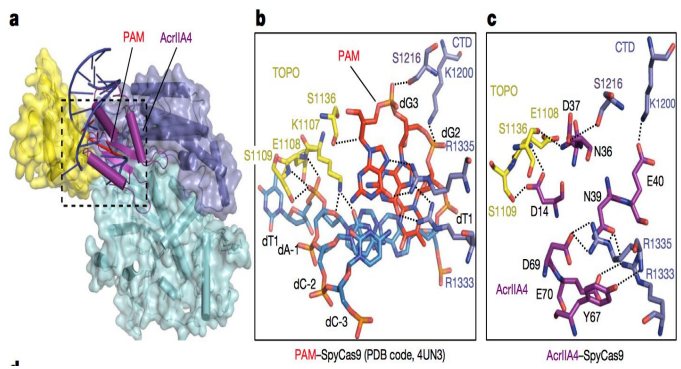
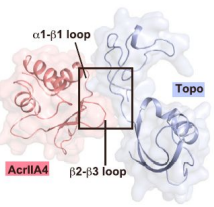
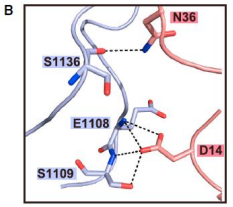
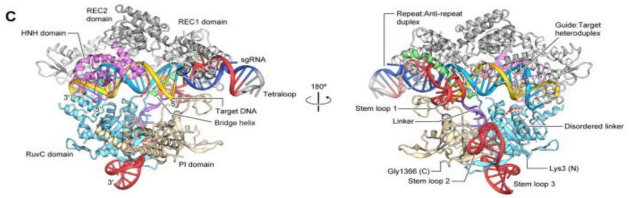
BIOCHEMISTRY: DNA: deoxyribose sugar; RNA: ribose sugar; aminoacyl-tRNA-synthetases: attach amino acids to tRNA; mass of Amino Acid: 110 daltons; A - T two hydrogen bonds, C - G three hydrogen bonds; Purines = A,G, double ring, Pyrimidines = C, T, U, single ring; **Intron:** noncoding region, interrupts coding regions; **Exon**: coding region; **Functional Groups:** Carboxyl: C=O & O-H; Amine: NH2; Sulfhydryl: SH; Carbonyl: C = O; Phosphate: PO4; Alpha carbon: carbon connecting R-group with hydrogen and peptide chain; **Bonds:** Ionic: electrostatic attractions between positive and negative ions, strongest bond, electrons are transferred (metal and non-metal); Covalent: sharing electrons; **Protein synthesized 5' to 3' (N to C)**; **Peptide bonds**: COOH & NH2, **double bond character, provides protein rigidity**; Disulfide bonds: covalent bonds between sulfhydryl groups; Hydrogen: attraction between partially positive and negative atoms in covalent bonds; Van der Waals: attraction caused by concentrated electron dispersion, causing areas of partial positives and negatives; **Denaturation**: mainly caused by *increasing* temperature/pH, *decreasing* pH, NOT decreasing temperature; can also be done with urea. **Primary**: peptide chain (unique chain of amino acids); **Secondary**: hydrogen bonding between peptide backbone; **Alpha Helices:** MALEK amino acids; hydrogen bonding between 1st and 4th amino acids (coil shape), 3.6 acids per turn, 5.4 Angstroms, phi = -60 degrees, psi = -45 degrees. *L orientation (enantiomer) of amino acids allows for helices to form right-handedly*; **Beta-pleated Sheet:** FIVTYW amino acids; parallel hydrogen bonds (zig-zag shape), bonds between carbonyl and amine groups, Parallel: 2 C-Terminuses opposite each other, 3.25 Angstroms, less stable due to angle; Anti-Parallel: C-Terminus opposite N-Terminus, 3.47 Angstroms; Turns: located on protein surface, contain polar and charged residues; **Tertiary:** side chains interact with each other; hydrogen bonds, salt bridges (ionic bonds), hydrophobic interactions (Van der Waals), disulfide bridges; **chaperonins**: proteins that assist in folding during tertiary structure; Thermodynamically favorable for hydrophobic amino acids to pack inwards **Motifs**: Heptad Repeat: repeating pattern of seven amino acids; Greek key: 2 beta sheets folded over; Beta hairpin: 2 antiparallel beta sheets with a turn; Omega loop: turn of proteins (omega sign); Helix-loop-helix: 2 alpha helices with a turn; Zinc finger: 2 antiparallel beta strands with alpha binded to zinc; Helix-turn-helix: 2 alpha helices with amino acids (gene expression); has a Zn2+ ion to stabilize fold Nest: 3 consecutive amino acids capable of binding with anions; Niche: nest, but for cations; **Amino Acids**: Glycine - the only amino acid that is not optically active (does not have a Left [L] or Right [D] configuration); Proline/Glycine - "helix breakers"; Acidic solution: carboxylic acid donates H+; Basic solution: amine gains H+; Alpha amino acids: amino group directly next to alpha carbon - PROLINE HAS NO ALPHA AMINO ACID (IMINO ACID) Peptide Bonds: **double bond character**, causes planarity and provides structure; **Essential Amino Acids**: histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine; **Glycosylation**: attachment of sugar moieties to proteins, is a post-translational modification (PTM) that provides greater proteomic diversity than other PTMs, critical for a wide range of biological processes, including cell attachment to the extracellular matrix and protein–ligand interactions in the cell, various glycosidic linkages, including N-, O- and C-linked glycosylation, glypiation (GPI anchor attachment), and phosphoglycosylation; **Types of proteins**: glycoproteins, lipoproteins, nucleoproteins, chromoproteins (contains cofactors); **Fibrous proteins**: structural, do not aggregate crystalline, form long