

Biomolecular chemistry

4. From amino acids to proteins

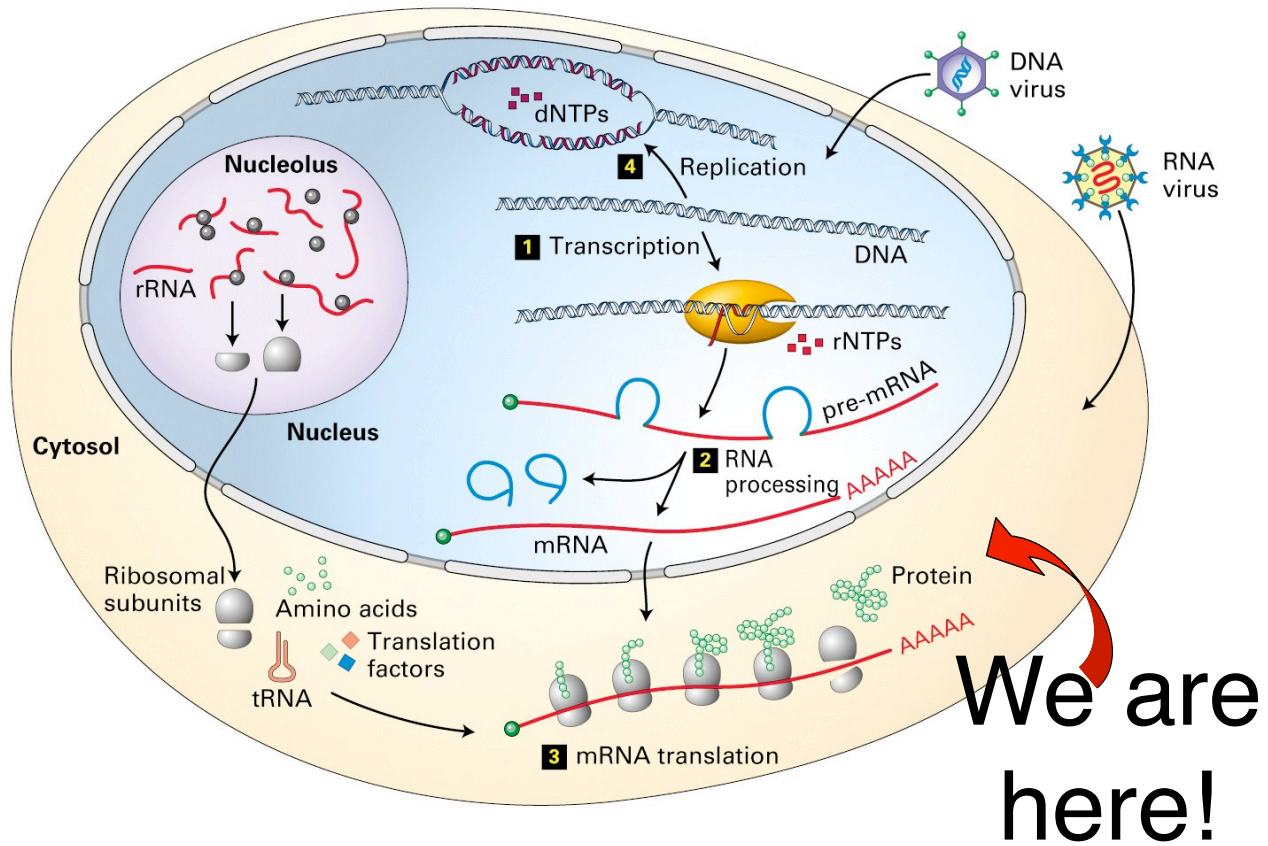
Suggested reading: **Sections 14.5 to 14.8 and Sections 2.1 to 2.4** of
Mikkelsen and Cortón, *Bioanalytical Chemistry*

Primary Source Material

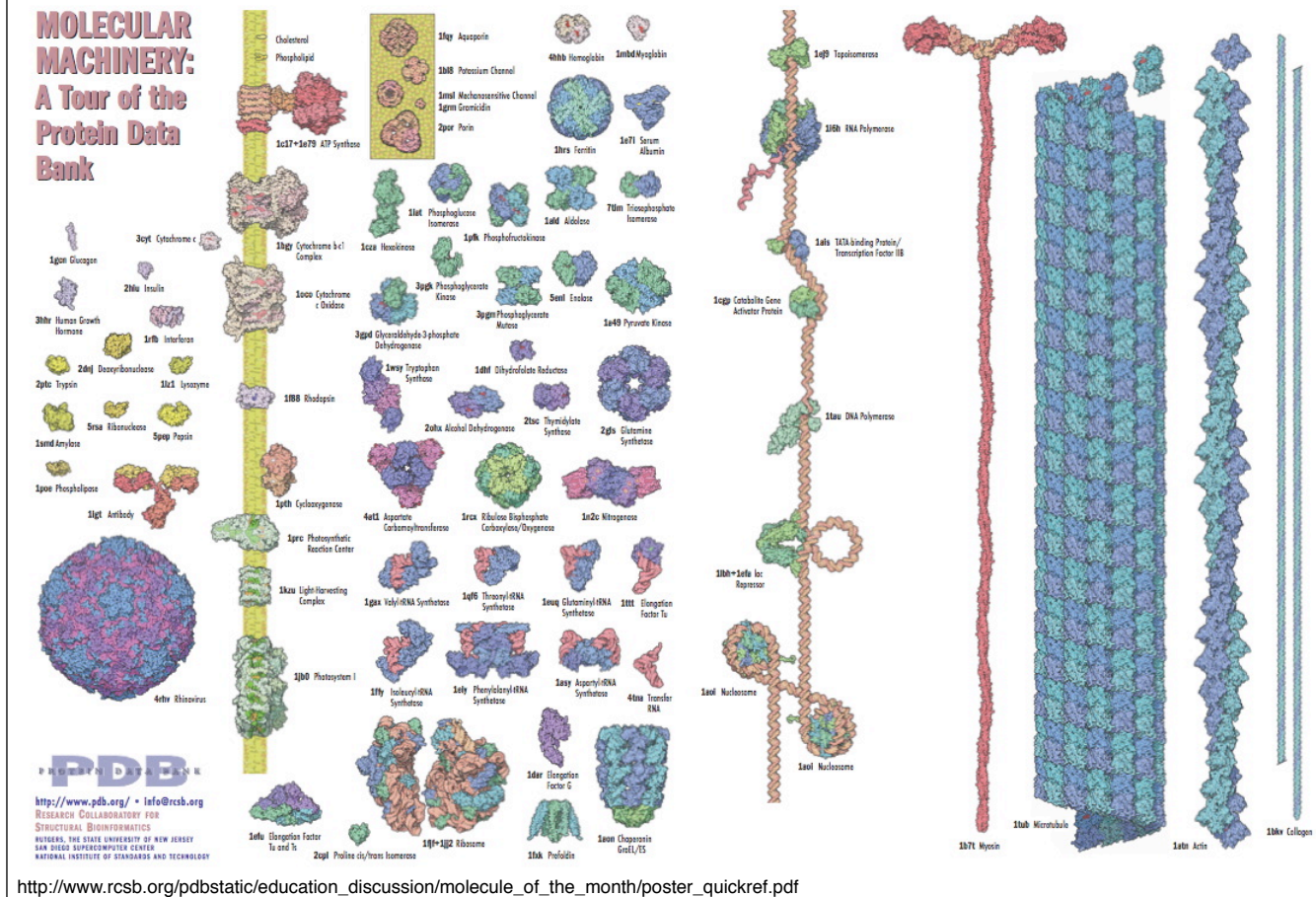
- Chapters 4 and 12 of Introduction to Genetic Analysis Anthony: J.F. Griffiths, Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, William M. Gelbart (courtesy of the NCBI bookshelf).
- Chapters 4, 4 and 6 of Biochemistry: Berg, Jeremy M.; Tymoczko, John L.; and Stryer, Lubert (courtesy of the NCBI bookshelf).
- Chapters 3 and 7 of Molecular Cell Biology: Lodish, Harvey; Berk, Arnold; Zipursky, S. Lawrence; Matsudaira, Paul; Baltimore, David; Darnell, James E. (courtesy of the NCBI bookshelf).
- ExPASy: online course on Principles of Protein Structure
- Many figures and the descriptions for the figures are from the educational resources provided at the Protein Data Bank (<http://www.pdb.org/>)
- Most of these figures and accompanying legends have been written by David S. Goodsell of the Scripps Research Institute and are being used with permission. I highly recommend browsing the Molecule of the Month series at the PDB (http://www.pdb.org/pdb/101/motm_archive.do)

Where are we and how did we get here?

78

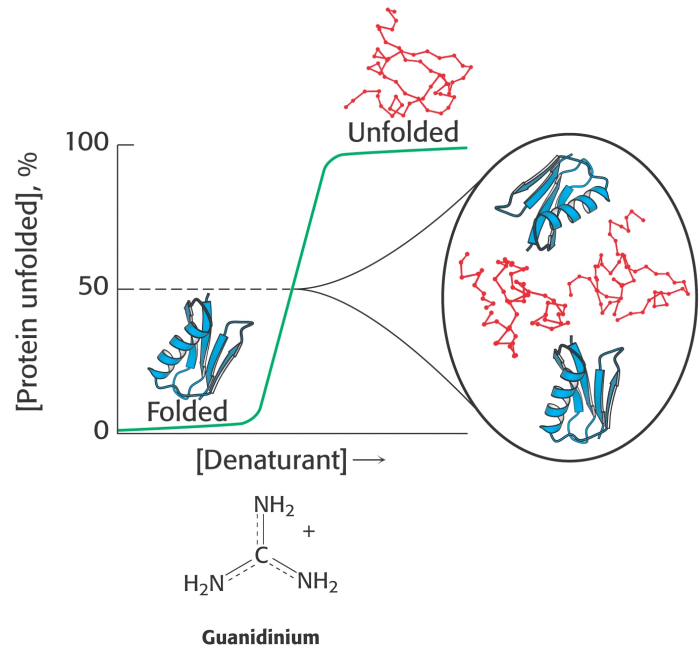


- We are done with the Central Dogma and now we move into the realms of protein structure and function. The Central Dogma only relates to the flow of genetic information, not to the function of biological macromolecules.



- Proteins are diverse and versatile 'nano' structures and machines
- Large number of potential combinations
 - There is a relatively large number number of amino acids (a.a.) which you can use to construct a protein.
 - Includes 20 common a.a.'s plus numerous post-translational modifications.
 - 200 amino-acid protein could have 20 to the 200th power possible sequences.
- Structurally versatile
 - Polypeptide backbone can adopt a variety of conformations
 - Many conformers of side chains
 - Secondary structural elements can pack together in a wide variety of orientations
 - Various states of homo- and hetero- oligomerization
 - Proteins can bind prosthetic groups or cofactors (non-protein)
 - Heme
 - Metal ions
 - flavins
- Structurally dynamic
 - Allosteric activation
 - Active and inactive forms

80



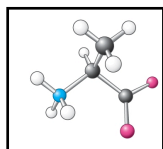
Ribonuclease

<http://www.users.csbsju.edu/~hjakubow/classes/rasmolchime/01ch331finproj/Rnase/templateprot.htm>

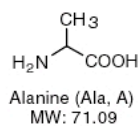
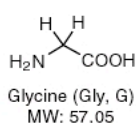
- The classic work of Christian Anfinsen in the 1950s on the enzyme ribonuclease revealed the relation between the amino acid sequence of a protein and its conformation. For this work he was awarded the Nobel Prize in Chemistry in 1972. Anfinsen discovered that:
 - Ribonuclease is a single polypeptide chain consisting of 124 amino acid residues cross-linked by four disulfide bonds.
 - Agents such as urea or guanidinium chloride effectively disrupt the noncovalent bonds.,
 - The disulfide bonds can be cleaved reversibly by reducing them with a reagent such as β -mercaptoethanol.
 - When ribonuclease was treated with β -mercaptoethanol in 8 M urea, the product was a fully reduced, randomly coiled polypeptide chain devoid of enzymatic activity. In other words, ribonuclease was denatured by this treatment.
- Anfinsen then made the critical observation that the denatured ribonuclease, freed of urea and β -mercaptoethanol by dialysis, slowly regained enzymatic activity. All the measured physical and chemical properties of the refolded enzyme were virtually identical with those of the native enzyme.
- These experiments showed that the information needed to specify the catalytically active structure of ribonuclease is contained in its amino acid sequence.

The 20 common amino acids

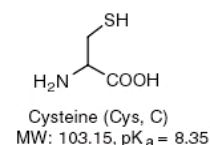
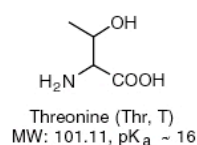
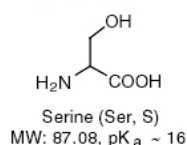
Ala showing L-stereochemistry



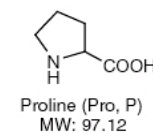
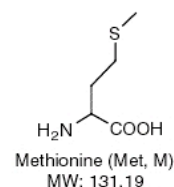
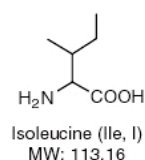
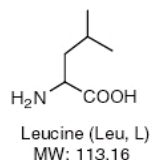
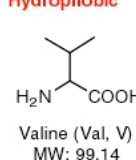
Small



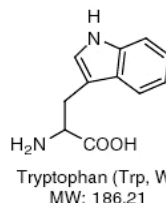
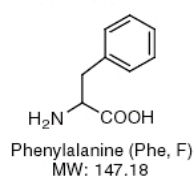
Nucleophilic



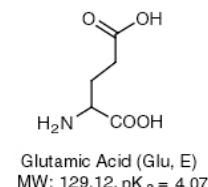
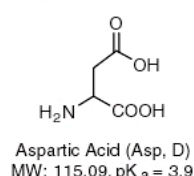
Hydrophobic



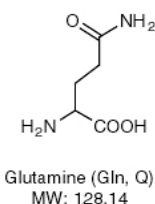
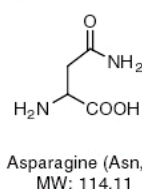
Aromatic



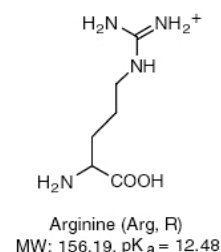
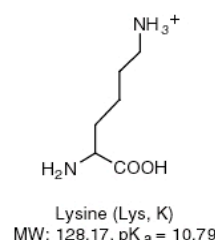
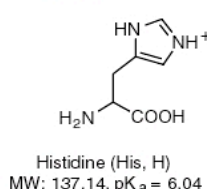
Acidic



