Drug Discovery and Design A

Proteins: Fundamentals

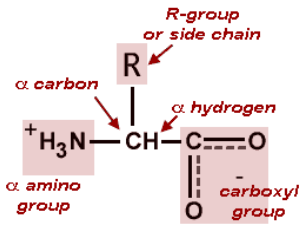
- Proteins are hetero-polymers (polypeptides) of covalently linked α-amino acids as well as the addition of co-factors (metal ions), coenzymes, prosthetic groups or other modifications.
    - Cofactor: general functional non-amino acid component –metal ions, organic molecules
    - Coenzyme: used to designate an organic co-factor –NAD^+ in lactate dehydrogenase
    - Prosthetic groups: covalently attached co-factors –heme in myoglobin
- Proteins have varied functions:
- In catalysis*:* such as enolase in glycolytic pathway, DNA polymerase in DNA replication
- In transport: haemoglobin transports O_2, lactose permease transports lactase across cell membranes
- In structure: such as collagen connective tissue or keratin in nails, hair, feathers and horns.
- In motion: including myosin (muscle tissue) and actin (muscle tissue and cell motility)
- Important to know where drug/ligand binds to a protein and why binding occurs at one protein but not another (differences). Also important to know why a protein is stable or why it aggregates or why it is soluble/insoluble.

Protein structure:

- Primary structure: the linear sequence of amino acids/amino acid residues. Each amino acid has a unique structure (due to uniqueness of side chain). Structure and chemistry is available in building functionality.
- Secondary structure: certain sequences form H bonds that cause the region to fold into a spiral (α helices), a sheet (β sheets/strands) and loops/coils. β sheets must have two strands.
    - Ramachandran plot: phi and psi angles form cooperative patterns called the secondary structure and are relatively free to rotate. The plot visualises energetically allowed regions for backbone dihedral angles phi and psi of amino acids residues in protein structure
- Introns and exons: eukaryotic genes are composed of exons (protein coding sequences) and introns (non-coding sequences). Introns removed after primary transcription by splicing. RNA is stabilised by 7-methyl guanosine (5 cap) and poly A tail.
    - Alternate splicing: using same gene to create different proteins –exons can be joined together in different combinations and hence different proteins are formed with varied functions.
- Post translational modification (proteolytic processing) –pre-sequence is the signal peptide, the pro-sequence is to be removed to generate the mature protein
- Tertiary structure: the arrangement of secondary structure elements.
- Quaternary structure: molecules of proteins associate to form multimers. Bonds include: covalent (backbone and disulphide bridges), non-covalent (VDW), hydrogen bonds and salt bridges.

→ *Structure and naming of amino acids*

- General structure of an amino acid:



- Amino acids have properties that are well suited to carry out a variety of biological functions because of their capacity to polymerise, useful acid-base properties, varied physical properties and varied chemical functionality.
- Peptide bonds are planar; planarity is associated with delocalised electrons (partial double bond)
- Amino acids in 5 classes:
    - Non-polar, aliphatic (hydrophobic): glycine, alanine, proline, valine, leucine and methionine.
    - Aromatic groups absorb UV light: phenylalanine, tyrosine and tryptophan.
    - Polar (hydrophilic) form H/S-S bonds: serine, threonine, cysteine, asparagine and glutamine
    - Positively charged (basic): lysine, arginine and histidine
    - Negatively charged (acidic): aspartate and glutamate (aspartic acid and glutamic acid)

- Uncommon a.as in proteins are not incorporated by ribosomes. They arise by post-translational reversible modifications of proteins especially phosphorylation (important in regulation and signalling)

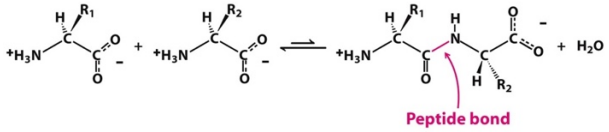
→ *Ionisation behaviour of amino acids and peptides*

  - At acidic pH the carboxyl group is protonated and the amino acid is in the cationic form
  - At neutral pH the carboxy group is deprotonated but amino group is protonated. The net charge is zero and the ion is a zwitterion.
  - At alkaline pH the amino group is neutral, the carboxylic acid is in the anionic form.

