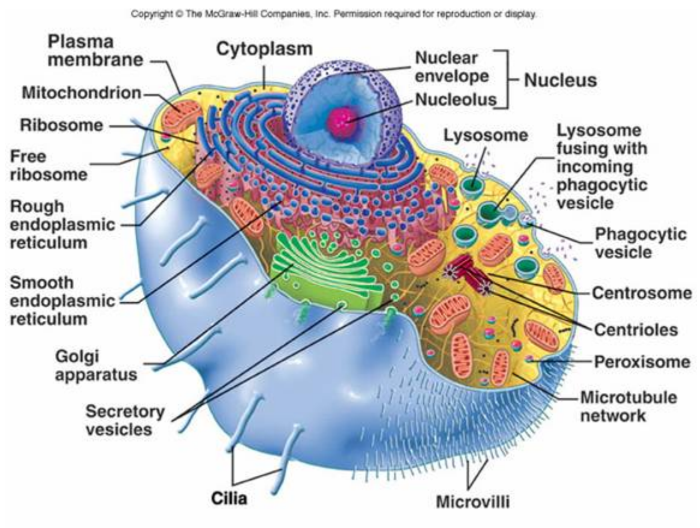
1 Central Dogma of Biology

1 Definition (The Cell)



2 Definition (The Cell)

- single cell organisms
- multi cell organisms

3 Definition (The Cell)

- prokaryotes
- eukaryotes

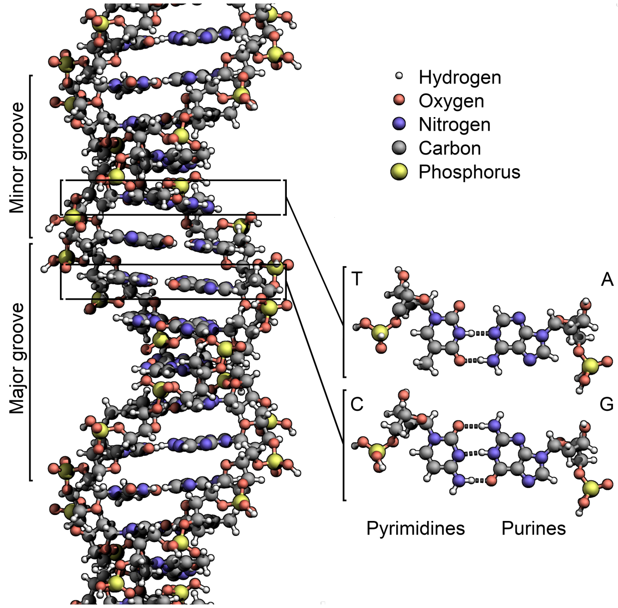
4 Definition (Nucleic Acids)

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

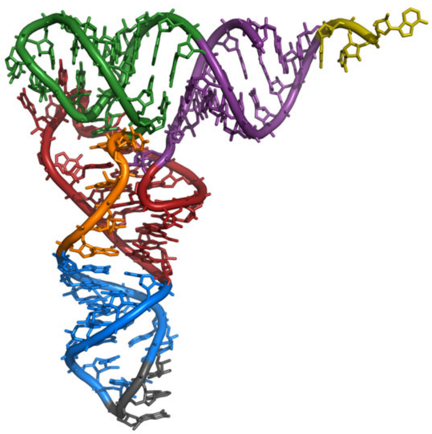
5 Definition (Amino Acids)

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

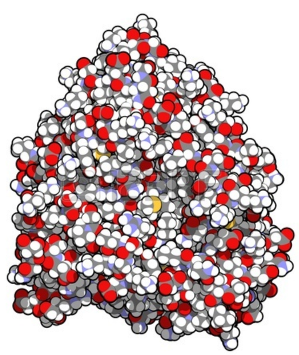
6 Definition (DNA Molecule)

