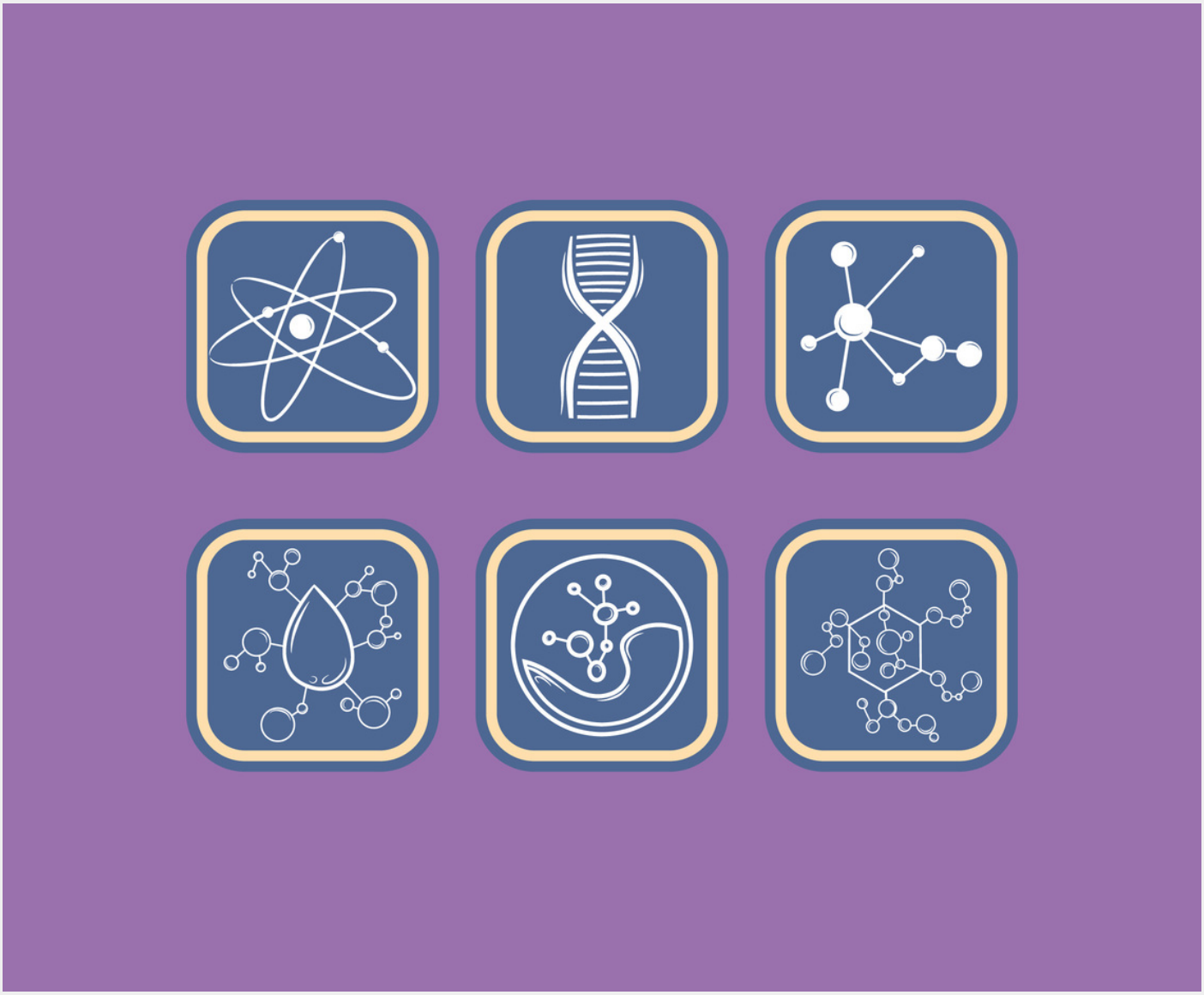
Biochemistry II

Notes from lectures no textbook required



Bioinformatics

A.A. 2021 Semester II

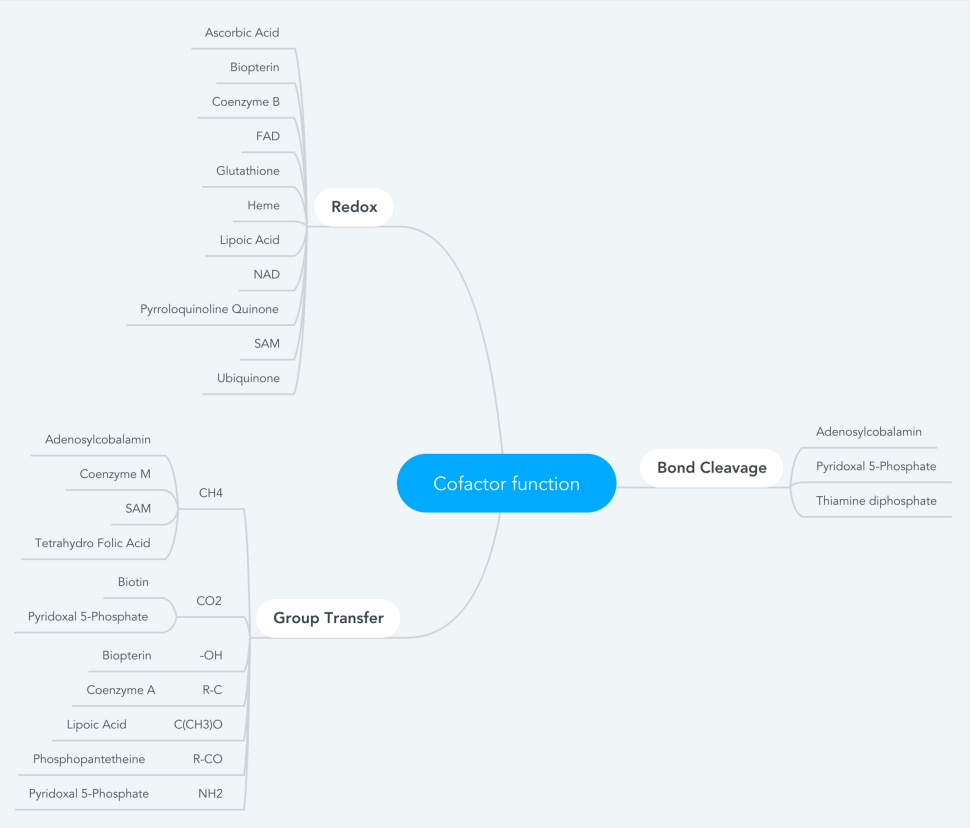
Prof. Alessandro Paiardini

Table Of Contents

Syllabus

- ❏ Amino Acids - physico-chemical properties and structure
- ❏ Visualizing Macromolecules with molecular viewers
- ❏ Protein Sequence Databases
- ❏ Protein Sequence Alignments
- ❏ The Peptide Bond in vitro and in Vivo
- ❏ Secondary Structure
- ❏ Tertiary Structure - Protein Stability and Folding
- ❏ Domains, Motifs, Modules and Repeats - Quaternary Structure
- ❏ Electron Microscopy and X-ray crystallography
- ❏ Nuclear Magnetic Resonance of Proteins
- ❏ Protein Structure Prediction
- ❏ Molecular Dynamics and Drug Design
- ❏ Selected examples of Protein mechanisms

Exam will be oral with written exercises on PC BUT there will be two exceptions, which are written exercises, if we pass both, no oral, or optional oral to improve, exemption grades don't expire