- The chemical environment affects pKa values –α-carboxyl group much more acidic than in carboxylic acids, α-amino group is slightly less basic than in amines. Perturbations in pKa may occur from nearby interactions.
- Amino acids may act is buffers. Amino acids with uncharged side chains (e.g. glycine) have two pKa values and hence can act as buffers in two pH regimes. Zwitterions predominate at pH values between the pKa values of the amino acid and carboxyl group.
- The isoelectric point (pI) of an amino acid is the point in which the net charge is 0, the amino acids are least soluble in water and will not migrate in an electrical field.
  - For amino acids without ionisable side chains: $pI = pKa_1 + pKa_2 / 2$
  - Ionisable side chains can also be titrated but titration curves become more complex. pK_a values are discernible on the curve if they are more than two pH units apart.
  - For amino acids with ionisable side chains to find pI:
    - Identify the species that carries a net 0 charge
    - Identify the pKa value that defines the acid strength of this zwitterion
    - Identify the pKa value that defines the base strength of the zwitterion
    - Take the average of the two values (average of pK_a and pK_R)
- All amino acids (except glycine) are chiral. The α carbon has 4 different substituents and is tetrahedral – each amino acid has a unique substituent (R group), in glycine the R group is a hydrogen.
- All amino acids (except proline) have an acidic carboxyl group, a basic amino group and an α hydrogen connected to the α carbon.
- Stereoisomers are non-superimposable chemical isomers that have identical covalent structure and exist for all chiral amino acids. There are two classes of stereoisomers: enantiomers & diastereomers.
  - Enantiomers have either an L or D rotation
  - CORN rule –clockwise COOH → R → NH is L and anticlockwise is D
  - L-alanine is the naturally occurring form of alanine

→ *Structure and properties of peptides*

- Small condensation products of amino acids (mw < 10kDa)

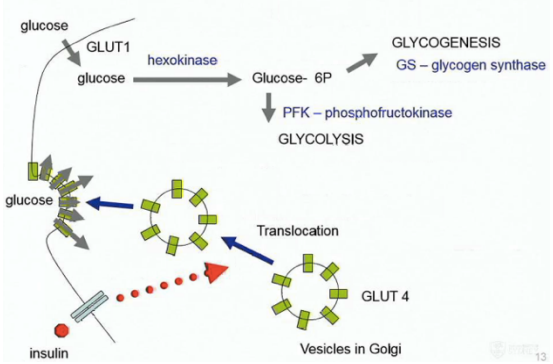


- Peptides have a variety of functions that include:
  - Hormones and pheromones: e.g. insulin, oxytocin, sex peptide.
  - Neuropeptides: e.g. substance P (pain mediator)
  - Antibiotics: e.g. polymyxin B for gram negative and bacitracin for gram positive
  - Protection: toxins e.g. amanitin (mushrooms), conotoxin (cone snails), cholortoxin (scorpions)

Fates of Dietary Carbohydrates

- Nearly all dietary carbohydrates are starch (polymer of glucose). There are two forms of starch:
  - Amylose: linear, forms helices, difficult to digest
  - Amylopectin: branched, easy to digest, post-prandial response like pure glucose
- Disaccharides:
  - Lactose: galactose + glucose, consequences of lactase deficiency (lactose intolerance)
  - Sucrose: fructose + glucose
  - Maltose: glucose + glucose
- Monosaccharides:
  - Glucose
  - Fructose: fruit and honey
- **Insulin:** Synthesised in the endocrine part of the pancreas by β-cells of the islets of Langerhans
  - When glucose is ingested, blood glucose rises and this stimulates the pancreas to release insulin. There are also other stimulants.
  - Insulin causes blood glucose to fall because it causes glucose to be taken up by the tissues. It also supresses the liver's ability to make glucose
- Glucose responses: aim to maintain blood glucose concentration at ~5mM. Levels less than 3mM would lead to coma. Prolonged high glucose concentration may lead to glycosylation of proteins.
- Consequences of intolerance:
  - Post-prandial hyperglycaemia: if it occurs after every meal and persists for several hours then there will be problems. The person will rarely be eugylcemic, and may lead to complications of hyperglycaemia and protein glycosylation.
    - Root cause may be insulin resistance: impaired ability to respond to insulin (underlies type 2 diabetes)
  - Control of glucose intolerance: consumption of slowly absorbed starches
- The glycemic index: describes post-prandial glucose response (area under the test food glucose curve divided by the area under a reference food glucose curve). Reference food is normally 50g glucose and the test food is given in an amount that will give the same amount of digestible carbohydrate.
- Ratio is expressed as a percentage: GI of modern, processed amylopectin foods is > 80%, GI of old world grains and legumes is < 30%. By definition, GI of glucose would be 100%.
- The GI is useful knowledge for controlling blood glucose especially in relevance to diabetes (↑ BGL). Quality of a carbohydrate (its GI) is as important as the total amount of the carbohydrate.
  - Sugary (sucrose) foods have a low GI because half the carbohydrate is fructose. Fructose containing foods are low GI.
  - Dairy foods are low GI because half the carbohydrate is galactose (protein elicits insulin secretion). GI is assessed as response to 50g of absorbable test CHO (not all of which is glucose).