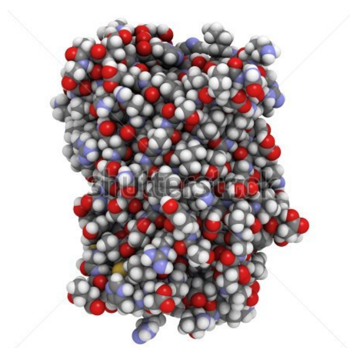


7 Definition (tRNA Molecule)

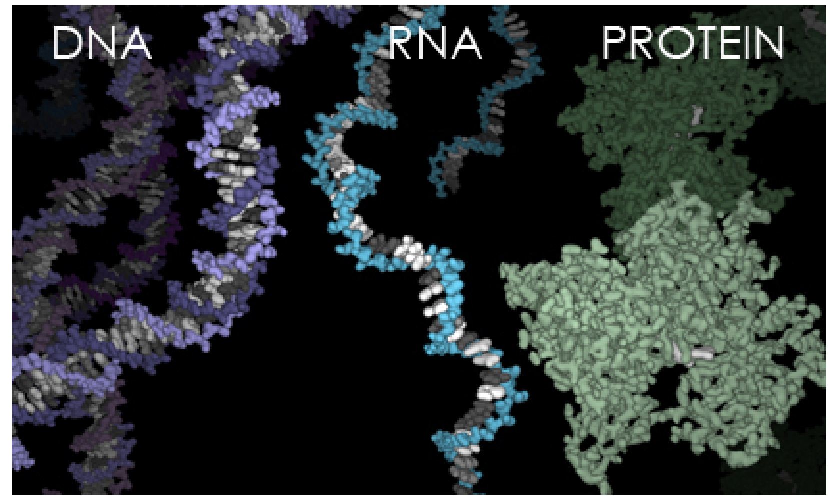


8 Definition (Protein Molecules)





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	T		C		A		G		Third Position
				Second 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).

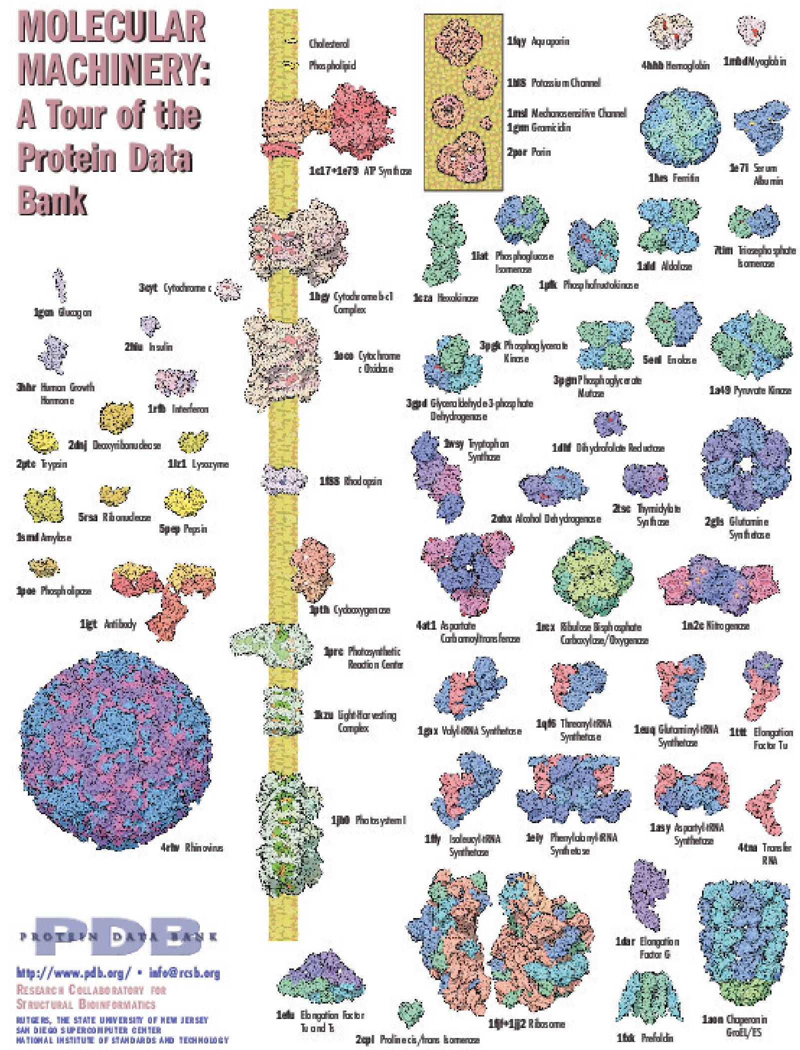
(b) the file `insulin_human_cDNA.txt` which contains the sequence for the cDNA for human insulin.