Amino Acids

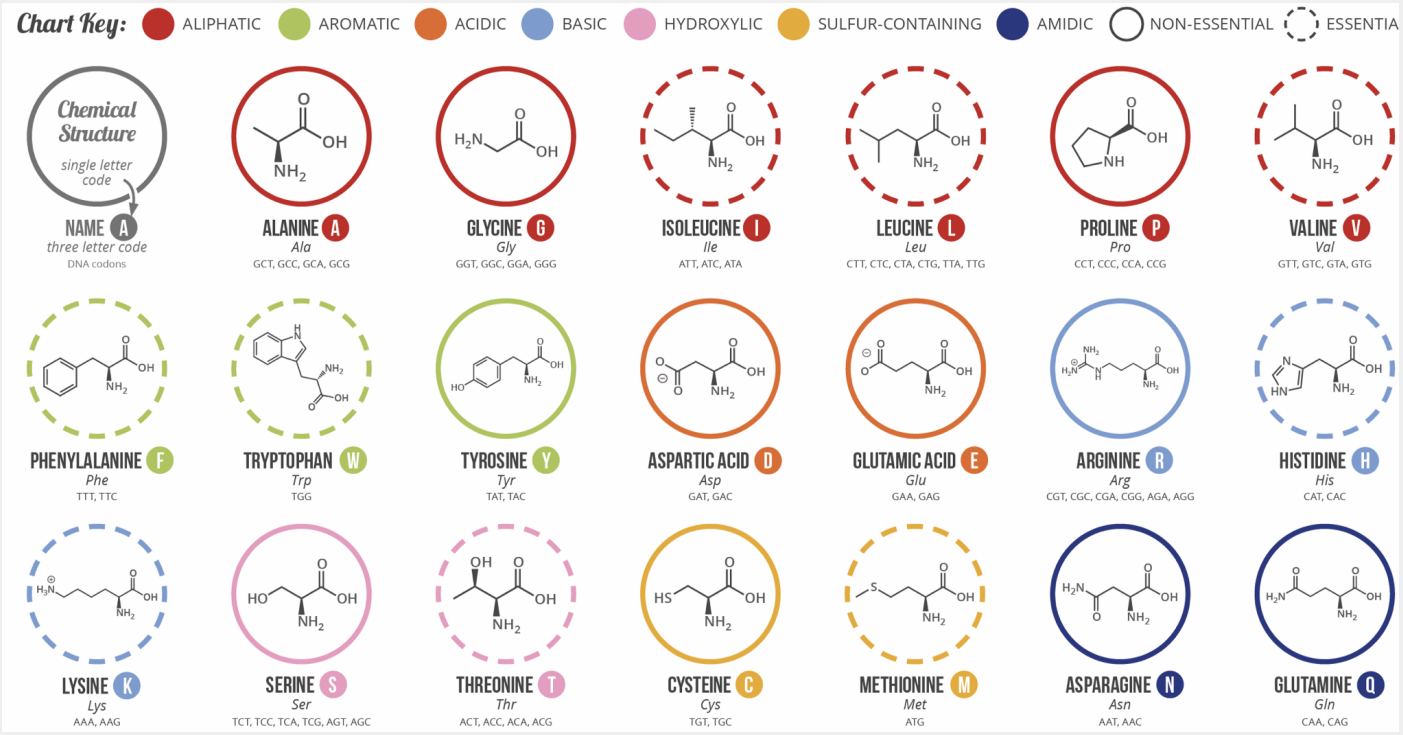
Amino acids have a carboxyl(-COOH) and an amino group (-NH2) around a chiral carbon. In biological systems we have the carboxyl group deprotonated (COO- pK 3) and the protonated amino (-NH3 pK 9). This means in neutral pH we find mainly positively charged amino acids from the amine group. While there are 20 proteinogenic amino acids, there are actually more than 500 involved in various processes in the cell such as carnitine, GABA ect.

Classifications

The alpha carbon of amino acids is chiral and non-superimposable. The only non-chiral amino acid is glycine. In fisher nomenclature L-Amino acids have the amino group on the left when the carboxyl group is up, while D-AA have the amino on the right. Instead in Kahn-Ingol-Prelog nomenclature we define the chiral center, and assign the side groups priority based on molecular weight. We put the lowest priority group to the back, then draw a circle from 1-2-3 priority groups. If the rotation is counterclockwise, we have the sinistra movement to the left. If the movement is clockwise, the isomer is R for rightward. For amino acids, Fisher L is always S and fisher D is always R (except Cystine). Proteins are composed of L amino acids, so the L series is proteinogenic.

Proteinogenic AA can also be classified in two other ways, either through nutritional needs as essential (PVT. TIM HALL) or non-essential, or by their R group. R-group classification can be Non-Polar, Polar uncharged, aromatic, positively charged, negatively charged, sulfur containing and amidic.

Functionality of Amino Acids



It's important to learn the interactions, properties of the amino acids and cofactors. The common exam question is to explain which amino acids are interacting and what they are doing.

Glycine: (Gly, G) Simplest possible amino acids, and most flexible and non-essential. It is the only AA that is not chiral. Glycine plays a key role due to its flexibility as a molecule in protein binding (RGG binding motif to bind nucleotides), and is used as a building block for proteins, biosynthetic intermediates and neurotransmitters. Glycine is responsible for the G-quadruplex structure in DNA found in telomeres by forming a 4 stranded DNA region which used Hoogsteen base pairing. G-Quadraplexes are also used as a transcription regulation through translation termination in RNA. Quadruplexes are usually stabilized by a K ion in the middle. In RGG regions, Arginine interacts with the phosphate backbone and forms the stacking interaction with nearby base pairs. We also find glycine, which is flexible enough, and small enough to be able to intercalate the DNA allowing the arginine to enter.

Glycine is also important in the methylation processes of DNA, which it serves as a carbon storage which can be used for methylation by the enzyme glycine n-methyl-trasferase from S-adenosylmethionine (SAM) to form S-adenosythhomosysteine (SAH) and adding the methyl group to DNA or Proteins. The glycine that served as the methyl acceptor in the reverse reaction now forms sacrosine. Another methylation of sacrosein forms betaine. Instead the SAH is converted into Homocysteine, where it can become either methionine and recycle the methionine storage SAM or become cysteine.

