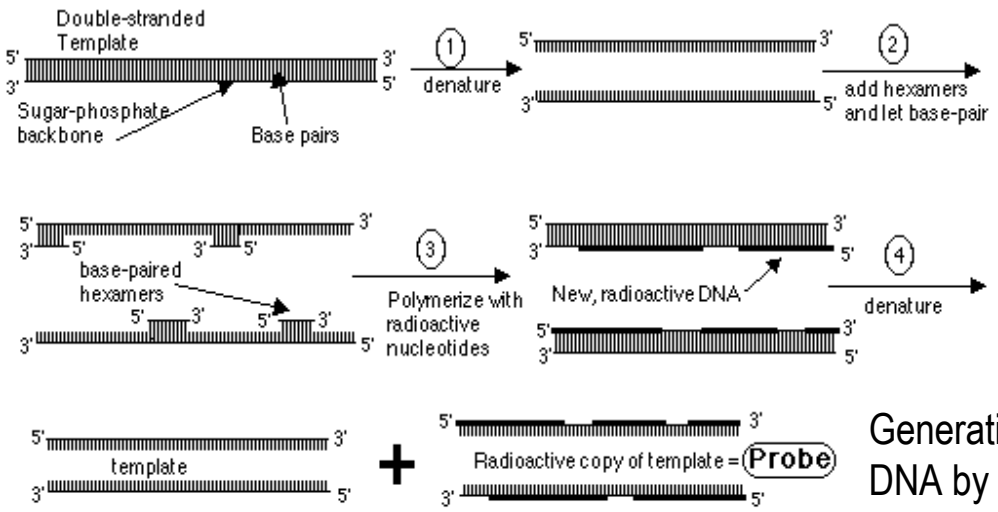
3) Protein-Protein. An antibody probe molecule (antibodies are proteins) can form a complex with a target protein molecule if the antibody's antigen-binding site can bind to an epitope (small antigenic region) on the target protein.

The two most important features of hybridization:

1) Hybridization reactions are highly specific

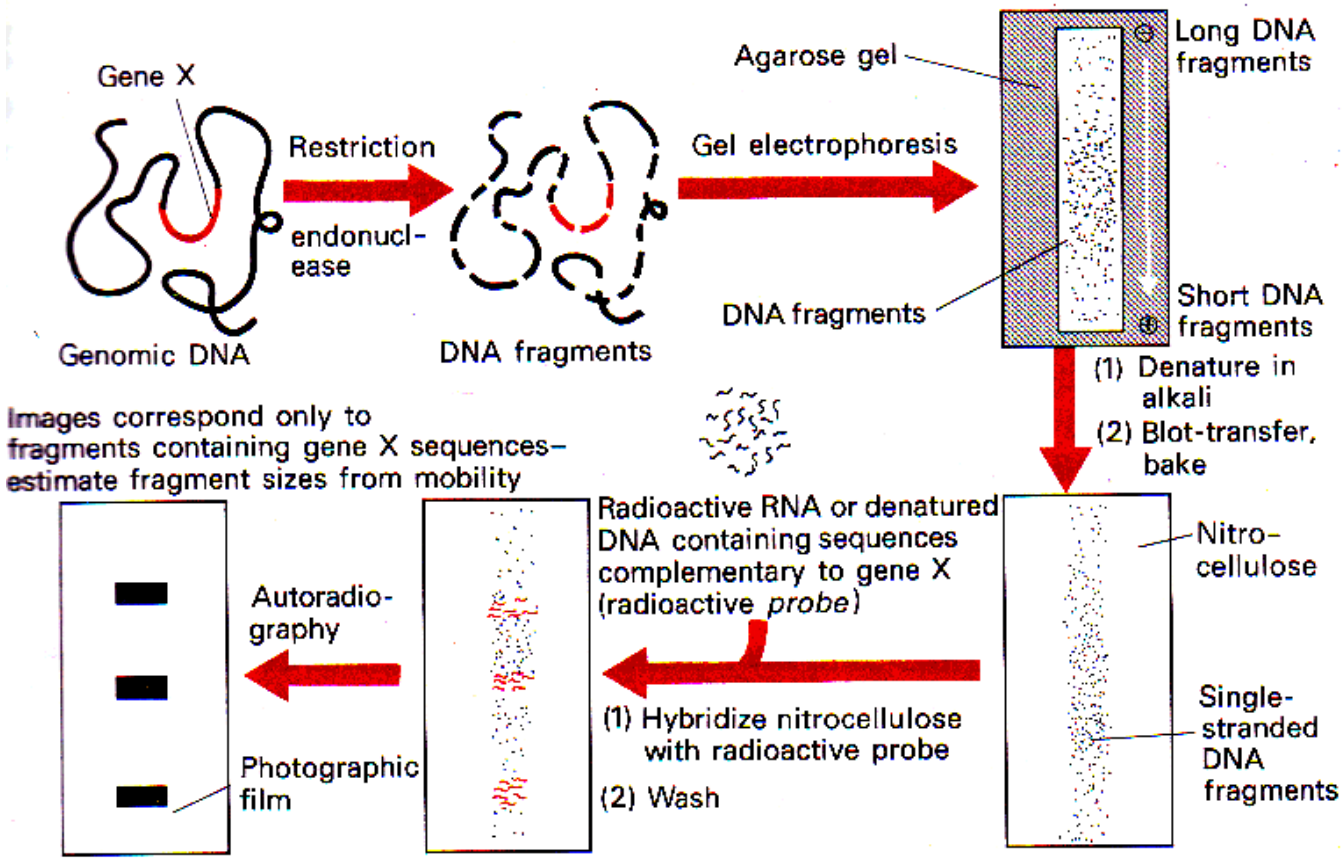
2) Hybridization reactions will occur in the presence of large quantities of molecules that are similar but not identical to the target.

- Cells contain thousands of genes, mRNAs, proteins
- Hybridization can be use to detect specific macromolecules in a complex mixture of other macromolecules

