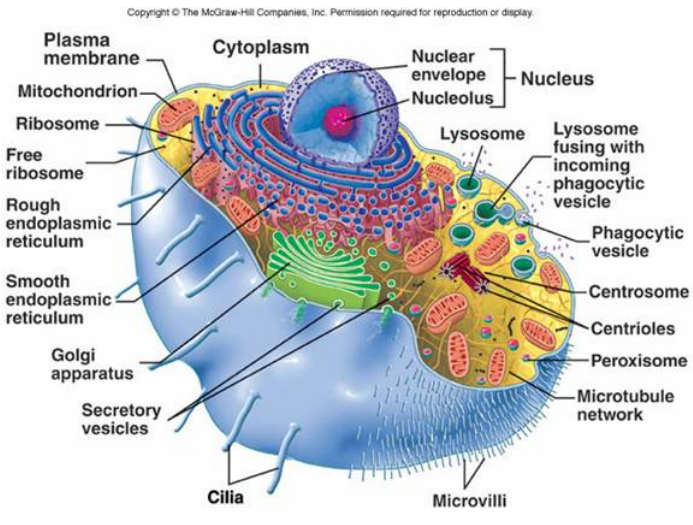
1 Central Dogma of Biology

1 Definition (The Cell)



2 Definition (The Cell)

- single cell organisms
- multi cell organisms

3 Definition (The Cell)

- prokaryotes
- eukaryotes

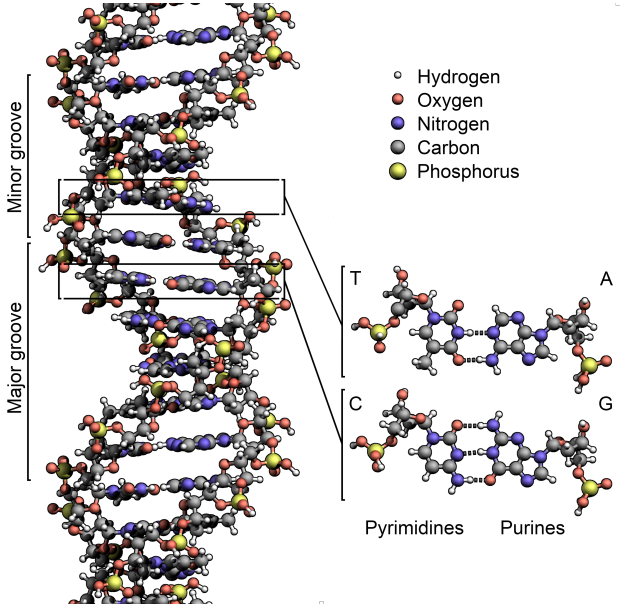
4 Definition (Nucleic Acids)

- DNA: A T G C
- RNA: A U G C

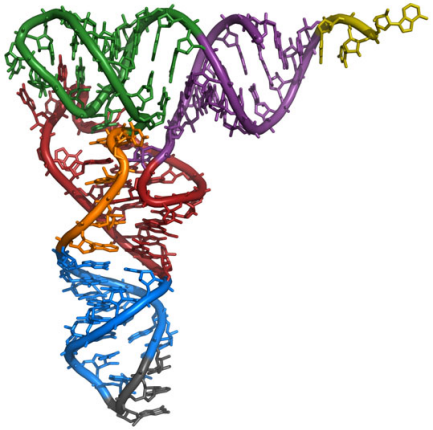
5 Definition (Amino Acids)

Ala	Arg	Asn	Asp	Cys	Glu	Gln	Gly	His	Ile
A	R	N	D	C	E	Q	G	H	I
Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
L	K	M	F	P	S	T	W	Y	V

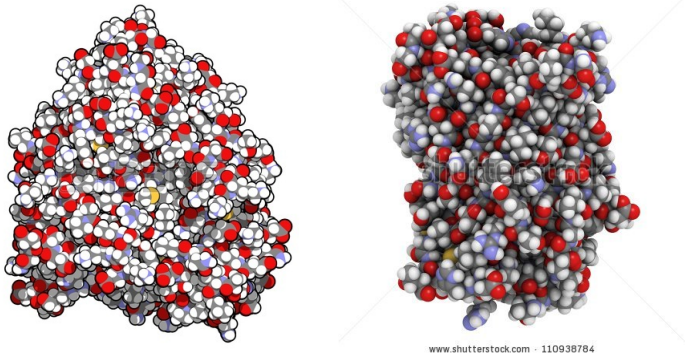
6 Definition (DNA Molecule)