Amide



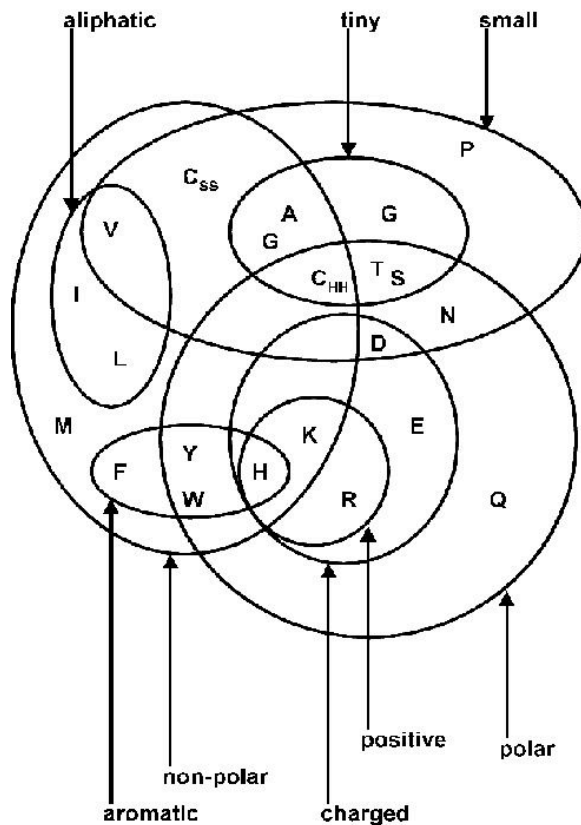
Basic



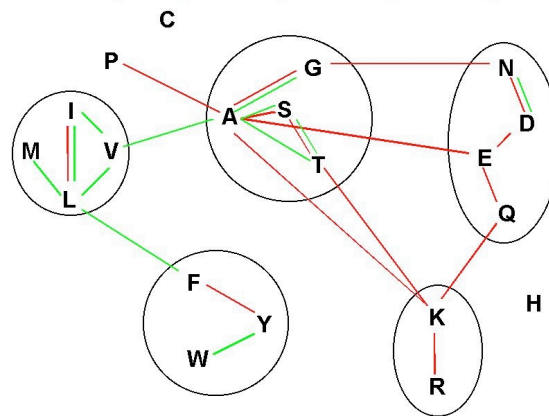
- 20 different common amino acids only differing in side chain
- Note that stereochemistry at C α has not been indicated in this figure.
- All natural a.a.'s are in L-configuration
- A more general system of stereochemical designation is the R/S system. The L-configuration nearly always corresponds to S in the R/S system. The exception is L-cysteine which is R.
- You might want to keep this sheet handy as a reference.
- I will often use the one letter codes and you should learn these.
- Most are easy, but I find D, E, N and Q the most tricky to remember
- *Q: Do we need to memorize the structure and names of amino acids on the test?*
- *A: Yes. You should know the structure, name, 3 letter abbreviation, and 1 letter code of all the common amino acids.*

Amino acid classification by property

82



Suggested Amino Acid Substitutions
 solvent exposed ($SEA^* > 30 \text{ \AA}^2$) / interior ($SEA^* < 10 \text{ \AA}^2$)



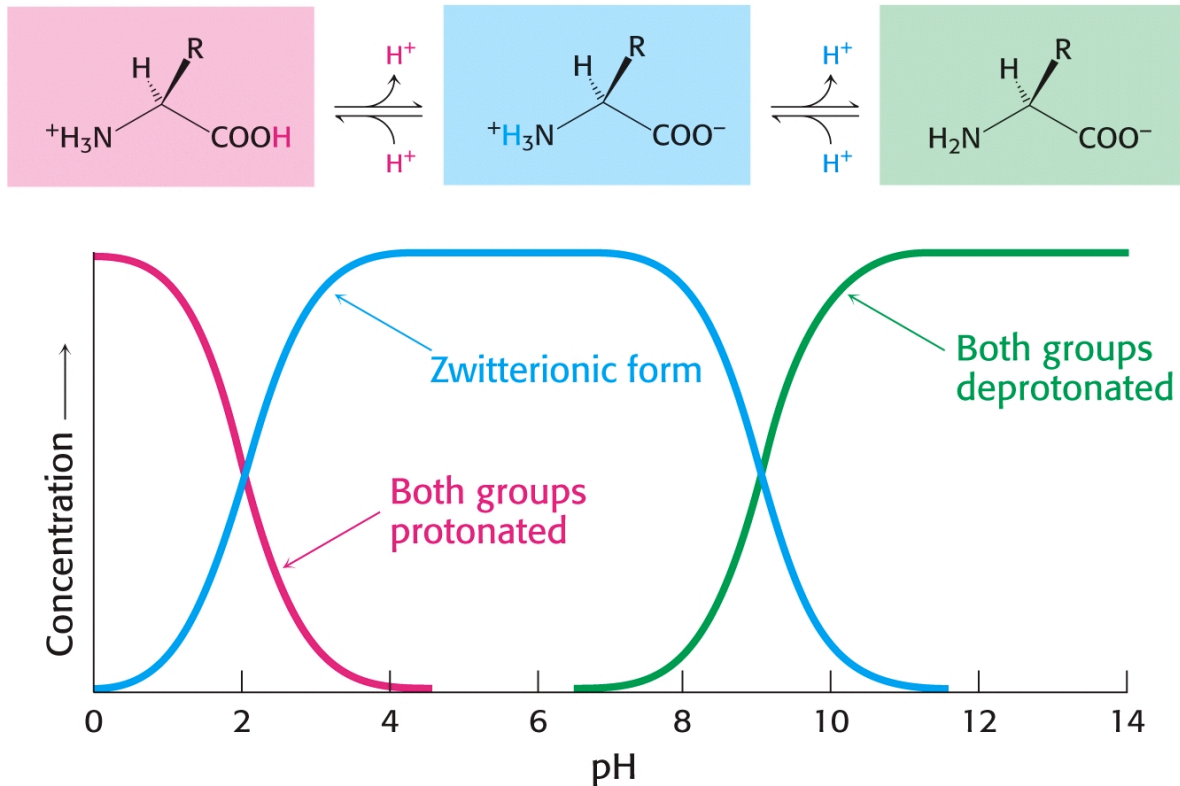
Amino acids connected by a solid line can be substituted with 95% confidence
 (D. Bordo and P. Argos, J. Mol. Biol. 217(1991)721-729)

* SEA =solvent exposed area

http://www.imb-jena.de/image_library/GENERAL/aa/mut1.jpg
http://www.imb-jena.de/image_library/GENERAL/aa/chemprop.jpg

- Various simple textbook classifications for a.a.'s
 - e.g. small, nucleophilic, hydrophobic, aromatic, acidic, amide, basic
 - e.g. aliphatic, non-polar, aromatic, polar, charged -ve, charged +ve
- However, no simple classification can properly capture the diversity of a.a. interactions and properties.
 - the same amino acid in different charge states can go from polar to nonpolar (H or K for example)
 - Different portions of the same amino acid can have different properties (aliphatic chain vs. guanidinium of arginine)
- Generally find aliphatic/hydrophobic residues inside proteins and polar/charged on the surface of proteins
- Notes:
 - Cysteine is special because it is best nucleophile, is most easily oxidized, and can form disulphide bonds.
 - Proline has a tertiary as opposed to secondary amide nitrogen and induces bend in polypeptide chain.
 - Theonine and Isoleucine have chiral carbons in side chain

Free amino acids are almost always zwitterions⁸³



Commentary on the topic of zwitterions: <http://bip.cnrs-mrs.fr/bip10/zwitter.htm>

- Amino acids in solution at neutral pH exist as dipolar ions (zwitterions).
- In the dipolar form, the amino group is protonated ($-NH_3^+$) and the carboxyl group is deprotonated ($-COO^-$).
- Under almost any conceivable physiologically relevant conditions, the amino and carboxylate group of a free amino acid will be in its charged state.
- This is also true of a polypeptide chain: the N-terminus and the C-terminus will be in the charged states
- Possible exceptions
 - Groups buried in the interior of proteins or lipid bilayers
 - Proteins in the stomach

pKa values of protein functional groups

84

TABLE 3.1 Typical pK_a values of ionizable groups in proteins

Amino acid	pK _a values (25°C)		
	α-COOH group	α-NH ₃ ⁺ group	Side chain
Alanine	2.3	9.9	
Glycine	2.4	9.8	
Phenylalanine	1.8	9.1	
Serine	2.1	9.2	
Valine	2.3	9.6	
Aspartic acid	2.0	10.0	3.9
Glutamic acid	2.2	9.7	4.3
Histidine	1.8	9.2	6.0
Cysteine	1.8	10.8	8.3
Tyrosine	2.2	9.1	10.9
Lysine	2.2	9.2	10.8
Arginine	1.8	9.0	12.5

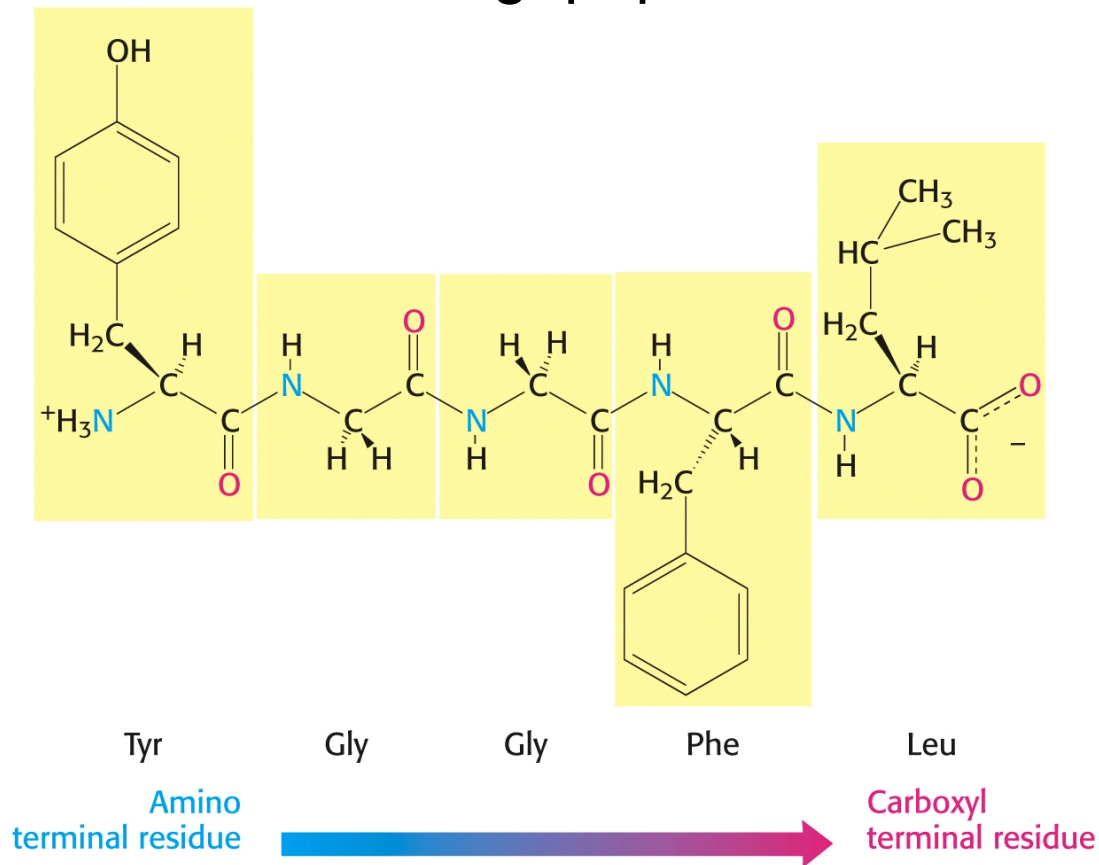
After J. T. Edsall and J. Wyman, *Biophysical Chemistry* (Academic Press, 1958), Chapter 8.

Group	Acid	Base	Typical pK _a *
Terminal α-carboxyl group			3.1
Aspartic acid Glutamic acid			4.1
Histidine			6.0
Terminal α-amino group			8.0
Cysteine			8.3
Tyrosine			10.9
Lysine			10.8
Arginine			12.5

*pK_a values depend on temperature, ionic strength, and the microenvironment of the ionizable group.

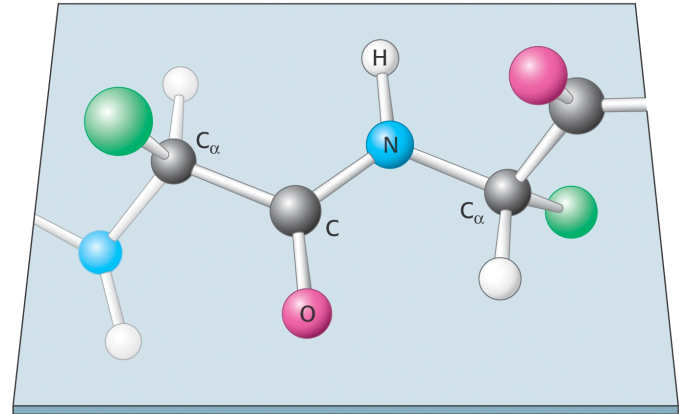
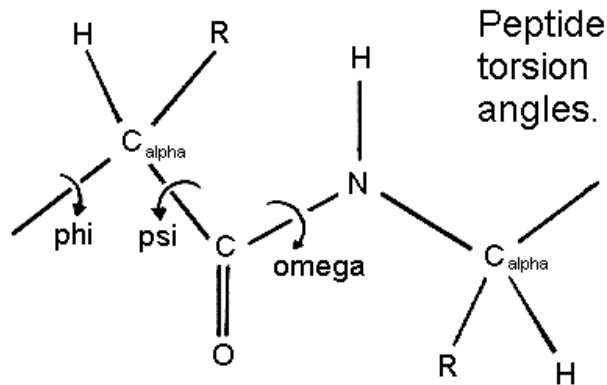
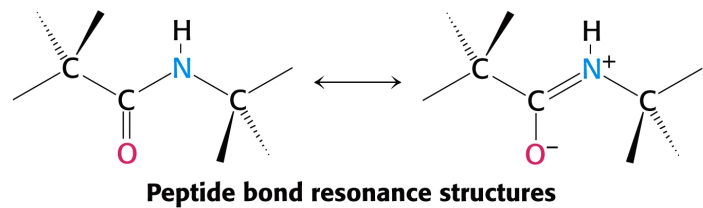
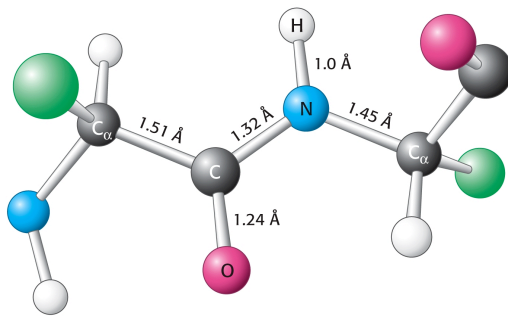
- Seven of the 20 amino acids have readily ionizable side chains. These 7 amino acids are able to donate or accept protons to facilitate reactions as well as to form ionic bonds.
- The above table gives equilibria and typical pK_a values for ionization of the side chains of tyrosine, cysteine, arginine, lysine, histidine, and aspartic and glutamic acids in proteins.
- Two other groups in proteins—the terminal α-amino group and the terminal α-carboxyl group—can be ionized.
- You should know the approximate values for all of these ionizable groups. It is safe to say that all carboxylic acids in proteins have a pK_a of about 3-4.
- Q: What is so special about Histidine? It has a pka of ~6, but did you mention that it does not react with anything much?*
- A: Histidine is very good at donating and accepting protons at physiological pH. This is a very important part of many enzyme mechanisms. I may have mentioned that histidine is not such a good nucleophile. For enzyme mechanisms that involve a nucleophilic attack on the substrate, cysteine would be the best amino acid, followed by lysine.*

An oligopeptide



- Oligopeptide: A compound made up of the condensation of a small number (typically less than 20) of amino acids
- Polypeptide: A compound made up of the condensation of more than ~20 amino acids
- Each type of protein differs in its sequence and number of amino acids. It is the particular sequence of the various side chains that makes each protein distinct.
- The two ends of a polypeptide chain are chemically different: the end carrying the free amino group (NH_3^+ , sometimes incorrectly written as NH_2) is the amino, or N-terminus, and that carrying the free carboxyl group (CO_2^- , sometimes incorrectly written as CO_2H) is the carboxyl, or C-terminus.
- The amino acid sequence of a protein is always presented in the N to C direction, reading from left to right. This corresponds to the 5' to 3' direction in which genes are read.

