Gene Expression and Protein Synthesis

Objective:

To construct using three dimensional models or diagrams the structure and role of a “gene” in protein synthesis, including the concept of gene expression, promoter sequence, exons, and introns, at the level of 85% proficiency for each student.

In order to achieve this objective, you will need to be able to:

1. Examine and identify the structure of nucleotides, nucleic acids, RNA and DNA.
2. Define coding DNA *vs* template DNA, pre-messenger RNA, messenger-RNA, transfer-RNA, ribosomal-RNA, codon and anti-codon.
3. Define “gene expression.”
4. Describe the role of DNA and RNA in the process of protein synthesis.
5. Explain the significance of the organization of a "gene", including the regions for the Promoter, the Exons and the Introns.

Materials:

Group Supplies

Computer with JAVA and JMol installed.
https://java.com/en/download/index.jsp
https://sourceforge.net/projects/jmol/files/Jmol/
(These have been pre-installed onto the computers in the physiology lab. Jmol version 14.29.17 or newer is recommended)

Structures of DNA and Nucleic Acids were obtained from
https://www.rcsb.org/
https://cactus.nci.nih.gov
https://pubchem.ncbi.nlm.nih.gov

The nucleotide sequence of the human “AVP” gene was obtained from:
https://www.ncbi.nlm.nih.gov
https://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide&cmd=Search&term=M11166&doptcmdl=GenBank

Methods and Results:

DNA Structure, and Nucleotide Base Pairing

The instructor may choose to do the JMol activities below as a demonstration

Double click the shortcut named Jmol to open Jmol, an Open-source Java program for viewing chemical structures in 3D.

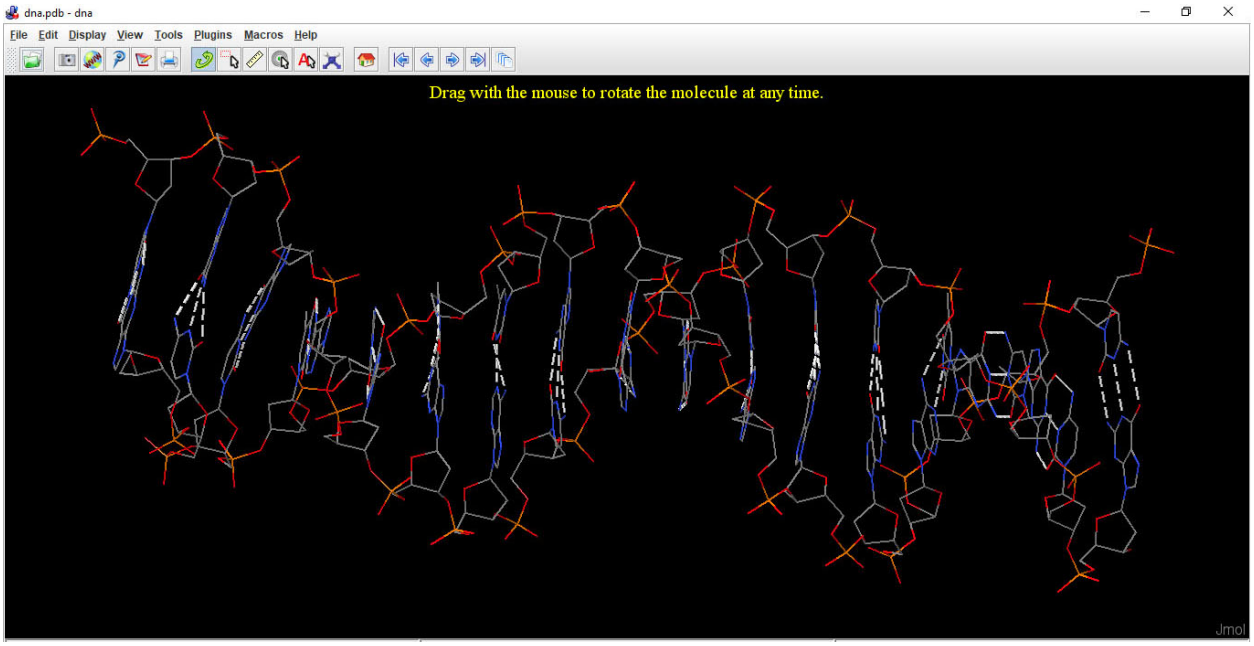
In this activity, files are viewed offline by Jmol by selecting Open from the File menu.

You probably downloaded these files in the previous Lab. If not, download them from my website.: https://www.dgward.com/pdf/physo101/pdf_other_physo101.htm
Click on "Download –updated Jmol biomolecules" and a ".zip" file should download automatically. After downloading, "Open", select "Extract All", and save to the Desktop.