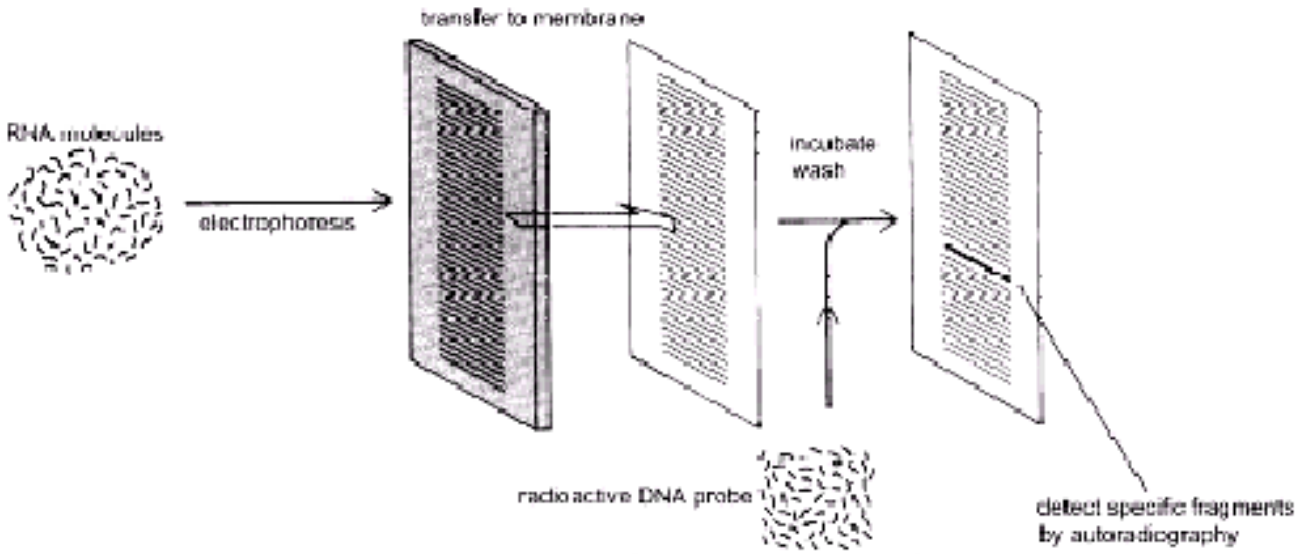


Generating radioactively labeled DNA by random priming

Southern Blotting



Northern Blotting



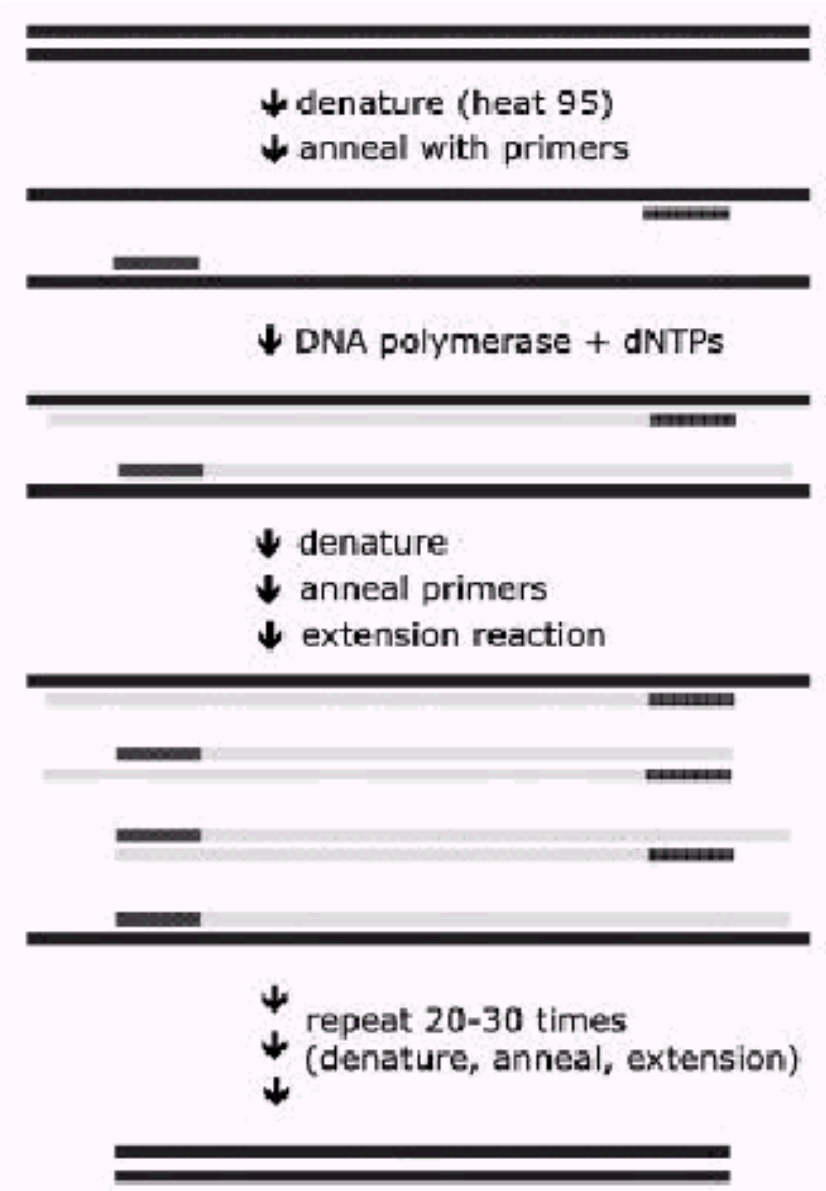
Southerns vs. Northerns vs. Westerns

Southern Blot	Northern Blot	Western Blot
1) Extract DNA from cells	1) Extract RNA from cells	1) Extract protein from cells
2) Cut with restriction enzyme		
	2) Denature with formaldehyde	2) Denature with SDS
3) Run on gel (usually agarose)	3) Run on gel (usually agarose)	3) Run on gel (usually polyacrylamide — called "SDSPAGE")
3.5) Denature DNA with alkali		
4) Transfer to nitrocellulose (usually by capillary action)	4) Transfer to nitrocellulose (usually by capillary action)	4) Transfer to nitrocellulose (usually by electrophoresis)
5) Block with excess DNA	5) Block with excess RNA	5) Block with excess protein
6) Hybridize with labeled DNA probe	6) Hybridize with labeled DNA probe	6) Hybridize with labeled antibody probe
7) Wash off unbound probe (use controlled stringency)	7) Wash off unbound probe (use controlled stringency)	7) Wash off unbound probe
8) Autoradiograph	8) Autoradiograph	8) Autoradiograph or develop with chromogenic substrate

	Southern	Northern	Western
What is separated by molecular weight? (target)	DNA cut with restriction enzymes	RNA denatured with formaldehyde	Protein denatured with SDS
Probe	radioactive gene X DNA	radioactive gene X DNA	Antibody against protein X, labeled with radioactivity or enzyme
What do you learn from it?	Restriction map of gene X in chromosome	-how much gene X mRNA is present? -how long is gene X mRNA?	-how much protein X is present? -how big is protein X?

The Polymerase Chain Reaction

- enzymatic amplification of DNA sequences
- need oligonucleotide primers bracketing the target region of interest
- Thermostable DNA polymerase; dNTPs
- Thermal cycling: denaturing; annealing; extension



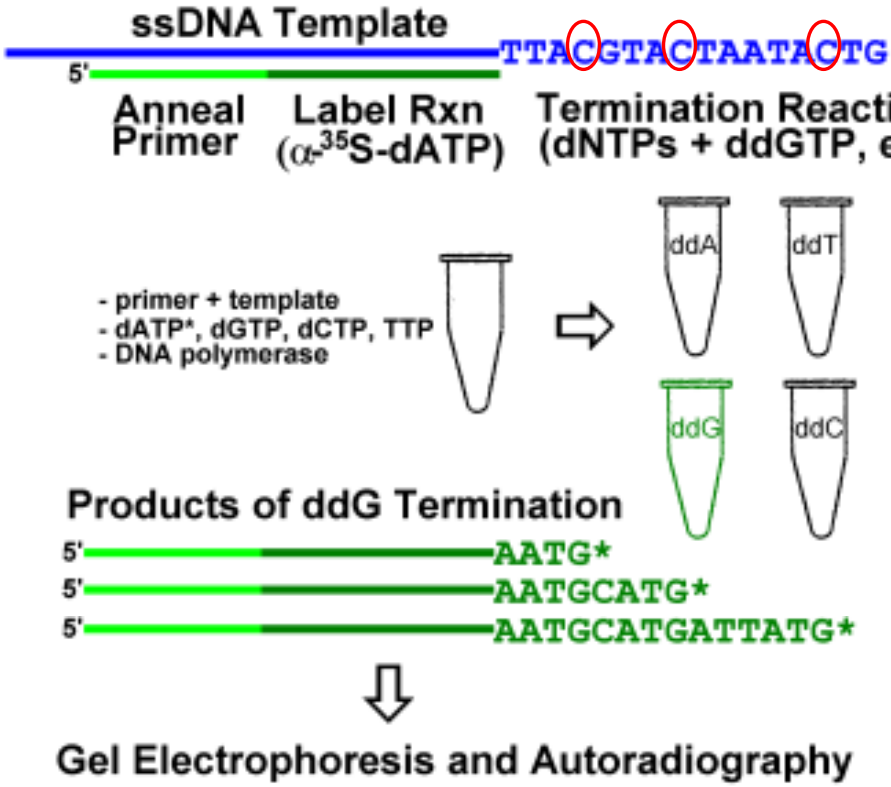
After 30 cycles

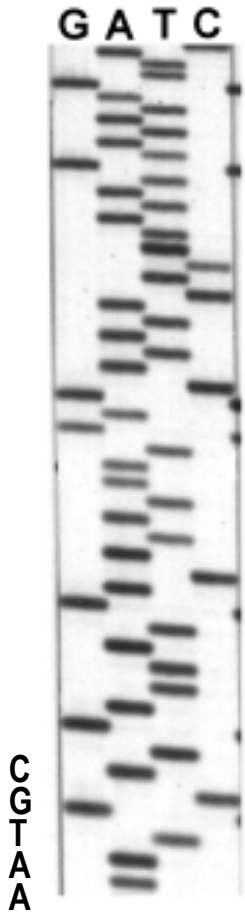
- the amplification is 2^{30}
- $2^{10} \sim 1000$, $2^{30} \sim 10^9$

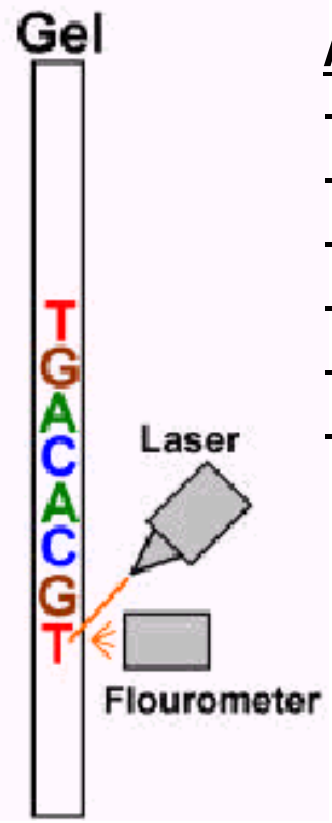
~ billion-fold amplication

- Products from one cycle can become targets in the next cycle
- Exponential amplification of the target region !!!!!
- Amplification without cloning: quick and easy

DNA Sequencing

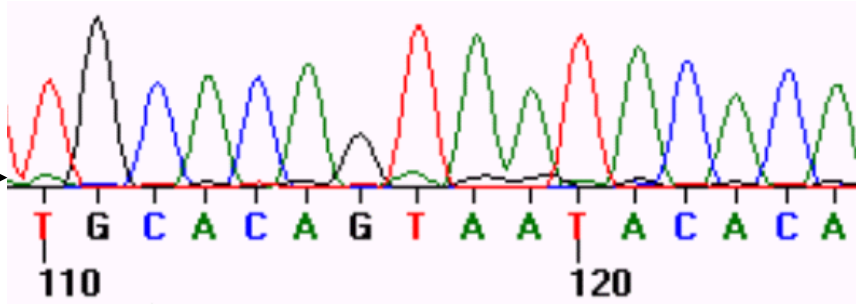




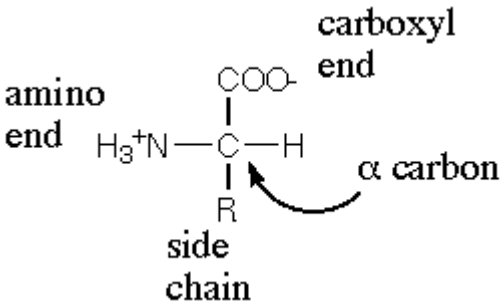


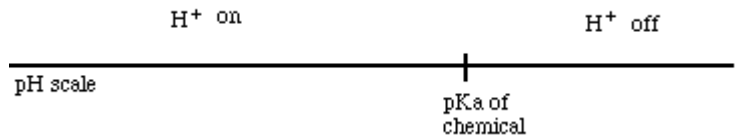
Automated Sequencing:

- same chemistry
- fluorescent dye terminators
- all 4 termination reactions in one tube
- fragments of increasing size pass by detector
- chain terminator detected in real time
- sequence decoded in real time

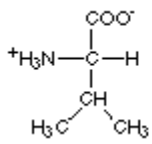


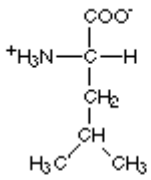
Amino-Acid Structure

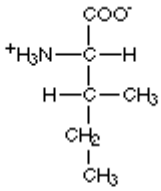


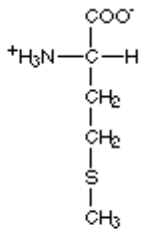


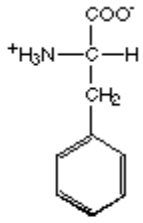
Hydrophobic Side-chain


Valine (val)

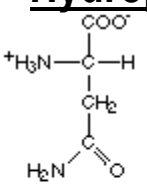

Leucine (leu)

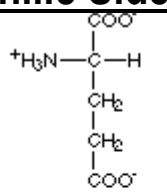

Isoleucine (ile)

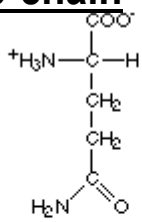

Methionine (met)

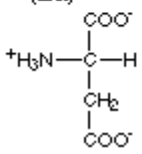

Phenylalanine (phe)