fibers, strengthened by disulfide bridges **Globular proteins**: dynamic, L-amino acids often move, stack in layers, do not contain double helices; **Protease**: enzyme that digests/breaks down other proteins **ANFINSEN'S DOGMA**: protein's structure is determined by amino acids and optimal environment (sometimes the lowest energy confirmation requires protein to go over an energy hump) **LEVINTHAL'S PARADOX**: every amino acid link can go in 3 angles, 198 total bonds (peptide + psi + phi), (3^{198}), impossible for protein to reach final forms in milliseconds - paradox solved by **anfinsen's dogma + chaperonins**; **DNA**: polymerase works from 5' to 3', but works along strand that is 3' to 5'; ABZ DNA, B DNA is most common, different duplex structures between all three, 10 Angstroms per turn; **CRISPR Cas9:** (Clustered Regularly Interspaced Short Palindromic Repeats), Cas = CRISPR Associated; **Classes**: Class 1 = uses multiple Cas proteins to degrade nucleic acids, Class 2 = uses one large Cas protein (Cas9); **Types**: 90% of archaea, 50% of bacteria use CRISPR; three types are readily distinguishable by virtue of the presence of three unique signature genes: Cas3 in type I systems (bacteria + archaea), Cas9 in type II (bacteria), and Cas10 in type III (bacteria + archaea); Cas9 bends DNA helix 30 degrees, forms R-loop of DNA and RNA hybrid; **Inactive Cas9 Activated by guide RNA, considered an endonuclease** Function: genome editing: CRISPR array contains previous viral infection spacers separated by repeating GTT nucleotide sequences; **Mechanism: spacer acquisition, crRNA biogenesis, and target inference;** Cas1 and Cas2 proteins remove viral insertion (protospacer) and add to CRISPR array; **Palindromic Repeats**: allow for Cas1/2 to identify where to insert protospacer and allow for transcription of sgRNA, 30 bases; RNA sequence is transcribed from DNA of CRISPR array; Transacting RNA (tracrRNA) pairs with CRISPR RNA transcription through base pairing, forms sgRNA (single guide RNA), sgRNA is distinct in 5'/N terminus due to viral spacer, palindromic repeats are the same in all sgRNA in 3'/C terminus, Cas9 forms search complex; If RNA matches potential viral DNA, cuts DNA; Uses (Protospacer Adjacent Motif) for distinction from CRISPR array, 2-6 base pair sequence different from CRISPR array's GTT spacer, adjacent to tDNA, (-GG); Cas9 verifies CRISPR RNA match only if there is a PAM in viral DNA; Many errors occur in replacing double stranded breaks (DSBs) as a result of cuts; double-stranded break joined by Non-Homologous End Joining (NHEJ) in absence of repair template or Homology Directed Repair (HDR) w/ repair temp.; Origin - **E. Coli/Streptococcus Pyogenes**; **Structure: Helical-I and Helical-II, RuvC, HNH, Topo, and CTD domains;** Apo state - inactive, molecular interactions must occur for conformation to stabilize into active form; lack of binding to nucleic acids, HNH section is stabilized in a conformation rotated around 180 degrees, must reorient to bind, inactive in the absence of guide RNA; **RuvC/HNH = NUC lobe, endonucleases**, HNH - cleaves DNA strand complementary to the sgRNA sequence (target strand tDNA); 2 stranded antiparallel beta sheets, alpha helices on each side; Mg^{2+} ion and His840 are necessary for cleavage; Mutation of His840 to Ala840 results in inactivity of HNH domain; Single metal ion required for cleavage; RuvC - cleaves non-complementary strand (non-target ntDNA); Asp10 and Mg^{2+} ions are critical for non-complementary strand cleavage; RNase H fold where active site is, cleave RNA in RNA/DNA fusion; Two metal ions required for cleavage; TOPO - residues 1136spy to 1200spy, resemble topoisomerase, Greek Key Motif (β - β - α - β); CTD - $\alpha\beta$ region in Carbon terminal, protrudes off NUC lobe; **REC1/REC2 = REC lobe, recognition, alpha helices**; Initially bind to complementary DNA, later bind tDNA to guide RNA; REC1 - elongated, 25 alpha helices, 2 beta sheets; Critical for cas9 activity; Responsible for binding guide RNA to tDNA, forming duplex; REC2 - inserted within REC1; 6 alpha helices; Not necessary for activity; **Bridge helix - combines NUC and REC lobes**; Arg66,70,74 form multiple salt bridges w sgRNA; Arginine abundance allows for dsDNA cleavage; REC and NUC lobes fold into positively charged groove at their interface where negatively charged sgRNA can bind; **PI (PAM interacting domain): recognizes noncomplementary PAM sequence** different from crRNA - **made of TOPO & CTD**; Topo - topoisomerase - unwinds/relaxes DNA, CTD - C-Terminal Domain. 