

Restriction Endonucleases

- restriction endonucleases: proteins that can **cut DNA at a specific point in a specific sequence**, allowing genome editing.
- termed "restriction enzymes" because they **restrict the infection of bacteriophages**.
 - bacteria are under constant attack by bacteriophages (e.g. bacteriophage phiX174)
 - To protect themselves, many types of bacteria have developed a method to **chop up any foreign DNA**, such as that of an attacking phage.
 - bacteria build an endonuclease (an enzyme that cuts DNA) which is allowed to circulate in the bacterial cytoplasm, waiting for phage DNA.
- Each type of restriction enzyme seeks out a single DNA sequence and precisely cuts it in one place.
 - Example: EcoRI, cuts the sequence GAATTC, cutting between the G and the A.
- Roving endonucleases can be dangerous, so bacteria **protect their own DNA by modifying it with methyl groups**.
 - These groups are added to adenine or cytosine bases (depending on the particular type of bacteria) in the major groove.
 - methyl groups **block the binding of restriction enzymes but don't block the normal reading and replication** of the genomic information stored in the DNA.
 - DNA from an attacking bacteriophage won't have protective methyl groups and will be destroyed.
 - Each particular type of bacteria has a restriction enzyme (or several different ones) that cuts a specific DNA sequence, paired with a methyl-transferase enzyme that protects this same sequence in the bacterial genome.

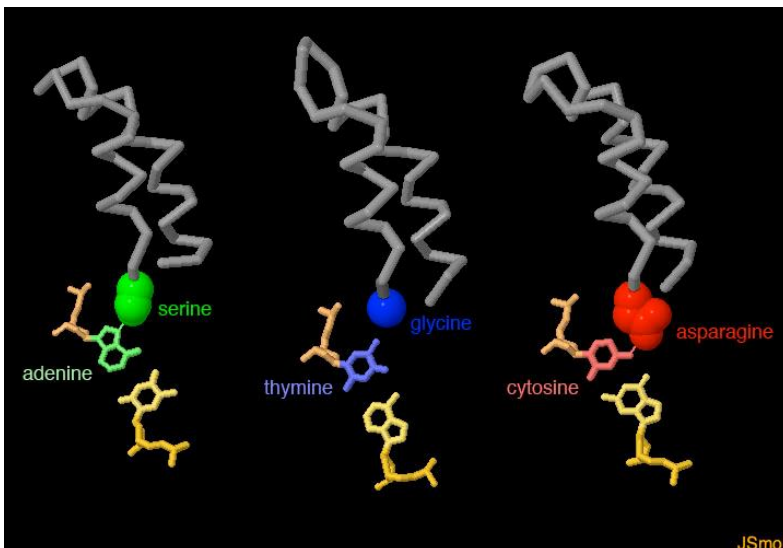
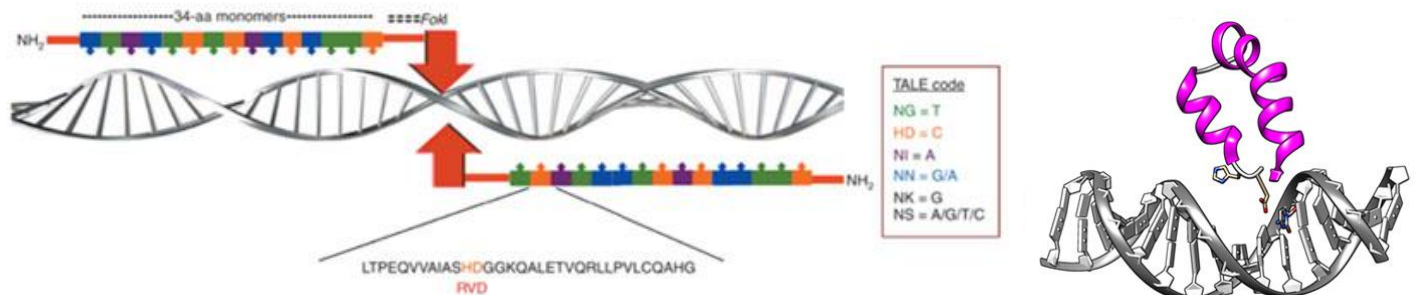
Endonuclease FokI

- The specific nuclease FokI occurs naturally in bacteria as a defense mechanism against invading viruses.
 - enzyme derived from **Flavobacterium okeanokoites** (or Planomicrobium okeanokoites)
- This protein, like other restriction enzymes, has two domains (functional parts):
 - the **cleavage domain (nuclease)**
 - the **DNA-binding domain, composed of zinc fingers**.
- commonly used in designing genome editing nucleases
 - the **nuclease** of the FokI is typically **removed from its natural DNA binding domains** and attached to new binding domains, to create a new specialized restriction enzyme.
- the nuclease **functions solely as a dimer**, meaning it requires two copies (one attached to each strand of DNA) in order to successfully cleave the DNA.
- Winged helix DNA binding domain: The winged helix motif has of two wings, three alpha helices and three beta-sheets arranged in the order H1-S1-H2-H3-S2-W1-S3-W2.
- The DNA-recognition helix makes sequence-specific DNA contacts with the major groove of DNA, while the wings make different DNA contacts, often with the minor groove or the backbone of DNA. Several winged-helix proteins display an exposed patch of hydrophobic residues thought to mediate protein-protein interactions.
- It has a **cofactor: Mg²⁺**
- can recognize specific DNA sequences (**5'GGATG3'** and **5'CATCC3'**) and cuts or cleaves it on both DNA strands 14 bases after the first bolded and underlined G and 13 bases before the bolded and underlined C.



TAL Effectors

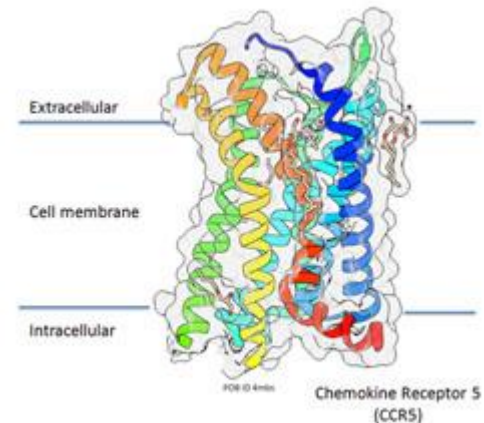
- Transcription activator-like (TAL) effectors
- First identified in species of *Xanthomonas* bacteria that cause diseases in plants such as rice and cotton.
- These bacteria inject several factors (including the TAL effectors) in the plant cells activating plant genes that allow the bacteria to thrive. In effect, these proteins act as weapons against the plant cells.
- The TAL effectors have a repeat motif composed of ~35 amino acids arranged into two helices connected by a loop.
- Each TAL effector module is composed of two alpha helices with a small kink in one, which improves the packing between the two.
- At the tip of the loop (position 12 and 13 of the repeat) there are two specific amino acid residues, called Repeat Variable Diresidues (RVDs) that vary in each repeat. These residues are responsible for each repeat being able to recognize specific base pairs in the DNA sequence.
- Understanding the structure and function of these repeats have helped design TAL effectors that can recognize specific sequences in the genome as part of genome editing nucleases called TAL effector nucleases (TALENs).
- These nucleases are also sequence specific DNA binding proteins. The interaction between RVDs in the TAL effectors and the target DNA base pairs define the specificity of DNA binding.
- TALEN attach to the DNA-cutting domain of FokI nuclease to one end of a TAL effector. FokI needs to form a dimer to cut DNA, so the TALEN only becomes active when two of them bind to the proper DNA target sequence. Then, it breaks both strands.
- This can be used to knock out a specific gene, or to stimulate the natural DNA repair methods that are present in cells, which may be coaxed into inserting a new engineered gene while they are making the repair.