The peptide bond is planar

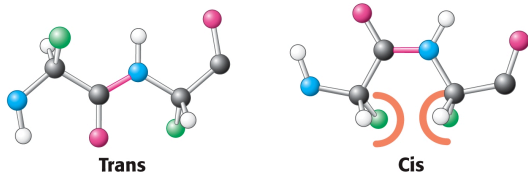


- Linus Pauling and Robert Corey analyzed the geometry and dimensions of the peptide bonds in the crystal structures of molecules containing one or a few peptide bonds.
- This analysis led Pauling to correctly predict the existence and structure of the alpha helix and beta sheets (for which he was awarded the 1954 Nobel Prize in Chemistry)
- The take home message is that the secondary structure elements of proteins can be predicted by looking at the structure of an individual amino acid. That is, an amino acid in an alpha helical or beta sheet conformer is also in a minimal energy conformer because its bonds are staggered and the peptide bond is planar.
- Note that the C-N bond length of the peptide is 10% shorter than that found in usual C-N amine bonds. This is because the peptide bond has some double bond character (40%) due to resonance which occurs with amides.
- As a consequence of this resonance all peptide bonds in protein structures are found to be almost planar. This rigidity of the peptide bond reduces the degrees of freedom of the polypeptide during folding.
- The planarity of the peptide bond is described using the angle 'omega'. This is the dihedral angle between the C α -carbonyl bond and the N-C α bond.

The peptide bond is almost always trans

87

All amino acids except proline



Proline

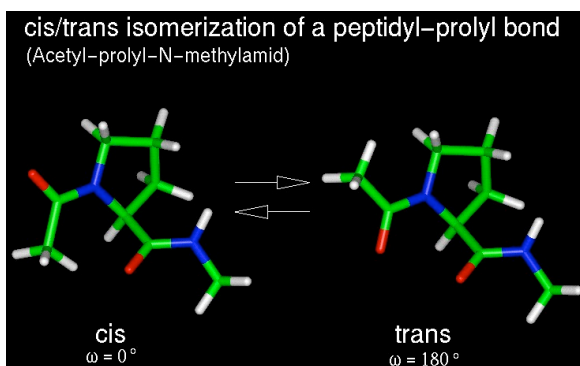
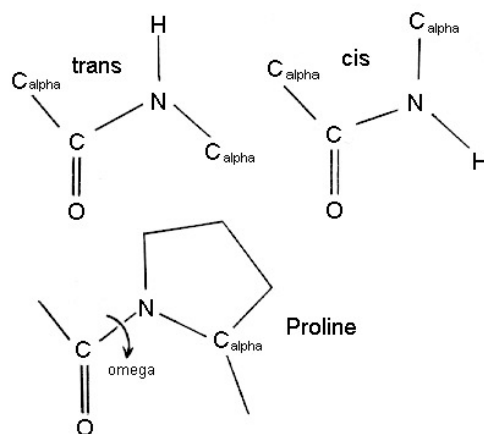
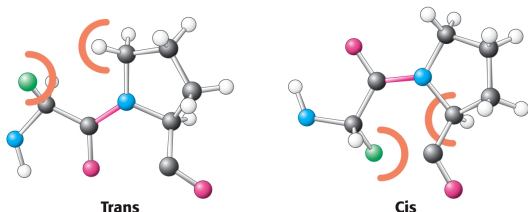
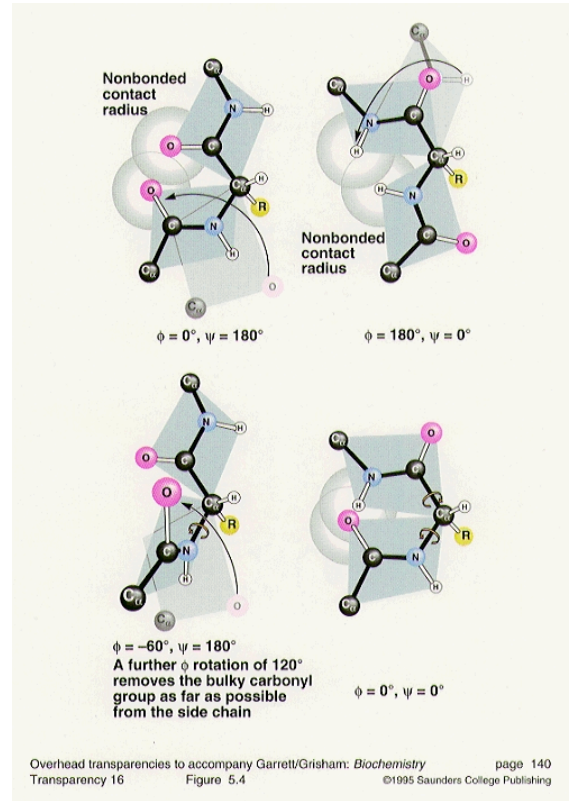
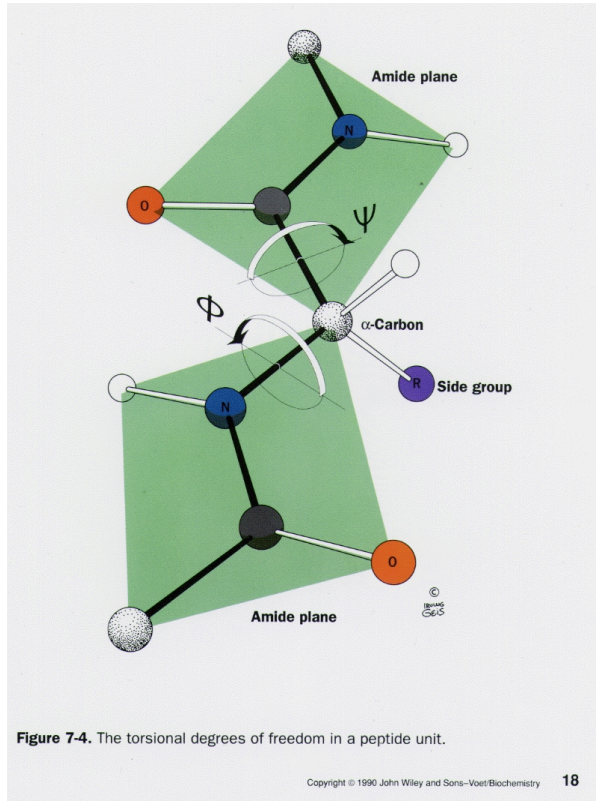


image credit: <http://www.imb-jena.de/IMAGE.html>.

- The omega (ω) angle is almost always 180° (trans) though sometimes (extremely rarely) it is 0° (cis).
- Note that both the cis and trans form are planar.
- Of the cis-peptide bonds found in proteins, almost all involve proline residues.
 - The overall atom geometry in cis proline is very similar to the trans-proline case. Energetically, the trans proline structure is not markedly more favourable than its cis-proline counterpart since much the same spatial conflicts are present in both cases.
- Approximately 1% of prolines in proteins are cis.
- A cis-peptide bond induces a very sharp kink in the polypeptide chain.

Certain combinations of ϕ and ψ angles are preferred ⁸⁸



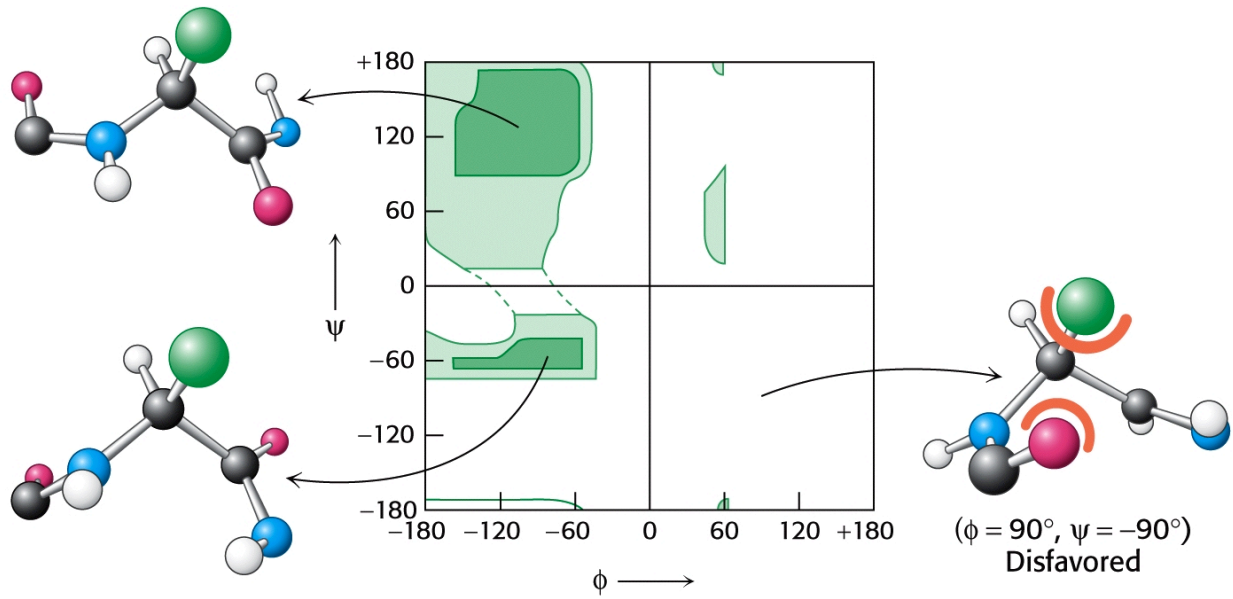
Scans downloaded from: <http://www.nd.edu/~aseriann/cou.html>

- A polypeptide can be thought of as a series of planar units (peptide bonds) joined by flexible hinges (C_α -atoms).
- Each C_α -atom has two rotatable bonds, the C-N bond (ϕ , phi) and the C-C bond (ψ , psi)
- Only certain combinations of ϕ and ψ angles are allowed due to steric clashes between the adjacent residues.

The Ramachandran Plot (ϕ vs. ψ)

89

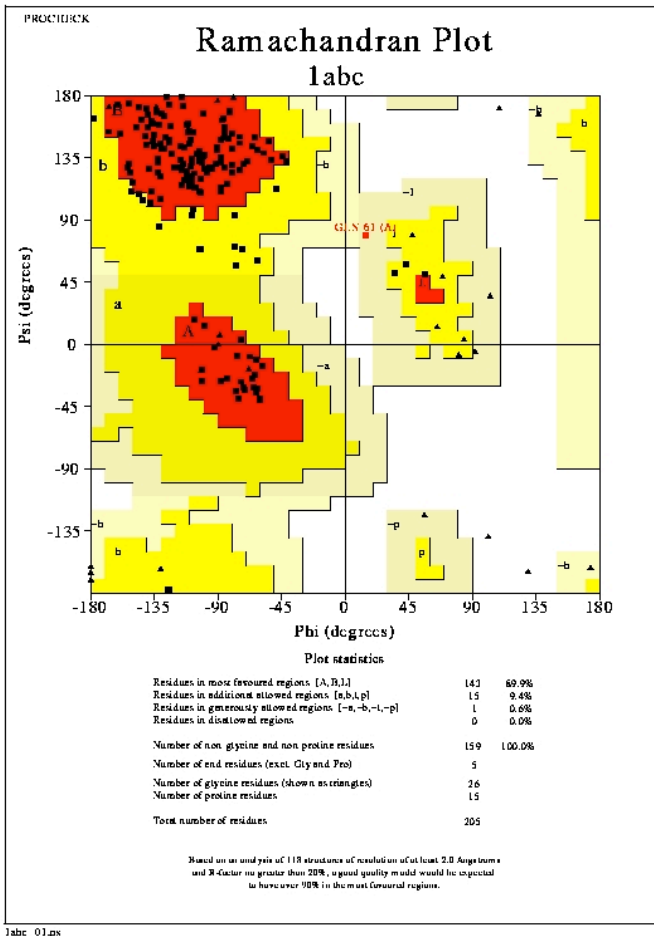
β -strand conformation



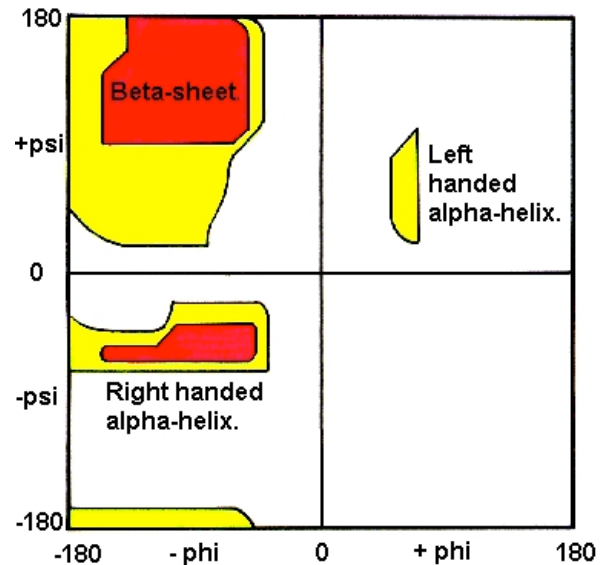
- A graph of ϕ angle vs. ψ angle vs. occurrence in proteins is called a Ramachandran plot.
- There are actually only a few conformations that are strongly preferred and these give rise to the common elements of secondary structure.

The Ramachandran⁹⁰ Plot of a typical protein

(as output by the
program PROCHECK)



The Ramachandran Plot.