Glycine is also important in heme formation through the activity of ALAS enzyme or to form glutathione (antioxidant) with the fusion of glycine and cysteine. Therefore, glycine is involved in mitochondria function (heme), detoxification (Glutathione), Methylation (SAM/SAH homeostasis), food intake signal (NMDA receptor formation) and hormonal and cytokine response (GlyRs).

Proline: The most rigid amino acid due to its cyclic structure formed by the amino group and side chain. It is often used as a hinge due to its ability to convert between cis and trans confirmation in peptide bonds. Proline is a helix stopper, when its present the helix wont form. Proline is posttranslationally modified by the hydroxylation to form hydroxyproline, a key component of collagen as well as its role in HIF. HIF is the main O2 sensor in our body. High concentrations of oxygen are required for formation of hydroxyproline on the HIF. This hydroxylated proline is able to interact with VHL, which moves the protein to degradation. In hypoxia, hydroxyproline is not formed, and VHL cannot bind, so the signal remains active and induces hypoxia inducible genes to deal with the high concentration of reactive oxygen species.

Proline biosynthesis is a good example of feedback regulation and is under control of the amino acid starvation pathway. Proline synthesis happens in the mitochondria, and comes from glutamate. It can also come from outside the cell through proline transporters. In these conditions, proline is used to charge tRNA, usually for the synthesis of collagen. In proline starvation instead we see the activation of the transcription factor ATF4 by the increased level of uncharged tRNA. ATF4 activates expression of various genes including amino acid transportes. When proline is increased again, tRNA is charged and the AAR-ATF4 pathway is shut off.

Hydrophobic Amino Acids: Hydrophobic amino acids make up the core of proteins, and are in fact aliphatic. Alanine, valine, leucine, isoleucine and methionine are linear molecules, and play important roles in the hydrophobic effect that stabilizes proteins. An example is the leucine zipper motif found in some DNA binding proteins. The leucine zipper has many of these aliphatic, hydrophobic residues

facing the other chain and bonding together.

Physiologically, we see the importance of methionine in the methionine cycle composed of SAM and SAH, and the buffering function of glycine. Here, methionine is used in order to form the methyl donor SAdenysyl-Methionine. Methionine is also seen in the methyl folate trap in those which lack vitamin b12, and thus folate is trapped as M5-methyl-folate, restricting the ability to form other folates, making the cells unable to make DNA, forming dragline megaloblastic erythrocytes.

Aromatic Amino Acids: Phenylalanine, Tyrosine, and Tryptophan are often found in the core of proteins due to their hydrophobic nature, and are also important in stacking interactions. Stacking, and cation-pi interactions results from the uneven distribution of electrons in the p-orbital of aromatic rings, forming an attraction between these four rings called a quadrupole. Pi cation rogers to attraction between an aromatic and a cation forming a dipole where the electrons are pulled to one side of the ring, and also serves in stabilizing.

Physiologically, aromatic amino acids serve an important role as precursors to many neurotransmitters. Adrenaline, noradrenaline, dopamine and serotonin are all aromatic neurotransmitters, while gaba, acetylcholine, glutamate and endorphins are non-aromatic. Tyrosine forms dopamine first by forming DOPA, followed by the decarboxylation of DOPA to dopamine. Tryptophan forms serotonin through decarboxylation, similarly to glutamate forming GABA from glutamate. This highlights the importance of ascorbic acid in the brain for the hydroxylases, B6 forms PLP the important cofactor in decarboxylases.

Small Polar Amino Acids: Serine, threonine, and cysteine are often found on the surface of protein, Ser and Thr are responsible for the hydrogen bonding interactions while Cys is able to form disulfide bonds releasing two proteins and two electrons. The formation of disulfide brindes is regulated by the presence of reactive oxygen species. Reducing species such as glutathione in the cell which can help cope with reactive oxygen species also are able to block the formation of disulphide bond formation in order to promote the proper native folding of the proteins.

Serine and glycine are important in cancer metabolism. The warburg effect details how cancer cells favor the formation of lactate even more than acetyl-CoA compared to healthy cells. The warburg effect is the basis of PET (positron emission tomography) where a glucose analogue is released which emits gamma rays, therefore we can see which tissues are likely malignant cancer cells. Warburg effect comes from the metabolic reprogramming of cancer cells, mainly in glycolysis in order to increase antioxidants and biosynthetic intermediates for rapid cell growth. This reprogramming heavily involves the activation of the serine/glycine one carbon (SGOC) metabolism, a network based on chemical reactions of folate compounds.

Charged Amino Acids: Aspartate, Glutamate, lysine, arginine and histidine are often found on the outside of amino acids as they are able to form hydrogen bonds, stacking and electrostatic interactions.