7 alpha helices, 3 stranded antiparallel beta sheet, 5 stranded antiparallel beta sheet, 2 stranded antiparallel beta sheet; If there's no PI domain, prevents cleavage activity; Initiates binding to target DNA, has PAM specificity; Target DNA phosphate backbone interacts w REC1 and PI domains; Some form VDW interactions with REC1 and RuvC; This is how Cas9 can differentiate between DNA and RNA; Stem loops 2-3 in sgRNA are recognized by NUC lobe, stem loop 1 is recognized by NUC and REC lobes; Uses: Cas9 can transport specific enzymes onto DNA which can do diff processes to DNA; Transcriptional activators can bind to deactivated Cas9 (dCas9); Activators can cause transcription of DNA; KRAB domain inactivates transcription by physically blocking gene (gene silencing); Fluorescent proteins can be attached to see where particular DNA sequences are found **GENETIC ENGINEERING APPLICATIONS:** gRNA - 17 to 20 bps; **Dead CRISPR (CRISPRi)** - catalytically inactive, unable to cut dsDNA, prevents transcription of particular gene by covering it, reversible; **CRISPR Activation (CRISPRa)** - selective activation of certain genes using activator protein; **NHEJ Repair**: insertion/deletion of nucleotides following cleavage; **Knockout**: deactivation of a gene due to frameshift mutation caused by indels; **HDR (Homology Directed Repair)**: use of DNA template to knock in a desired gene following cleavage, requires inserted DNA template; **Zinc Finger Nickases**: recognizes a 3- to 4-bp DNA sequence, recognizes two sequences flanking the site, one on the forward strand and the other on the reverse strand. Upon binding of the ZFNs on either side of the site, the pair of FokI domains dimerize and cleave the DNA at the site; **TALEN (transcription activator-like effectors nuclases)**: restriction enzymes that are larger versions of ZFNs; Anti-CRISPR: Function: **inhibition of Cas9**: Inhibition mechanism: AcrIIA2 and AcrIIA4 proteins utilize **crystal structure** to target regions that prevent PAM recognition and RuvC endonuclease activity in opposite strand cleavage, **Arg** residues recognize PAM in Cas9; AcrIIA4 interacts with **positively charged surface of Cas9** through hydrophilic groups, AcrIIA4 has **negative charge**; Side chains (Asp14 and Asn36) in AcrIIA4 are involved in hydrogen bonding with amino acids in Cas9; Ser1136 in Cas9 is involved in contact with PAM nucleotides; **competitively occupies both PAM-inter-acting and non-target DNA strand cleavage catalytic pockets**; Mimics PAM, occupies interaction site, prevents PAM recognition, prevents endonuclease cleavage of dsDNA (double stranded DNA); **Structure:** AcrIIA4 creates tertiary complex through **binding with TOPO, CTD, RuvC-III face of Cas9**; Triple-stranded b sheet with three a helices positioned along one face of b sheet; Asp14 & Asn36 hydrogen bond with Glu1108, Ser1109, and Ser1136; Cas9 residues normally stabilize PAM nucleotide base Ternary complex of AcrIIA4, Cas9, & sgRNA is almost identical to original binary complex of Cas9 & sgRNA; Interacts with positive face of TOPO, CTD, RuvC face in NUC lobe through hydrophilic contacts; AcrIIA4 tertiary complex takes place of dsDNA/PAM duplex; Separation of DNA requires stabilization of phosphate lock, occupied by AcrIIA4; Arg1333 & Arg1335, critical in reading of NGG sequence in PAM are also inhibited, recognized by Tyr67, Asp69, Glu70, and Asn39; Asn25 hydrogen bonds with Cas9 key residues Asp10, Glu762, His983, blocks active site, **prevents access to non-target DNA for verification, prevents RuvC endonuclease activity**; Ser20 to