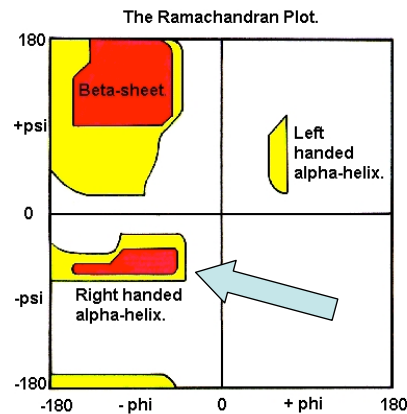
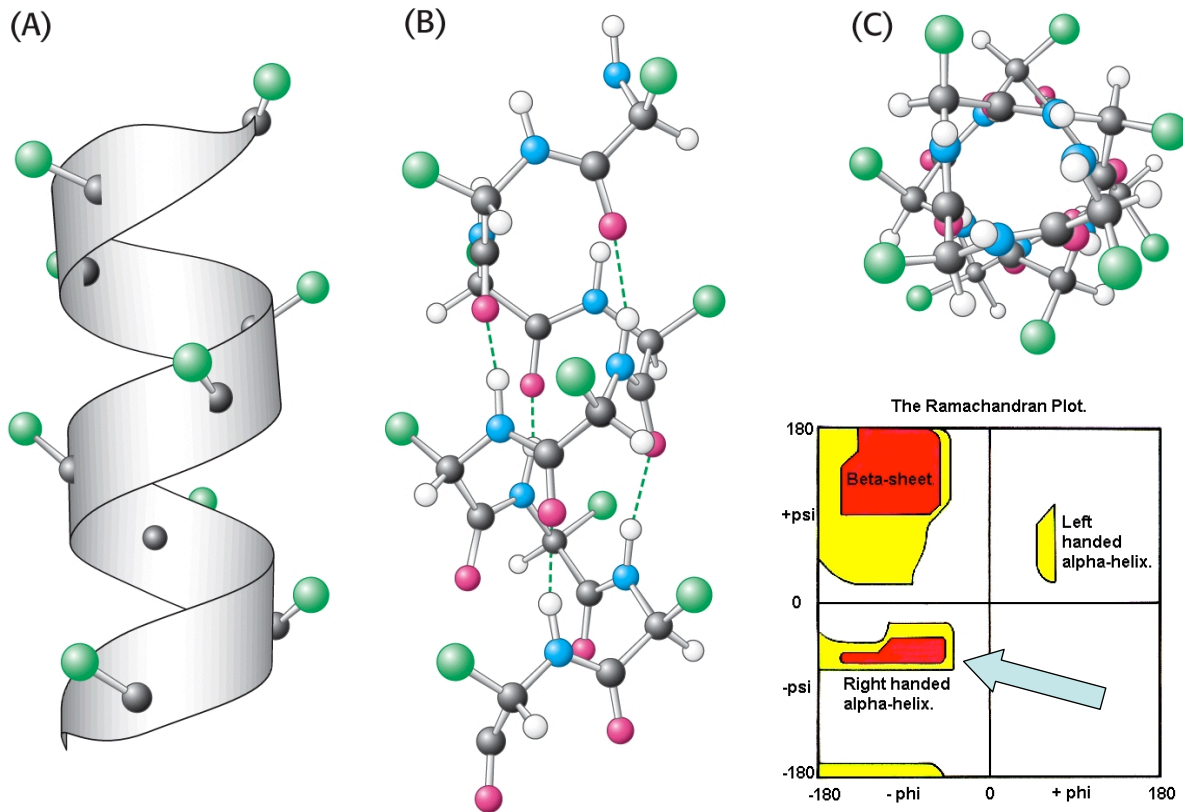


<http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html>
<http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html>

- The Ramachandran plot for a particular protein shows the phi-psi torsion angles for all residues in the structure
- By looking at how well the angles match up with expected distribution, the quality of a structure can be assessed.
- Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other sidechain types.
- The coloring/shading on the plot represents the various levels of favorability: the darkest areas (here shown in red) correspond to the "core" regions representing the most favorable combinations of phi-psi values.
- A properly folded protein will have over 90% of the residues in these "core" regions.
- The percentage of residues in the "core" regions is one of the better guides to stereochemical quality for assessing experimental protein structures.
- An ideal Ramachandran plot can be generated computationally using known atomic radii and bond distances.

alpha-helices: an 'island' of preferred conformation

91

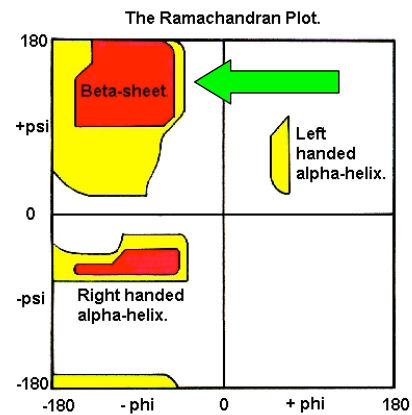
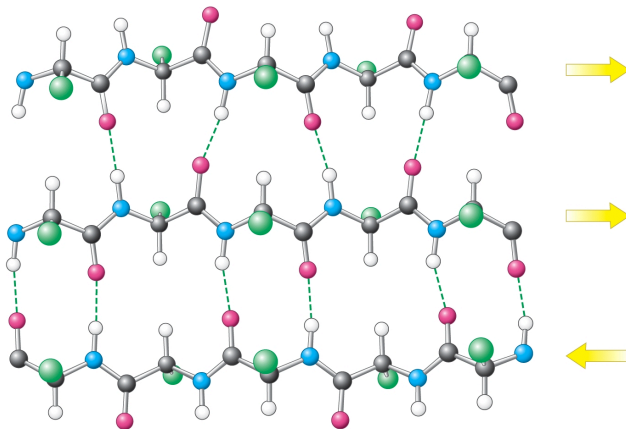
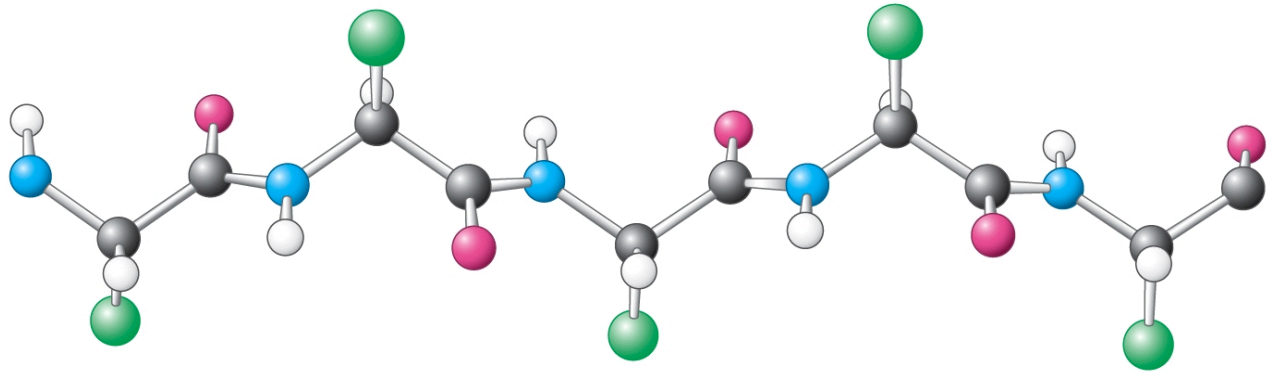


<http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html>

- As mentioned earlier, Pauling and Corey twisted models of polypeptides around to find ways of getting the backbone into regular conformations which would agree with experimental diffraction data (much like the way the structure of DNA was determined). The most simple and elegant arrangement is a right-handed spiral conformation known as the 'alpha-helix'.
- The structure repeats itself every 5.4 Angstroms along the helix axis, i.e. we say that the alpha-helix has a pitch of 5.4 Angstroms. Alpha-helices have 3.6 amino acid residues per turn, i.e. a helix 36 amino acids long would form 10 turns. The separation of residues along the helix axis is $5.4/3.6$ or 1.5 Angstroms, i.e. the alpha-helix has a rise per residue of 1.5 Angstroms.
- Every mainchain C=O and N-H group is hydrogen-bonded to a peptide bond 4 residues away ($O(i)$ to $N(i+4)$). This gives a very regular, stable arrangement.
- The peptide planes are roughly parallel with the helix axis and the dipoles within the helix are aligned. That is, all C=O groups point in the same direction and all N-H groups point the other way. This alignment of C=O and N-H bonds gives the alpha-helix a permanent dipole with a partial positive charge at the amino-terminus and a partial negative charge at the carboxy-terminus.
- Side chains point outward from helix axis and are generally oriented towards its amino-terminal end.
- All the amino acids have negative phi and psi angles, typical values being -60 degrees and -50 degrees, respectively

beta-strands: another 'island' of preferred conformation

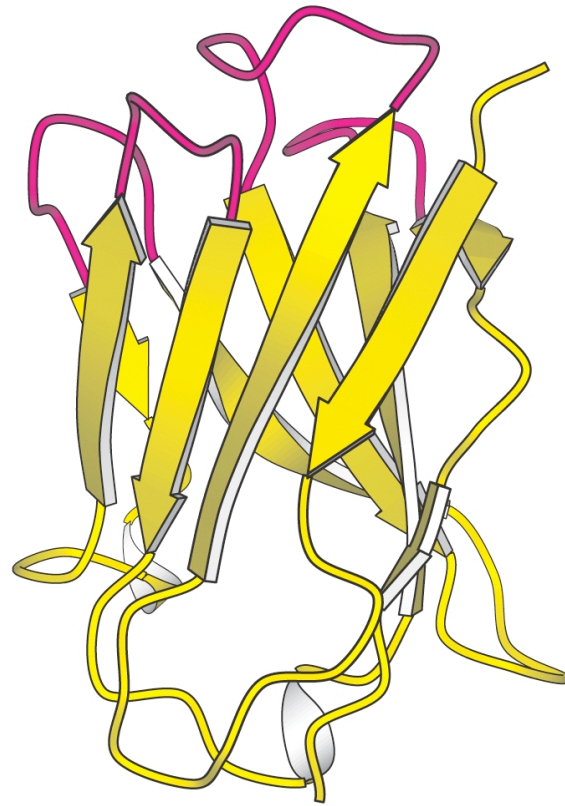
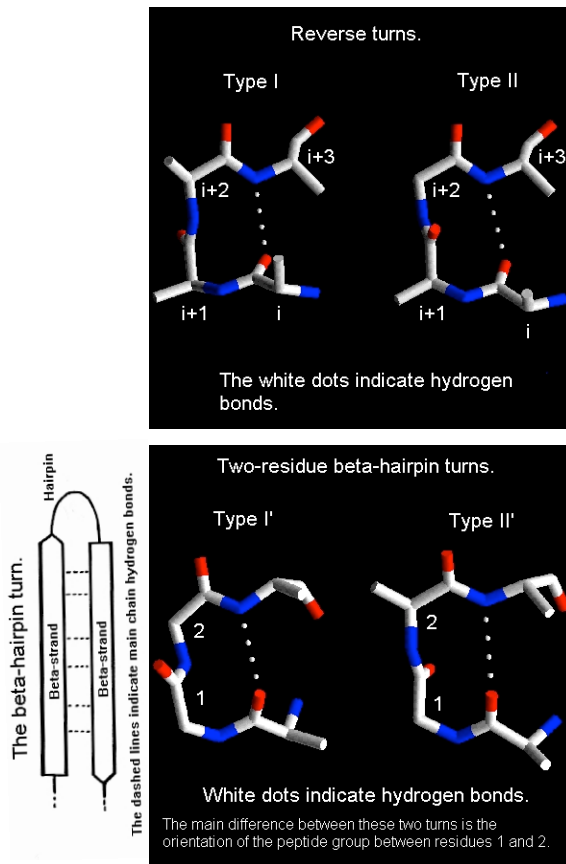
92



<http://www.cryst.bbk.ac.uk/PPS2/course/section3/rama.html>

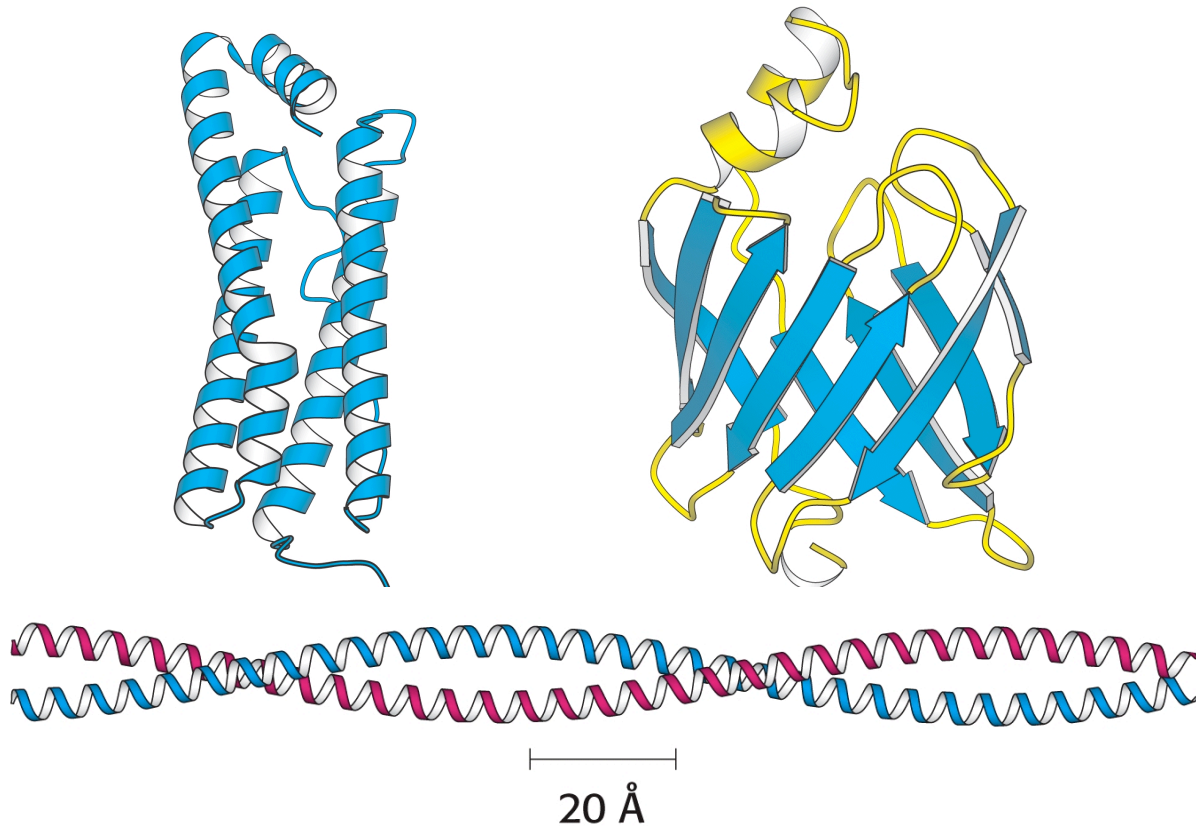
- In addition to the alpha helix, Pauling and Corey discovered another periodic structural motif which they named the β -pleated sheet (β because it was the second structure that they elucidated, the α helix being the first).
- The β -sheet differs markedly from the rodlike α -helix. A polypeptide chain, called a β -strand, in a β -sheet is almost fully extended rather than being tightly coiled as in the α -helix. A range of extended structures are sterically allowed. The side chains of adjacent amino acids point in opposite directions.
- A β -sheet is formed by linking two or more β -strands by hydrogen bonds. Adjacent chains in a β -sheet can run in opposite directions (antiparallel β -sheet) or in the same direction (parallel β -sheet).
- In the antiparallel arrangement, the NH group and the CO group of each amino acid are respectively hydrogen bonded to the CO group and the NH group of a partner on the adjacent chain.
- In the parallel arrangement, for each amino acid, the NH group is hydrogen bonded to the CO group of one amino acid on the adjacent strand, whereas the CO group is hydrogen bonded to the NH group on the amino acid two residues farther along the chain.
- Many strands, typically 4 or 5 but as many as 10 or more, can come together in β -sheets. Such β -sheets can be purely antiparallel, purely parallel, or mixed.
- β -sheets can be relatively flat but most adopt a somewhat twisted shape.