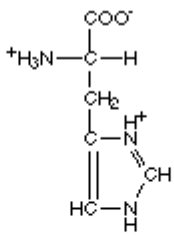
Hydrophilic Side-chain

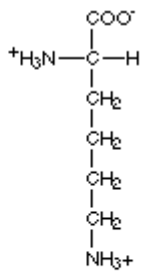

Asparagine (asn)

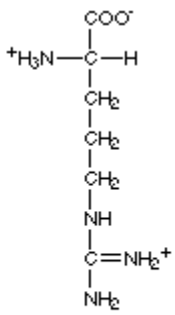

Glutamic acid (glu)


Glutamine (gln)

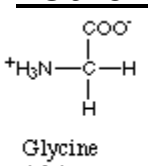

Alanine (ala)

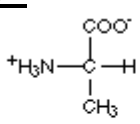

Histidine (his)

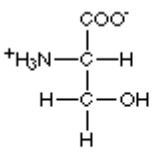

Lysine (lys)

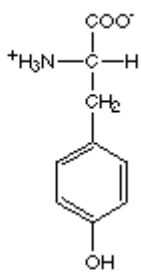

Arginine (arg)

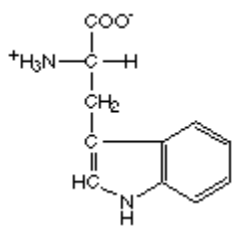
"Others"

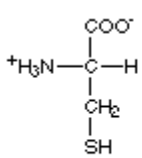

Glycine (gly)

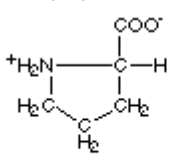

Serine (ser)


Threonine (thr)

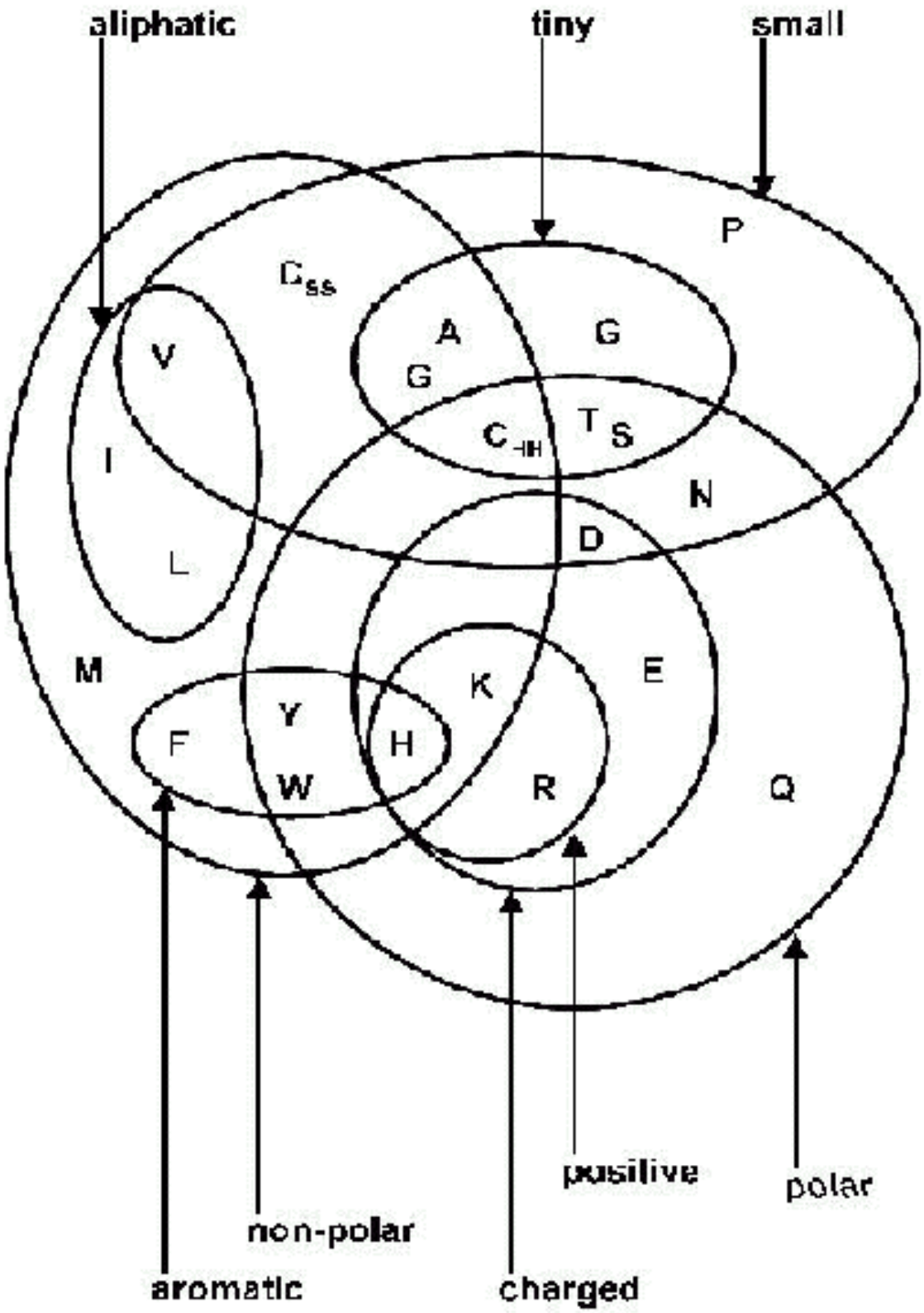

Tyrosine (tyr)


Tryptophan (trp)

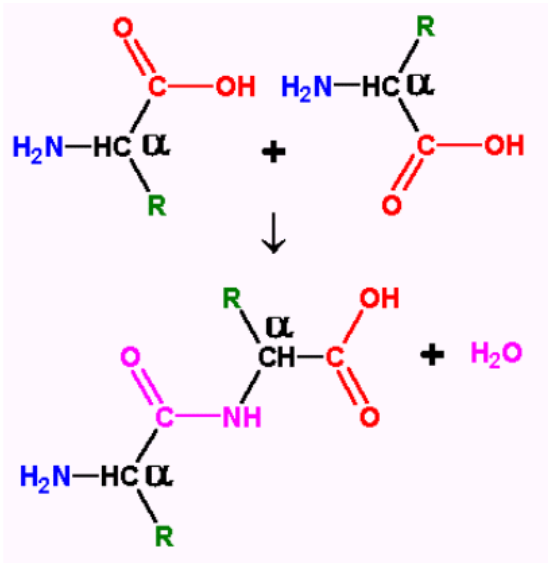

Cysteine (cys)

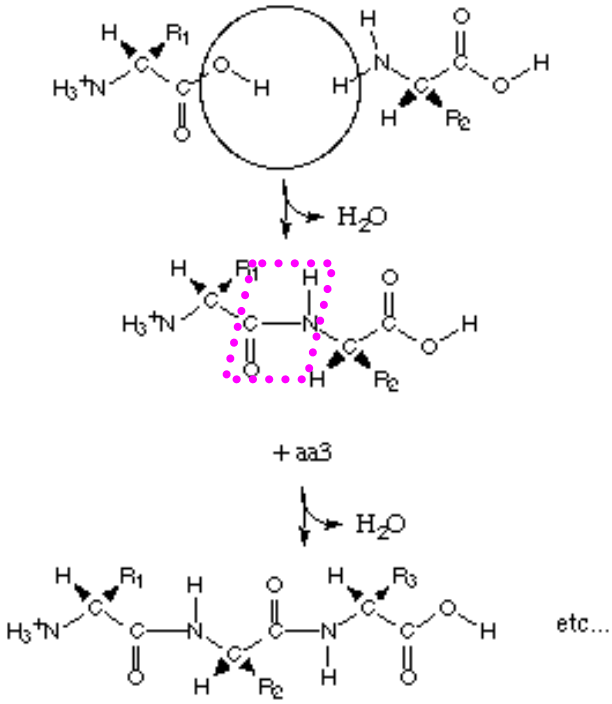

Proline (pro)

Amino-Acids don't like to be pigeon-holed...

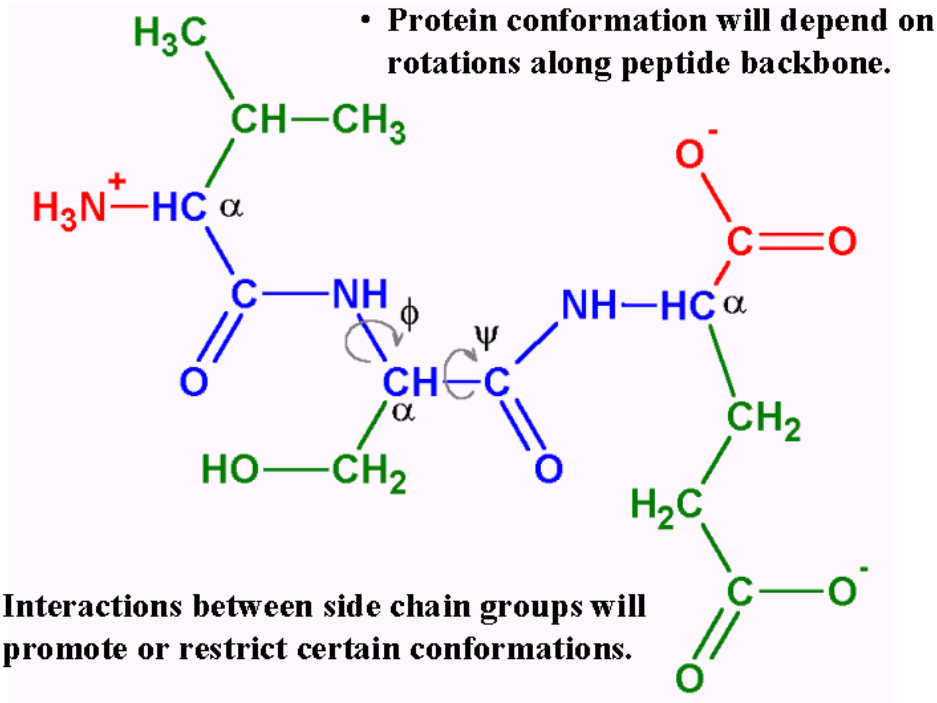


The Peptide bond





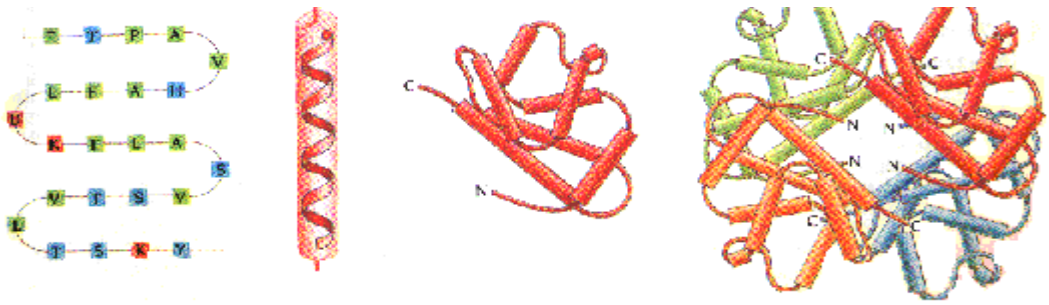
- dehydration reaction produces **peptide bond**
- proteins are heteropolymers of amino acids



Protein Structure

Levels of Protein Structure

Primary	Refers to the amino acid sequence and the location of disulfide bonds (i.e., covalent bonds).
Secondary	Refers to interactions between amino acids that are close together (eg., α-helix, β-sheet, β-turn, random coil).
Tertiary	Refers to interactions between amino acids that are far apart (eg., motifs, domains).
Quaternary	Refers to interactions between two or more polypeptide chains (i.e., protein subunits).