Glucose uptake in muscle and white adipose tissue:



- GLUT-1 and GLUT-4 bring the glucose into the cells.
- GLUT-1 is not under insulin control and GLUT-4 is.
- GLUT-4 is present in the cell as vesicles of Golgi and under the influence of insulin, it translocates to the to the cell surface to facilitate the entry of glucose.

- Inside the cell, hexokinase phosphorylates the glucose (G6-P). This requires ATP.
- The glucose 6-phosphate may then enter the glycolysis pathway when energy is needed via the rate limiting enzyme –phosphofructokinase
- If energy is not needed, the glucose will be used in glycogenesis (making glycogen) facilitated by the glycogen synthase enzyme
- If glucose 6-P builds up, it will inhibit the hexokinase so that phosphorylation stops and glucose can easily exit the cell
- Hexose metabolism –glycogen formation: glucose is phosphorylated (G 6-P), then rearranged to form glucose 1-phosphate. Then using UTP (releases PP) forms a UDP glucose (activated glucose). PP hydrolysis pulls reaction to completion. The UDP glucose then adds on to the glycogen chain via glycogen synthase.
- Glycogen is a polymer of glucose with 12-14 glucose residues. Body cannot store a lot of glycogen because it is branched and hence a heavy molecule.
- UDP needs to be regenerated to UTP = UDP + ATP → UTP + ADP (cost 1 ATP to add one glucose to the glycogen chain)
- Glycogen synthase: Glycogen synthase with a phosphate attached is inactive. It is regulated by reversible phosphorylation (covalent modification) –a phosphate group is removed via the enzyme protein phosphatase I (PPI). Insulin stimulates this (stimulates glycogen synthesis)
    - To deactivate the activated enzyme, it is phosphorylated back in a process catalysed by glycogen synthase kinase.
- Phosphofructokinase: catalyses the second energy investment stage of glycolysis. It is regulated allosterically –stimulated by low energy charge (when high ADP, AMP and low ATP). Small change in ATP, ADP causes a large relative change in AMP.
    - AMP molecules bind at a site away from the active site (allosteric binding site)
- Coupling: stimulation of glycogen synthase by insulin creates an energy demand. Glycogenesis is anabolic so trapping and activation of glucose requires ATP.
    - Drop in energy charge stimulates PFK hence insulin has indirectly stimulated PFK and glucose oxidation
    - Anabolic pathway requires catabolic pathway i.e. signal to store fuel also causes fuel burning
- Liver glucose uptake:
    - GLUT-2 is used to take up glucose from blood in the liver and pancreas –very high activity and very abundant ([glucose] in blood = [glucose] in liver)
    - Glucokinase (variant of hexokinase) rapidly convert G → G6P. it is not inhibited by build-up of product. High K_M (10mM) for glucose –not saturated by high levels of liver glucose so [G6P] rapidly increases as [glucose] in blood rises
    - G6P in liver can stimulate inactive glycogen synthase in liver –even phosphorylated GS
- Glycogenesis:
    - In liver: *the push mechanism*
        - [G6P] can get high enough to stimulate GS
        - Glycogenesis responds to blood glucose without the need of insulin
        - Although insulin will stimulate glycogenesis further
    - In muscle: *the pull mechanism*
        - [G6P] never gets high enough to stimulate GS
        - insulin stimulates GS and pulls glucose into glycogen
    - In both liver and muscle:
        - 2 ATPs required for the incorporation of a glucose into glycogen chain
        - branching enzyme needed to introduce a1→6 branch points
        - transfers a segment from one chain to another
        - limit the size of glycogen molecule –branches become too crowded even if they become progressively shorter
- Lipogenesis: process of converting acetyl CoA back into fatty acid. It is stimulated when insulin binds to its receptor on the cell surface and inhibited by dietary fat (high intake). The process consumes ATP and a reductant (produced in glycolysis). NADH released in the Krebs cycle ultimately produces the ATP that is

consumed in lipogenesis. Following lipogenesis, three fatty acids join to a glycerol to form a triglyceride in an energy requiring process called esterification.