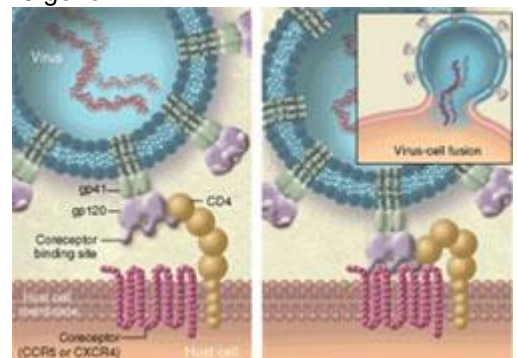
This structure includes an engineered TAL effector bound to a short piece of DNA, colored to highlight the amino acids that make specific contacts with the DNA. Aspartate is shown with red spheres, which makes specific interactions with cytosine bases; glycine is shown with a blue sphere, making contact with thymine in the DNA; and serine is shown with green spheres, interacting with adenine. Each module also includes a lysine and a glutamine that form non-specific interactions with the DNA backbone. The stabilizing residues are glutamine, glutamine, and histidine, respectively.

CCR5 (Chemokine Receptor 5)

- CCR5 is a membrane receptor protein found in human immune cells that is used by HIV to enter the host cell.
 - is an HIV co-receptor; cooperates with the host cellular CD4 primary receptor to allow the initial docking of the HIV virus onto T-cells, and subsequent infection.
 - The CD4 bound HIV envelope spike protein use this molecule as a co-receptor to enter and infect host cells.
 - In some instances HIV uses another similar chemokine receptor CXCR4 as the co-receptor for entry into host cells.
- A naturally occurring deletion in this protein enables a cell to become resistant to the HIV virus since it is unable to properly bind and insert its genetic information.
- Normally 353 amino acids long, and folds up into a structure composed of 7 transmembrane alpha helices with structural homology to the family of G protein-coupled receptors (GPCRs).
- Primarily, the CCR5 gene is involved in the receiving of chemical signals called chemokines and recruiting other immune cells to help the immune system function.
- However, this variation is homozygous recessive, meaning it requires both recessive alleles in order to express its resistant properties.
- In some ethnic groups (Caucasians) a 32 nucleotide deletion in the gene results in a corresponding deletion in the mRNA.
 - Because the genetic code is a triplet code, and 32 isn't a multiple of 3, the deletion results in
 - 1) the deletion of 11 amino acids
 - 2) a switch in the translational reading frame resulting in a scrambled amino acids sequence even after the deletion site.
 - 31 additional amino acids are added as a result of the deletion before a stop codon is met by the ribosome. This prematurely terminated CCR5 protein is 215 amino acids long.
- CCR5 normally dimerizes and is phosphorylated in the endoplasmic reticulum and is then efficiently trafficked through the Golgi to the cell membrane.
 - In contrast, 32CCR5 is not phosphorylated, and is not trafficked to the cell membrane.
 - 32CCR5 retains its ability to dimerize with wild type CCR5 leading to a transdominant negative effect on the delivery of the functional CCR5 to the cell surface.
- Approximately 15-20% of the northern European population is heterozygous for a naturally occurring 32 base pair deletion in their CCR5 gene – making them resistant to HIV infection.
- Approximately 1% of European caucasians are homozygous for this mutation – and resistant to HIV infection.
- Based on the functional cure of the Berlin patient it appears that introducing the CCR5 delta 32 mutation may make host cells resistant to HIV.
 - Using an engineered nuclease, such as a zinc finger nuclease, and specifically targeting the CCR5 gene in HIV patients to isolate and deactivate the CCR5 protein will make the patient's endogenous T-cells resistant to further infection.
- Since HIV infection is persistent, making the host cells resistant may provide a functional cure for HIV infected individuals.
- Sangamo Biosciences (a biotech company specializing in the development of therapeutic zinc finger nucleases) has developed a zinc finger nuclease that is targeted to disrupt the CCR5 gene.
 - currently being tested in a Phase 2 clinical trial with HIV/AIDS patients by Sangamo Biosciences in collaboration with groups from the University of Pennsylvania School of Medicine and the Albert Einstein College of Medicine.



The structure of the Chemokine Receptor CCR5 shown here is displayed within the context of the T-helper cell membrane. The PDB entry [4mbs.pdb](#) is that of an engineered molecule fused to rubredoxin and in complex with a fusion inhibitor drug bound to the extracellular face of the molecule.



Amino Acid Sidechain List

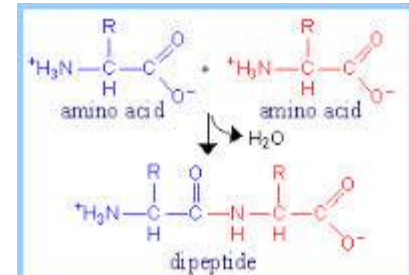
Dual Color Scheme: 1. Name and background shading indicates hydrophobic amino acids (yellow); hydrophilic non-charged amino acids (white); + charged amino acids (blue); - charged amino acids (red); cysteine (green).
2. Atom type indicates carbon (gray), oxygen (red), nitrogen (blue), and sulfur (yellow).

Name	Amino Acid	Sidechain	Name	Amino Acid	Sidechain	Name	Amino Acid	Sidechain	Name	Amino Acid	Sidechain
Alanine	Ala	A	Glutamine	Gln	Q	Leucine	Leu	L	Serine	Ser	S
Arginine	Arg	R	Glutamic Acid	Glu	E	Lysine	Lys	K	Threonine	Thr	T
Asparagine	Asn	N	Glycine	Gly	G	Methionine	Met	M	Tryptophan	Trp	W
Aspartic Acid	Asp	D	Histidine	His	H	Phenylalanine	Phe	F	Tyrosine	Tyr	Y
Cysteine	Cys	C	Isoleucine	Ile	I	Proline	Pro	P	Valine	Val	V

The Standard Genetic Code

	U	C	A	G	
U	UUU → Phe F	UCU → Ser S	UAU → Tyr Y	UGU → Cys C	U
	UUC → Phe F	UCC → Ser S	UAC → Tyr Y	UGC → Cys C	C
	UUA → Leu L	UCA → Ser S	UAA → Stop	UGA → Stop	A
	UUG → Leu L	UCG → Ser S	UAG → Stop	UGG → Trp W	G
C	CUU → Leu L	CCU → Pro P	CAU → His H	CGU → Arg R	U
	CUC → Leu L	CCC → Pro P	CAC → His H	CGC → Arg R	C
	CUA → Leu L	CCA → Pro P	CAA → Gln Q	CGA → Arg R	A
	CUG → Leu L	CCG → Pro P	CAG → Gln Q	CGG → Arg R	G
A	AUU → Ile I	ACU → Thr T	AAU → Asn N	AGU → Ser S	U
	AUC → Ile I	ACC → Thr T	AAC → Asn N	AGC → Ser S	C
	AUA → Ile I	ACA → Thr T	AAA → Lys K	AGA → Arg R	A
	AUG → Met M	ACG → Thr T	AAG → Lys K	AGG → Arg R	G
G	GUU → Val V	GCU → Ala A	GAU → Asp D	GGU → Gly G	U
	GUC → Val V	GCC → Ala A	GAC → Asp D	GGC → Gly G	C
	GUA → Val V	GCA → Ala A	GAA → Glu E	GGA → Gly G	A
	GUG → Val V	GCG → Ala A	GAG → Glu E	GGG → Gly G	G

 translation start codon
 hydrophobic amino acids
 negatively charged amino acids
 cysteine
 translation stop codon
 hydrophilic non-charged amino acids
 positively charged amino acids



Genome Editing

- requires bio-molecular tools (proteins and/or nucleic acids) that can specifically recognize a target sequence and cut the genome in a precise and predictable way.
 - The gene sequence that is cut out may be replaced with another gene sequence using homology dependent repair (HDR) or joined back together with intentionally introduced errors inactivating the gene using non-homologous end joining (NHEJ).
- Nuclease Domain: The portion of the protein responsible for cutting the DNA sequence once it is bound at the correct location.
- DNA Binding Domain: The portion of the protein responsible for finding a specific DNA sequence or gene and binding to it at the correct location.