Primary Secondary Tertiary Quaternary

Primary Structure

- the order of the amino acids (ie. covalent bonds !!!!)
- contains all of the information necessary for folding the peptide chain into its "native" structure
- displayed from amino- to carboxy-terminus

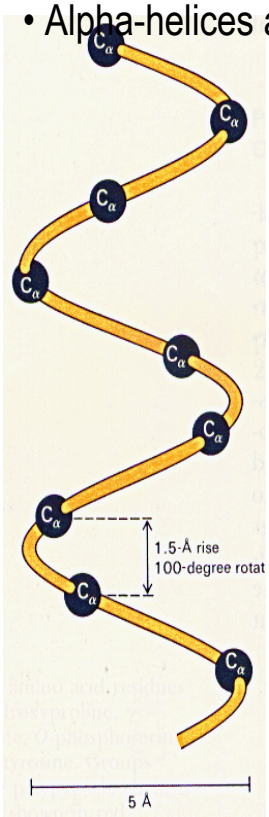
```
              5         10        15        20        25        30
  1 A A S X D X S L V E V H X X V F I V P P X I L Q A V V S I A
 31 T T R X D D X D S A A A S I P M V P G W V L K Q V X G S Q A
 61 G S F L A I V M G G G D L E V I L I X L A G Y Q E S S I X A
 91 S R S L A A S M X T T A I P S D L W G N X A X S N A A F S S
121 X E F S S X A G S V P L G F T F X E A G A K E X V I K G Q I
151 T X Q A X A F S L A X L X K L I S A M X N A X F P A G D X X
181 X X V A D I X D S H G I L X X V N Y T D A X I K M G I I F G
211 S G V N A A Y W C D S T X I A D A A D A G X X G G A G X M X
241 V C C X Q D S F R K A F P S L P Q I X Y X X T L N X X S P X
271 A X K T F E K N S X A K N X G Q S L R D V L M X Y K X X G Q
301 X H X X X A X D F X A A N V E N S S Y P A K I Q K L P H F D
331 L R X X X D L F X G D Q G I A X K T X M K X V V R R X L F L
361 I A A Y A F R L V V C X I X A I C Q K K G Y S S G H I A A X
391 G S X R D Y S G F S X N S A T X N X N I Y G W P Q S A X X S
421 K P I X I T P A I D G E G A A X X V I X S I A S S Q X X X A
451 X X S A X X A
```

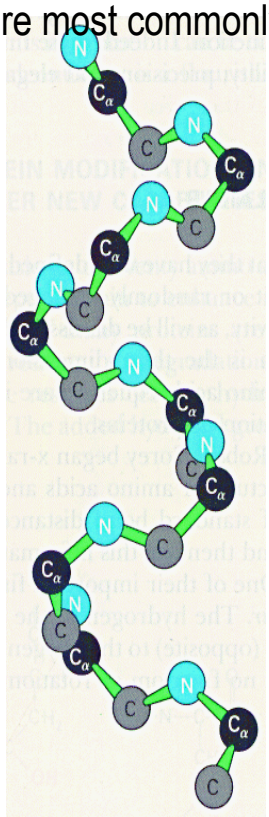
Secondary Structure

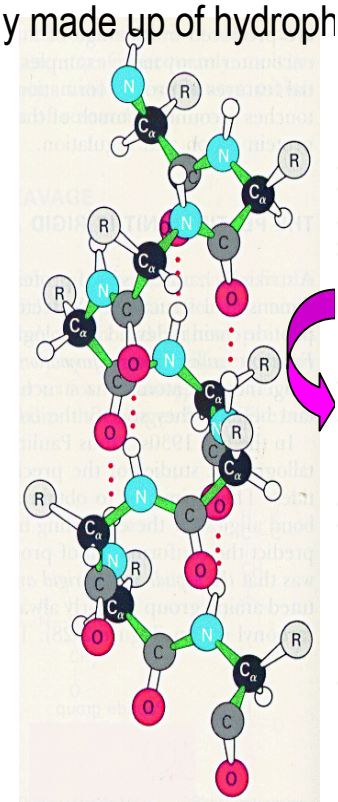
- Refers to certain common repeating structures found in proteins
- Lots of Hydrogen bonding !!!!
- Two types of secondary structures: alpha-helix and beta-pleated sheet.

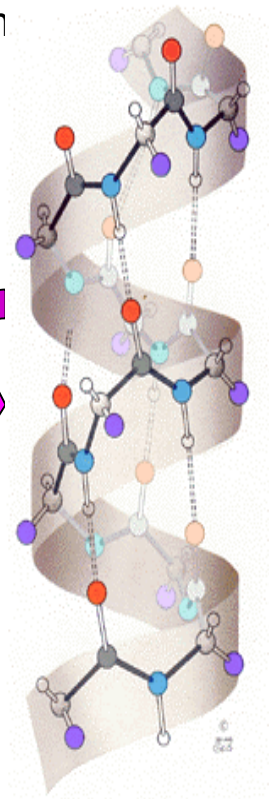
The Alpha-Helix

- Tight helix formed out of the polypeptide chain
- Polypeptide main chain makes up the central structure
- Side chains extend out and away from the helix.
- Amide Carboxyl Oxygen of one amino acid (n) is hydrogen bonded to the Hydrogen on a Nitrogen four residues away (n +4)
- H-bonding between and Hydrogen on a Nitrogen three residues away
- Every C- and N-group of the backbone is hydrogen bonded
- Alpha-helices are most commonly made up of hydroph



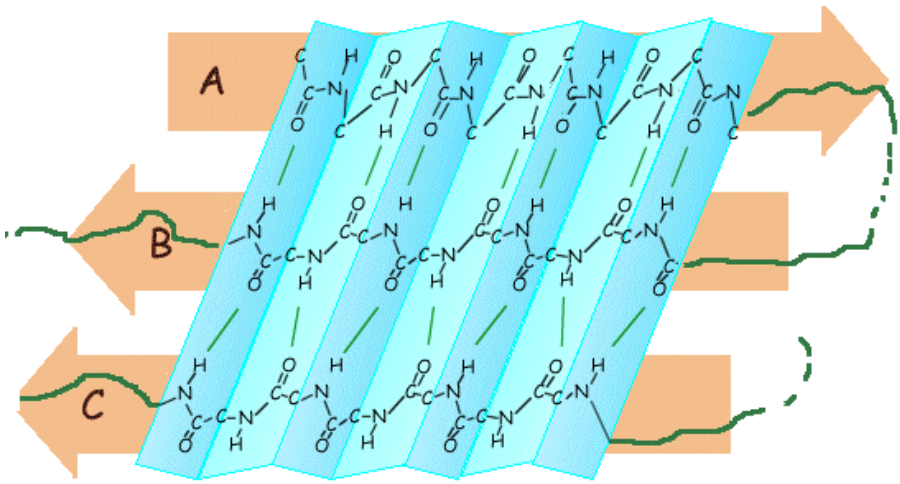






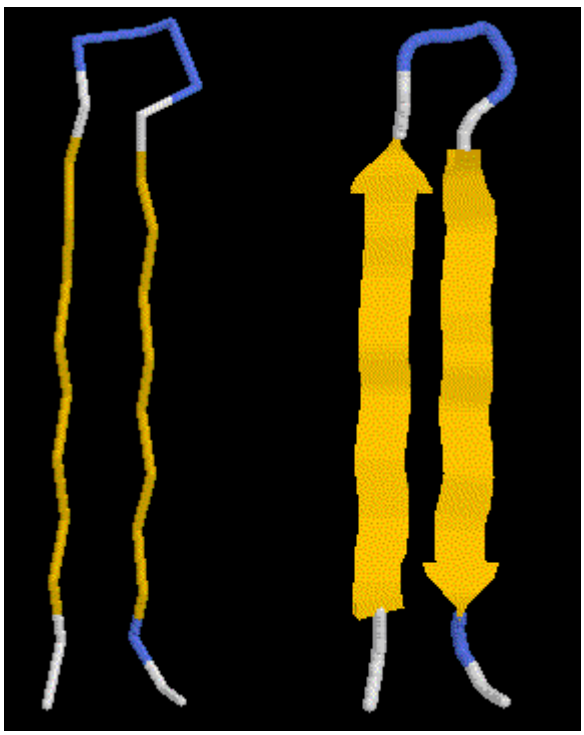
The Beta-Pleated Sheet (beta-strands)

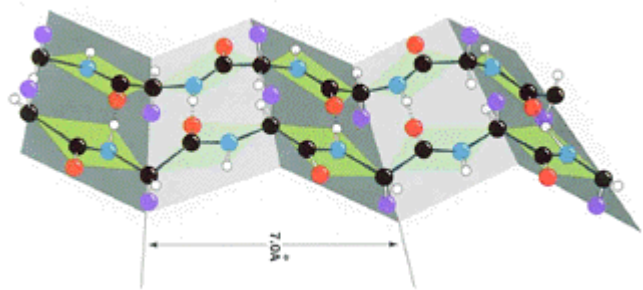
- Hydrogen bonding between the Hydrogen on a Nitrogen and the Amide Carbonyl Oxygen in another peptide bond that is quite a distance away
- The H-bonding atoms are quite a distance apart along the sequence of the polypeptide (ie. the chain folds on itself to bring them close together)
- Bonds between chains (rather than within as seen in alpha-helix)
- R-groups project above and below the plane of the sheet