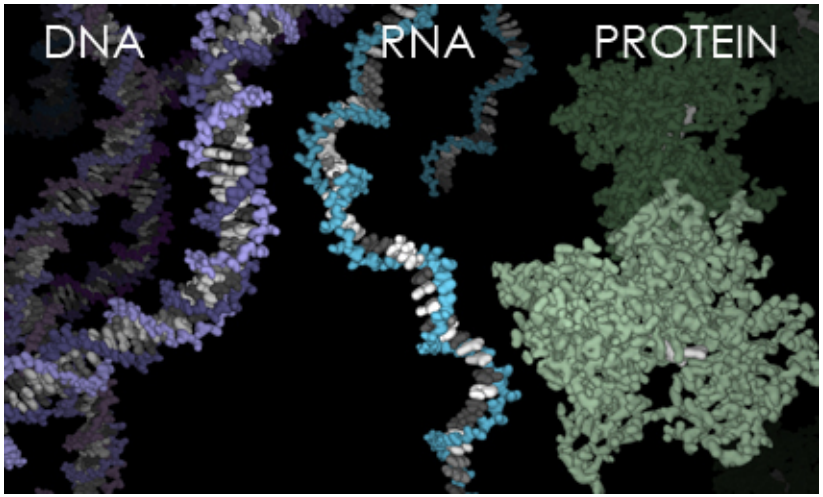
7 Definition (tRNA Molecule)



8 Definition (Protein Molecules)



9 Definition (DNA vs RNA vs Protein Molecule)



1 Lesson (Molecules of Life)
Discuss how the four types of molecules, DNA, mRNA, tRNA and proteins are:

(a) similar.

(b) different.

2 Lesson (Genetic Code)
How can the 20 different amino acides of protein sequences be coded using only the four different nucleic acids of DNA?

10 Definition (Genetic Code)

Genetic Code

First Position	Second Position T		C		A		G		Third Position
T	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys	T
	TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys	C
	TTA	Leu	TCA	Ser	TAA	Stop	TGA	Stop	A
	TTG	Leu	TCG	Ser	TAG	Stop	TGG	Trp	G
C	CTT	Leu	CCT	Pro	CAT	His	CGT	Arg	T
	CTC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser	T
	ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	ATG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly	T
	GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

11 Definition (Genes)
Genes are segments of DNA that are transcribed and translated into a protein sequence. Splicing may be required.

- codon: group of three nucleic acids that code for a single amino acid.
- introns: portions of a gene that are removed before translation.
- exons: portions of a gene that are spliced before translation.

3 Lesson (cDNA)
What is cDNA?

4 Lesson (DNA Transcription and Translation)
The following cDNA sequence contains the beginning of the gene that codes for the human insulin protein.

```
TGCCTGTCTCCCAGATCACT
GTCCTTCTGCCATGGCCCTG
TGGATGCGCCTCCTGCCCC
```

(a) Determine the sequence of cDNA that codes for the first nine letters of the insulin protein sequence.

Solution:

(b) Determine the sequence of mRNA that codes for the first nine letters of the insulin protein sequence.

Solution:

(c) Determine the first nine letters of the insulin protein sequence.

Solution:

(d) Use the Uniprot database to check your answer to part (c) by looking up the protein sequence for the human insulin protein.

Solution:

5 Lesson (Transcription and Translation)
Use the link below to look at an animation of the transcription and translation of DNA and discuss what you see.

`http://www.dnalc.org/resources/3d/central-dogma.htmli`

Because it takes three letters of a DNA sequence to translate to a single letter of a protein sequence, where we start the translation in the DNA sequence has a big effect on the resulting protein sequence. In other words, we have to choose the correct **reading frame** before we attempt to translate a DNA sequence into a protein sequence. Reading frames which start with the start codon `ATG` are called **open reading frames.**

6 Lesson (Jemboss (Translation))
Install the free bioinformatics software package `Jemboss` and use it to check your answers to Lesson 4 part (c). Use the commands

`NUCLEIC, TRANSLATION, transeq`

7 Lesson (Jemboss (Open Reading Frames))

(a) Determine three of the six possible translations of the following segment of DNA.