Turns and loops connect strands and helices ⁹³



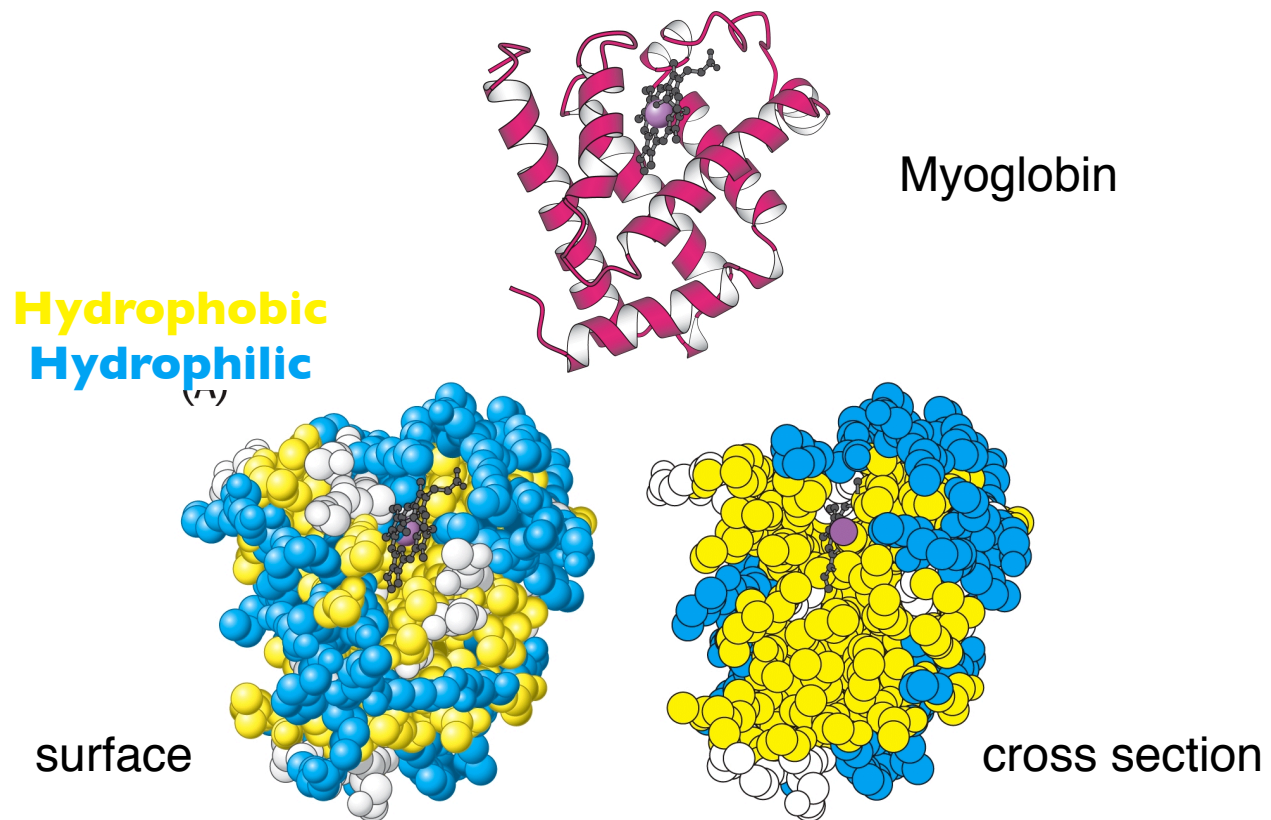
- Most proteins have compact, globular shapes, requiring reversals in the direction of their polypeptide chains. Many of these reversals are accomplished by reverse turns and hairpins.
- A reverse turn is region of the polypeptide having a hydrogen bond from one main chain carbonyl oxygen to the main chain N-H group 3 residues along the chain (i.e. O(i) to N(i+3)). Helical regions are excluded from this definition and turns between beta-strands form a special class of turn known as the beta-hairpin.
- Reverse turns are very abundant in globular proteins and generally occur at the surface of the molecule. It has been suggested that turn regions act as nucleation centers during protein folding.
- Reverse turns are divided into classes based on the phi and psi angles of the residues at positions i+1 and i+2. Types I and II shown in the figure are the most common reverse turns, the essential difference between them being the orientation of the peptide bond between residues at (i+1) and (i+2). The torsion angles for the residues (i+1) and (i+2) in the two types of turn lie in distinct regions of the Ramachandran plot.
- 2-residue beta-hairpin turns occur between two antiparallel beta-strands as shown in the figure.
- The residues forming these two-residue turns have torsion angles in characteristic regions of the Ramachandran plot.
- For type I' turns, residue 2 is always glycine whereas for type II' turns residue 1 is always Gly. This is because amino acids other than glycine would cause steric hindrance involving the residue's side chain and the main chain.
- In other cases, more elaborate structures are responsible for chain reversals. These structures are called loops or sometimes Ω loops (omega loops) to suggest their overall shape. Unlike α -helices and β -strands, loops do not have regular, periodic structures. Nonetheless, loop structures are often rigid and well defined. Turns and loops invariably lie on the surfaces of proteins and thus often participate in interactions between proteins and other molecules.
- For example, a part of an antibody molecule has surface loops (shown in red) that mediate interactions with other molecules.
- *Question: Does reverse turn only exists among alpha-helix? And beta-hairpin only exists among beta-sheet?*
- *Answer: As the name implies, the beta hairpin is most commonly found as a connector between strands of an antiparallel beta sheet. The reverse turn is a bit more general and can be found in loops that connect both helices and strands.*

Proteins generally composed of α -helices and/or β -sheets connected by turns and loops



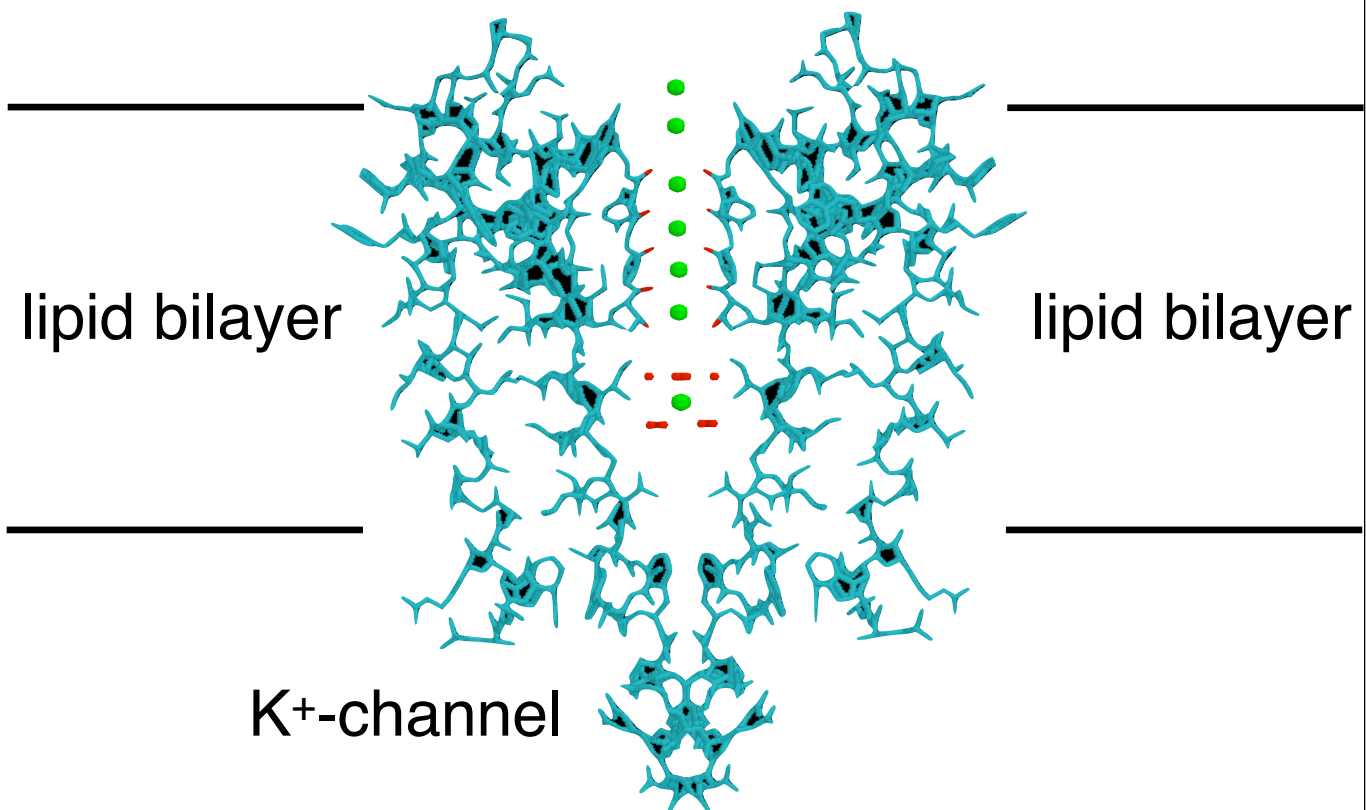
- The α -helical content of proteins ranges widely, from nearly none to almost 100%. For example, about 75% of the residues in ferritin, a protein that helps store iron, are in α -helices. Single α -helices are usually less than 45 Å long. However, two or more α -helices can entwine to form a very stable 'coiled coil' structure, which can have a length of 1000 Å (100 nm, or 0.1 μ m) or more. Such α -helical 'coiled coils' are found in myosin and tropomyosin in muscle, in fibrin in blood clots, and in keratin in hair. The helical cables in these proteins serve a mechanical role in forming stiff bundles of fibers, as in porcupine quills. The cytoskeleton (internal scaffolding) of cells is rich in so-called intermediate filaments, which also are two-stranded α -helical coiled coils. Many proteins that span biological membranes also contain α -helices.
- The β -sheet is an important structural element in many proteins. For example, fatty acid-binding proteins, important for lipid metabolism, are built almost entirely from β -sheets.

Protein folding is largely driven by hydrophobic interactions



- Myoglobin, the oxygen carrier in muscle, is a single polypeptide chain of 153 amino acids. The capacity of myoglobin to bind oxygen depends on the presence of heme, a prosthetic (helper) group consisting of protoporphyrin IX and a central iron atom.
- The folding of the main chain of myoglobin, like that of most other proteins, is complex and devoid of symmetry. A unifying principle emerges from the distribution of side chains. The striking fact is that the interior consists almost entirely of nonpolar residues such as leucine, valine, methionine, and phenylalanine. Charged residues such as aspartate, glutamate, lysine, and arginine are absent from the inside of myoglobin. The only polar residues inside are two histidine residues, which play critical roles in binding iron and oxygen.
- The outside of myoglobin, on the other hand, consists of both polar and nonpolar residues. This contrasting distribution of polar and nonpolar residues reveals a key facet of protein architecture. In an aqueous environment, protein folding is driven by the strong tendency of hydrophobic residues to be excluded from water.
- The polypeptide chain therefore folds so that its hydrophobic side chains are buried and its polar, charged chains are on the surface.
- The secret of burying a segment of main chain in a hydrophobic environment is pairing all the NH and CO groups by hydrogen bonding. This pairing is neatly accomplished in an α -helix or β -sheet.
- The ability to predict whether or not a given polypeptide sequence will fold into a given tertiary structure remains one of the 'grand challenges' of science.
- In nature, protein fold either independently or with the help of other proteins known as chaperones.

Membrane proteins have grease on the outside⁹⁶

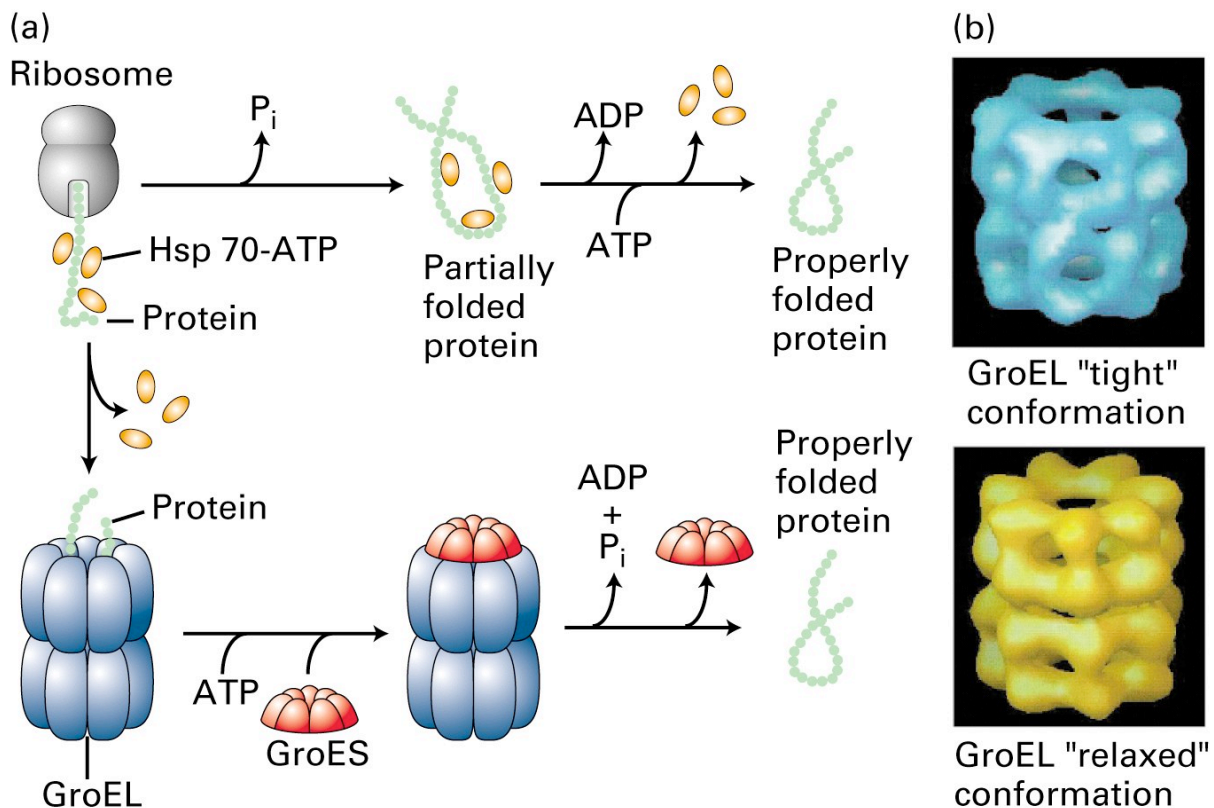


David S. Goodsell: The Molecule of the Month appearing at the [PDB](#)

- Some proteins that span biological membranes are “the exceptions that prove the rule” regarding the distribution of hydrophobic and hydrophilic amino acids throughout three-dimensional structures. For example, ion channels are covered on the outside largely with hydrophobic residues that interact with the neighbouring alkane chains. The inner channel is quite polar and there are many specific interactions with the ion being transported.
- The remarkable ability of the potassium channel to pass only potassium ions is accomplished by a selectivity filter at one end of the pore, as shown here from PDB entry 1k4c. For clarity, only two of the four protein molecules, one either side of the channel, are shown in a stick representation. The green spheres are potassium ions passing through the selectivity filter.
- Normally potassium ions float around encased in a cushion of water, like the one at the bottom of the stack of ions. Notice that it is surrounded by eight water molecules, shown as red spheres. In order to pass through the selectivity filter, each potassium ion has to shed these water molecules. This is how the selectivity filter works: the dimensions of the channel are designed to mimic this shell of water. Protein oxygen atoms that line the pore (colored in red) are oriented toward the center of the channel. Eight of these oxygen atoms surround each potassium ion, and act as a perfect replacement for the normal layer of water molecules.
- The 2003 Nobel Prize in Chemistry was awarded for work in the area of channels
 - Roderick Mackinnon pioneered x-ray crystallography of ion channels.
 - Peter Agre discovered water channels.
- Water channels facilitate the rapid transport of water across cell membranes in response to osmotic gradients. These channels are believed to be involved in many physiological processes that include renal water conservation, neuro-homeostasis, digestion, regulation of body temperature and reproduction. Members of the water channel superfamily have been found in a range of cell types from bacteria to human.