How to introduce a protein into a cell:

- In order for a cell to make a protein, the genetic code (DNA) must be present within the host cell therefore the researchers must be able to give the cell the necessary DNA to make the appropriate protein.
 - Genes that are inserted directly into a cell usually do not function.
- Carriers, called vectors, are genetically engineered to deliver the gene into the target cell.
 - Some viruses are modified to be a vector for the gene of interest because viruses can infect the cell.
 - The virus integrates the genetic material into a chromosome in the host cell
 - it can be transcribed and then translated to generate a protein, as if part of the host cell's own genetic code.

Transcription Factors:

- The binding of certain proteins to DNA in order to regulate what is being transcribed into mRNA.

Protein Structure

- Glycine and alanine can combine together with the elimination of a molecule of water to produce a dipeptide.
 - possible for this to happen in one of two different ways - so you might get two different dipeptides.
 - In each case, the linkage in the structure of the dipeptide is known as a peptide link.
 - In chemistry, this would also be known as an amide link.
- If you joined three amino acids together, you would get a tripeptide. If you joined lots and lots together (as in a protein chain), you get a polypeptide.
- A protein chain will have somewhere in the range of 50 to 2000 amino acid residues. You have to use this term because strictly speaking a peptide chain isn't made up of amino acids.
- The end of the peptide chain with the -NH₂ group is known as the N-terminal [Amino], and the end with the -COOH group is the C-terminal [Carboxyl].
- 3D structure of a protein defines not only its size and shape, but also its function.
- One characteristic that affects function is the hydrophobicity of a protein, which is determined by the primary and secondary structure.
 - For example, membrane proteins. Membranes contain large amounts of lipids, which are hydrophobic.
 - The membrane-spanning regions of membrane proteins are typically alpha helices, made of hydrophobic amino acids.
 - These hydrophobic regions interact favorably with the hydrophobic lipids in the membrane, forming stable membrane structures.
- Most protein sequences instantaneously and reliably fold into its stable and functional shape, mostly without any assistance.
 - we have learned some rules about protein folding, but are still not able to predict protein structures accurately using computational structure prediction algorithms alone.
 - From studying the more than 100,000 experimentally determined structures in the Protein Data Bank, we can see that globular proteins have a hydrophobic core and most polar or charged amino acid side chains are located on the surface of these proteins.
 - there are a finite number of different protein domains.
 - Small changes in these domains (through evolution in nature or by engineering) can produce changes in its specific function.
 - Also combining the protein domains in various ways can produce a large variety in protein functions.

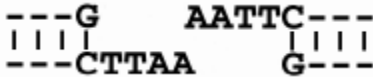
Sticky Ends

- The booming field of biotechnology was made possible by the discovery of restriction enzymes in the early 1950's.
 - With them, DNA may be cut in precise locations.
- A second enzyme--DNA ligase--may then be used to reassemble the pieces in any desired order.
- Together, these two enzymes allow researchers to assemble customized genomes.
 - For instance, researchers can create designer bacteria that make insulin or growth hormone or add genes for disease resistance to agricultural plants.
- An interesting property of restriction enzymes simplifies this molecular cutting and pasting. Restriction enzymes

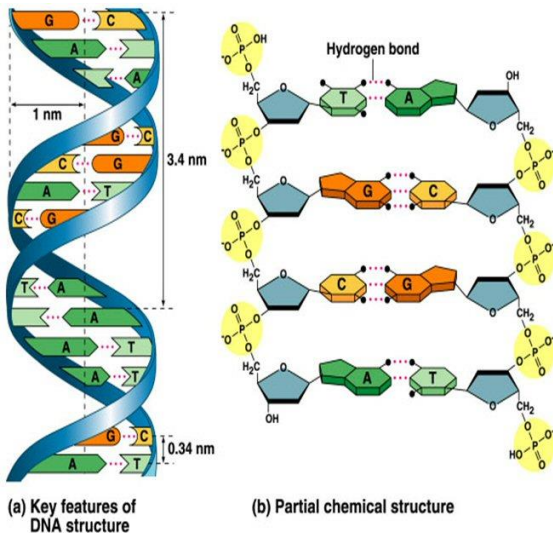


typically recognize a symmetrical sequence of DNA, such as the site of EcoRI:

- The top strand is the same as the bottom strand, read backwards.
- When the enzyme cuts the strand between G and A, it leaves overhanging chains:



- These are termed "sticky ends"
 - the base pairs formed between the two overhanging portions will glue the two pieces together, even though the backbone is cut.
 - Sticky ends are an essential part of genetic engineering, allowing researchers to cut out little pieces of DNA and place them in specific places, where the sticky ends match.



DNA

In general, the density of DNA is indicative of the frequency of transcription. **Octameric protein complexes** called nucleosomes are responsible for the amount of supercoiling of DNA, and these complexes can be temporarily modified by processes such as **phosphorylation** or more permanently modified by processes such as **methylation**.

Methylation of DNA is a common method of gene silencing. DNA is typically methylated by methyltransferase enzymes on cytosine nucleotides. Analysis of the pattern of methylation in a given region of DNA (which can be a promoter) can be achieved through a method called **bisulfite mapping**. Methylated cytosine residues are unchanged by the treatment, whereas unmethylated ones are changed to uracil.

Histone acetylation is also an important process in transcription. Histone acetyltransferase enzymes (HATs) dissociate the DNA from the histone complex, allowing transcription to proceed. Often, DNA methylation and histone deacetylation work together in gene silencing.

Activator - protein that binds to an enhancer (or activator binding region) and activates transcription from nearby promoter.

Baseline - a measure of the gene expression level of a gene or genes prior to a perturbation in an experiment, as in a negative control. Baseline expression may also refer to the expected or historical measure of expression for a gene.

Cellular reprogramming - conversion of a cell from one tissue-specific cell type to another. This involves a conversion to a pluripotent state, otherwise known as dedifferentiation.

Cofactor - non-protein compound that is bound to an enzyme. Cofactors are required for the initiation of catalysis.

Conditional gene expression - controlled inducible expression of transgene either in vitro or in vivo.

Constitutive gene or constitutive expression - a gene that is transcribed continually compared to a facultative gene which is only transcribed as needed.

Cis-dominant - mutations (e.g., of an operator) that alter the functioning of genes on that same piece of DNA. It arises because the operator represents a site on the DNA rather than a gene that encodes a product.

Coregulator - Transcription coregulators are proteins that work with transcription factors to regulate gene expression.

Distance measures - used to measure the dissimilarity between the expressions of different genes.