Amide-containing Amino Acids: Asparagine and glutamine are non-toxic transporters of ammonia in the blood circulation. Glutamate donates one amide group to form carbamoyl phosphate, while the second amide for urea comes from aspartate. Asparagine and glutamine serve as the ammonia carries to the liver from other tissues, keeping the blood ammonia level very low, as even slightly increased levels lead to hyperammonemia which is toxic to the central nervous system. This is because

glutamate in the brain serving as a neurotransmitter becomes glutamine in the presence of ammonia, losing its functionality. Asparagine is important in the onset of leukemia, and therefor a asparaginase therapy is useful for acute myeloid leukemia allowing cells make asparagine from aspartate.

Selenocysteine: Selenocysteine -called the 21st amino acid- contains a selenium group instead of the sulfur group of a cysteine. It is necessary in enzymes such as glutathione peroxidases to remove ROS. selenocysteine is translated from the stop codon UGA, only when the mRNA being translated includes a SECIS element, causing the UGA to encode selenocysteine instead of a stop codon. This complex In archaea and eukaryotes is found in the 3' UTR in bacteria immediately after.

Pyrrolysine: The extra pyrrolidine ring is incorporated into the active site of several methyltransferases in archaea, a job done by Folate in Eukaryotes and bacteria. This reaction starts with a nucleophilic attack by a nitrogen in the substrate to the pyrrolidine ring, similar to the job of pyridoxal phosphate in eukaryotes. Pyrrolysine is coded by a stop codon UAG in archaea, and like selenocysteine requires the PYLIS sequence in the 3' UTR

N-Formylmethionine (fMET, fM): formylmethionine is very similar to methionine but contains a formyl motif. Since fMet resembles a peptide, bacterial ribosomes prefer tRNA with peptide-like ligands at the p-site, increasing the interaction with the peptide. In humans, circulating fMet is recognized by immune cells as forign material, or signals mitochondria damage and activating apoptosis. fMet is synthesized from methionine from a post translational modification by the formyltransferase-I, an enzyme depending on 10-formyl-tetrahydrofolate. 10-fTHF is found only on the inner matrix of the mitochondria. Gcn2 plays a role in activating Fmt1 to form the isoform fMet-tRNA from Met-tRNA when stress conditions such as cold, starvation ect. fMet plays a role in eukaryotes on the end terminal of proteins which need to be ubiquitinated and broken down.

Peptide Bond

Peptide bond is a covalent bond between two amino acids, formed from the condensation reaction between the amino group and the carboxyl group, forming an amide bond. Breaking of a peptide bond is a hydrolysis reaction. The peptide bond is most stable hydrolyzed, but due to the very large energy barrier makes the peptide bond kinetically stable unless able to overcome this barrier. Proteases are able to lower this energy barrier, and disassemble proteins.

Chemistry of peptide bond

The peptide bond is 1.32 A, between a single and double bond due to the resonance where the electron can be placed on the C-N bond, or diverted to the C-O carboxyl group nearby. This means there is no rotation around the peptide bond, and is defined as the omega dihedral angle. Omega is usually trans, that is the two a-carbons of the two amino acids are on opposite planes, because in the cis we see large steric hinderance between side chains. Cis confirmation can be seen between two Gly, or Pro because the side chain of Pro is equally sterically hindering in both confirmations (hinge) .

Translation

Phase I: Activation

Activation of Amino Acid

Amino acid reacts with ATP to form amino acid -AMP complex and pyrophosphate. The reaction is catalyzed by a specific amino acid-activating enzyme called aminoacyl- tRNA synthetase in the presence of Mg2+. There is a separate aminoacyl – tRNA synthetase enzyme for each kind of amino acid. Much of the energy released by the separation of phosphate groups from ATP is trapped in the amino acid — AMP complex. The complex remains temporarily associated with the enzyme. The amino acid-AMP-enzyme complex is called an activated amino acid. The pyrophosphate is hydrolysed to 2Pi, driving the reaction to the right.

The amino acid-AMP- enzyme complex joins to the amino acid binding site of its specific tRNA, where its -COOH group bonds to – OH group of the terminal base triplet CCA. The reaction is catalyzed by the same aminoacyl-tRNA synthetase enzyme.