DNA Double Helix

Click on "File", "Open", "Look In: Desktop", select "updated_Jmol_biomolecules"," PDB", and "dna.spt ". Other file options will be discussed if necessary.

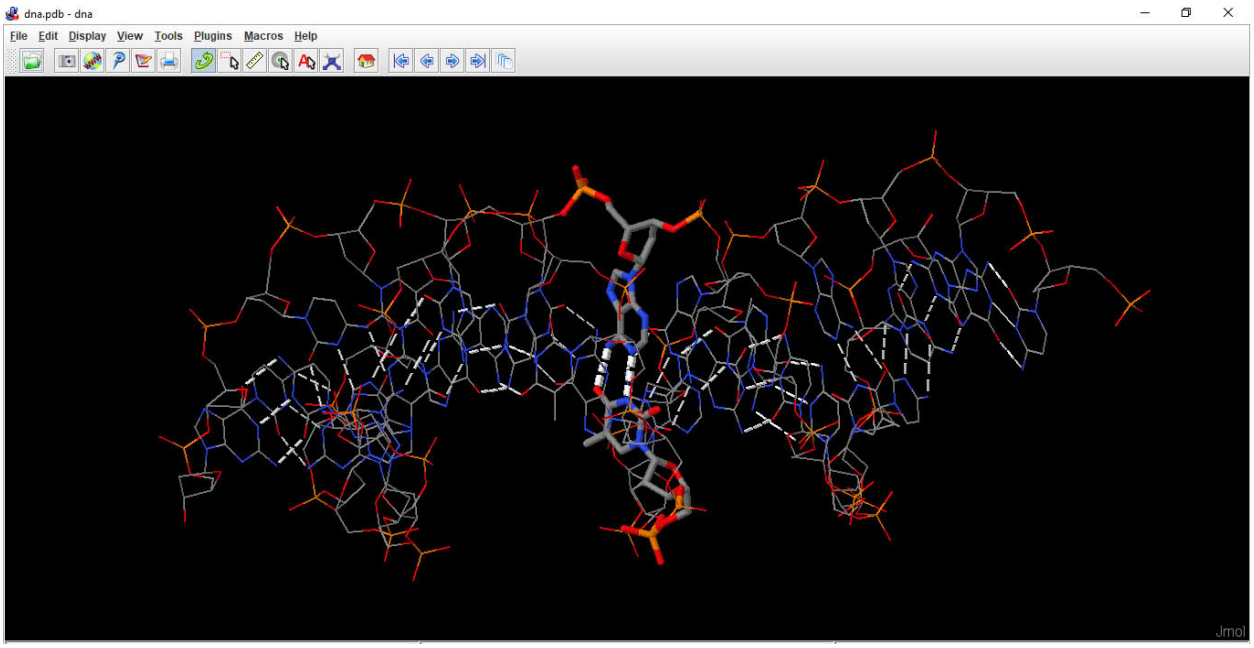
A 3D model of a section of DNA double helix will appear, as shown below.