- Pyruvate dehydrogenase: converts pyruvate to acetyl CoA in the mitochondria. Insulin stimulates PDH phosphatase. Coenzyme A alone can't cross membranes.
- Fate of acetyl CoA: depends on the energy need (charge) and stimulus driving lipogenesis (e.g. insulin).
    - Catabolic: oxidation, when there is little energy in the mitochondria it is burnt in the Krebs cycle. Carbon is fully oxidised to CO_2 and NADH is produced to generate ATP.
    - Anabolic: transport, when there is a lot of energy it goes to lipogenesis. Carbons are moved to the cytoplasm and activated for fatty acid synthesis.
    - First step in both is citrate formation because the big CoA group can't leave the mitochondria
- ATP-citrate lyase (ACL):
    - Once in the cytoplasm, citrate is cleaved using CoA to generate acetyl CoA and oxaloacetate
    - Reaction requires conversion of ATP → ADP+Pi and is catalysed by ATP citrate lyase (ACL)
    - ACL is inhibited by hydroxyl-citrate (OHCit)
- Acetyl CoA carboxylase (ACC): activates acetyl CoA and primes it for lipogenesis. It is unusual in that it fixes carbon dioxide in the form of bicarbonate (carboxylation): acetyl CoA + CO_2 → malonyl CoA
    - Carboxylation reaction requires conversion of ATP to ADP and Pi
    - It is the rate limiting step for lipogenesis. Requires biotin as a co-factor.
    - Stimulated by insulin –malonyl CoA is committed to lipogenesis → reversible phosphorylation
    - Regulated allosterically –stimulated by citrate (polymerisation of the enzyme) and inhibited by long-chain fatty acyl CoA. (eventual product of lipogenesis)
- Fatty acyl synthase: makes fatty acids. Requires reductant (NADPH). Gene expression is stimulated by insulin → mRNA. There are two free SH groups on an acyl-carrying protein which keeps the intermediates in the right position for interaction with the right active sites –each new 2C unit is added onto the carboxy end.
- Addition sequence: each round of 2 carbon addition requires 2 NADPH but no ATP. Release of CO_2 that went on during the production of malonyl CoA (carboxylation of acetyl CoA doesn't fix CO_2).
- Fatty acids then get released as FA-CoA when chain length is greater than ~ 14 carbons. Desaturation (adding double bonds) is done after FA synthase.
- Desaturated fatty acids: double bonds are in cis (H on same side) form. Trans fats in foods can be harmful because they can't be processed by enzymes
- Essential fatty acids: cannot be synthesised by humans because the body lacks the required desaturase enzymes, hence are essential in a diet. Include omega 3 and 6 fatty acids.
- Pentose phosphate pathway: provides reductant (NADPH) for lipogenesis:
    - NADPH form of NADH involved in the anabolic reactions
    - NADH production by PPP is proportional to demand
    - Early reactions oxidise and rearrange G6P, generating NADPH and 5C sugar phosphate
    - Later reactions oxidise and rearrange carbons to allow re-entry to glycolysis
- PPP is also used to make ribose-5-phosphate for DNA, RNA, NAD, CoA and ATP.
- NADPH also acts as an antioxidant –red blood cell deficiency in first enzyme in PPP can cause anaemia
- Esterification: formation of fat (glycerol + 3 fatty acids). Glycerol needs to be in the glycerol 3-phosphate form from reduction of glyceraldehyde 3-phosphate for glycolysis. Glycolysis is important for production of both acetyl CoA and glycerol.
    - Esterification enzymes use FA-CoA –fatty acids added one at a time
    - Both esterification enzyme and FAS are unregulated by insulin, regulated by insulin stimulated gene expression and protein synthesis.
    - Fatty acids down regulate when there is an abundance of fat