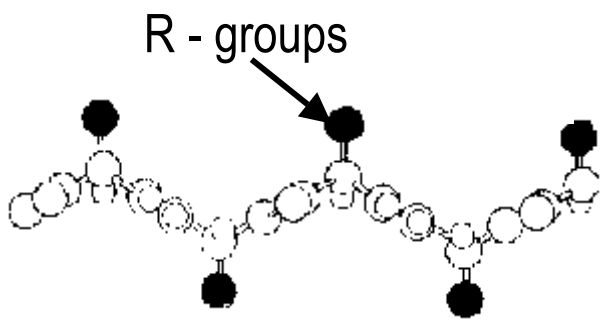


Anti-Parallel
(parallel H-bonds)

Parallel
(oblique H-bonds)

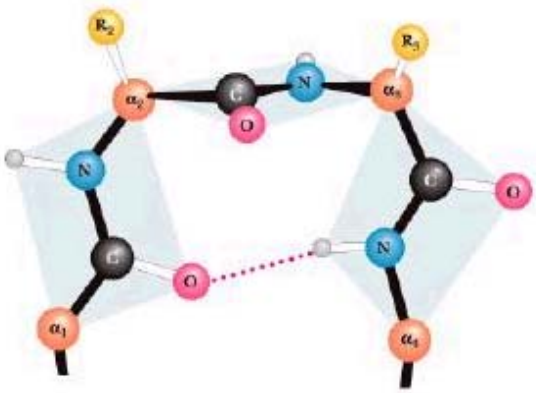


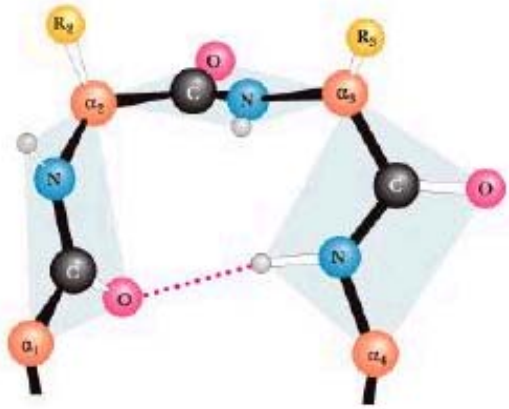


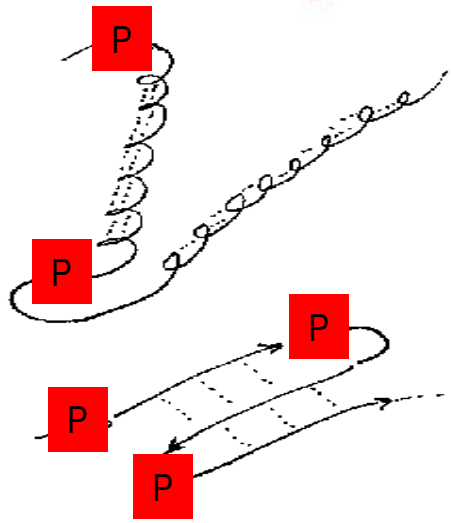


The Beta-Turn (or beta-hairpin)

- Allows a peptide chain to reverse direction by folding onto itself
- H-bonding between Amide carboxyl oxygen and Hydrogen on a Nitrogen three residues away
- Often find Proline and Glycine residues
    - Pro: rigid (helix-breaker)
    - Gly: small and flexible

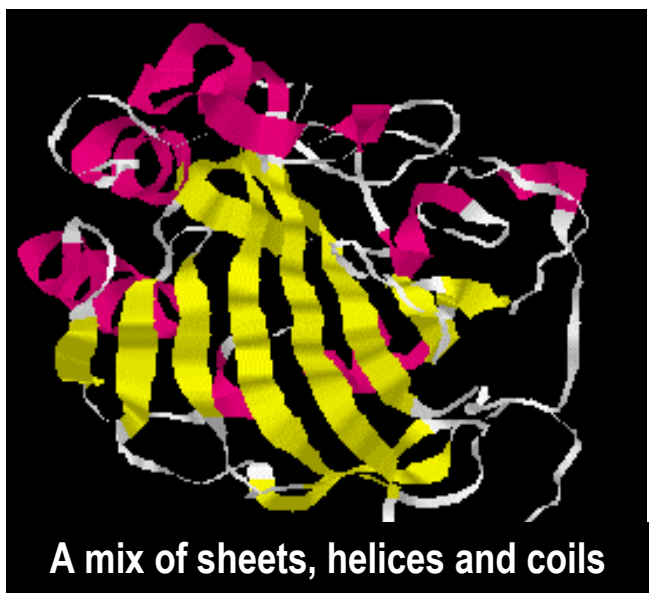




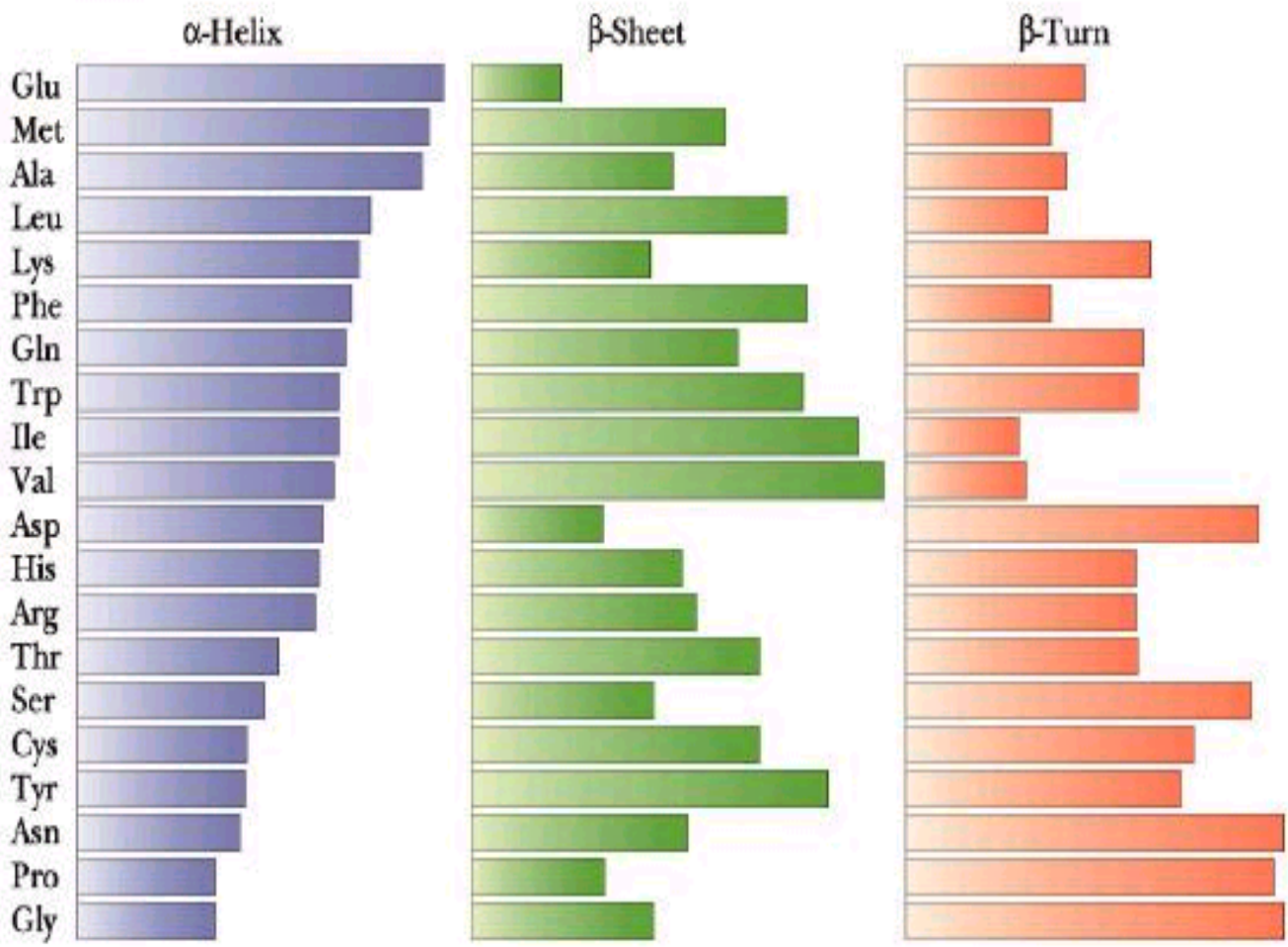


Random Coils

- Proline is quite inflexible
    - breaks up secondary structures
- Often found at the ends of alpha-helices and beta-strands
- Absence of sheets and helices gives rise to “random coils”
- Not really random, just hard to define