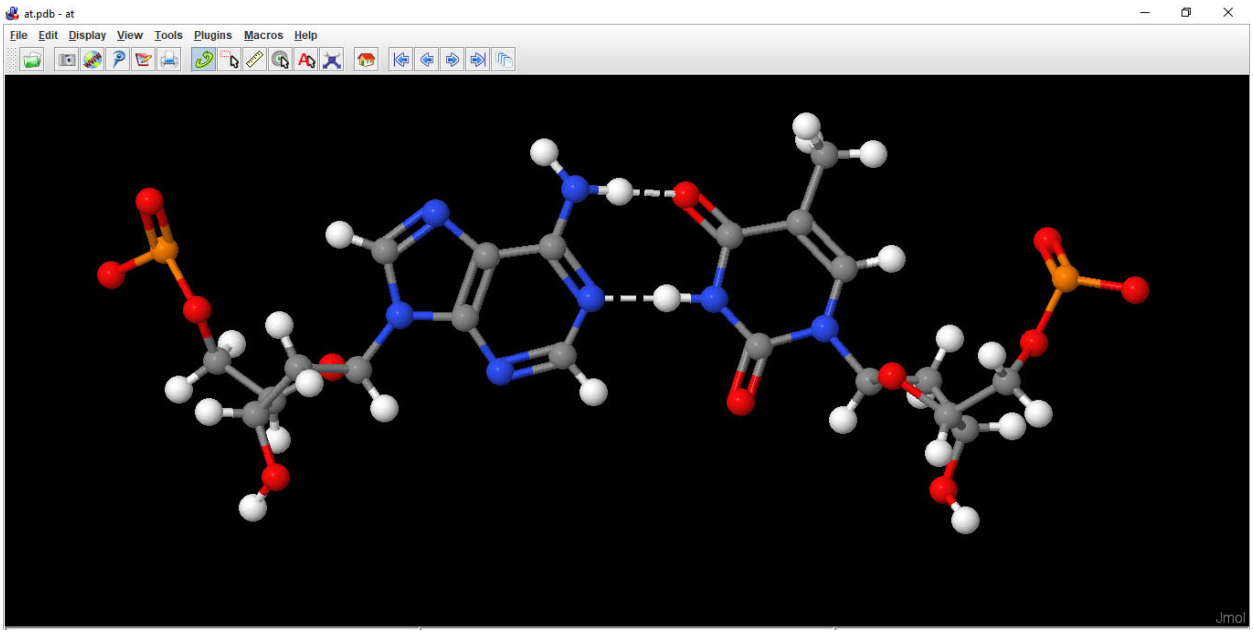
DNA Nucleotide Base Pairing

Adenine-Thymine Base pairing within the double helix

Click on "File", "Open", "Look In: Desktop", select "updated_Jmol_biomolecules"," PDB", and "dna_at.spt". A 3D model of a section of DNA double helix with an Adenine nucleotide and a Thymine nucleotide highlighted appears, as shown below. Note the orientation of the nucleotides.



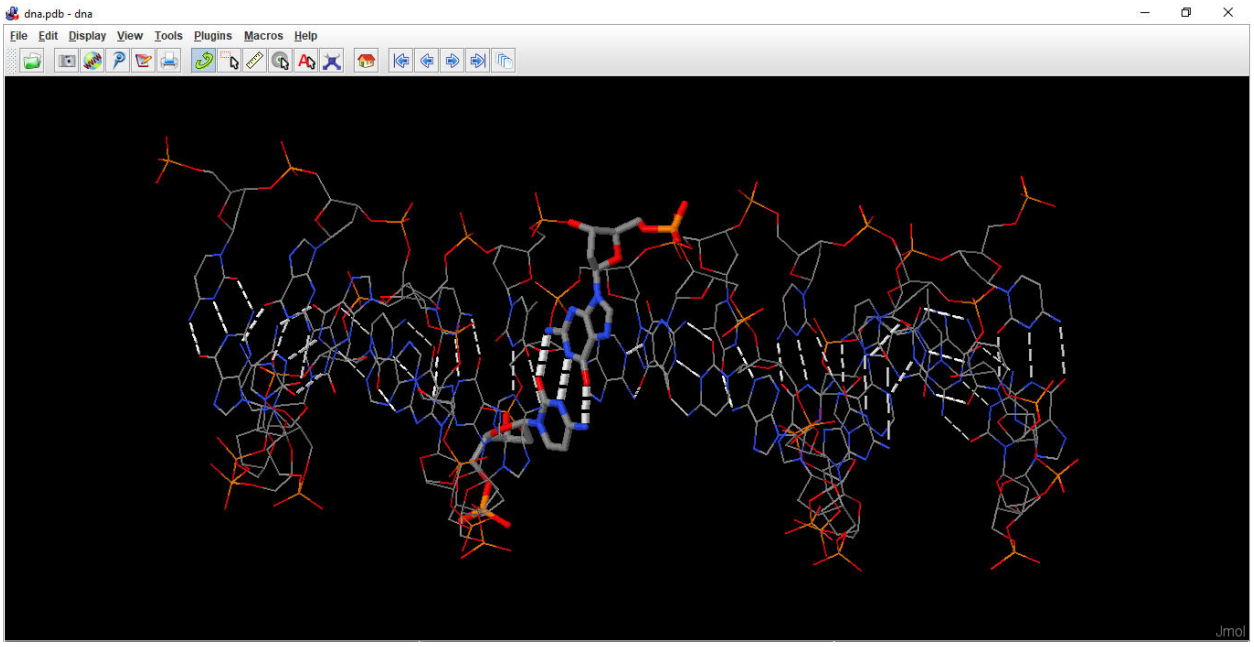
Click on "File", "Open", "Look In: Desktop", select "updated_Jmol_biomolecules"," PDB", and "at.spt". A 3D model of an Adenine nucleotide and a Thymine nucleotide appears, as shown below. Note the <u>two hydrogen bonds</u> between these two nucleotides.