DNA Microarray - A DNA microarray is a high-throughput technology used to measure expression levels of mRNA transcripts or to detect certain changes in the nucleotide sequence. It is an array of series of thousands of microscopic spots of DNA oligonucleotides, called features, each containing picomoles of a specific DNA sequence. This can be a short section of a gene or other DNA element that are used as probes to hybridize a cDNA, cRNA or genomic DNA sample (called target) under high-stringency conditions.

Down-regulated - describes a gene which has been observed to have lower expression (lower mRNA levels) in one sample compared to another sample (usually a control).

Emergenesis - quality of genetic traits that result from a specific configuration of interacting genes, rather than simply their combination

Epistasis - the collective action of multiple genes that interact during expression. A form of gene action, epistasis can either be additive or multiplicative in its effects on specific phenotypic traits.

Gene expression - the transcription of DNA into messenger RNA by RNA polymerase. The messenger RNA is then translated into protein by the ribosome. In gene expression analysis, expression level refers to the amount of mRNA detected in a sample.

Facultative gene - a gene which is only transcribed as needed compared to a constitutive gene.

Gene knockdown - refers to techniques by which the expression of one or more of an organism's genes is reduced, either through genetic modification (a change in the DNA of one of the organism's chromosomes) or by treatment with a reagent such as a short DNA or RNA oligonucleotide with a sequence complementary to either an mRNA transcript or a gene.

Genetic regulatory network (GRN) - a graph that represents the regulatory complexity of gene expression. The vertices (nodes) are represented by various regulatory elements and gene products while the edges (links) are represented by their interactions. These network structures also represent functional relationships by approximating the rate at which genes are transcribed.

Housekeeping gene - typically a constitutive gene that is transcribed at a relatively constant level across many or all known conditions. The housekeeping gene's products are typically needed for maintenance of the cell. It is generally assumed that their expression is unaffected by experimental conditions. Examples include actin, GAPDH and ubiquitin.

Hybridization - refers to the process by which the single-stranded DNA or RNA preparation is added to the array surface, in solution, and potentially anneals to the complementary probe.

Inducible gene - a gene whose expression is either responsive to environmental change or dependent on the position of the cell cycle.

Insulator - a stretch of DNA that prevents a gene from being influenced by the activation or repression of nearby genes.

Library - a collection of DNA or oligonucleotide probes, often stored in a microtiter plate, which are transferred to the array during fabrication. NBF or Neutral Buffer Formalin - a common solution used in the fixation of cells or cellular tissue for marking and study.

Oligo - short for oligonucleotide, a small chain of nucleic acid residues.

- When used to detect the presence of larger mRNA molecules, see probe or reporter.
- When assembled into a two-dimensional array, see oligonucleotide array or microarray.

Operon - a genetic unit for coordinating expression of different genes.

Probe - a term to describe a reagent used to make a single measurement in a gene expression experiment. See reporter, probe-set.

Probe-set - a collection of two or more probes that are designed to measure a single molecular species. For example, several oligonucleotides designed to hybridize to various parts of the mRNA generated from a single gene.

Promoter - a region of DNA that facilitates the transcription of a particular gene.

Promotion - increasing the rate of gene expression.

Putative genes - a nucleotide sequence thought to be a gene based on the identification of its open reading frame (ORF). The gene is "putative" in the sense that no function has been assigned to its products.

Replication - a technique to estimate technical and biological variation in experiments for statistical analysis of the microarray data. Replicates may be:

- technical replicates, such as dye swaps or repeated array hybridizations, or
- biological replicates, using biological samples from separate experiments that test the effects of the same treatments.

Reporter - a term to describe a reagent used to make a single measurement in a gene expression experiment.

Repression - decreasing the rate of gene expression.

Repressor - a DNA-binding protein that regulates the expression of one or more genes by binding to the operator and blocking the attachment of RNA polymerase to the promoter, thus preventing transcription of the genes.

RNA splicing - modification of an RNA strand where exons (the coding regions of a transcribed gene) are retained and the introns are removed. Sometimes the exons are recombined either in vivo or experimentally to form alternative splicings, which have various functional effects.

Signal transduction pathway - a set of biochemical reactions and biomolecular interactions that convert a stimulus into a metabolic response. A signaling molecule initially binds to a receptor on the surface of the cell, which stimulates the activation of immediate early genes and second messenger molecules. This further activates enzymes that carry out larger-scale processes such as phosphorylation and gene expression. The process just described is known as a signaling cascade, during which the initial stimulus is amplified to have a far-reaching effect on the cellular environment.

Signature - refers to the set of expression measurements which satisfy a certain arbitrary threshold criteria, such as 1.5 fold change with a significance $p < 0.01$. Analysts may define a signature for a given experiment or compare gene signatures across many experiments.

Spatially-restricted gene expression - genes that are expressed only in a specific anatomical region or tissue, often in response to a paracrine signal. The boundary between two spatially-restricted genes can set up a sharp gradient, often expressed phenotypically as striping patterns.

Suppression - decreasing the rate of gene expression.

Tissue-specific expression - Gene function and expression which is restricted to a particular tissue or cell type. For example, the glycoprotein hormone alpha subunit is produced only in certain cell types of the anterior pituitary and placenta, not in lungs or skin; thus expression of the glycoprotein hormone alpha-chain gene is said to be tissue-specific. Tissue specific expression is usually the result of an enhancer which is activated only in the proper cell type.

Transcriptional bursting - transcription (and also translation) occurs in "bursts" or "pulses", with periods of gene activity separated by irregular intervals.

Up regulation - increasing the rate of gene expression.

Up-regulated - describes a gene which has been observed to have higher expression (higher mRNA levels) in one sample compared to another (usually the control).

phi -- designate rotation around the alpha carbon and amino N

psi -- designate rotation around the alpha carbon and carboxyl C

HIV

The Human Immunodeficiency Virus (HIV) is an RNA virus that can infect specific immune cells in our body, called T helper cells. The RNA genome of HIV is encased in a capsid, which is in turn covered by an envelope derived from the host cell membrane. The structures and functions of most of HIV's proteins are now known. We are still learning about the accessory and regulatory proteins of HIV that exploits the host cell's machinery for its own advantage.

Life Cycle

- Attachment: The HIV spike or envelope protein, gp120, attaches to the host cell protein CD4 on specific types of T-cells.
- Fusion and entry: Binding of gp120 and CD4 rearranges their structures allowing the complex to bind another host cell receptor, the chemokine receptors, called CCR5. In some cases an alternate receptor called CXCR4 may replace CCR5 in this interaction. This in turn facilitates the stock of the HIV spike (the protein gp41) to penetrate the host cell membrane and fuse the viral envelope with the host cell membrane.
- Reverse transcription: Upon entry, HIV sheds its capsid and the 2 single strands of viral RNA are converted to a double stranded DNA by a special viral enzyme called Reverse transcriptase.
- Integration: The double stranded DNA, or proviral DNA, enters the host cell nucleus and is integrated in the cell's genome by another special viral enzyme called Integrase.
- Transcription and translation: The proviral DNA is transcribed and translated like any other host cell gene using host cell machinery (RNA polymerase, Ribosomes etc.)
- Assembly and budding: The various viral proteins and RNA come together to assemble the virus. At this stage some of the viral proteins are still linked to each other as part of the polyprotein synthesized by the virus. Various HIV proteins and RNA are packaged into an immature viral particle that buds off from the host cell encased in its membrane.
- Maturation of viral particle: With action of the viral protease the various HIV proteins are cut and separated, free to perform their specific functions. This rearrangement or maturation helps the HIV become a mature infectious particle ready to infect another cell. All the steps of the viral lifecycle are presented in the HHMI Biointeractives animation, narrated by HHMI investigator, Bruce Walker, MD.