T A T A G G G A C T C A

(b) Check your answer with `Jemboss`. Use the commands `NUCLEIC`, `TRANSLATION`, `sixpack`

Solution:

8 Lesson (Jemboss — Open Reading Frame)
Use Jemboss to translate the human insulin gene DNA sequence into the protein sequence for insulin using:

(a) the GenBank entry J00265.1. (Hint: Use Google to search for GenBank and select `Nucleotide` and search for J00265.1).

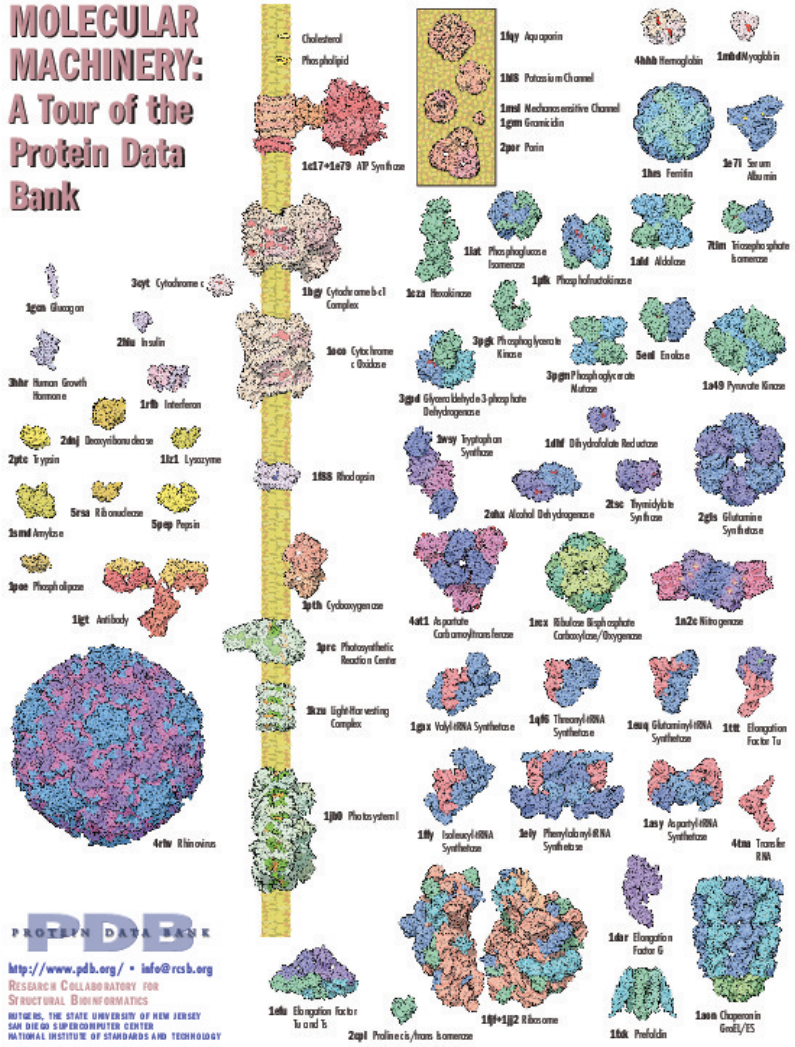
(b) the file `insulin_human_cDNA.txt` which contains the sequence for the cDNA for human insulin.