Ser26 and Lys18 interact with Glu762, His983, Asp10 on Cas9, prevents entrance of non-target DNA; Application: regulation of gene editing; CRISPR Cas9's cutting function is inaccurate with replacement of DNA, may cause mutation/inactivation; Allows for researchers to turn off CRISPR, avoid unwanted modifications; Origin: **Listeria Monocytogenes Prophage** Activity of AcrIIA4 is concentration-dependent. Cas9 must bind to sgRNA in order for AcrIIA4 to bind to Cas9. **YANG:** **Size Exclusion Chromatography (SEC)**: Graphs show that AcrIIA4 will not bind to Cas9 unless already bound to sgRNA, **prefers to bind to Cas9-sgRNA over Cas9-sgRNA-dsDNA or dCas9-sgRNA-dsDNA complex**, Preference of Cas9-sgRNA-AcrIIA4 is due to the lack of dsDNA, prevents competition, allows for the formation of a tight ternary complex; Absorption = amount of protein, volume = amount of eluent, smaller molecules move slower, require more eluent, more excess AcrIIA4 = less binding ; **Molecular weight of AcrIIA4** - 10.2-10.4 kDa, exists as a monomer in solution. AcrIIA2/4 cannot remove dsDNA from its interactions in ternary complex **DONG:** dCas9 = Cas9 without endonuclease activity; Spy belongs to type II-A CRISPR subtype; 3 Alpha Helices bundle + 2 Beta Sheets; conformations of PAMinteracting domain and RuvC of SpyCas9 interacting with both PAM and AcrIIA4 are nearly identical. By contrast, the conformations of REC lobe and HNH domains are markedly different due to binding of the other part of the target DNA; Crystal structure determined at 3.0 A resolution; AcrIIA4 has negatively charged single domain; **Technologies in Biochemistry:** BLAST - algorithm used to compare amino acid sequences/DNA/RNA sequences to a database. Size Exclusion Chromatography (SEC): Passes proteins through a column with small molecular sieves. Larger molecules move faster through the column and appear closer to the y-axis on a absorption vs. elution volume graph. Peaks indicate similarly-sized proteins/complexes. Gel Electrophoresis: Uses an electric current to pull molecules through agarose gel, Smaller molecules move faster. If the molecules are negatively charged, then the anode (+) is placed on the bottom, vice versa. Different gels are used in biochem, esp. Proteins. Could use a denaturing polyacrylamide gel, esp. For proteins. X-ray Crystallography: X-rays are incident on a protein/molecule; the molecule causes the X-rays to diffract. Crystallographer then measures the angles/intensities of diffracted rays; produces an electron density map. Based on this, can determine location of atoms/bonds. Western Blot: separates proteins based on polypeptide length, separated on polyacrylamide gel, transferred to protein sheet, exposed to radiolabel antibodies that detect desired protein Cryogenic Electron Microscopy (Cryo-EM): Uses frozen substances and an electron beam to produce an image. Similar to Transmission Electron Microscopy (TEM) but uses a gentler electron beam and more sophisticated imaging technology. Polymerase Chain Reaction (PCR): Used to clone genes/copy DNA. Needs a primer (complimentary to target sequence). Uses thermal cycling. First **Denatures** (96C) DNA - unwinds, then **Anneals** (55C) - allows primer to bond to single-strand template, then **Extends** (72C) - allows primer to extend; **AMINO ACIDS:** **Alanine(Ala,A,89mm)** - α-AA, L-ala in proteins, D-ala in bacterial cell wall; liked with TII diabetes with altered glucose-alanine cycle between tissues and liver, substitutes (sub) small amino acid, has normal C-beta carbon, part non-critical protein contexts, cannot adopt many conformations, function non-reactive, can play role in substrate recognition; **Arginine(Arg,R,174)** - α-AA, precursor to NO biosynthesis=regulates BP, essential to infants, cell division, reduces healing time, removes ammonia from body, immune function, release of hormones, semi essential (conditionally); multiple hydrogen bonds with phosphates since +charge delocalized w/ double bond+3 N LPs, sub other (+) amino acids, amphipathic part near side chain hydrophobic, end positively