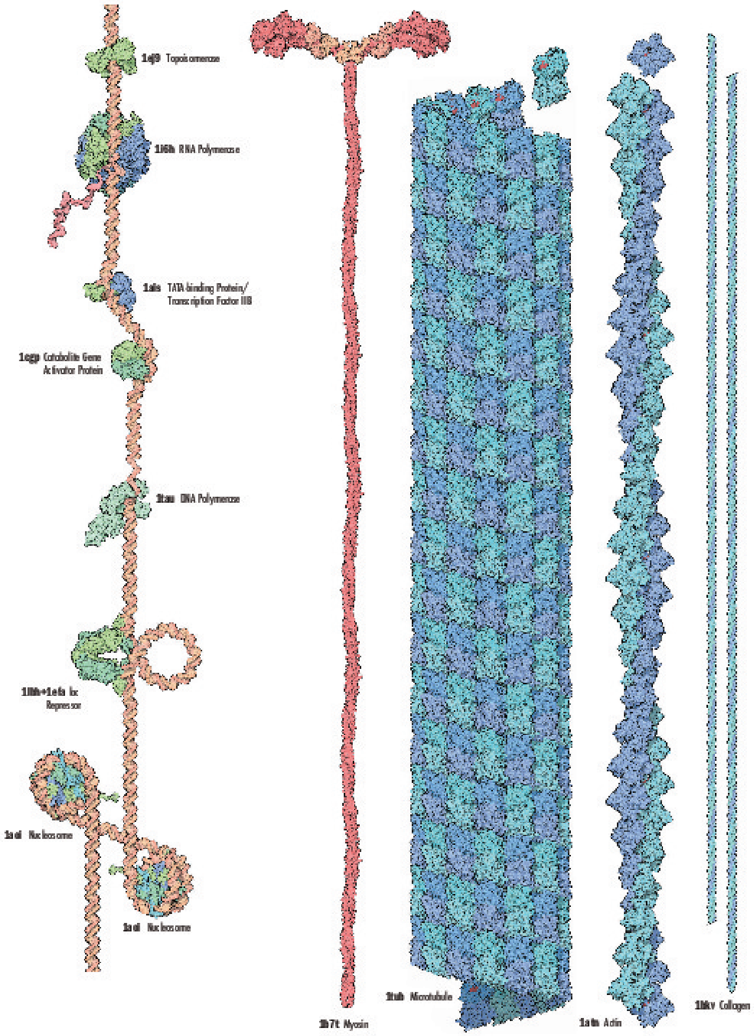
(c) Compare the protein sequence from part (a) and part (b). The first 62 letters of the sequences should be the same. Why are the rest different?

(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)?