DNA Repair

- Repair of DNA is vital to the survival of species. Unlike RNA, DNA has the capability to repair itself. The extra copy provides the template and elaborate repair mechanisms have evolved to correct corruptions. Many errors are corrected by the 3' → 5' exonuclease activity of DNA pols I and III. There are corruptions to the sequence which occur after replication. 130 genes which encode proteins are responsible for repair in the human genome.
- When DNA is exposed to UV radiation, it forms pyrimidine dimers. C5 and C6 bonds of two consecutive pyrimidines form covalent bonds with each other (cyclobutyl ring). Cyclobutyl bonds are ~1.6 Å so are much shorter than distance between bases (~3.4Å). This causes a bulge in the DNA double helix which prevents transcription and replication. There are ways to repair the dimer.
    - One way is to cleave the bonds directly with photolyase. This enzyme uses the energy from visible light to cleave the cyclobutyl bond. Photolyase is present in many prokaryotes and eukaryotes but not in humans.
    - Another type of repair is nucleotide excision repair. The pyrimidine dimer and adjoining nucleotides are cut out by excision enzymes. The enzymes seek out the bulge and cut it out. DNA pol I then fills out the missing sections using the second strand as a template. The gap is sealed by DNA ligase.
- Uracil is created by spontaneous deamination of cytosine, and does not belong in the DNA. A set of excision enzymes (base excision repeair) will cleave the base at the glycosidic bond (depurination) leaving an apurinic or apyrimidinic site. Deoxyribose + adjoining nucleotides are then excised by an AP endonuclease. The whole section is filled with DNA pol I and sealed with ligase.
- A number of anti-cancer drugs that induce DNA damage and apoptosis (e.g. Doxorubicin). Anti-cancer drugs used to treat brain tumors interfere with replication. Ability of cells to repair this damage limits the effectiveness of the drug. Success of these therapies will depend on the activity of repair enzymes.
- Mutations: a mutant gene has a different sequence to the wild type gene. The change is inheritable, however the mutation may or may not cause a change in phenotype. For a mutation to be inherited, it must be present in the germ line cells. Mutations in somatic cells or in the RNA code will not be inherited.
- Static mutations: change in the code becomes a stable incorporation into the genome of the germline cells as well as all somatic cells in the organism. The change is transferred to the next generation. Examples include PKU, cystic fibrosis and sickle cell anaemia. Expression in the phenotype depends on genetic information from both parents and epigenetic factors.
- Dynamic mutations: mutations increase in severity with each generation, varies between tissues of the same organism. Examples include Trinucleotide repeats (TNR). These mutations lead to genetic anticipation. The following generation will be more affected and/or have an earlier onset.
    - Can occur whenever DNA is being copied
    - Formation of aberrant loop structure when DNA strands are separated
- Types of mutations:
    - *Transversion*: purine replaced by pyrimidine or vice versa, implications for 3D shape of helix
    - *Transition*: pyrimidine for pyrimidine or purine for purine
    - *Silent mutations*: altered codon still codes for the same amino acid (cause code redundancy)
    - *Frameshifts*: shifts reading frame by adding/deleting bases, leading to non-functional protein
    - *Neutral mutation*: altered codon codes for a functionally similar amino acid hence has no effect on the functionality of the protein
    - *Point mutations*: single base pair change, can be a substitution, deletion or addition
    - *Missense*: altered codon for functionally different amino acid. Can be lethal.
    - *Nonsense mutation*: produces a stop codon and causes a truncated protein. Very dangerous.
    - *Splice mutation*: produces or removes a splice site –only in eukaryotes
    - *Temperature sensitive*: mutation causes changes in function of temperature sensitive protein Protein functions normally at lower temperature but is inactive at high temperatures.
    - *Reversion*: a mutation which reverts to wild type, referred to as revertant
    - *Leaky mutation*: doesn't affect organisms in normal conditions only in stressed conditions.
- Mutagen: physical or chemical agent that causes mutation to occur at a higher rate or frequency.

- Natural/spontaneous mutagenesis: Occurs at a normal background rate all the time. Alternative tautomer of the base –if base changes during replication, the wrong nucleotide will attach and by two rounds of replication there will be a base pair switch. An error in replication that is not picked up by proof-reading activity of DNA pol III will result in a mutation.
    - Deamination of cytosine or adenine to hypoxanthine.
    - If the adenine flips to the alternative tautomer at the time of replication, just as the new incoming nucleotide is selected, it will base pair to cytosine rather than thymine as the H donor and acceptor are the other way around.
    - Within another generation, there will be a GC instead of an AT
    - These can be corrected by specific repair mechanisms
- Induced mutagenesis:
    - Intercalators: planar, ring structures with slide in between the pairs causing a disruption to normal base stacking (e.g. ethidium bromide)
    - Alkylating agents: methylate or ethylate bases result in altered base pairing during replication
    - Base altering agents: e.g. nitrous acid converts amino (donors) to keto (acceptors) groups
    - Base analogues: e.g. 5-bromouracil replaces thymine but base pairs to guanine

Testing mutagenesis: Ames test.

- Quick screening test for potential mutagenic compound
- Strain of salmonella with defect in histidine biosynthetic pathway is plated out on a medium containing minimal histidine (enough to keep cells alive, but not for proliferation)
- Looking for a reversion mutation as bacteria is already a mutant
- Compound of interest applied to a disk in the centre of a plate and incubated over night
- Different plates with increasing amounts of the compound are put up
- If compound is mutagenic, it will cause a number of cells to revert to wild type and grow on the medium. Linear dose-response curve.
- The more colonies forming around the disk, the more mutagenic it is
- Non-mutagenic compounds will have a few colonies scattered over the whole plate