charged, normally charged part outside, involved in salt bridges, frequent protein active/binding site, has complex guanidinium group geometry; **Asparagine(Asn,N,132)** - α-AA, req in nervous system (brain) develop/function, ammonia synthesis, non-essential; found in sheets, surface of proteins; when on inside, help to stabilize, prefer sub other polar residues, involved protein active/binding sites frequently, usually found near beginning and end of alpha helices asx turns/motifs, "caps" hydrogen bond interactions, provides key sites for N-linked glycosylation, inhibits HIF1-mediated gene factor, enters Krebs cycle as oxaloacetate; **Aspartic Acid(Asp,D,133)** (pH 3.9)- neurotransmitter, non-essential, sub Gln and polar amino acids, found in surface of proteins, form salt bridges when inside protein, involved in active/binding site, shorter side chain = more rigid, stronger preference to be involved in active sites, Asp-His-Ser catalytic triad; **Cysteine(Cys,C,121)** - antioxidant properties, hydrophilic, semi-essential, generally doesn't sub, involved in disulfide bonds in TERTIARY STRUCTURE - Cys oxidized/stabilize structure, can tolerate other small amino acids, extracellular proteins, intracellular environment sulfhydryl side chain good bind metals (zinc), common protein active/binding sites, nucleophile (donates electrons), reactive center of enzymes; **Glutamine(Glu,Q,146)** - regulates acid-base balance in kidney by producing ammonium, cellular energy (next to glucose), typically non-essential but may become essential during certain gastrointestinal disorders; found in helices, prefer surface of protein, sub other polar residues, involved protein active/binding sites; **Glutamic Acid(Glu,E,147)** (pH 4.2)- most abundant excitatory neurotransmitter, plant growth, non-essential, prefers to sub Aspartate but can sub polar amino acids, generally on surface, involved in active/binding sites; **Glycine(Gly,G,75)** - inhibitory neurotransmitter, non-essential; flexible, bonds to phosphates, conformational flexibility due to hydrogen in side chain - reside tight turns, sidechain-less backbone bind phosphates - conserved glycine change other amino acids = impact. Only achiral amino acid, NOT ALWAYS IN L CONFORMATION ; **Histidine(His,H,155)** - important in hemoglobin, precursor for histamine, essential; common in active sites, doesn't sub anything well, easy to move protons on/off side chains - ambiguous about whether it's inside or outside core, ideal residue protein functional centers, charge relay systems, common metal binding sites, ONLY ACID THAT DOES NOT ABSORB NEAR-UV; **Isoleucine(Ile,I,131)** - essential (not produced by body, must be ingested), aliphatic, prefer sub w/ same type, prefer protein core, C-beta branched - 2 non-hydrogen substituent - lots bulkiness near backbone - more restricted in conformations, difficult adopt alpha helix, prefer beta sheets; non-reactive - rarely involved, except for substrate recognition, beta branched; **Leucine(Leu,L,131)** - essential, aliphatic, prefer core of protein, part of alpha helix, nonreactive, rarely involved except for substrate recognition, beta branched; **Lysine(Lys,K,146)** - essential, reduces anxiety, conjugates cause cancer cells to destroy themselves (with phototherapy), prefer sub Arginine or other polar, amphipathic as part side chain near backbone hydrophobic and end is positively charged, generally on surface, involved in salt bridges, frequent protein active/binding site, form hydrogen bonds; **Methionine(Met,M,149)** - used by plants for synthesis of ethylene, coded by AUG - "start" codon, essential, lack linked to senile graying of hair, aliphatic, subs hydrophobic AA, in protein core, non-reactive except recognition, can be involved bind metals, has methyl group, limited compared cysteine; **Phenylalanine(Phe,F,165)** - precursor for norepinephrine (noradrenaline), epinephrine (adrenaline), and melanin (skin pigment), aromatic, prefer sub same type esp Tyr, prefer protein core, involved stacking interaction w/ aromatic side chains, non reactive