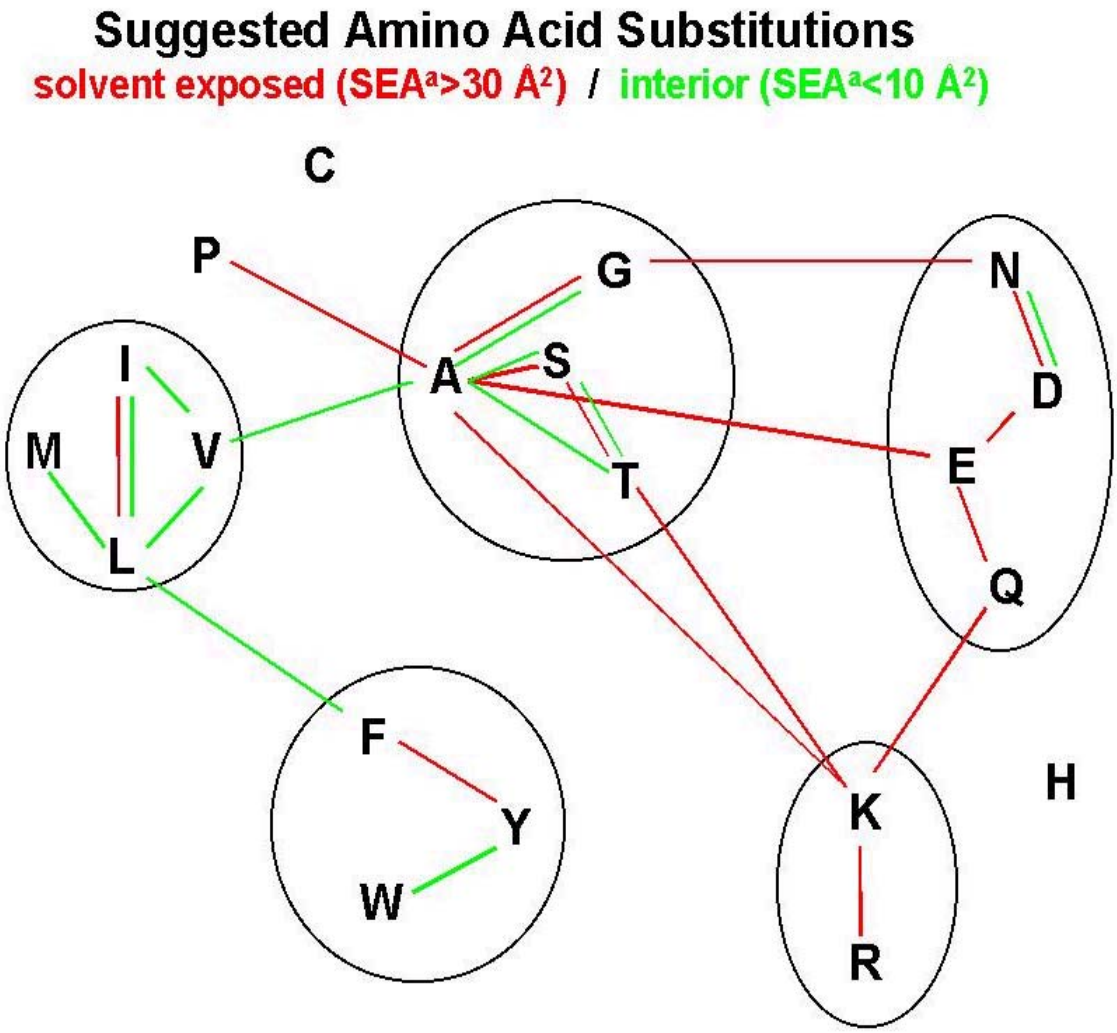


What if we ask Amino Acids what they like to do ? (part I)



What if we ask Amino Acids what they like to do ?

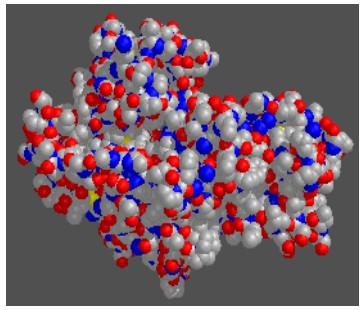
- Some amino acids like impersonating one another
- others don't!



Amino acids connected by a solid line can be substituted with 95% confidence
(D. Bordo and P. Argos, J. Mol. Biol. 217(1991)721-729)

[a]SEA=solvent exposed area

Tertiary Structure

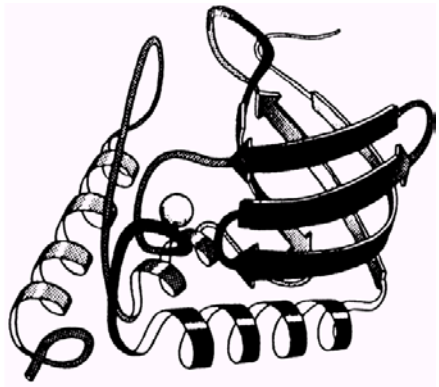


- Tertiary structure is the full 3-D folded structure of a polypeptide chain
- Long distance interactions of the Amino Acid side chains

A) Ionic bonds ("Salt Bridges"): Polar, charged AA side chains

Negative -- Glu Asp

Positive -- Lys Arg His



B) Hydrogen bonds: Polar, Non-charged AA side chains

Alcohols -- Ser Thr Tyr

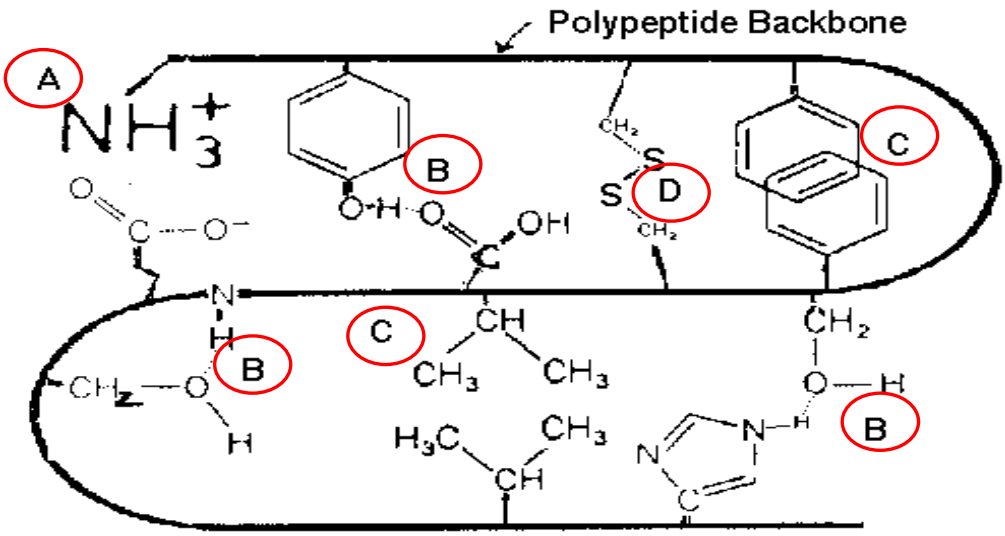
Amides --- Asn Gln

C) Hydrophobic interactions: Non-Polar amino acids

Hydrocarbon -- Ala Val Leu Ile Pro Met

Aromatics ----- Phe Trp (Tyr)

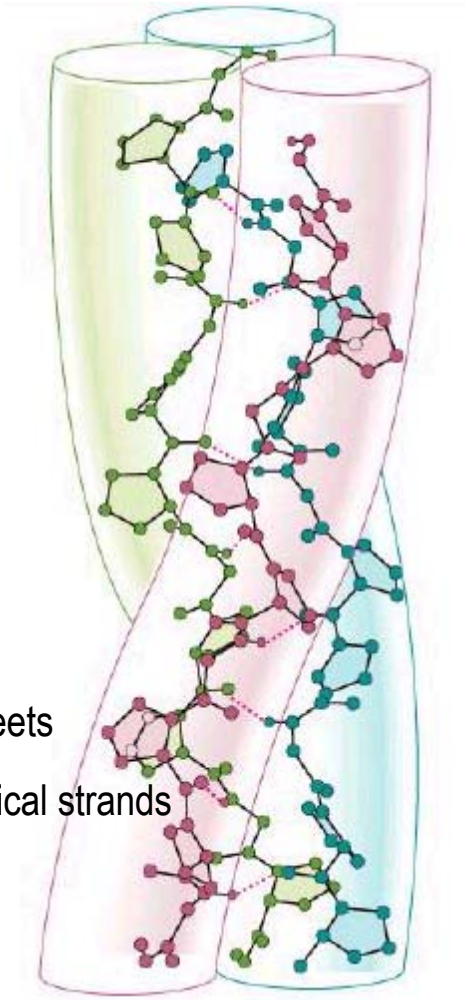
D) Disulfide bonds: Weak covalent bond between 2 Cys (R-Cys-S-S-Cys-R)



- Backbone links between elements of secondary structure usually short and direct
- Proteins fold to maximize stability (make H-bonds and minimize solvent contact)
- Close packing of sheets and helices

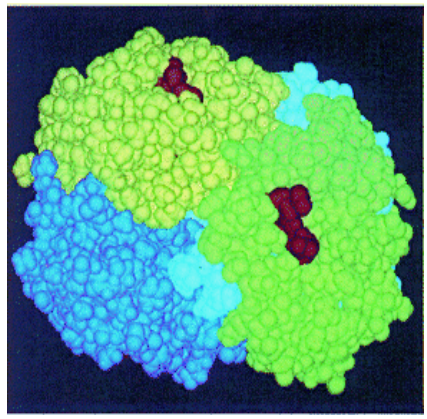
Unusual Structures: The Collagen triple helix

- Unusual AA composition:
    - Nearly one residue out of three is Gly
    - Proline content is unusually high
    - Unusual amino acids found:
        - 4-hydroxyproline
        - 3-hydroxyproline
        - 5-hydroxylysine
    - Pro and HyPro make 30%
- AA composition unsuited for alpha-helices OR beta-sheets
- Suited for the collagen triple helix: three intertwined helical strands
- Long stretches of Gly-Pro-Pro/HyP



Higher order structure can be dictated by primary structure and amino acid composition

Quaternary Structure



- Formed between separate polypeptides
- Entropy gain due to burying of hydrophobic groups - very favorable!
- Multiple polypeptides **(subunits)** form a complex
- Non-covalent bonding between 2 polypeptides
- Complexes can be made-up from 2 or more polypeptides
  - dimers (2), trimers (3), tetramers (4)...
- Complex can be made-up of :
  - single type of unit: eg. Homo-dimer
  - more than one type of unit: e.g. Hetero-dimer

Commonly found:

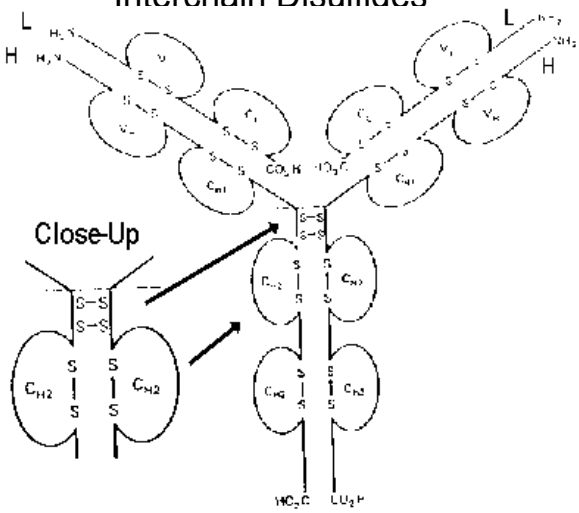
Ionic Bonds

Hydrogen Bonds

Less Commonly found:

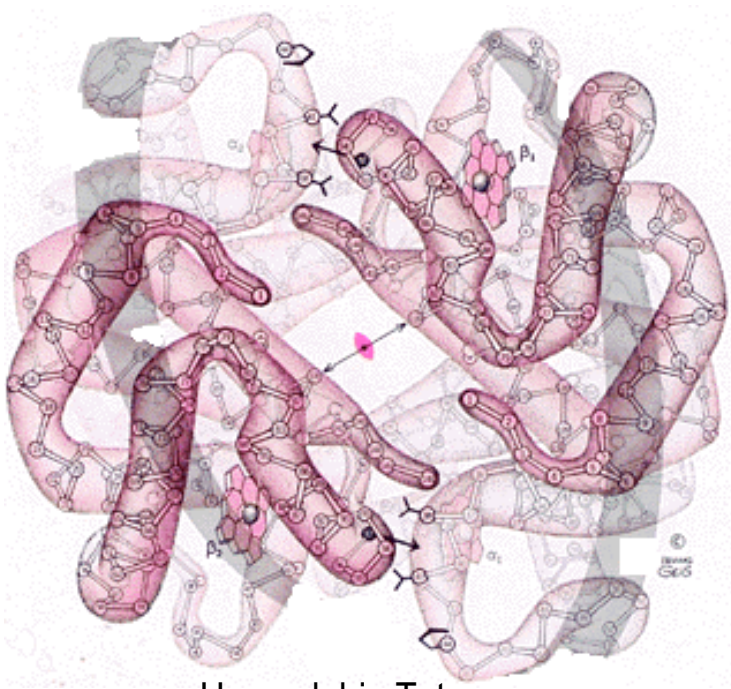
Hydrophobic interfaces

Interchain Disulfides



Antibody Tetramer

- disulfide bonds between and within polypetides



Hemoglobin Tetramer

- non-covalent ionic bonds

Some Structural Concepts

The Forces That Drive Folding:

- Peptide chain must satisfy the constraints inherent in its own primary structure
- Peptide chains fold burying hydrophobic side chains, minimizing their contact with water
- Peptide chains have a right-handed twist inherent with their L-amino acid composition

Layered Structures:

- Helices and sheets often pack in layers
- Hydrophobic residues found between the layers
- Outside layers made up of mostly hydrophilic residues

Protein Modules

- Many proteins are constructed as a composite of two or more "modules" or domains
- Each of these is a recognizable domain that can also be found in other proteins
- Sometimes modules are used repeatedly in the same protein
- There is a genetic basis for the use of modules in nature
    - **“exon shuffling”**

Some Structural Concepts (cont)

Globular Proteins

- Most polar residues face the outside of the protein and interact with solvent
- Most hydrophobic residues face the interior of the protein and interact with each other
- Packing of residues is close
- there is some empty space within: small pockets

Classes of Globular Proteins

- Antiparallel alpha helix proteins
  - short connecting loops and antiparallel packing
- Parallel or mixed beta sheet proteins
  - nonpolar residues on both sides of the beta sheet
  - both faces of the sheet must be protected from solvent
- Antiparallel beta sheet proteins
  - nonpolar residues on only one face of the sheet
  - only one face must be protected from solvent
  - antiparallel beta sheet proteins may contain as few as two layers
- Metal- and disulfide-rich proteins
  - usually less than 100 residues
  - conformations usually heavily influenced by metals and/or disulfide bridges
  - usually unstable if the metals are removed or the disulfides are reduced

Everything I need to know about protein engineering……..

I learned from mother nature

We can learn lots about protein engineering and design by observing what's out there

- Proteins have been conserved through the course of evolution
  ie. mother nature has been performing protein engineering forever
- Conserved proteins have similar structural or catalytic activity, difference in primary structure
- Comparing protein sequences:
  - there is a conserved backbone interspersed with sequence differences
  - we can compare primary structures to determine "important bits"
    - conserved residues in catalytic centers
    - conserved structural motifs
  - we can get a feel for what types of substitutions are OK
  - we can see how evolutionary forces have shaped the protein structure
    - adaptation to environment:
    - differences in kinetic parameters
    - differences in activity

Many protein engineering goals mirror these evolutionary forces !!!!

- Recombinant DNA techniques can be used to "engineer" changes in protein structure
- Recombinant DNA techniques can be used to mass produce proteins of interest

Protein Engineering: the goals

Changes in catalytic properties

- Increase in V_{max}
- Decrease in K_m
- Change in pH optimum
- Elimination of an inhibition site
- Alteration of specificity of reaction
- Elimination of a residue that confers instability

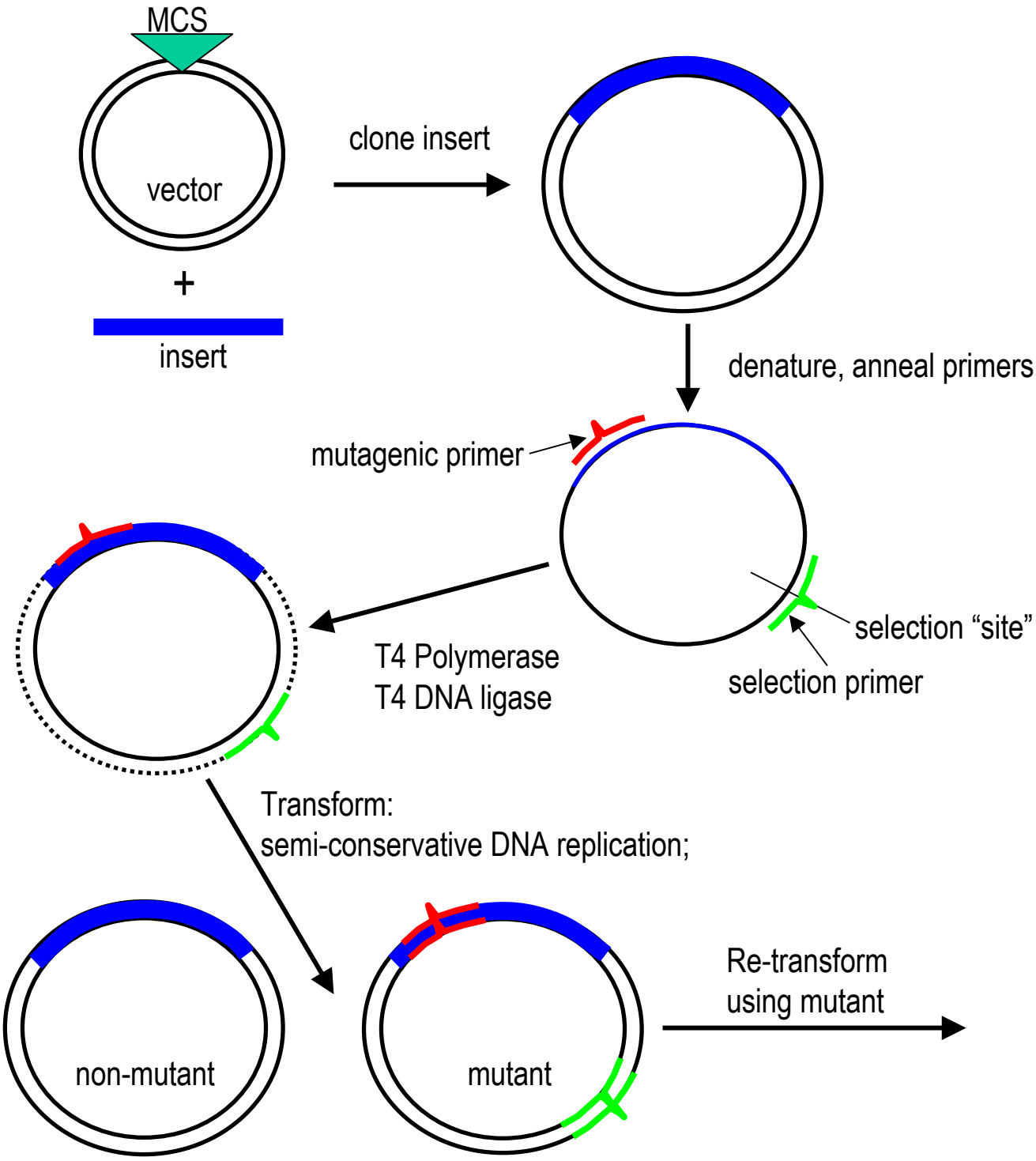
Changes in structural properties

- Improvement in thermostability
- Improvement in stability in organic solvents
- Changes in physiochemical properties
- Modification of ligand binding specificity

Creation of new systems

- Chimeric and multifunctional proteins
- Addition of tags for purification
- Improvement of effectiveness of therapeutic agents
- Addition of signal peptides for extracellular expression

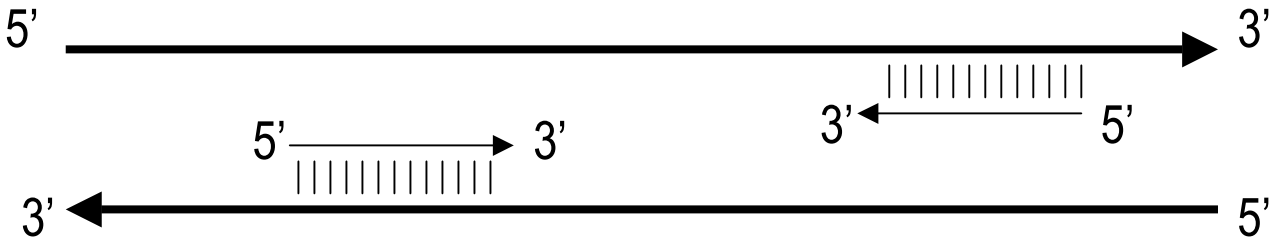
Protein Engineering begins at the DNA level: Site-Directed Mutagenesis



Mutant Selection

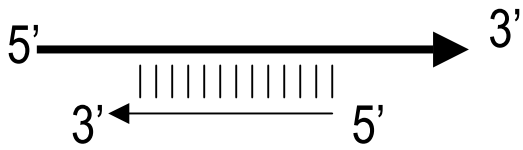
- restriction site elimination (negative selection)
- restoration of an antibiotic resistance gene (positive selection)
- many other variants exist (eg. Methylation/DpnI digestion)