---

2 Proteins





13

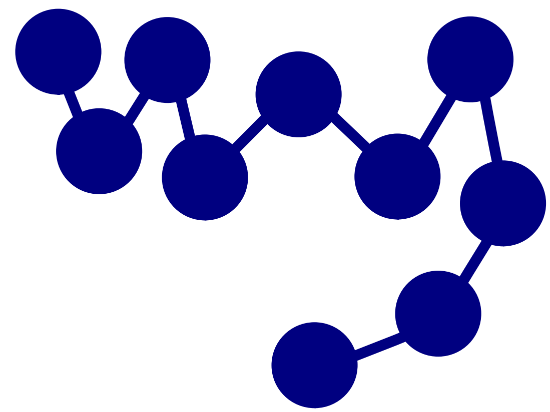
---

14 Definition (Protein Structure)

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

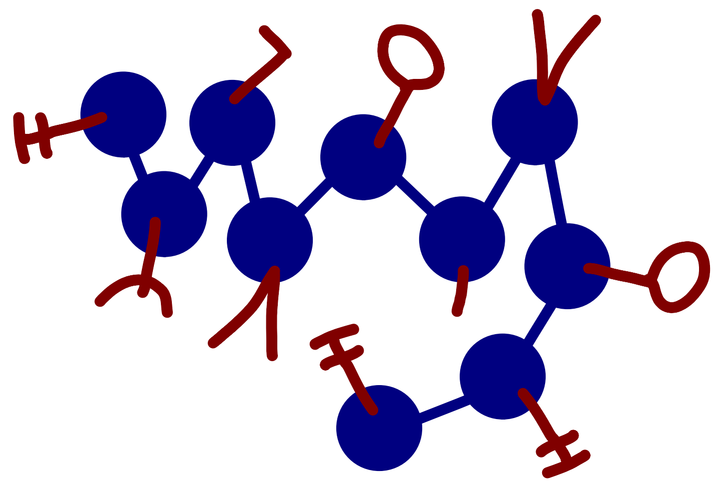
---

15 Definition (Primary Structure—Backbone)



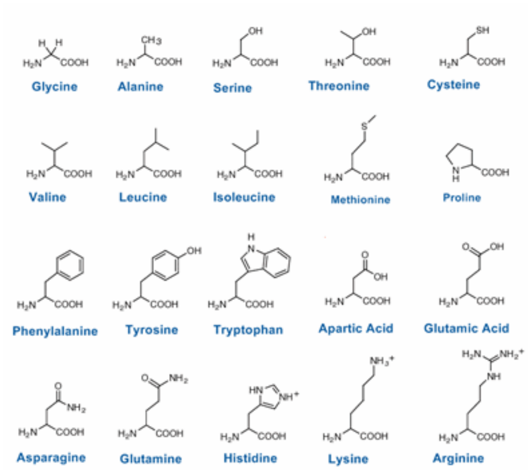
---

16 Definition (Primary Structure—Sidechains)

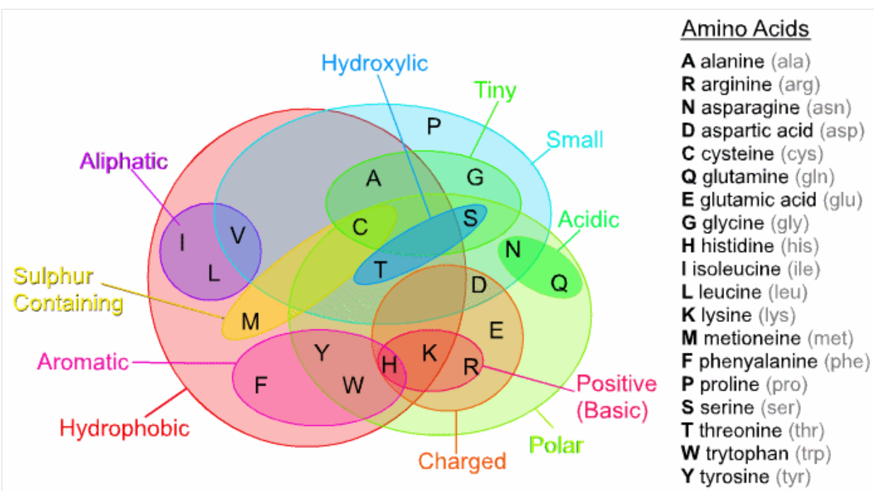


---

17 Example (Amino Acid Structures)



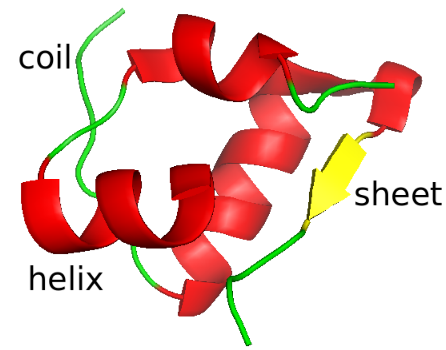
18 Example (Amino Acid Properties)



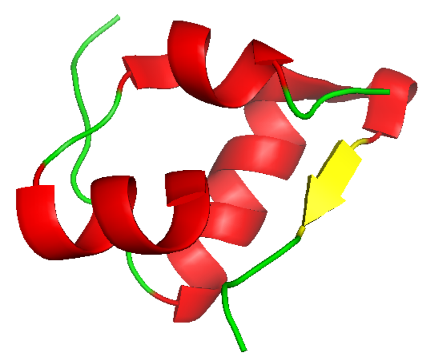
19 Example (Insulin Primary Sequence)

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

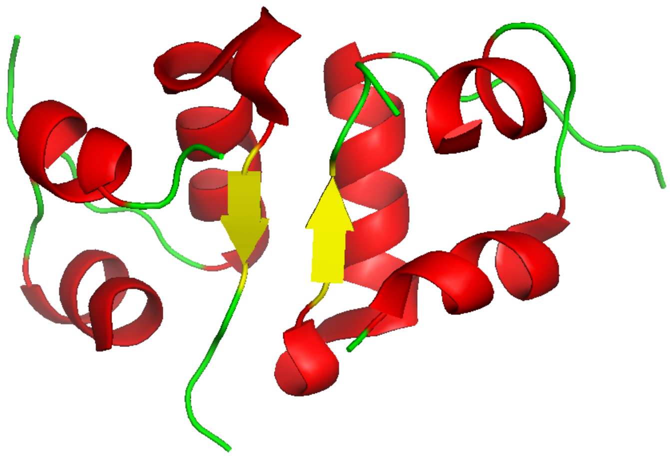
20 Example (Insulin Secondary Structure)