The resulting tRNA-amino acid complex is called a charged tRNA. AMP and enzyme are freed. The freed enzyme can activate and attach another amino acid molecule to another tRNA molecule. The energy released by change of ATP to AMP is retained in the amino acid-tRNA complex. This energy is later used to drive the formation of peptide bond when amino acids link together on ribosomes

There are many families of tRNA synthetases that evolved. Amino AcyltRNA synthases must maintain high fidelity, there is no way for the ribosome to interpret an attached amino acid as correct or not. Most amino acids are easy to differentiate from each other, due to their very different confirmations, but AAs such as valine and isoleucine which are structurally similar happen around 1/150 times. For this reason, isoleucyl-tRNA synthetase overcomes this by a second active site which preforms an editing reaction. The newly synthesized tRNA is moved tot he editing site, if too large such as isoleucine it can leave the complex. Valine instead can enter the editing pocket, leading to its hydrolysis.

Error check

Most tRNA binding events are by non-matching tRNA, and therefore the ribosome needs to reject them as fast as possible. In the initial phases of codon recognition and GTP activation, the ribosome can sense an energetically more stable confirmation from the ‘correct’ codon. The energetic landscape of this forward and backward reaction reveals the release of a wrong codon is spontaneous, while the release of a correct codon is not easy.

Phase II: Initiation

Pre-initiation Complex

The small and the large subunits of ribosomes must be joined together for protein synthesis. This is brought about by the mRNA chain. The latter joins the small ribosomal subunit (40S) by first codon through base pairing with appropriate sequence on rRNA. The combination of the two is called initiation complex. The large subunit (60S) later joins the small subunit, forming an active ribosome. Activation of ribosomes by mRNA requires proper concentration of Mg2+.

Cap dependant Initiation factors

Cap dependent initiation of translation usually involves the interaction of certain key proteins, the initiation factors, with a special tag bound to the 5'-end of an mRNA molecule, the 5' cap, as well as

with the 5' UTR.

eIF1 and **eIF1A** both bind to the 40S ribosome subunit-mRNA complex. Together they induce an "open" conformation of the mRNA binding channel, which is crucial for scanning, tRNA delivery, and start codon recognition. In particular, eIF1 dissociation from the 40S subunit is considered to be a key step in start codon recognition eIF1 and eIF1A are small proteins (13 and 16 kDa). eIF1 binds near the ribosomal P-site, while eIF1A binds near the A-site, so only the E-site is open.

eIF2 is the main protein complex responsible for delivering the initiator tRNA to the P-site of the preinitiation complex, as a ternary complex containing Met-$tRNA_i^{Met}$ and GTP (the eIF2-TC). eIF2 has specificity for the methionine-charged initiator tRNA, which is distinct from other methionine-charged tRNAs used for elongation of the polypeptide chain. The eIF2 ternary complex remains bound to the P-site while the mRNA attaches to the 40s ribosome and the complex begins to scan the mRNA. Once the AUG start codon is recognized and located in the P-site, eIF5 stimulates the hydrolysis of eIF2-GTP, effectively switching it to the GDP-bound form via gated phosphate release. The hydrolysis of eIF2-GTP provides the conformational change to change the scanning complex into the 48S Initiation complex with the initiator tRNA-Met anticodon base paired to the AUG. After the initiation complex is formed the 60s subunit joins and eIF2 along with most of the initiation factors dissociate from the complex allowing the 60S subunit to bind. eIF1A and eIF5-GTP remain bound to one another in the A site and must be hydrolyzed to be released and properly initiate elongation.

eIF3 independently binds the 40S ribosomal subunit, multiple initiation factors, and cellular and viral mRNA. In mammals, eIF3 is the largest initiation factor, made up of 13 subunits (a-m). It controls the assembly of the 40S ribosomal subunit on mRNA that have a 5' cap or an IRES. eIF3 may use the eIF4Fcomplex, or alternatively during internal initiation, an IRES, to position the mRNA strand near the exit site of the 40S ribosomal subunit, thus promoting the assembly of a functional pre-initiation complex.

The **eIF4F** complex is composed of three subunits: **eIF4A, eIF4E**, and **eIF4G**. Each subunit has multiple human isoforms and there exist additional eIF4 proteins: eIF4B and eIF4H.

eIF4G is a scaffolding protein that interacts with eIF3 and the Poly(A)-binding protein (PABP), as well as the other members of the eIF4F complex. eIF4E recognizes and binds to the 5' cap structure of mRNA, while eIF4G binds PABP, which binds the poly(A) tail, potentially circularizing and activating the bound mRNA. eIF4A – a DEAD box RNA helicase – is important for resolving mRNA secondary structures.

eIF4B contains two RNA-binding domains – one non-specifically interacts with mRNA, whereas the second specifically binds the 18S portion of the small ribosomal subunit. It acts as an anchor, as well as a critical cofactor for eIF4A. It is also a substrate of S6K, and when phosphorylated, it promotes the formation of the pre-initiation complex.

eIF5 is a GTPase-activating protein, which helps the large ribosomal subunit associate with the small subunit. It is required for GTP-hydrolysis by eIF2 and contains the unusual amino acid hypusine. **eIF5A** is the eukaryotic homolog of EF-P. It helps with elongation and also plays a role in termination. **eIF5B** is a GTPase, and is involved in assembly of the full ribosome. It is the functional eukaryotic analog of bacterial IF2. **eIF6** performs the same inhibition of ribosome assembly as eIF3, but binds with the large

subunit.