The approaches currently used to treat HIV infections include:

1. Viral Enzyme inhibitors: block the actions of some critical enzymes in the HIV lifecycle.
 - Reverse transcriptase inhibitors (RTI): block initial conversion of viral RNA to proviral DNA that is integrated in the host cell genome
 - By mimicking the enzyme substrate and directly binding to the active site (nucleoside RTIs)
 - By binding to a site near the enzyme active site and blocking its function (non-nucleoside RTIs)
 - Integrase inhibitors: block integration of proviral DNA into the host cell genome preventing permanent infection of the host cells
 - Protease inhibitors: block cleavage of viral polyprotein, preventing maturation of HIV to infectious particles
2. Entry inhibitors: block interaction of the CD4-gp120 complex with the chemokine co-receptor preventing entry of HIV in the host cell
3. Fusion inhibitors: block the structural changes in the stock of the HIV spike (gp41) that are needed for the viral envelope and host cell membranes to fuse

Upcoming Approaches:

Making the host cells resistant to HIV:

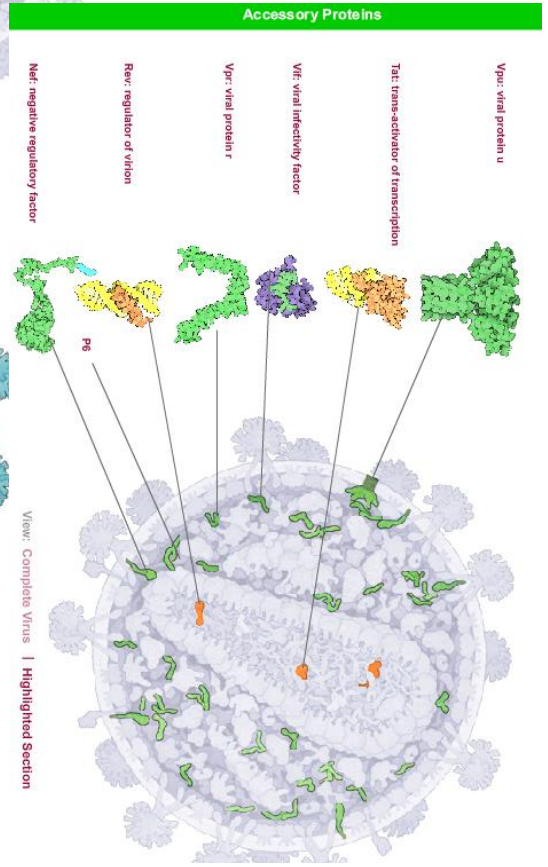
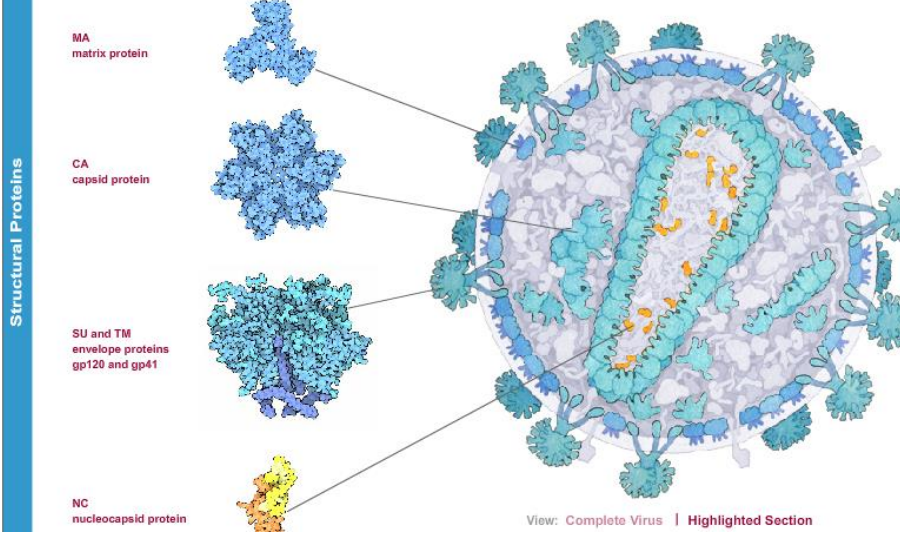
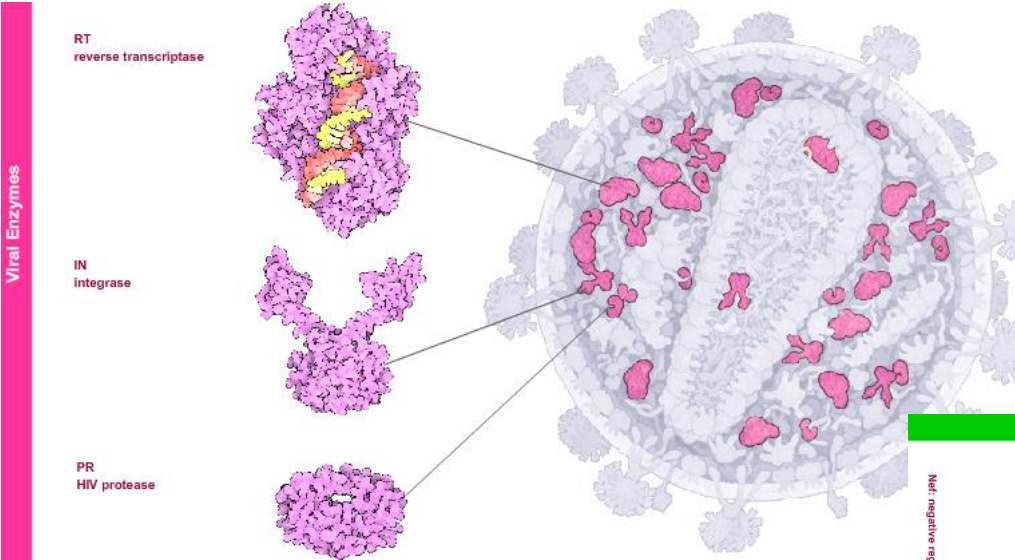
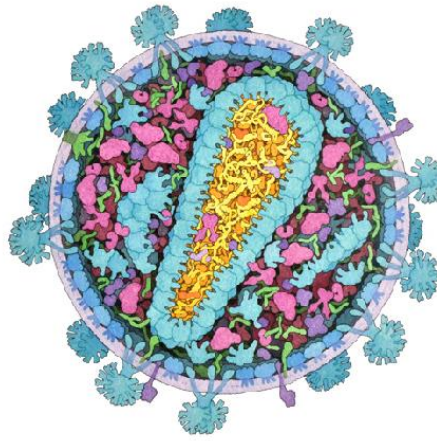
Currently researchers are using Zinc finger nucleases to target the CCR5 gene in stem cells that give rise to blood cells and introduce a deletion or disruption in the gene. As a result these cells are unable to make a functional CCR5 protein and become resistant to HIV infection. A treatment protocol using approach is currently in a Phase II clinical trial conducted by a group from the University of Pennsylvania School of Medicine, the Albert Einstein College of Medicine and Sangamo Biosciences (a biotech company specializing in the development of therapeutic zinc finger nucleases).

Seek out and destroy all the integrated proviral DNA:

A recent research report has suggested the possibility of using a gene therapeutic approach to specifically identifying and editing out the integrated proviral HIV-1 DNA. While there is a long way before this can even be tested as a treatment option it offers the hope that gene therapy can be used for dealing with tough diseases like HIV/AIDS.