except recognition (lipids),can bind poly-proline containing peptides; **Proline(Pro,P,115)** - non-essential, can bind twice to backbone, sub small amino acids, rigid, tight turns protein structure, introduce kinks alpha helix/beta strands, found on surface, important molecular recognition, bind surface w/ aromatic residues, non-reactive. Has 2ndary amino group. Limits flexibility of protein chains with N in ring. Usually a cis more than trans.; **Serine(Ser,S,105)** - non-essential, important in metabolism, subs other polar/small amino acids, can be inside or on surface, common tight turns, hydroxyl oxygen form hydrogen bond w/ backbone, mimic proline, common functional centers, highly reactive/form hydrogen bonds, known Asp-Cys-Ser catalytic triad found in hydrolyses, cut other molecules, intracellular proteins, phosphorylation= kinases attach phosphates to serine facilitate signal transduction process **Threonine(Thr,T,119)** - essential, subs polar or small AA, can be on inside or surface, C-beta branched, 2 non-H substituent, restrictive/prefer beta sheets, common functional centers, hydroxyl group reactive, intracellular proteins, phosphorylation **Tryptophan (Trp,W,204)** - essential, improper absorption (linked to lactose intolerance) can lead to depression, aromatic, prefer sub same type, buried in protein core, involved in stacking, contain non-carbon atom (N) in ring system, more reactive than Phe, less reactive than Tyr, nitrogen can bind non-protein atoms, rare, can be involved in non-protein ligands, poly-proline containing peptides, melatonin; **Tyrosine(Tyr,Y,181)** - non-essential, precursor to neurotransmitters, hormones, and pigments, aromatic, prefer sub same type esp Phe, partially hydrophobic, prefer protein core, involved in stacking, create reactive hydroxyl group (more likely involved in interaction of non-protein atoms), involved non-protein ligands, poly-proline, intracellular proteins phosphorylation **Valine(Val,V,117)** - essential, causes abnormal red blood cell shape in sickle-cell anemia, aliphatic, prefer sub other AA of same type, prefer protein core, C-beta branched, prefer beta sheets, non-reactive, rarely involved function except substrate recognition, beta branched **HISTORY:** Jennifer Doudna and Emmanuelle Charpentier published paper in *Science* in April 2012 on Cas9. CRISPR sequences were initially discovered in the **E. coli genome in 1987**, but their function as a safeguard against bacteriophages was not elucidated until 2007. Acr origin is in Listeria Monocytogenes Prophage. Francisco Mojica found palindromic sequences in Archaea, coined CRISPR acronym, analyzed palindromic repeats using BLAST. Philippe Horvath analyzed specific DNA sequences in bacterial immune system, discovered Cas9. John van der Oost developed artificial CRISPR Array. Emmanuelle Charpentier discovered tracrRNA.Virginijus Siksnys applied CRISPR in vivo. Doudna and Charpentier created artificial sgRNA. Feng Zhang used CRISPR to edit mammalian cells precisely **RAMACHANDRAN ANGLES:** dihedral (defined by 4 points), Psi Torsion: around Alpha Carbon - Carbon bond; Phi Torsion: around Alpha Carbon - Nitrogen bond; Omega (peptide): double bond character due to nitrogen's lone pair; omega bond does not twist due to double bond character, psi and phi twist; beta sheets have a positive psi, negative phi; alpha helices have positive psi and phi; **MODELING TIPS:** correct position and direction of end caps; correct number of helices and sheets; right-handed helices; properly folded/labeled sheets; correct order of secondary structures; correct topology; correct angles; parallel/antiparallel beta sheets; correct position and direction of sidechains. Color beta strands with sharpie if time allows. **PEPTIDE pH: carboxyl: 2.19, amine: 9.67, arginine: glutamic acid: 4.2, aspartic acid: 3.9**; if **pka > pH solution, protonated**, if **pka < pH, deprotonated**; at pH of 2, 7, 12: Glutamic acid 0 -1 -1 Aspartic acid 0 -1 -1 Lysine +1 +1 0 Arginine +1 +1 +1 (50%) Histidine +1 +1 (25%) 0 Tyrosine 0 0 -1 Cysteine 0 0 -1;