Cap-Independent Initiation

Cap Independent translation is important in conditions that require translation of specific mRNA during cellular stress, when overall translation is reduced. These mRNA contain an internal ribosome binding site, where translation can initiate independently of the CAP protein, viruses such as hepatitis C cleave eLF4G so cap dependent translation is inefficient, and instead only viral mRNA containing IRES can be translated.

uORF

About half mammalian genes encode mRNA having at least one short ORF upstream (uORF) of the main ORF. In this case, ribosomes translate the uORF and resume scanning and reinitiate downstream. uORFs can regulate eukaryotic gene expression Translation of the uORF typically inhibits downstream expression of the primary ORF.

Circularization

Circularization of mRNA is thought to promote cycling of ribosomes on the mRNA leading to more efficient translation, and may function to ensure only intact and not 'too old' mRNA are translate as the length of the poly-A tail is limited. In eukaryotic mRNA circular molecules are formed due to interaction between eIF4E and poly-A tail binding proteins which bind to both eLF4G forming protein mRNA bridges.

Phase III: Elongation

Elongation factors are a set of proteins that function at the ribosome, during protein synthesis, to facilitate translational elongation from the formation of the first to the last peptide bond of a growing polypeptide.

eEF-1 subunit α mediates the entry of the aminoacyl tRNA into a free site of the ribosome. eEF-1 subunit $\beta\gamma$ serves as the guanine nucleotide exchange factor for eEF-1, catalyzing the release of GDP from eEF-1. eEF-2 catalyzes the translocation of the tRNA and mRNA down the ribosome at the end of each round of polypeptide elongation.

At the end of the initiation step, the mRNA is positioned so that the next codon can be translated during the elongation stage of protein synthesis. The initiator tRNA occupies the P site in the ribosome, and the A site is ready to receive an aminoacyl-tRNA. During chain elongation, each additional amino acid is added to the nascent polypeptide chain in a three-step microcycle. The steps in this microcycle are (1) positioning the correct aminoacyl-tRNA in the N site of the ribosome, which is brought into that site by **eIF2**, (2) forming the peptide bond and (3) shifting the mRNA by one codon relative to the ribosome.

Phase IV: Termination

Termination of elongation depends on eukaryotic release factors. In eukaryotes, there is a universal release facto eRF1, that recognizes all three stop codons. Upon termination, the ribosome is disassembled and the completed polypeptide is released. eRF3 is a ribosome-dependent GTPase that

helps eRF1 release the completed polypeptide. The human genome encodes a few genes whose mRNA stop codon are surprisingly leaky: In these genes, termination of translation is inefficient due to special RNA bases in the vicinity of the stop codon.

Drugs and translation

Due to the large differences between eukaryotic and prokaryotic translation processes, many antibiotics are useful for inhibiting bacterial translation. Thermorubin inhibits binding of fMet- tRNA to the P-site in an initiation factor- dependent manner. Although the binding site of thermorubin does not overlap with the P-site tRNA, the drug induces conformational changes in the rRNA nucleotides A1913 and C1914, and this is proposed to perturb IF1 (Bacterial Initiation Factor) activity.

Ataluren is a novel small molecule compound that is used to treat genetic disorders stemming from nonsense mutations. The drug is thought to make the ribosome less sensitive to premature stop codons by masking nonsense codons, but is unable to distinguish between premature stop and natural stop leading to many side effects.

Secondary structure

A common test question is why do we only see two secondary structures in nature? The answer is only these two formations allow for optimization of the rotational angels, in order to fall in line with Rahmen shander plots (don't know what its called) and secondly, the ability to optimize hydrogen bond interactions.

Alpha Helices

Leucine zippers bind DNA through a dimmer, one of the most simple motifs. Helix-turn-helix is one of the most ancient DNA binding motifs. Zinc fingers are found in nuclear proteins which transduce a signal and can bind to DNA.

Beta sheets:

Beta sheets can be antiparallel or parallel. They form hydrogen bonds intramolecularly, that is they bond to surrounding sheets, not with themselves such as in a-helices. The B-barrel is an extremely common architecture (related to the organization in space of the secondary structures). Each of these architectures can be formed by different topologies (the way secondary structures are connected together). In the case of B-barrels, these topologies can be hairpins, greek keys (crystallin-like), or jelly rolls (google).