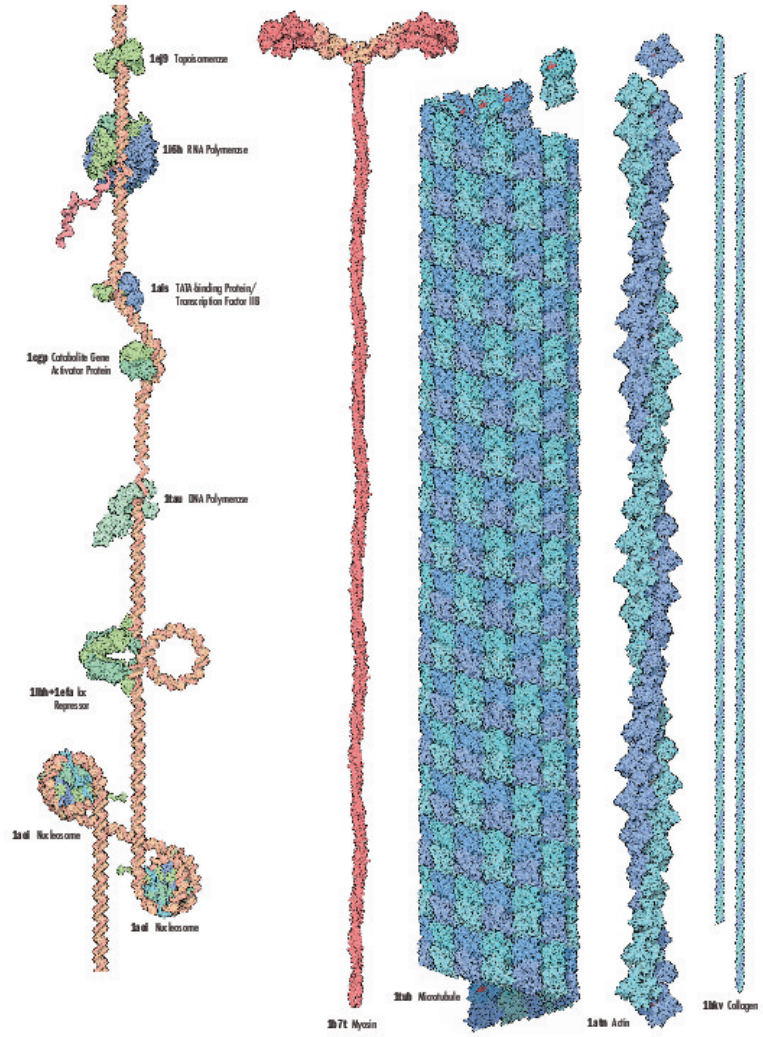
(c) Compare the protein sequence from part (a) and part (b). The first 62 letters of the sequences should be the same. Why are the rest different?

(d) The section `CDS` in the GenBank insulin entry J00265.1 describes special aspects of the insulin gene, e.g. where the exons are. Use this section to determine the length of the first protein segment corresponding to the first exon of the insulin gene. Does this length agree with part (c)?

12

2 Proteins





13

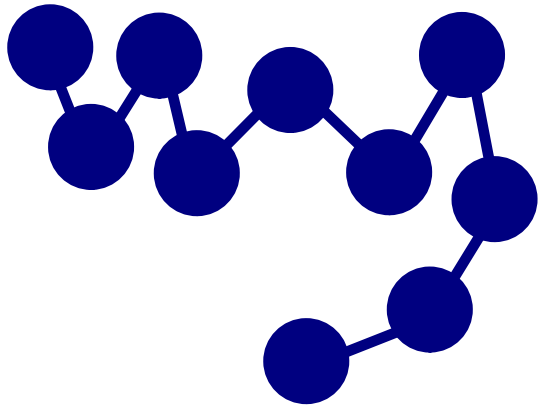
---

14 Definition (Protein Structure)

- primary structure
- secondary structure
- tertiary structure
- quaternary structure