Novel Applications of PCR: Site-directed mutagenesis

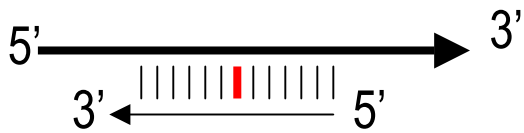


Primer / Template interactions:

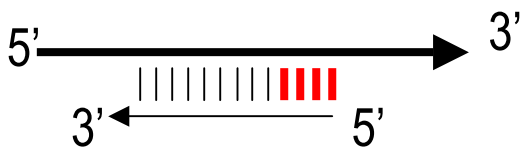
- 3' base-pairing is most important
  - 9-12 bp in the 3' are all that is necessary to have stable enough interaction for polymerase extension
  - 5' base-pairing not as crucial



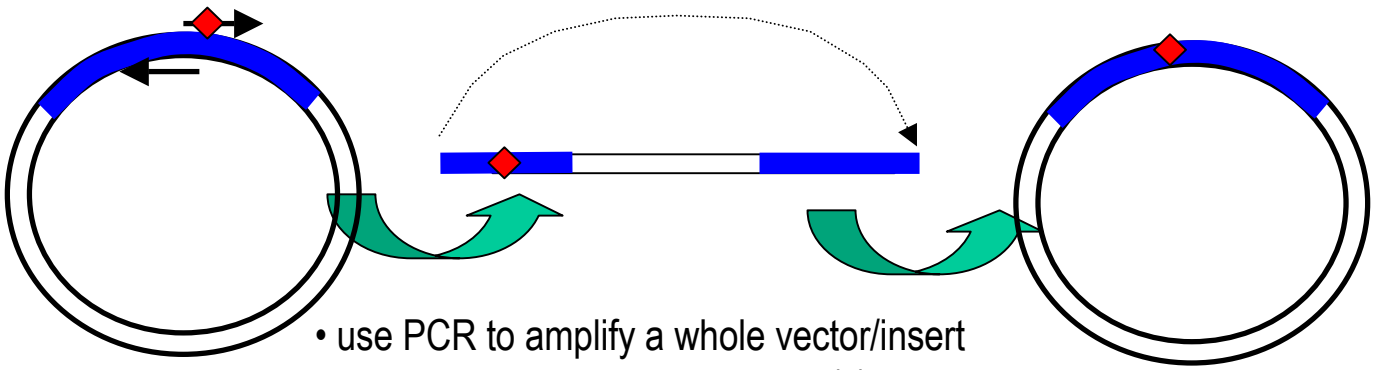
- On a long enough primer (> 30 bases)
  - mismatches in the middle are tolerated
    - introduction of a 1 bp mutation



  - mismatches in the 5' end of the primer are highly tolerated
    - introduction of a restriction site



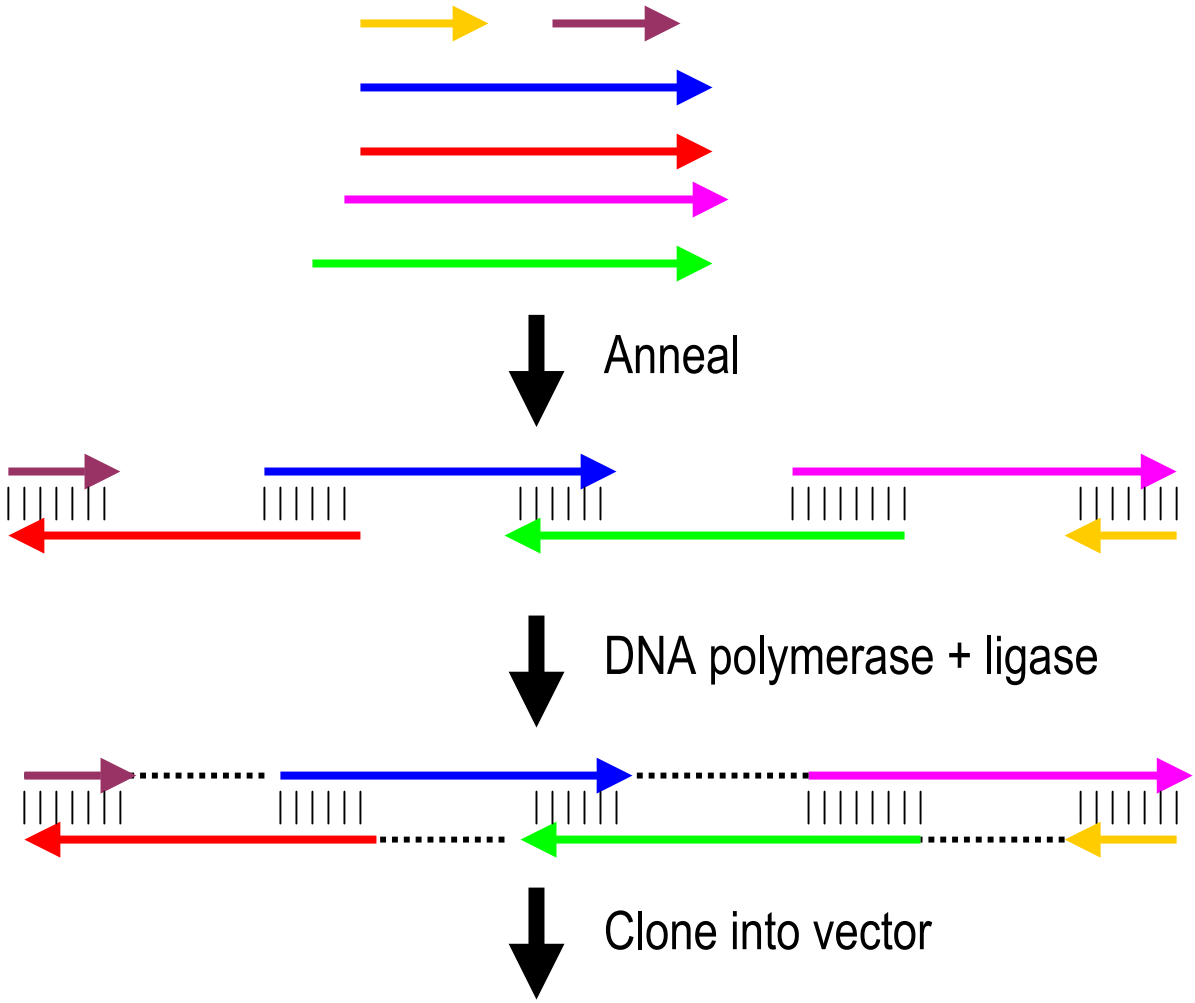
- "Errors" can be introduced into the resulting PCR products by using primers that have mismatches with respect to template
- Easy to introduce mutations !!!



- use PCR to amplify a whole vector/insert
- use "outward" mutagenic primer(s)
- ligate PCR product to itself to regenerate plasmid

Synthetic Genes

- Automated synthesis of short DNA molecules (oligonucleotides) is routine
  - Primers for PCR and sequencing
  - Oligo synthesizers are available (~$40per oligo)
  - Oligo synthesis services are also available (~60 cents / base)
- State of the art oligo synthesis can generate oligos up to ~ 70-90 bp
- Synthetic genes can be assembled from multipleoverlapping “long oligos”
  - “stitch together” by using DNA polymerase & ligase



Synthesis of an entire viral genome (Polio) was recently described !!!

Protein Engineering of a Xylanase

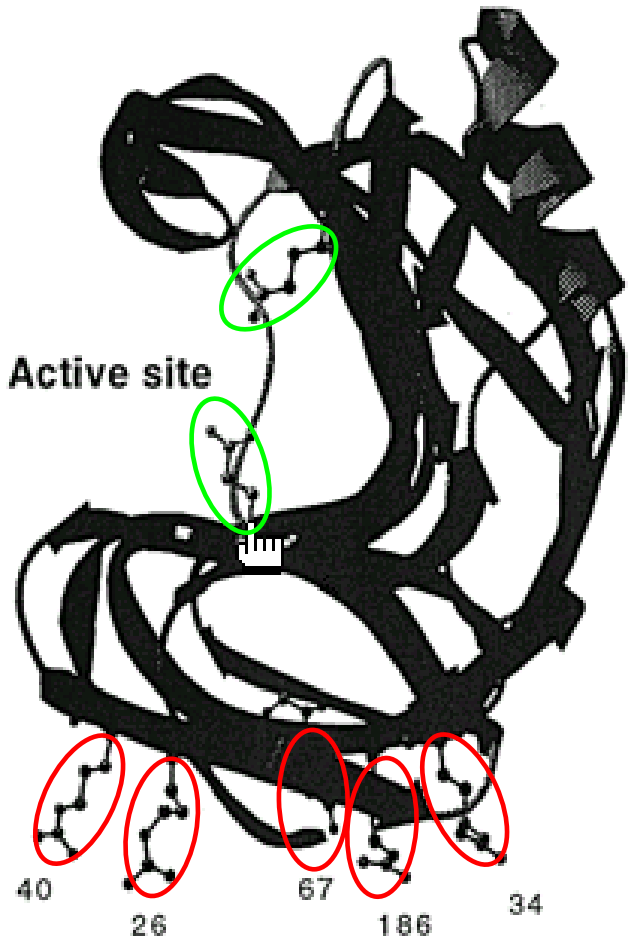
- xylanase hydrolyzes xylan, one of the most abundant polysaccharide fibres in nature
    - important cell wall component
- among the most important industrial uses: enviro-friendly pulp bleaching
- engineering of disulfide bridges has already been used to increase protein stability
    - NRC plug: Sung et al. (1994) in collaboration with IOGEN
        - examined protein structure for places that were in close 3-D proximity
        - 1998 and 2000 patents
        - widespread use in canadian pulp and paper mills: less chlorine use

More Improvements ?

Turunen et al. (2002)

Surface Rs:

- can increase stability
    - observed in nature !!!!
    - stronger hydrogen bonding ?
- can give adaptation to high pH



- tested effects of surface R addition
    - PCR-based mutagenesis
- added up to 5 Rs on one side
- checked for:
    - half-life (ie. stability)
    - pH dependent activity

Results:

- addition of 5 Rs on one side of double-layered beta-sheet
    - shifted activity profile by ~ +1 pH unit
    - increased enzyme activity at high temperature
    - increased enzyme stability 4-5 fold in the presence of substrate(60-65 degrees)
    - **decreased** enzyme stability in the absence of substrate (50-55 degrees)