HIV (human immunodeficiency virus) is composed of **two strands of RNA**, **15 types of viral proteins**, and **a few proteins from the last host cell it infected**, all surrounded by a **lipid bilayer membrane**. Together, these molecules allow the virus to infect cells of the immune system and force them to build new copies of the virus. Each molecule in the virus plays a role in this process, from the first steps of viral attachment to the final process of budding.

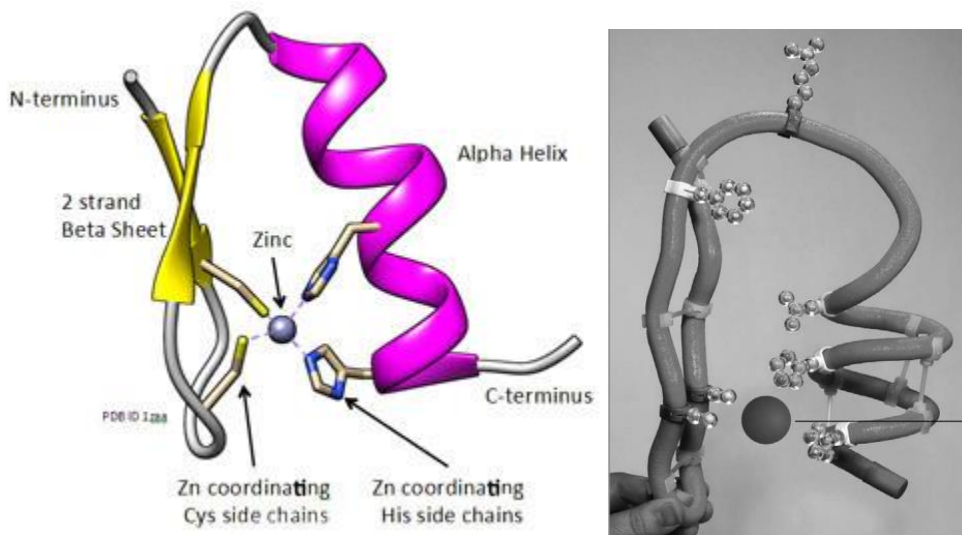
Since 1986, research on the structural biology of HIV have revealed the atomic details of these proteins. These structures are all publicly available in the Protein Data Bank (PDB) archive. Using these data, researchers have designed new treatments for HIV infection, including effective drug regimens that halt the growth of the virus. The structures also provide new hope for development of a vaccine.



Year	Major Event	Treatment Strategies
1981	CDC reports a rare form of pneumonia	
1982	CDC introduces the term Acquired Immune Deficiency Syndrome, or AIDS	
1983-84	HIV is established as the cause for AIDS	
1985	FDA approves the first HIV antibody test and blood banks begin screening for HIV	
1987	FDA approves Retrovir as the first drug to treat HIV	First reverse transcriptase inhibitor
1995	FDA approves saquinavir to treat HIV	First protease inhibitor
1996	Combination therapy for HIV treatment proposed	Combination therapy
1996	FDA approves Viramune (Nevirapine)	First Non-Nucleoside RTI
2002	OraQuick Rapid HIV test is approved, allowing HIV antibody testing in as little as 20 minutes using blood from a finger prick	
2003	FDA approves Fuzeon (enfuvirtide)	First HIV fusion inhibitor
2006	First one-pill-a-day HIV med available	HIV treatment with 1 pill a day
2007	FDA approves Isentress (raltegravir); FDA approves CCR5 blocker Selzentry (maraviroc) First integrase inhibitor	First integrase inhibitor and first entry inhibitor
2010	"The Berlin Patient", Tim Brown, a man living with HIV is classified as cured of his HIV	Transplant (in 2007) involving HIV-resistant (CCR5 delta-32) stem cells for the treatment of leukemia
2013	Phase I Clinical trial of CCR5-specific zinc finger protein nuclease (SB-728-T)*	Clinical trial for gene therapy to make HIV resistant cells*
2014	Report shows viral suppression may bring HIV transmission risk close to zero. Research report showing use of gene therapy to disrupt the latent HIV-1 provirus*	Decreasing transmission risk by reducing viral load. Research report showing use of gene therapy to disrupt the latent HIV-1 provirus*

Zinc-Finger Proteins

- The zinc finger protein has a tetra-coordinated zinc at the core of the structure to stabilize its structure.
 - Some scientists experimented with the idea of replacing the zinc coordination with other interactions. This exercise led to the design of a peptide that could adopt the same shape and structure as the DNA binding zinc finger domain but had a completely different rationale for its stability.
- Zinc Finger Nucleases are **sequence specific DNA binding proteins**.
 - each finger binds three bases
- Each finger is composed of a **short alpha helix and a 2-stranded beta sheet**.
- Zinc fingers were first identified in a frog transcription factor (transcription factor IIIA).
 - this protein structure was found to bind both 5S RNA and its cognate DNA.
 - Over the years zinc fingers have been identified in many other proteins and is one of the most common protein domains that binds to specific DNA/RNA sequences.
- Each zinc finger domain has ~30 amino acids.
- In addition to its **hydrophobic core**, it is **stabilized by a Zinc ion coordinated by side chains of four Cysteines, four Histidines or a combination of these**.
- Most zinc finger containing proteins have a **series of these domains linked to each other**.
 - These domains **bind to the major groove of the DNA**.
 - Specific amino acid side chains reach out from these domains to **"read" the DNA sequence by interacting with specific DNA bases**.



The two cysteine and two histidine side chains will be oriented to simultaneously bind to a single zinc atom in the center of the structure.

The two hydrophobic amino acid side chains phenylalanine and leucine will be orientated toward the inside of the structure.

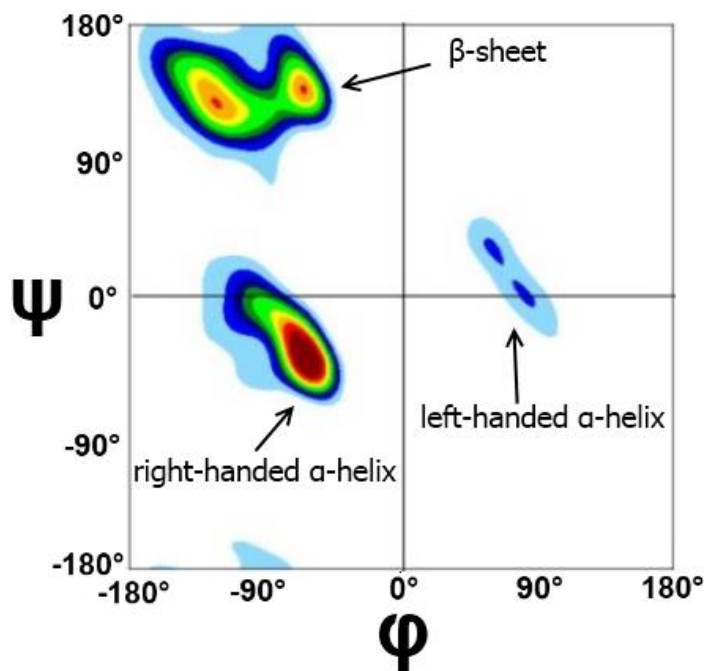
The positively-charged arginine side chain will be exposed at the top of the alpha helix, where it is available to bind to the negatively-charged phosphate backbone of DNA via an electrostatic interaction.

The nitrogen in the amino group and the oxygen in the carboxyl group share the hydrogen in these hydrogen bonds. DNA has a negatively-charged phosphate backbone.