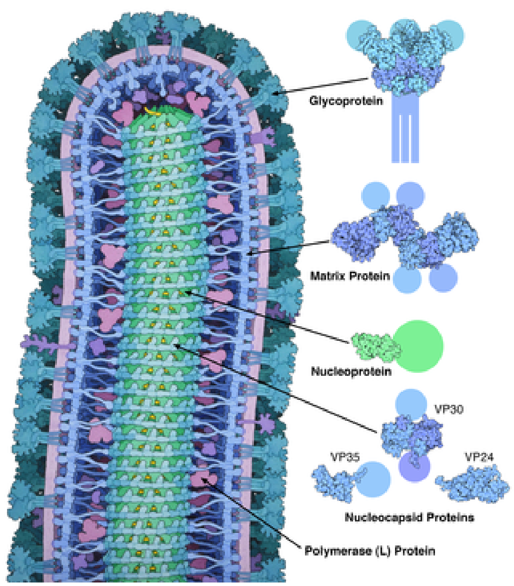
21 Example (Insulin Tertiary Structure)



22 Example (Insulin Quaternary Structure)



23 Example (Proteins in Ebola)



3 Biopython

Install the Anaconda Python package.

Startup the IPython notebook. If you are using linux, open a terminal and type

```
ipython notebook -pylab inline
```

If you are using Windows enter the command `%pylab inline` at the beginning of your notebook.

Use the following commands to plot a sine curve. To create 100 points equally spaced between 0 and 2π type

```
t = linspace(0,2*pi,100)
print t
```

Plot the sine curve

```
plot(sin(t))
```

The ipython notebook combines the best features of programs like Matlab and Maple.