B-Barrels:

B barrels are common in pore structures, notably aquaporin.

Beta-Helices:

A beta helix is a tandem protein repeat structure formed by the association of parallel beta strands in a helical pattern with either two[1] or three faces.

Protein Stability

Protein stability is the net balance in forces between the native folded conformation and the unfolded state. The equilibrium between folded and unfolded is Keq = N/U = Fn/(1-Fn) where N is the number of folded, and U is number in unfolded state. The thermodynamic stability of a protein is defined as the difference in Gibbs free energy between native and unfolded state. Although the energy difference between folded and unfolded state is very small, this difference still increases the ratio between folded:unfolded to around 2.7 million: 1. The two most important parts of these forces are the intra-protein forces, and the protein water forces, but water-water, protein-ion and water-ion are also present.

Intraportien forces can be ionic, hydrogen bonds, van-der waals forces, divided into Debye (permanent-induced), london (induced-induced) and keesom (permanent-permanent). London forces are the ones responsible for the hydrophobic effect. Van der waals radius is derived from lennard jones-interatomic potential equation, which shows equilibrium between strong repulsion when they are too close, and weak attraction when too far. This energy equilibrium lands at 1.12 sigma, and the van der waals radius is half of this distance.

Another feature of London forces is the hydrophobic effect, hydrophobic surfaces tend to come together to exclude water to minimize the area of contact between water and nonpolar molecules. This process is spontaneous, and entropy driven, that is the order of disorder of the system. When water is in contact with hydrophobic surfaces it forms cages called clathrates. The hydrophobic effect is very important for protein folding, in the unfolded state many clathrates form, while the folded state buries these hydrophobic residues, increasing stability. Hydrogen bonds, van der waals and electrostatic effect make up a small net change in Gibbs free energy compared to that of hydrophobic effect, which is by far the most important. This decrease in free energy needs to overcome the conformational entropy created by the folding process.The overall difference in conformational entropy and the hydrophobic, hydrogen bonds..ect are the total gibbs free energy, around 2 hydrogen bonds strength.

Because we only consider the initial free energy and final, increasing temperature decreases the formation of clathrates, therefore increasing entropy of the unfolded state, preventing the correct folding. pH instead is not related to the thermodynamics of the system, but instead related to the kinetics. Because protein surfaces are covered in charged residues, working at low pH increases the rate of protonated residues, resulting in a positively charged protein. Increasing the pH instead leads to the unprotonated form of these residues, producing a negatively charged protein. The isoelectric point of a protein is the balanced state between positive and negative change and represents the minimum solubility of the protein. This results in the same denaturation seen in temperature increase, but through a kinetic change instead of a thermodynamic change. We can also consider ion concentration, salting in shields the charged residues, while salting out blocks the interactions with water.

Thermodynamics deals only with the initial and final energy state, while kinetics can show how the protein folds in this way. Anfinsen's experiment set to prove that proteins native folding was completely reliant on their amino acid sequence. This was studied by measuring native folding activity, denaturing the proteins, then removing the denaturing agents to see if the proteins are able to refold at the same rate in vitro. The experiment found that the activity was restored, concluding that proteins such as ribonuclease are able to spontaneously fold into their final state, even when

denatured. The experiment also found that the driving force of the initial folding was the hydrophobic interactions, while the disulfide bridges served to stabilise the final confirmation.

Levinthal's paradox assumes proteins composed of 100 amino acids, and each amino acid has only 3 possible confirmations, leading to total possible confirmations over trillions. If 1 million confirmations were able to be tested, these proteins would require trillions of years, concluding that protein folding is not a random walk, but influenced by stability of intermittent confirmations. Protein folding is also not a deterministic walk in the conformational space, but in reality is described by the funnel landscape, where protein folding is via multiple routes going downhill rather than through a single pathway. Proteins folded progressively can be trapped in energy minima states, and are more likely to move down the theoretical hole.

One other famous model for the folding landscape is the nucleation model, in which interactions between local residues are the first and increase the tendency to correctly fold. This foldning process shows a hydrophobic collapse, when around half of the residues are in their native confirmation called the Molten Globule. After the Molten globule is filmed, hydrogen bonds, disulfide bridges are more important, especially forming the tertiary structure.

In vivo, intramolecular contacts can prevent the proper folding. For this reason, chaperones, and chaperonins help folding in vivo. Chaperones are usually associated with the exit channel of the ribosome and recognize potentially misfolded proteins, binding them and providing energy to shake and overcome the local minima to reach the native conformation. Chaperonins instead are able to inclose folding proteins in order to provide a perfect environment for folding without interacting with other molecules in the cell.