Secondary Structures

Alpha Helices:

- Right Hand Alpha Helix
 - Rodlike and involves only one polypeptide chain
 - The helical conformation allows a linear arrangement of the atoms involved in the H bonds, which gives the bonds max strength and makes helical conformation very stable
 - Pitch of a helix – Linear distance between corresponding points on successive turns
 - Each amino acid corresponds to a 100 degree turn
 - 3.6 residues per turn
 - .15 nm along the helical axis
 - N-H group forms hydrogen bond with C=O group four residues earlier
 - can be described as a 3.6_{13} helix, since the $i + 4$ spacing adds 3 more atoms to the H-bonded loop compared to the tighter 3_{10} helix, and on average, 3.6 amino acids are involved in one ring of α helix
- Pi Helix
 - Hydrogen bonds every 5 residues
- Left Handed Alpha Helix
 - Dihedral angle, present in all atoms exactly 3 bonds apart, usually prevents left handed helices.
 - Rare exceptions include bacteria making short peptides with D amino-acids (tolaasin)
 - Because a D-amino acid is a mirror image of an L-amino acid, the steric hindrances that favor right-handed helices and prevent left-handed ones in L-amino acids instead favor left-handed helices and prevent right-handed ones in D-amino acids.
 - More common exception is found in glycine residues. They may adopt the (ϕ, ψ) angles corresponding to left-handed helices and often do for just a single residue or turn, but could in theory form a longer left-handed helix if enough glycines were found in a row.
 - This is observed to a lesser degree for asparagine which has an amide group that forms a hydrogen bond with the bulky carbonyl to help compensate for the unfavorable (ϕ, ψ) angles.
 - Occasionally other interactions in the three-dimensional structure can force an amino acid into an unfavored region of the Ramachandran plot, including the left-handed helix. It is rare to see amino acids adopt the (ϕ, ψ) angles of a left-handed helix for more than a couple sequential residues.



-
- 3_{10} helix
 - Hydrogen bonds every 3 residues
- 2_7 helix
 - 2 residues per turn and 7 atoms in a ring made by a hydrogen bond
- 4.4_{16} helix

- 4.4 residues per turn and 16 atoms in a ring
- Disrupting a helical structure
 - proline because it creates a bend in the backbone because of its cyclic structure
 - rotation around the bond between the N and the alpha carbon is severely restricted
 - proline's alpha amino group cannot participate in intra-chain H bonding
 - strong electrostatic repulsion owing to the proximity of several charged groups of the same sign, such as groups of (+) charged lysine and arginine residues or groups of (-) charge glutamate and aspartate
 - crowding (steric repulsion) caused by the proximity of several bulky side chains\
- In the alpha helical conformation all of the side chains lie outside the helix. There is not enough room for them in the interior. The alpha carbon located just outside the helix, so crowding occurs if it is bonded to two atoms other than H.
- Unlike the alpha helix, the peptide backbone in the beta sheet is almost completely extended
- H bonds in beta pleated sheets can be formed between different parts of a single chain that is doubled back on itself - intrachain bonds or between different chains - interchain bonds.
- A parallel pleated sheet is formed if the peptide chains run in the same direction - if they are all aligned in terms of their N-terminal and C-terminal ends
- An anti-parallel pleated sheet formed when alternating chains run in opposite directions
- The H bonds in the beta pleated sheet are perpendicular, instead of parallel in regards to the direction of the protein chain, when compared to the alpha helix.
 - H bonding between peptide chains gives rise to a repeated zigzag structure
- A common non-repetitive irregularity found in anti-parallel beta sheets is the beta-bulge. It normally occurs between two normal beta-structure hydrogen bonds and involves two residues on one strand and one on the other

Motifs

- a repetitive supersecondary structure
- smaller motifs can be repeated and organized into larger motifs
- they tell us much about the folding of proteins
- The combo of alpha and beta strands produce various kinds of supersecondary structures in proteins
 - the simplest protein structural motif is a helical bundle
 - The most common supersecondary structure is the beta-alpha-beta unit -- two parallel strands of beta sheet are connected by a stretch of alpha helix
 - Another is the alpha-alpha unit (helix-turn-helix) -- consists of two anti-parallel alpha helices and in such an arrangement, energetically favorable contacts exist between the side chains in the two stretches of helix
 - beta-meander -- an antiparallel sheet is formed by a series of tight reverse turns connecting stretches of the polypeptide chain
 - greek key -- another kind of antiparallel sheet is formed when the polypeptide chain doubles back on itself
 - In the fold known as the **TIM barrel fold** (the name is based on the first protein where it was found, **T**riose phosphate **I**so**M**erose), probably one of the most widespread type of protein folds, the strands of the β -sheet are also parallel, but the connectivity between them is made by α -helices

Tertiary Structure

- 3D arrangement of all the atoms in the molecule: the conformations of the side chains and the positions of any prosthetic groups and arrangement of helical and pleated-sheet sections

Five different ways to denature proteins

- heat
 - an increase in temperature favors vibrations within the molecule, and the energy of these vibrations can become great enough to disrupt the tertiary structure
- extremes of pH
 - at least some of the charges on the protein are missing and so the electrostatic interactions that would normally stabilize the native, active form of the protein are drastically reduced
- detergents
 - detergents such as sodium dodecyl sulfate (SDS) disrupt hydrophobic interactions and if a detergent is charged, it can also disrupt electrostatic interactions w/in the protein
- urea and guanidine hydrochloride

- form H bonds with the protein that are stronger than those w/in the protein itself. they can also disrupt hydrophobic interactions in much the same way as detergents
 - beta-Mercaptoethanol
 - used to reduce disulfide bridges to two sulfhydryl groups; to facilitate unfolding of the protein and to increase the accessibility of the disulfides to the reducing agent

Quaternary Structure

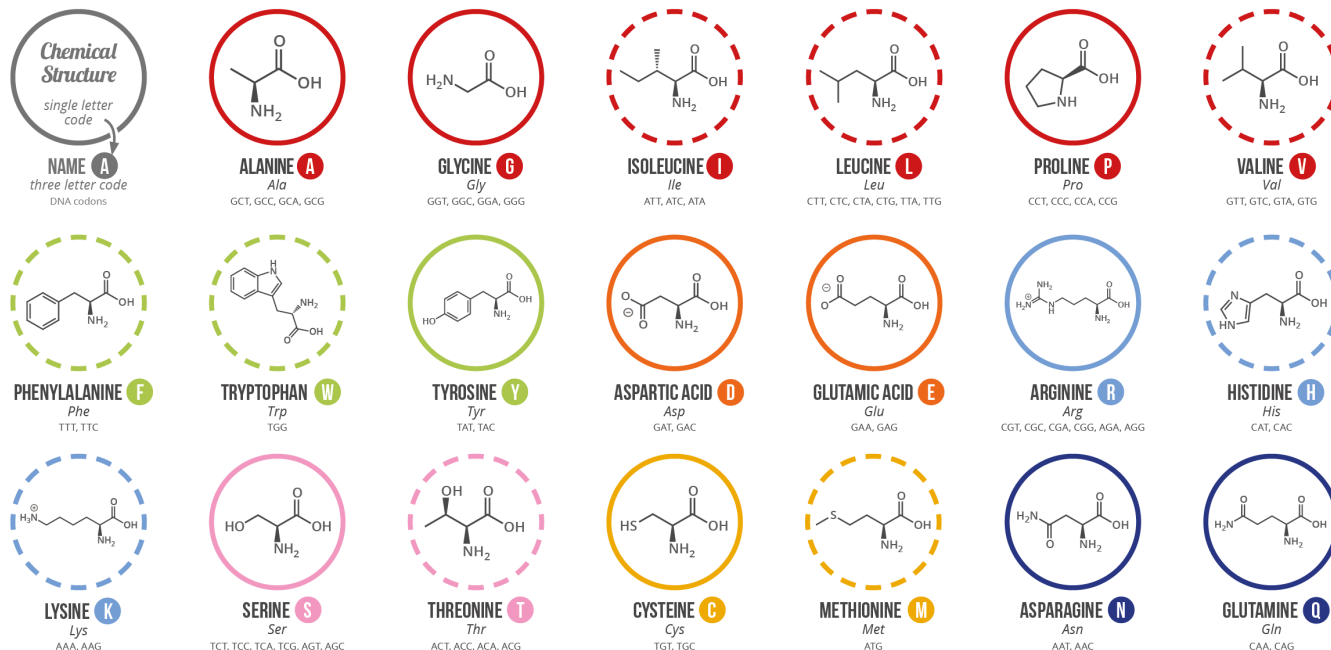
- those that consist of more than one polypeptide chain
- Each chain is called a subunit.
- Proteins with more than one chain
 - Dimers
 - Trimers
 - Tetramers
- Oligomer
 - Generic term for a structure, made up of a small number of subunits