Chaperone assisted protein folding

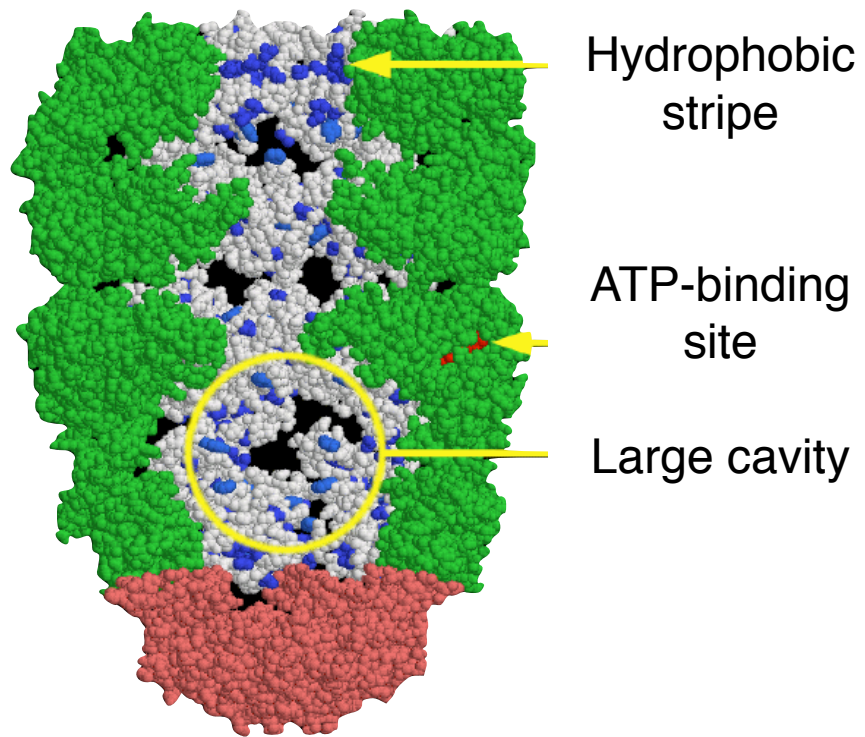
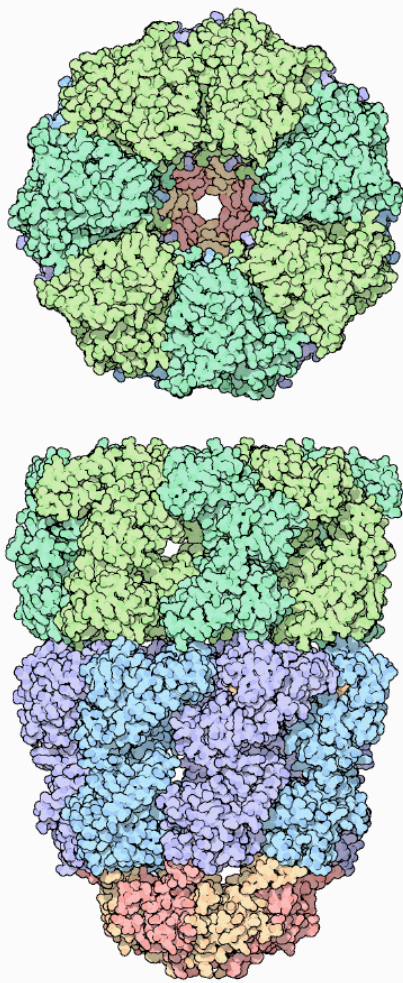
97



<http://www.users.csbsju.edu/~hjakubow/classes/rasmolchime/02ch331finproj/GroEL2/templateprotGROEL2.htm>

- Folding of proteins in vitro tends to be an inefficient process, with only a minority of unfolded molecules undergoing complete folding within a few minutes.
- More than 95 percent of the proteins present in cells are in their native conformation.
- The explanation for the cell's remarkable efficiency in promoting protein folding probably lies in chaperones, a family of proteins found in all organisms from bacteria to humans.
- There are two general families of chaperones: molecular chaperones, which bind and stabilize unfolded or partially folded proteins, thereby preventing these proteins from being degraded; and chaperonins, which directly facilitate their folding.
- Chaperonins are probably used for a specific and relatively small selection of proteins, whereas molecular chaperones are used for most, if not all, proteins.
- All chaperones have ATPase activity, and their ability to bind and stabilize their target proteins is specific and dependent on ATP hydrolysis.
- Molecular chaperones include the Hsp70 family of proteins. When bound to ATP, Hsp70 assumes an open form in which an exposed hydrophobic pocket transiently binds to exposed hydrophobic regions of the unfolded target protein. Hydrolysis of the bound ATP causes Hsp70 to assume a closed form, releasing the target protein. Molecular chaperones are thought to bind all nascent polypeptide chains as they are being synthesized on ribosomes.

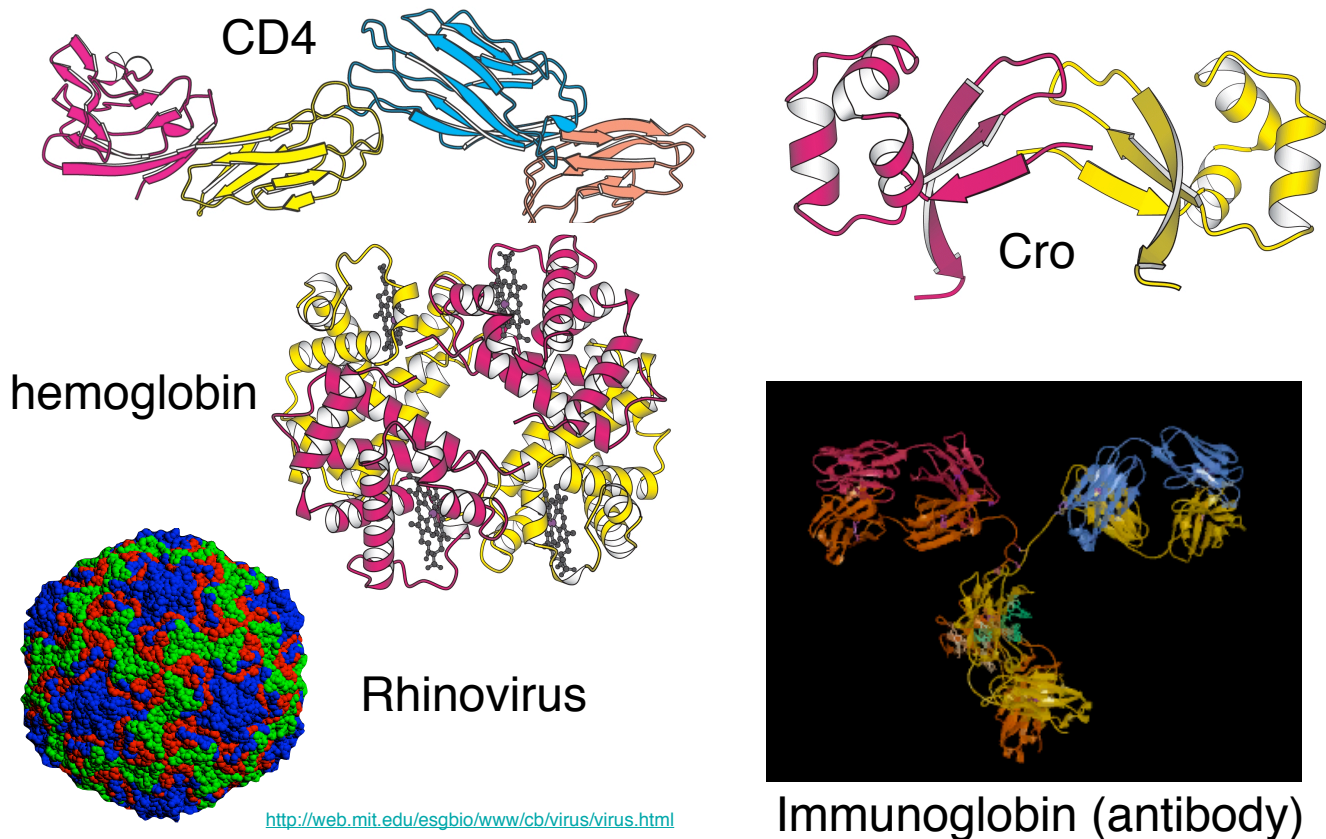
More on GroEL



David S. Goodsell: The [Molecule of the Month](#) appearing at the [PDB](#)

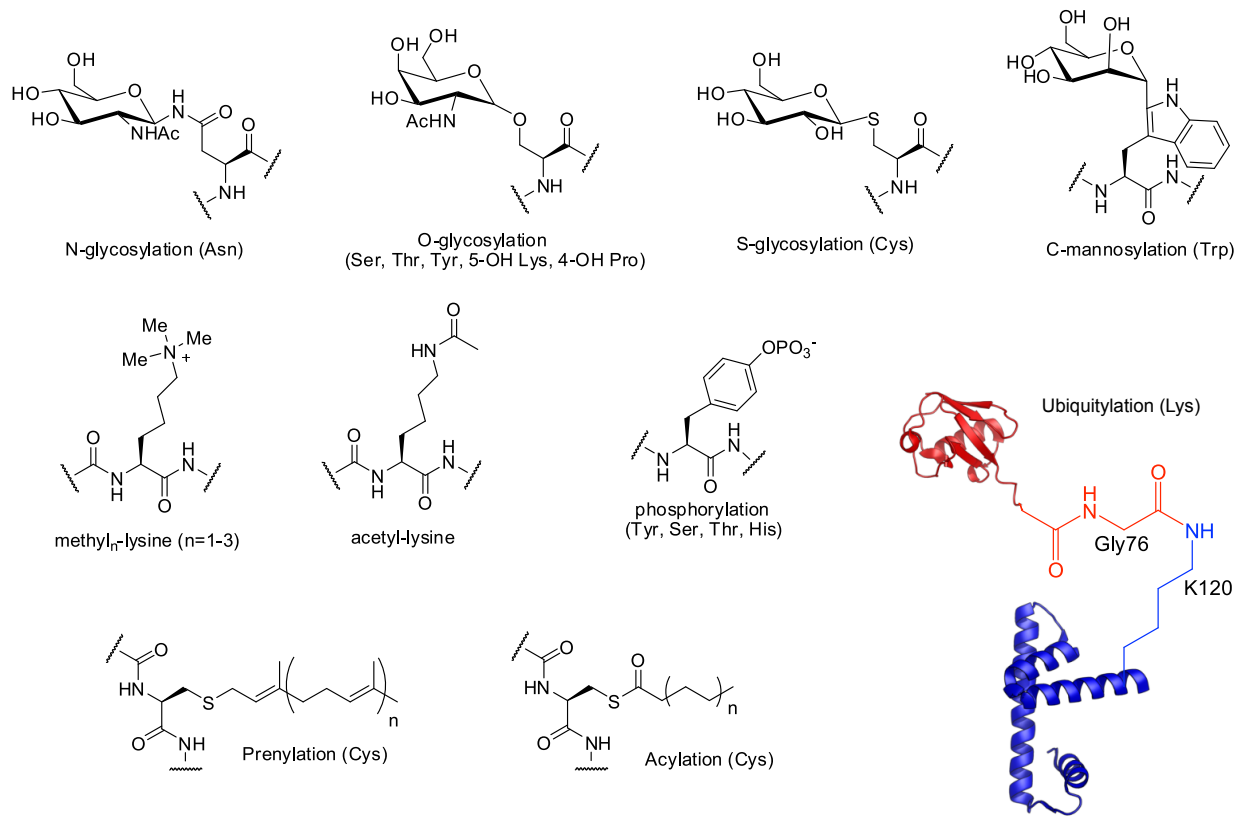
- Proper folding of a small proportion of proteins (e.g., the cytoskeletal proteins actin and tubulin) requires additional assistance, which is provided by chaperonins.
- Shown on this slide is the bacterial chaperonin, GroEL, which contains 14 identical subunits stacked in two concentric rings (green). GroES is shown at the bottom in pink.
- The large GroEL-GroES complex is available in PDB entry 1aon. In this picture, three of the subunits in each GroEL ring have been removed to show the interior, leaving four subunits in each ring. On the two in back, the hydrophobic amino acids, LEU, ILE, VAL, MET, PHE, TYR and TRP, are coloured blue.
- Notice the stripe of hydrophobic amino acids around the entry at the top. This will interact strongly with unfolded proteins by coaxing them into the upper cavity. Once the unfolded protein is bound, ATP and GroES bind to GroEL. This causes a conformational change that forces the protein into the larger lower cavity that is much more hydrophilic than the upper cavity.
- Now that the protein is in a hydrophilic environment, it will be forced to fold in order to minimize they unfavourable interactions between its hydrophobic portions and its hydrophilic environment.
- After the protein has folded, ATP is hydrolyzed and GroES (the lid on the cavity) is released along with the newly folded protein.
- *Q: When use chaperonin to help proteins to fold, the GroES will bind to GroEL to the large cavity side or hydrophobic stripe side?*
- *A: I believe it can bind to both sides. Don't worry about the details.*