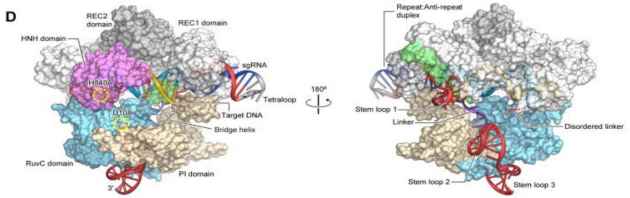


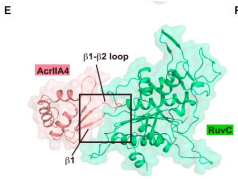
PAM–SpyCas9 (PDB code, 4UN3) AcrIIA4–SpyCas9

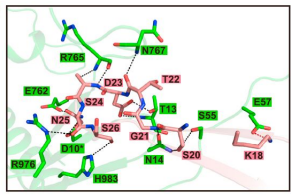


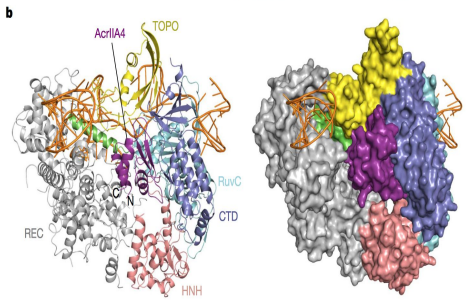


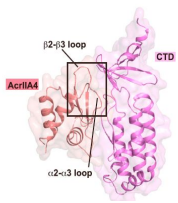


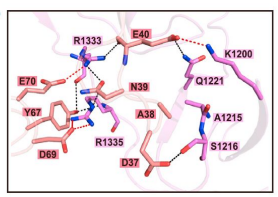


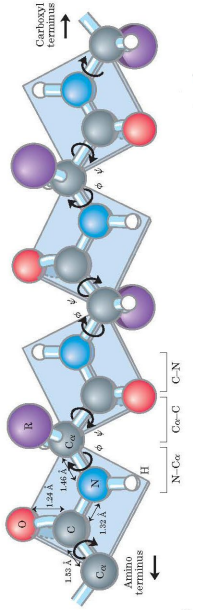










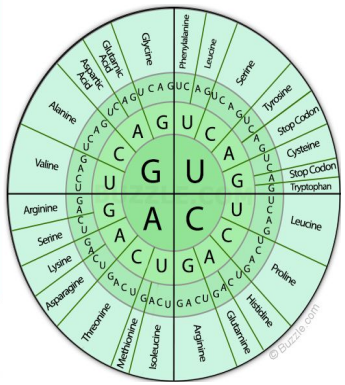


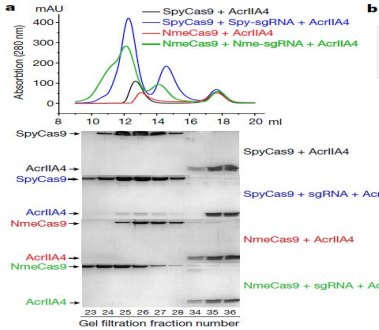
Name	Amino Acid	Side Chain	Name	Amino Acid	Side Chain	Name	Amino Acid	Side Chain	Name	Amino Acid	Side Chain
aliphatic Alanine Ala A	$C_3H_7NO_2$		acid Glutamine Gln Q	$C_5H_{10}N_2O_3$	C-E	aliphatic Leucine Leu L	$C_6H_{13}NO_2$	E	Serine Ser S	$C_3H_7NO_3$ hydroxyl-contain	
basic Arginine Arg R	$C_6H_{14}N_4O_2$	C-E	amide Glutamic Acid Glu E	$C_5H_9NO_4$	E	basic Lysine Lys K	$C_6H_{14}N_2O_2$	E	Threonine Thr T	$C_4H_9NO_3$ hydroxyl-contain	E
amide Asparagine Asn N	$C_4H_8N_2O_3$		aliphatic Glycine Gly G	$C_2H_5NO_2$	C-E	sulfur-contain Methionine Met M	$C_5H_{11}NO_2S$	E	aromatic Tryptophan Trp W	$C_{11}H_{12}N_2O_2$	E