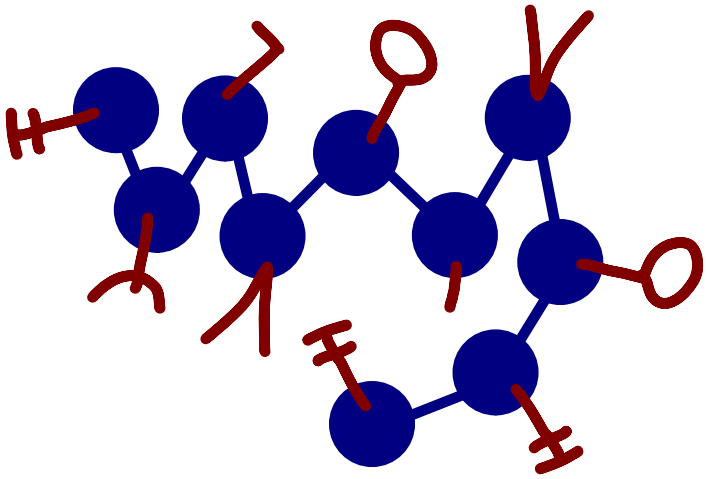
---

15 Definition (Primary Structure—Backbone)



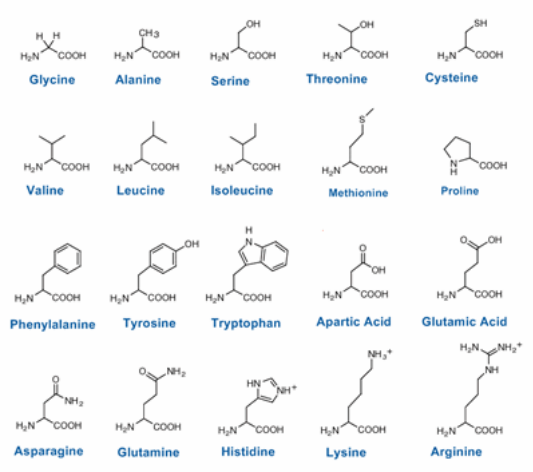
---

16 Definition (Primary Structure—Sidechains)

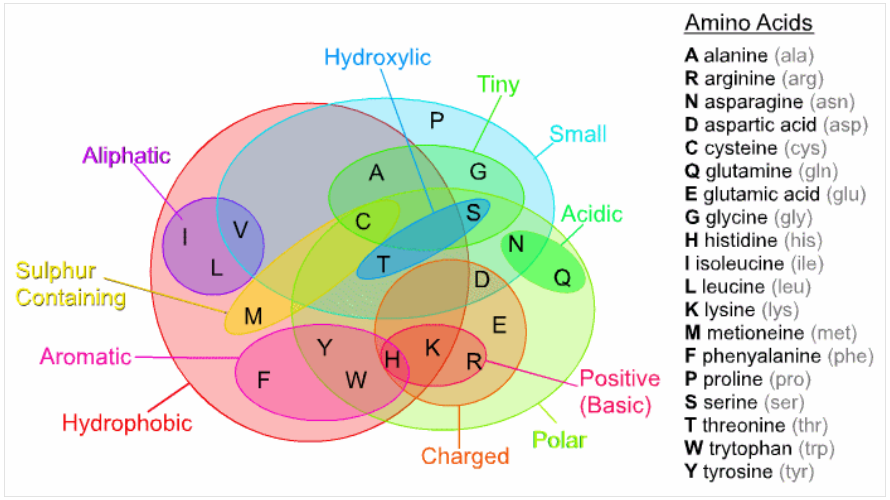


---

17 Example (Amino Acid Structures)



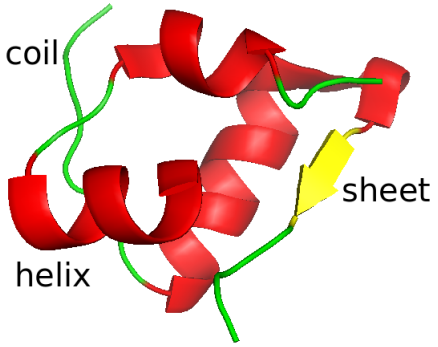
18 Example (Amino Acid Properties)



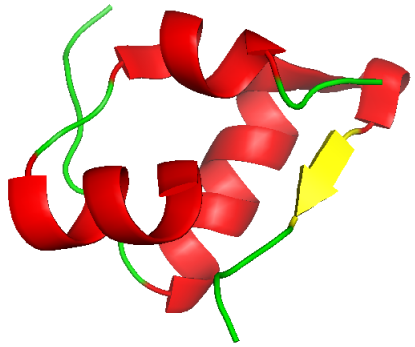
19 Example (Insulin Primary Sequence)

```
MALWMRLLPL LALLALWGPD PAAAFVNQHL CGSHLVEALY LVCGERGFFY    50
TPKTRREAED LQVGQVELGG GPGAGSLQPL ALEGSLQKRG IVEQCCTSIC   100
SLYQLENYCN                                               110
```

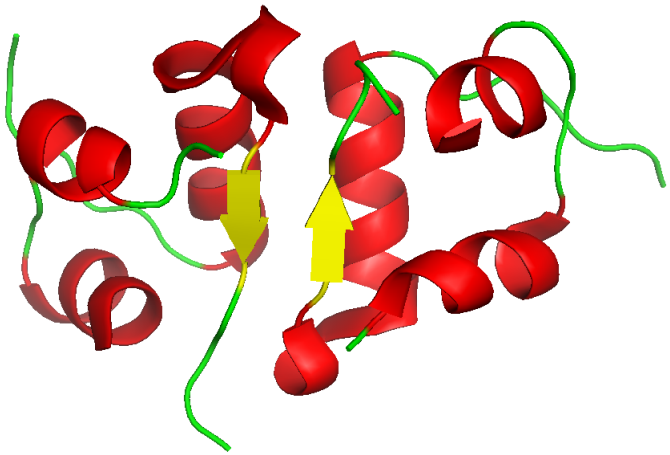
20 Example (Insulin Secondary Structure)



21 Example (Insulin Tertiary Structure)



22 Example (Insulin Quaternary Structure)



23 Example (Proteins in Ebola)