Amino Acids

A GUIDE TO THE TWENTY COMMON AMINO ACIDS

AMINO ACIDS ARE THE BUILDING BLOCKS OF PROTEINS IN LIVING ORGANISMS. THERE ARE OVER 500 AMINO ACIDS FOUND IN NATURE - HOWEVER, THE HUMAN GENETIC CODE ONLY DIRECTLY ENCODES 20. 'ESSENTIAL' AMINO ACIDS MUST BE OBTAINED FROM THE DIET, WHILST NON-ESSENTIAL AMINO ACIDS CAN BE SYNTHESISED IN THE BODY.

Chart Key: ● ALIPHATIC ● AROMATIC ● ACIDIC ● BASIC ● HYDROXYLIC ● SULFUR-CONTAINING ● AMIDIC ○ NON-ESSENTIAL ○ ESSENTIAL



Note: This chart only shows those amino acids for which the human genetic code directly codes for. Selenocysteine is often referred to as the 21st amino acid, but is encoded in a special manner. In some cases, distinguishing between asparagine/aspartic acid and glutamine/glutamic acid is difficult. In these cases, the codes asx (B) and glx (Z) are respectively used.

© COMPOUND INTEREST 2014 - WWW.COMPOUNDCHEM.COM | Twitter: @compoundchem | Facebook: www.facebook.com/compoundchem
Shared under a Creative Commons Attribution-NonCommercial-NoDerivatives licence.

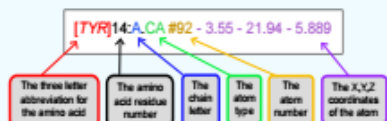


There are over 500 amino acids found in nature, yet, of these, the human genetic code only directly codes for 20. There has been some debate over whether or not another amino acid should be classified as the 21st. Selenocysteine is an amino acid which is found in a small number of human proteins; unlike the 20 pictured here, however, it is not coded for directly, but in a special manner. Yet another, pyrrolysine, is coded for in a similar manner, and considered the 22nd amino acid.

Jmol Commands

Mouse Movements

Clicking on an atom provides information in the console window. This information is explained in detail below.



Rotate on the X-Y axes:



Zoom in and out:



Translate the Molecule:



Rotate on the Z axis:



Display Formats

wireframe (displays stick bonds)

wireframe <value> (displays stick bonds with specific thickness)

example: `wireframe 1.0`

spacefill (displays atoms as spheres with atom radii equal to their Van der Waals radius)

example: `spacefill`

spacefill <value> (displays atoms as spheres with specific radius)

example: `spacefill 1.25`

backbone (displays alpha carbon backbone)

backbone <value> (displays backbone with specific thickness)

example: `backbone 1.5`

Color Formats

Method 1: `select <selection type>`

`color <color name>`

example: `select hydrophobic`
`color yellow`

Method 2: `color <selection type>`

`color <code[R,G,B]>`

example: `select helix`
`color [15,255,110]`

Default color mode: `color CPK`

Color secondary structures: `color structure`

For a full list of the predefined colors available in Jmol, visit: <http://jmol.sourceforge.net/jscolor/>

Bonds and Struts

Hydrogen Bonds:

`calculate hbonds` (adds hydrogen bonds to all selected areas)

`hbonds off` (removes all hydrogen bonds in a selected area)

`hbonds <number>` (displays hydrogen bonds with specific thickness)

`color hbonds <color>` (colors hydrogen bonds)

`set hbonds solid` (displays hydrogen bonds as solid lines)

`set hbonds backbone` (connects hydrogen bonds to the alpha carbon)

`set hbonds sidechain` (connects hydrogen bonds to the nitrogen and oxygen atoms)

To add or remove a single hbond, select only the two amino acids that that the hbond connects and use the `hbonds 1.0` or `hbonds off` command

example: `select 716 or 1341` example: `select 14 or 342`
`hbonds 1.0` `hbonds off`

Disulfide Bonds:

`ssbonds on` (adds disulfide bonds to all selected areas)

`ssbonds off` (removes disulfide bonds)

`ssbonds <number>` (displays with specific thickness)

`color ssbonds <color>` (colors disulfide bonds)

`set ssbonds backbone` (connects disulfide bonds to the alpha carbon)

`set ssbonds sidechain` (connects disulfide bonds to the nitrogen and oxygen atoms)

To add or remove a single ssbond, select only the two amino acids that that the ssbond connects and use the `ssbonds 1.0` or `ssbonds off` command

example: `select 716 or 1341` example: `select 14 or 342`
`ssbonds 1.0` `ssbonds off`

Struts:

`calculate struts` (adds structural supports called struts to all selected protein areas)

`struts off` (removes struts)

`struts <number>` (displays with specific thickness)

`color struts <color>` (colors struts)

To add or remove a single strut, select only the two atoms that that the strut connects and use the `strut` or `strut off` command

example: `select atomno=716 or atomno=1341` example: `select atomno=14 or atomno=342`
`connect strut` `connect strut delete`
`strut 1.0`

Selection and Restriction

`select <selection type>` (selects part of the file)

example: `select helix`

`restrict <selection type>` (removes the display of everything except what was restricted)

example: `restrict water`

List of Common Selection Types:

<code>backbone</code>	<code>sidechain</code>	<code>hydrophobic</code>
<code>hydrophilic</code>	<code>charged</code>	<code>hetero</code>
<code>water</code>	<code>nucleic</code>	<code>protein</code>
<code>helix</code>	<code>sheet</code>	

`*<letter>` (for selecting by chain letter)

`<number>` (for selecting by residue number)

`<number>-<number>` (for selecting by residue range)

`atomno=<number>` (for selecting by atom number)

`atomno=><number>` and `atomno=<=<number>`
(for selecting by atom range)

`<atom type>` (for selecting by atom type)

Standard Sizes for SMART Team Models

<code>backbone 1.5</code>	<code>hbond 1.0</code>
<code>wireframe 1.0</code>	<code>strut 1.0</code>
<code>spacefill 1.25</code>	<code>ssbond 1.0</code>

Adding a "Clean" Sidechain:

To select and display only the atoms of the sidechain of a specific amino acid, you want to use the `select` command followed by the amino acid name/number and end with the `and (sidechain or alpha)` text.

```
select cys30 and (sidechain or alpha)
spacefill 1.25
wireframe 1.0
```

To remove an incorrectly displayed sidechain:

```
select cys30
spacefill off
wireframe off
```