Proteins often consist of multiple independent⁹⁹ domains and have 4° structure



- Some polypeptide chains fold into two or more compact regions that may be connected by a flexible segment of polypeptide chain, rather like pearls on a string.
- These compact globular units, called domains, range in size from about 30 to 400 amino acid residues.
- For example, the extracellular part of CD4 (shown at top), the cell-surface protein on certain cells of the immune system to which the human immunodeficiency virus (HIV) attaches itself, comprises four similar domains of approximately 100 amino acids each. Often, proteins are found to have domains in common even if their overall tertiary structures are different.
- Antibodies (immunoglobins) have a distinct domain structure in addition to quaternary structure. We will be taking a much closer look at antibody structure in the next section.
- Quaternary structure refers to the spatial arrangement of subunits and the nature of their interactions.
- The simplest sort of quaternary structure is a dimer, consisting of two identical subunits. This organization is present in the DNA-binding protein Cro found in a bacterial virus called λ .
- More complicated quaternary structures also are common. More than one type of subunit can be present, often in variable numbers. For example, human hemoglobin, the oxygen-carrying protein in blood, consists of two subunits of one type (designated α) and two subunits of another type (designated β). Thus, the hemoglobin molecule exists as an $\alpha_2\beta_2$ tetramer.
- Viruses make the most of a limited amount of genetic information by forming coats that use the same kind of subunit repetitively in a symmetric array. The coat of rhinovirus, the virus that causes the common cold, includes 60 copies each of four subunits. The subunits come together to form a nearly spherical shell that encloses the viral genome.
- *Q: It mentioned that the coat of rhinovirus includes 60 copies each of four subunits. But from the picture I only see three coloured subunits. What's wrong in this.*
- *A: There is a 4th protein that is inside and not visible from the outside.*

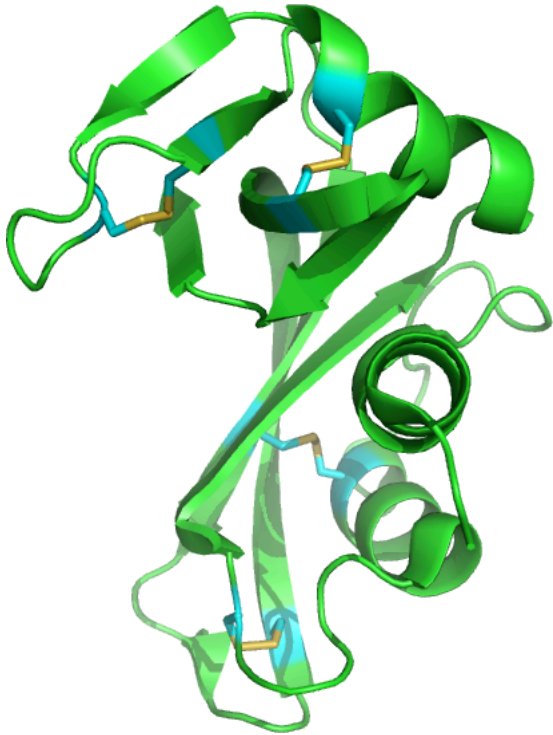
Post-translational modifications of proteins



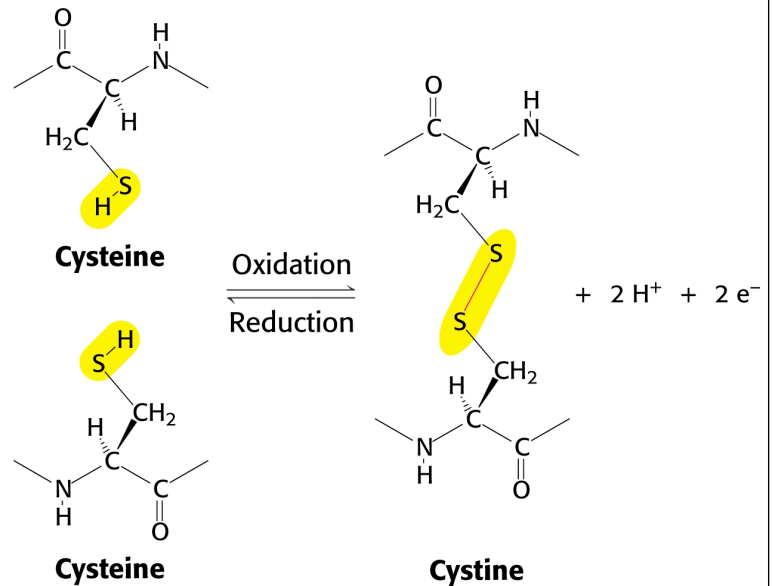
Courtesy of Spencer Alford

- Many proteins are covalently modified, through the attachment of groups other than amino acids, to augment their functions. Many proteins, especially those that are present on the surfaces of cells or are secreted, acquire carbohydrate units on specific asparagine residues. The addition of sugars makes the proteins more hydrophilic and able to participate in interactions with other proteins. Conversely, the addition of a fatty acid to an α -amino group or a cysteine sulfhydryl group produces a more hydrophobic protein that will be tightly associated with the membrane.
- Proteins can also be reversibly modified to regulate their activity. Perhaps the single most important of all modifications is phosphorylation and dephosphorylation of serine, threonine, and tyrosine residues. Regulation of protein activity by phosphorylation is basis for intracellular signalling. The enzymes that catalyze the addition of phosphate groups (from ATP donors) are called kinases (why kinases?). Enzymes that remove phosphate groups are called phosphatases.
- Histones—proteins that assist in the packaging of DNA into chromosomes as well as in gene regulation—are rapidly acetylated and deacetylated on specific lysine residues in vivo. More heavily acetylated histones are associated with genes that are being actively transcribed. A more permanent modification of lysines in histone proteins is methylation.
- The attachment of ubiquitin, a protein comprising 72 amino acids, is a signal that a protein is to be destroyed, the ultimate means of regulation.
- This slides shows only a few of the common examples. A number of additional post-translational modifications are known.

Disulphide bonds are another type of post-translational modification ¹⁰¹

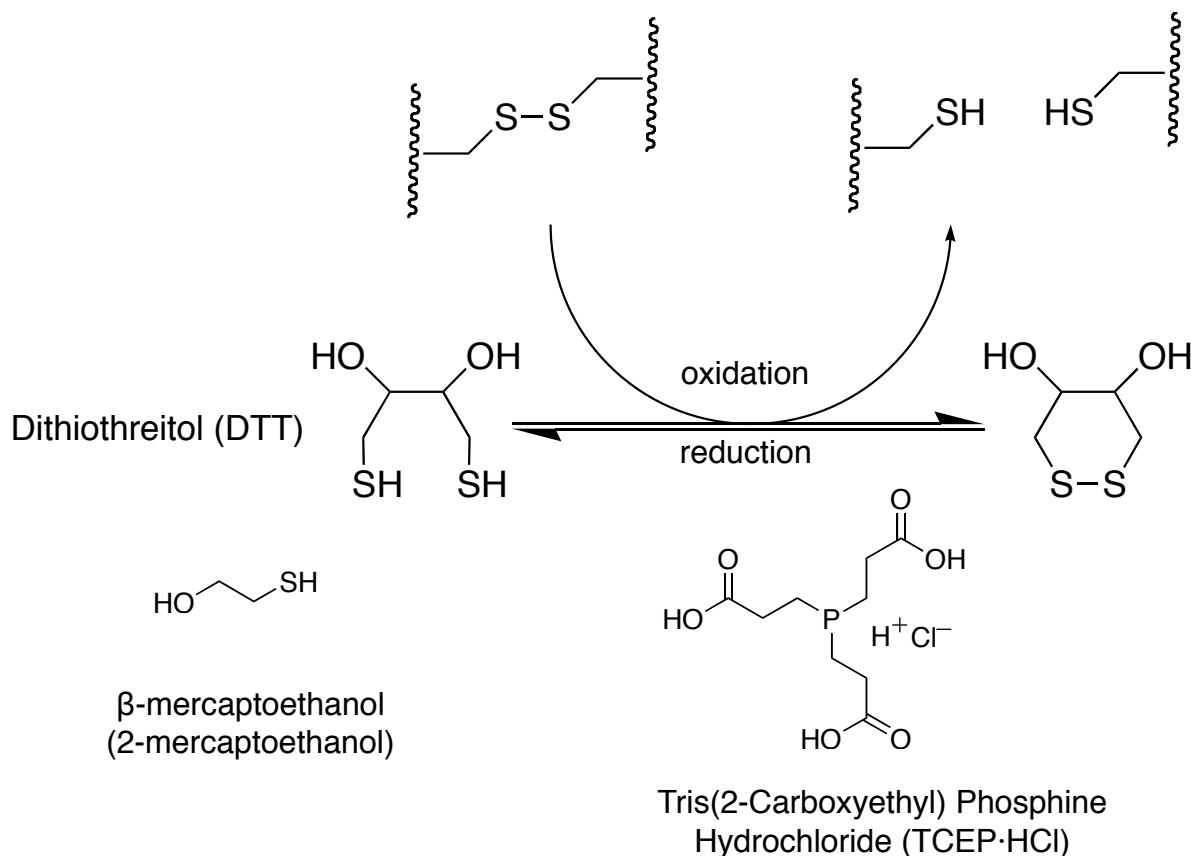


Ribonuclease
(PDB 5RSA)



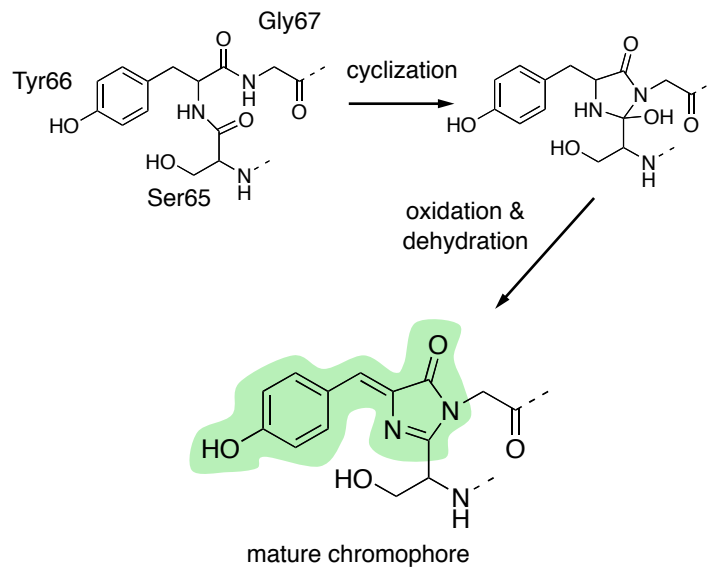
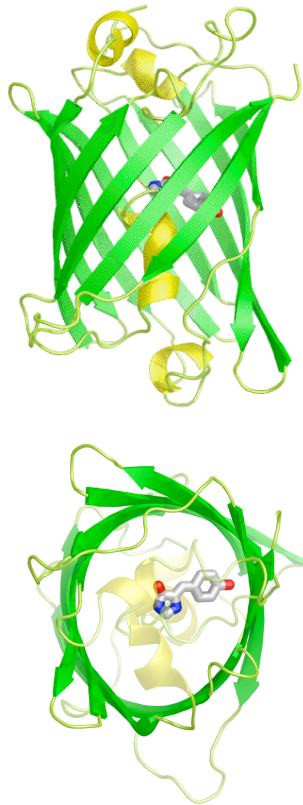
- Disulphide bonds generally only form under oxidizing conditions. They are very common in extracellular proteins such as those found in blood.
- Disulphide bonds generally do not form in the cytoplasm of living cells because this is a reducing environment (~5 mM glutathione, a thiol)
- A protein that has disulphide bonds will tend to be very stably folded relative to proteins that do not have disulphide since there is a covalent link holding two sections of chain together (as opposed to non-covalent interactions in a protein that lacks disulfide bonds). You would expect a protein with disulphide bonds to be more stable at elevated temperatures than a protein without disulphide bonds.
- To maintain the activity of a protein with free thiols in vitro, it is important to add reducing agents such as β -mercaptoethanol, TCEP, or DTT.

In vitro, reducing agents are necessary to maintain thiols in a reduced form



- Don't use β-mercaptoethanol (at least not in this building) if you can help it. It stinks
- DTT or TCEP are much better choices. These two reducing agents are similar in terms of effectiveness. TCEP can be slightly more expensive but it benefits from not having a free thiol (a good nucleophile). This is advantageous if you are working with electrophiles in solution.

The chromophore of green fluorescent protein¹⁰³ is a unique post-translational modification

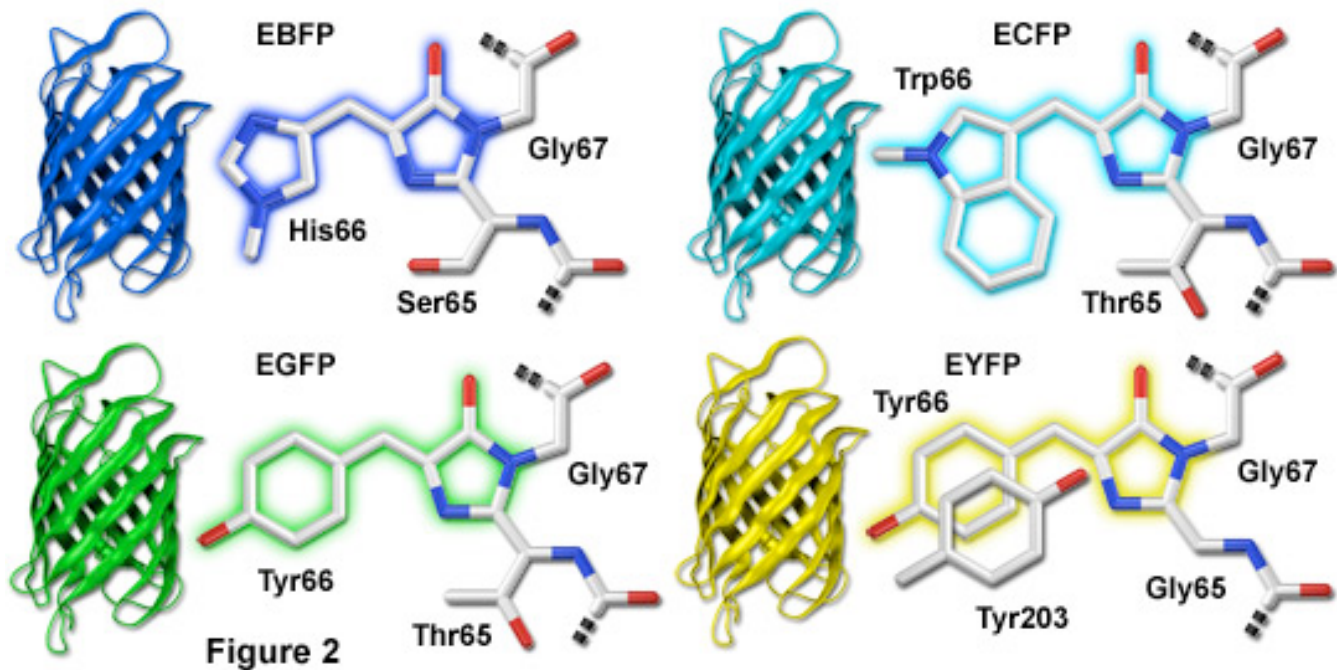


MSKGEELFTGVVPILVELDGDVNGQKFSVSGEGDATYGLTLKFICTTGKLPVPWPTL
 VTTFSYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFYKDDGNYKTRAEVKFEGLTLV
 NRIELKGIDFKEDGNILGHKMEYNYNSHNVIYIMADKPKNGIKVNFKIRHNIKDGSVQLAD
 HYQQNTPIGDGPVLLPDNHYLSTQSALS KDPNEKRDMILFEFVTAAGITHGMDELYK

Prasher et al. *Gene* (1992) 111: 229-233.
 Tsien, *Annu. Rev. Biochem.* (1998) 67: 509-544.

