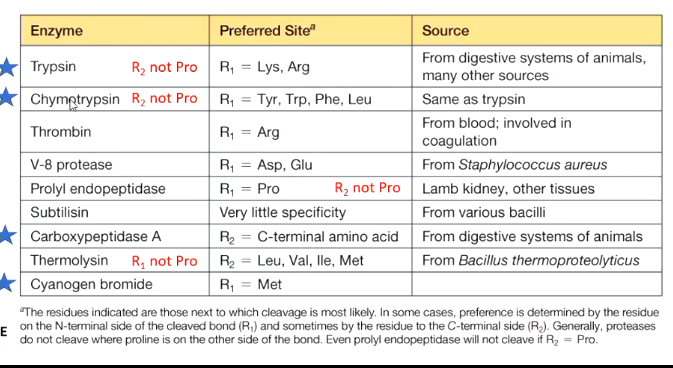
* Insulin has 2 N-termini and 2 C-termini – bind by disulphide bond
* Oxidise bond to break them
* Protect the disulphide position – cysteine residue – so that they don’t form disulphide bond and stick up again
* Sequence of entire insulin has 3 bonds
* Two chemical methods by which we can detect how many protein chains
  + Dinitrofluorobenzene – DNFB – add DNFB to get polypeptide (shorten it)
  + The aromatic ring can be cleaved out later to get the amino acid
  + Dansyl chloride – cleaves first aa off but the rest of polypeptide also chopped up into polypeptides – don’t get long chains – get free polypeptides
  + Pay attention to these methods
* Carboxyl terminal can be determined by using carboxypeptides – cleave off last residue – leaving new shortened amino acids
* Haemoglobin – both N-terminal of alpha and beta chains are the same – we need to know both chains
* Use beta-mercaptoethanol to break disulphide bond and then iodoacetate to keep them from sticking
* Use performic acid to cleave the bond in one step – form sulphate to prevent bonding
* After having protein chains, need to cut them into small sequences to sequence them
* Two different methods to assemble them back in unambiguous way
* Trypsin and chymotrypsin cleave proteins at different points
* Met-specific – use reagent cyanogen bromide (highly toxic) – only recognise methionine – cut there
* Trypsin cuts after lysine or arginine – only 2 aa that are positively charged in neutral Ph – cut after C=O and create O=C-O-
* Cannot cut residue of proline
* The star will be tested
* R1 is the side chain before peptide (C-N) bond – R2 is after
* R2 forms the new N-terminal
* R1 after cut will be the new C-terminal
* Which enzyme cuts after which site



* Edman used phenylisothiocyanate (PITC) to cleave off the first aa – aa is removed and identified – the first polypeptide chain is sent back and treated with PITC – recycled many times – one at a time – 1st aa, 2nd aa, etc.
* To determine exact sequences of aa in the peptide fregments – use MS-based proteomics – MS stands for mass spectrometry – give out mass values – how heavy is each fragment
* Look at mass values of each aa
* Digest protein with trypsin – cut them up – last fragine is usually argine and lysine – sent to mass spectrometry – come out as lines – sent to software computer to get the results – choices of what the sequences may look like
* Myoglobin contains only one chain
* Peptides still contain bonds that can be broken up
* Some peptides are sent for subsequence breakup
* Can locate disulphide bond using MS
* Real MS looks very messy – we zoom in to find bits of interest to us – fragment it futher to get subset
* Individual MS can be sent for second time to get more detailed information
* Fit peptide sequences together – chop in different places – look for regions of overlaps – how to put them together to get our desired sequence – help from softwares
* Overlap allows determination of aa sequence without doubt – use chemical reagent such as CNBr that looks for Met – trysin that cuts after arg and lys – matching of peptides using overlapping sequences give the long aa sequences
* GenBank is largest DNA bank
* Protein sequencing helps us understand what proteins do – large scale sequencing shows a lot of aa repeats – gene duplication
* Domains – similarly made proteins – have independent identity – in combination provide functionality
* A lot domains came from stealing from other organisms or common ancestors
* Domains mean parts of proteins that are also seen in other proteins – common bits of proteins that are seen in different proteins
* Try to summarise steps in sequencing proteins – why important to identify N-teminal – how? – explain why long polypeptides have to be broken up to small peptide fragments for sequencing – Prac 3 might be useful
* Protein structure – multiple levels
  + 3D is functional – form 3D shape that provides structure – come from small bits of local structures
  + Secondary structure – held by backbone hydrogen bonds
  + Quaternary structure – several subunits assemble – already folded up but assemble through non-covalent interaction – form functional units – last level of protein structure
* Haemoglobin
  + Primary structure is the aa sequence – 2 different chains – alpha and beta chains
  + Some parts of chains adopt repeatitive seq
  + The backbone curls up into spirals or springs – alphahelix – in globin (myoglobin etc) was predicted even before tested – responsible by hydrogen bond – got the name alpha because it was the first one
  + Helices are very compact – provide flexibility to structure
  + Tertiary structure – myoglobin? – helices assembled into a shape that facilitates binding of iron carrier group called haem – bound by certain aa to proteins – middle of flat assembly is the iron atom that holds oxygen
  + Quaternary – 2 alpha and 2 beta subunits
* Peptide bond is very unusual – neither single nor double – C-N is 1.49 A – C=N is 1.27 but peptide is 1.32 A
  + Partially double bond – because oxygen has 2 lone pairs but H has 1 lone pair not as strong as O at holding back electrons – lone pairs sometimes is between C and H – Oxygen binds with N – sp2
  + Usually see trans peptide bond – cis is very rare – Van de wal repulsion
* Rotation about alpha-C is the only free rotation
  + phi plane and psi plane
  + steric clash – when conformation of phi = psi = 0 – not allowed in polypeptides
  + naturally occurring secondary structures – no steric clash
  + right-handed alpha helix – left-handed is very rare in proteins
  + fully extended polypeptide chain – nothing is hitting anything else – beta sheet – 2 rows that can hydrogen bond each other – backbone also held by H bond
  + each sheet is composed of a single element called a strand – segment of polypeptide chains that is stretched to its max – fully extended
  + Proline – in such as collagen – structure protein – 310 helix forms forms tight spiral – mainly proline
  + Alpha – starting point is aa end – free N-terminus that has lone pairs – slightly positively charged – top has C group – negative ends – automatically dipolar – positive attracts to end of helix
* Peptide bond imposes planarity on the 6 backbone atoms of each peptide – free rotation only allowed in alpha C – dictate what types of assembly are possible – after proteins come out from ribosomes – automatically find a stable structure – without any external interaction - provide limited choices of how these sequences can fold up – both alpha and beta structures are stabilised by backbone H bond (non-covalent bond)