24 Example (IPython)
Download the `insulin_cDNA.txt` file. Explain what each of the following commands do:

```
handle = open('insulin_cDNA.txt')
handle?
handle.  (press tab key)
seq = handle.readline().strip()
print seq
len(seq)
seq[0]
seq[1]
seq[-1]
seq2 = seq.strip()
seq2[-1]
seq2. (press tab key)
seq.find?
seq.find('ATG')
seq[31:34]
seq.count('ATG')
```

Python uses object oriented programming.

25 Example (Object Oriented Programming)
For example, consider the python command

```
marker = MARKER(color = blue)
```

The function `MARKER(color=blue)` is a factory function which manufactures objects, in this case markers. The `color=blue` specifies that a blue marker should be manufactured.

Objects have attributes. For example, the color attribute `marker.color` should equal blue.

Objects also have methods. For example, the marker method `marker.change_color(red)` changes the color attribute of the marker from blue to red.

In IPython typing `marker.` followed by the tab key will list all the attributes and methods associated with the marker object. Typing `marker?` will provide information about the marker object.

26 Example (Biopython)
Explain what the following Biopython commands do:

```
from Bio.Seq import Seq
from Bio.Alphabet import IUPAC
handle = open('insulin_cDNA.txt')
seq = handle.readline().strip()
print seq
DNA = Seq(seq, IUPAC.unambiguous_dna)
DNA. (press tab key)
mRNA = DNA.transcribe()
print mRNA
protein = DNA.translate()
print protein
```

9 Lesson (DNA Transcription and Translation II)
Download the files:

`Human_insulin_gene_complete.txt`

`insulin_partial_genome_sequence.txt`

Translating the DNA in the first file to the insulin protein sequence is a much more challenging problem than the translation problem assigned for homework 1. Rather than asking you to translate the first DNA sequence, we have provided a segment of DNA from the first file that is easier to translate.

The DNA sequence contained in the file `insulin_partial_genome_sequence.txt` contains a code for the human insulin protein. For the following questions, cut-and-paste your answers into a text file.

(a) Determine the portion of the sequence of DNA that codes for the insulin protein sequence.

(b) Determine the letters of the insulin protein sequence.

(c) Does the amino acid sequence you obtain match the human insulin sequence in Uniprot? (Hint: Don't simply look at the first few amino acids.)

4 Sequence Alignment

DNA is subject to mutations. We will only consider insertions, deletions and substitutions.

27 Definition (Mutations)

```
orignal sequence   ATTGCTCC

orignal sequence   ATTG_CTCC
       insertion   ATTGGCTCC

orignal sequence   ATTGCTCC
        deletion   ATT_CTCC

orignal sequence   ATTGCTCC
    substitution   ATTTCTCC
```

28 Example (Sequence Alignment)
Consider the sequences:

```
TAGTA
ATAT
```

Before we can determine how similar the sequences are to each other, we must first align the sequences. Two optimal alignments obtained using *dynamic programming* are:

```
TAGTA   _TAGTA
_A_TAT  ATA_T_
```

29 Example (Dot Plots)
Use a dot plot to compare the following sequences:

```
TAGTA
ATAT
```

	T	A	G	T	A
A		○			○
T	○			○	
A		○			○
T	○			○	

10 Lesson (Dot Plots)
How similar are human, horse and chicken insulin? Use Jemboss to create dot plots comparing the insulin sequence for each.

- Go to www.uniprot.org.
- In the search field click on advanced.
- Select Gene name [GN] and type INS (for the insulin gene).
- Scroll down the results and click on the check box in the left column for human, horse and chicken insulin.
- Select download and a new window will appear containing the insulin sequences for human, horse and chicken in `fasta` format.
- Open Jemboss.
- Select ALIGNMENT, Dot Plots, polyplots.
- Cut and paste the `fasta` sequence data into Jemboss.
- Select pdf format for the output.
- Go to the Jemboss folder to retrieve the results.
- Interpret the plots.