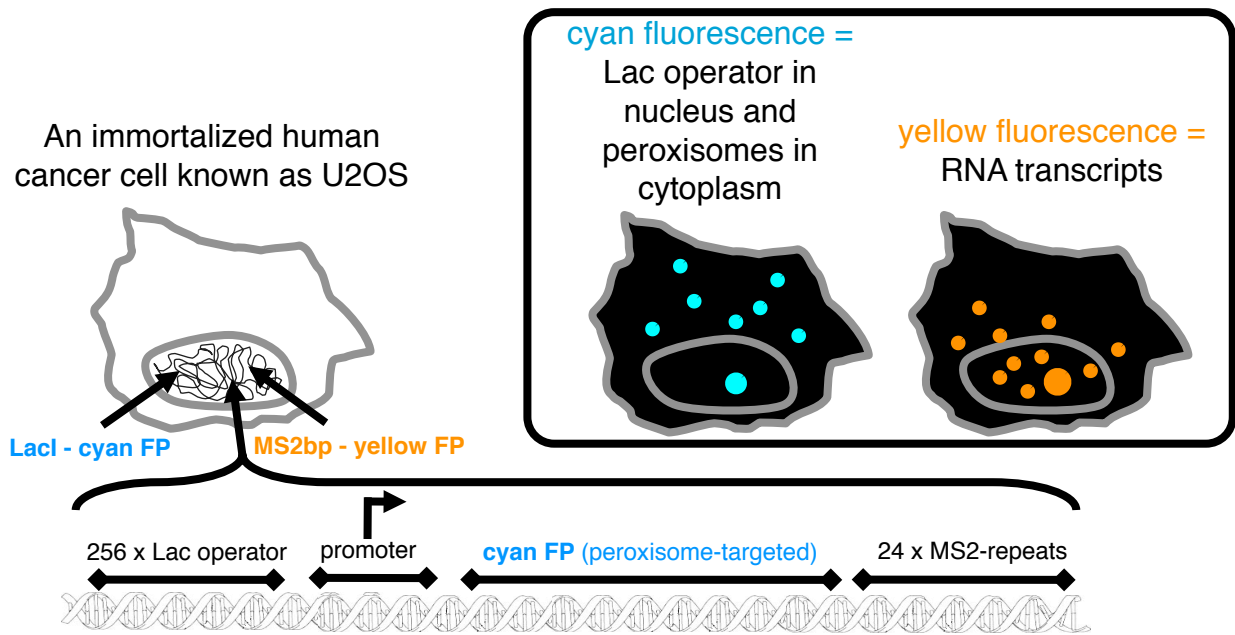
- Yet another type of post-translational modification are those that are generated by chemical rearrangements of side chains and, sometimes, the peptide backbone. For example, the *Aequorea Victoria* jellyfish produces a green fluorescent protein (GFP). Surprisingly, the chromophore is not a bound cofactor but rather a post-translationally modified sequence of amino acids.
- The chromophore (also okay to call fluorophore) is formed by the protein-promoted rearrangement and oxidation of the sequence Ser-Tyr-Gly within the center of the protein. The GFP is of tremendous utility to researchers as a marker within cells. The Nobel prize in Chemistry for 2008 was awarded to 3 researchers (Tsien, Chalfie, and Shimomura) who are considered pioneers in the development of GFP as a research tool.
- The spontaneous formation of the *Aequorea* green fluorescent protein chromophore within the folded beta-barrel protein structure must necessarily involve at least three key steps: cyclization of the main chain, loss of a molecule of water (dehydration), and oxidation with molecular oxygen. The exact order and mechanism of these steps is a matter of ongoing investigation.
- Chromophore formation is spontaneous only within the context of the fluorescent protein -barrel structure where steric constraints force the peptide into a tight turn conformation (Branchini et al. 1998) and the side chains of highly conserved residues, such as glutamate 222 and arginine 96, are positioned to facilitate the reaction.

Fluorescent proteins engineered to fluoresce at different wavelengths



- Note that, because the GFP chromophore is generated from amino acids, we can change the structure of the chromophore by introducing mutations in the gene that ultimately change the amino acids present in the chromophore forming tripeptide (or its immediate surroundings).
- For example, by changing the tyrosine of the GFP chromophore to other aromatic amino acids, new chromophore structures can be formed that fluoresce at different wavelengths from the wild-type protein.

Visualizing the central dogma in a live cell! ¹⁰⁵



This cell has 3 different genes artificially introduced into it:

- LacI - cyan FP, the gene for LacI fused to the gene for a cyan fluorescent protein
- MS2bp - yellow FP, the gene for MS2bp fused to the gene for yellow fluorescent protein
- A peroxisome targeted cyan FP followed by 24 copies of MS2 in the 3'UTR. This gene is preceded by 256 copies of the Lac operator which is not transcribed.

Definitions:

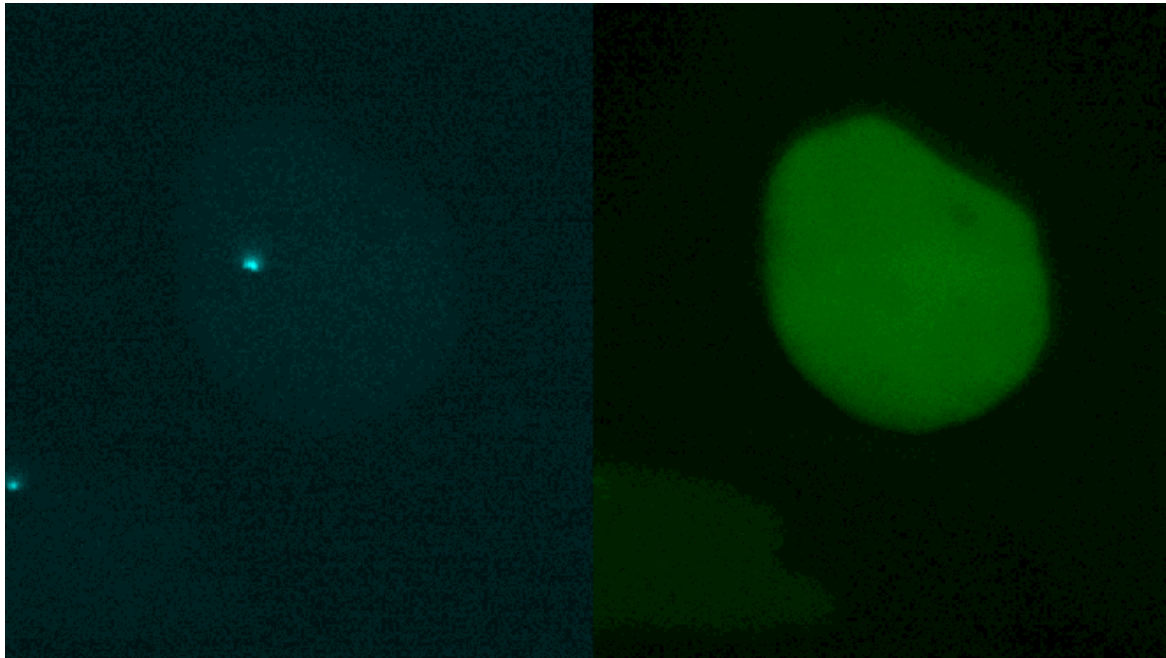
- U2OS cell line: a human cell line cultivated from the bone tissue of a fifteen-year-old human female suffering from osteosarcoma (the same cancer as Terry Fox had).
- LacI: a bacterial protein that binds tightly to the section of DNA known as the Lac operator.
- Lac operator: a DNA sequence that binds to LacI.
- MS2: an RNA sequence that forms a hairpin structure. This RNA hairpin binds tightly to the viral protein that I have called MS2bp.
- MS2bp (MS2-binding protein): a viral coat protein that binds tightly to the MS2 RNA sequence.
- cyan FP: the cyan fluorescent variant of the green fluorescent protein
- yellow FP: the yellow fluorescent variant of the green fluorescent protein
- peroxisomes: small organelles in the cytoplasm of eukaryotic cells. They have a role in destroying peroxides in the cell.
- This is the work of the labs of David L. Spector and Robert H. Singer
- *Question: This is about the topic of visualizing the central dogma in the class today. In the example of the sequence having 254 copies of Lac operator and 24 copies of MS2, the figure shows both the cyan and yellow fluorescence separately. But will there be more of the cyan fluorescence actually noticed since the sequence has 254 copies of Lac and 24 copies of MS2? But the figure actually shows more number of yellow spots compared to cyan in the nucleus. Could you please explain if there will be any differences?*
- *Answer: All of the copies of the Lac operator DNA sequence are located in the same place and there is only one 'copy' of the 254 copies in the cell (i.e., there is only one copy of the genome). There are many, many RNA transcripts created by transcription of this DNA sequence. As we saw earlier in the course, one DNA sequence can be read by many RNA polymerases over and over again in order to produce many mRNA molecules. Each of these transcripts carries the 24 copies of MS2 and thus shows up as a yellow fluorescent spot. You would expect each yellow spot (corresponding to one transcript) to be substantially dimmer than the cyan spot in the nucleus. Each transcript has 24 fluorophores attached to it, while the DNA sequence has 254 fluorophores attached to it.*

Visualizing the central dogma in a live cell! ¹⁰⁶

example 1

cyan fluorescence = Lac operator in nucleus and peroxisomes in cytoplasm

yellow fluorescence = RNA transcripts



- Imaging gene expression in single living cells *Nat Rev Mol Cell Biol* 5(10):855-862 (2004 October)
- Visualizing gene expression in living cells. The movie shows a cell with a stably integrated gene that also contains 256 lac operator repeats. This gene transcribes an RNA that contains both a coding sequence for the cyan fluorescent protein (CFP) protein (with a peroxisome-targeting sequence) and a stretch of MS2 stem-loops. In the beginning of the movie, the gene locus is visible as a result of tagging of the DNA with a CFP–lac-repressor protein. Once transcription is induced from this gene, the locus becomes structurally open and decondenses. The RNAs produced from the gene are tagged with yellow fluorescent protein (YFP)–MS2 and can be seen accumulating at the transcription site. The RNA is translated in the cytoplasm and at later times post-induction, CFP-labelled peroxisomes are detected. The cell was imaged every 2.5 min for a total of 4 hr and 22.5 min.
- http://singerlab.aecom.yu.edu/supplements/natrevmcb_v5p855/movies03.htm
- Q: The lac operator is connected to the DNA sequence of the cyan FP, and when transcription happens, cyan FP will be generated and the lac operator will not be transcribed? And the lac operator will not be replicated in the nucleus?
- A: Transcription starts from the promoter and so only things that come after the promoter, specifically the cyan FP (peroxisome targeted) and the 24 x MS2 repeats, will be transcribed. The 256 lac operator copies comes before the promoter, so it is not transcribed. The 256 lac operator are just there to serve as a binding site for Lac I- cyan FP, so that the location of the inserted DNA sequence can be visualized by fluorescence imaging. All of the DNA will be replicated when the cell divides, but otherwise there will be just one copy of the DNA in the cell.
- Q: The MS2 is an RNA sequence and it is connected to the DNA sequence of the yellow FP? What is the function of the MS2 here and can it be replicated?
- A: 'MS2' is an RNA sequence that forms a hairpin structure. 'MS2 binding protein' (MS2bp) is a protein that binds to the MS2 hairpin. The cell contains the gene for MS2bp fused to yellow FP. When this fusion protein is made in the cell, it will stick to the RNA molecules that contain the 24 x MS2 repeats and allow the RNA molecules to be visualized by fluorescence imaging.
- Q: Is one lac operator connected to one cyan FP?
- A: The lac operator isn't connected to anything. It is just a DNA sequence that is not transcribed. Lac Inhibitor (LacI) is a protein that binds to the lac operator. The cell contains the gene for LacI fused to cyan FP. When this protein is made, 256 copies of it will stick to the 256 repeats of the lac operator in the DNA.
- Q: Will the content of visualizing the central dogma in a live cell in lecture 4 be included in the exam?
- A: Everything that was covered in class and/or is in assigned reading could be included in the exam.

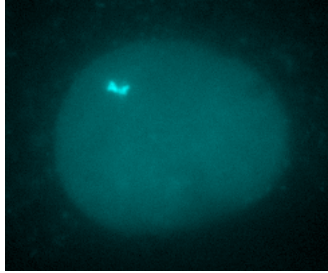
Visualizing the central dogma in a live cell! ¹⁰⁷

example 2

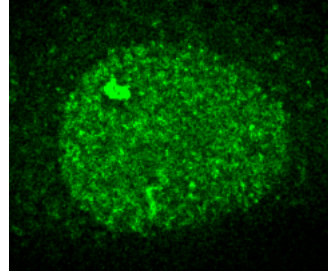
cyan fluorescence = Lac operator in nucleus and peroxisomes in cytoplasm

yellow fluorescence = RNA transcripts

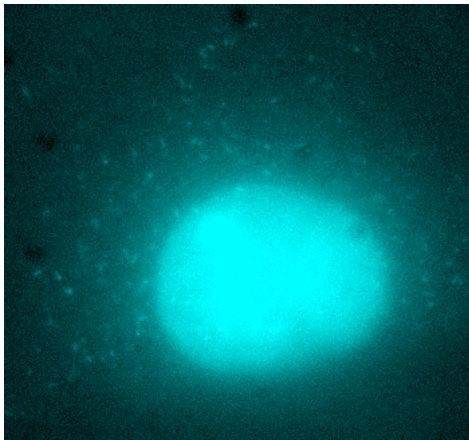
1



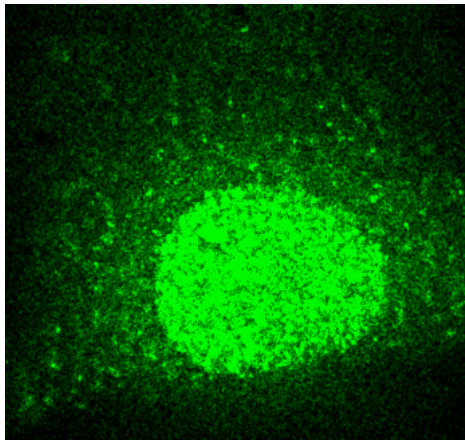
3



2



4



- Dynamics of Single mRNPs in Nuclei of Living Cells *Science* 304(5678):1797-1800 (2004 June 18)
- 1. Detection of open gene locus by CFP-lac repressor.
- 2. Detection of cytoplasmic CFP-peroxisomes (different plane than 1).
- 3. Detection of YFP-MS2 nuclear mRNPs and YFP-MS2 accumulation at the transcription site.
- 4. Different threshold of same cell showing cytoplasmic YFP-MS2 mRNPs.
- An important conclusion from this work is that RNA transcripts are freely diffusing inside of the nucleus. This is different than in the cytoplasm where they are actively transported.
- http://singerlab.aecom.yu.edu/supplements/science_v304p1797/movies.htm
- Q: In the example of "visualizing the central dogma in a living cell", you mentioned that we tag the MS2 RNA with the YFP. My question is that where does the YFP comes from? I found there is no DNA sequence that represent for the YFP
- A: The cell also contains a gene encoding MS2bp-YFP