acid Aspartic Acid Asp D	$C_4H_7NO_4$		basic aromatic Histidine His H	$C_6H_9N_3O_2$	E	aromatic Phenylalanine Phe F	$C_9H_{11}NO_2$	E	C9H11NO3 Tyrosine aromatic Tyr Y		C-E
sulfur-contain Cysteine Cys C	$C_3H_7NO_2S$	C-E	aliphatic Isoleucine Ile I	$C_6H_{13}NO_2$	E	cyclic Proline Pro P	$C_5H_9NO_2$	C-E	aliphatic Valine Val V	$C_5H_{11}NO_2$	E

Atom Color Key: Carbon, Oxygen, Nitrogen, Sulfur, Hydrogen

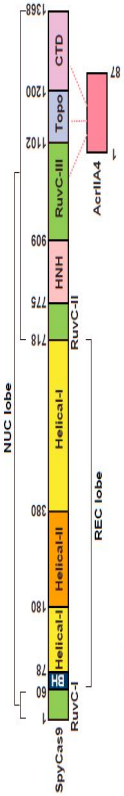
Amino Acid Property Key: Amino acid clip color and name color indicate property

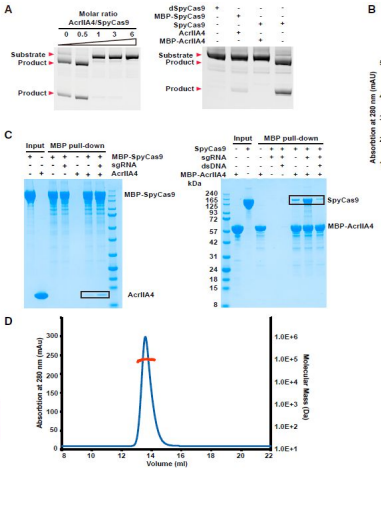
Negative Charge, Positive Charge, Hydrophilic, Cysteine, Hydrophobic

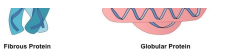




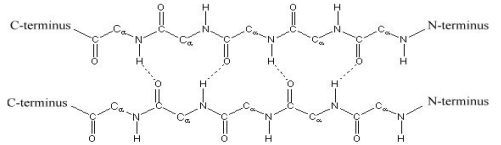
Fibrous	Globular
Long and narrow	Round / spherical
Structural	Functional
Repetitive amino acid sequence	Irregular amino acid sequence
Less sensitive to changes in pH, temperature, etc.	More sensitive to changes in pH, temperature, etc.
Collagen, myosin, fibrin, actin, keratin, elastin	Enzymes, haemoglobin, insulin, immunoglobulin
(Generally) insoluble in water	(Generally) soluble in water







Parallel β Sheet



Antiparallel β Sheet