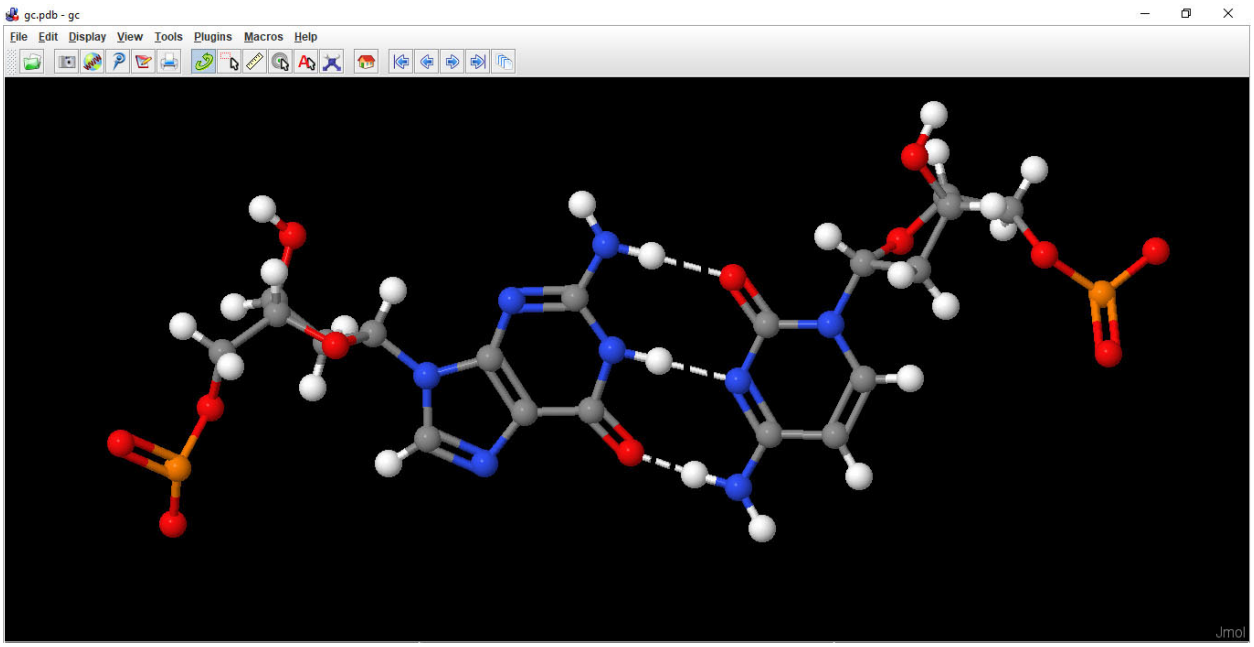
DNA Nucleotide Base Pairing

Guanine-Cytosine Base pairing within the double helix

Click on "File", "Open", "Look In: Desktop", select "updated_Jmol_biomolecules"," PDB", and "dna_gc.spt". A 3D model of a section of DNA double helix with an Guanine nucleotide and a Cytosine nucleotide highlighted appears, as shown below. Note the orientation of the nucleotides.



Click on "File", "Open", "Look In: Desktop", select "updated_Jmol_biomolecules"," PDB", and "gc.spt". A 3D model of a Guanine nucleotide and a Cytosine nucleotide appears, as shown below. Note the <u>three hydrogen bonds</u> between these two nucleotides.



DNA Structure, Transcription, and subsequent Translation

DNA structure

DNA, as we usually think about it, is composed of a double helix of DNA nucleotides. One strand of the double helix is the *template* strand of DNA (tDNA). The other strand of the double helix is the *coding* strand of DNA (cDNA). For example:

A T G C C T G A C	cDNA nucleotide sequence
T A C G G A C T G	tDNA nucleotide sequence

DNA duplication

DNA duplication occurs most commonly during the later stages of interphase, prior to mitosis. The process of DNA duplication uses DNA polymerase to separate the two strands of the double helix. Then each strand is used to make a complementary strand of itself. The result is two double helices, each with two strands of DNA. Determine the nucleotide sequence of the new strands.

A T G C C T G A C	cDNA nucleotide sequence
	tDNA nucleotide sequence

	cDNA nucleotide sequence
T A C G G A C T G	tDNA nucleotide sequence

When completed you will have two sets (pairs) of double helices. In the cell this means there will be two chromatids, rather than just one chromatid, in each chromosome,

DNA Transcription to pre-mRNA (pRNA)

Transcription occurs most commonly in the earlier stages of interphase, for the purpose of producing stands of RNA that will use the **nucleotide sequences** to code for and to orchestrate the synthesis of proteins, and of non-protein coding RNAs. The process of transcription uses RNA polymerase to separate the double strand of DNA and to make a strand of RNA that is the complement of the template strand of DNA (tDNA). The complementary strand of RNA is either called pre-messenger RNA (pre-mRNA) or precursor RNA (pRNA). Due to this process, the pre-mRNA will be very similar to the cDNA. Where DNA nucleotides have Thymine, RNA nucleotides have Uracil.

T A C G G A C T G	tDNA nucleotide sequence

Determine the nucleotide sequence of the precursor RNA (pRNA)

	Pre mRNA nucleotide sequence

Remember:

DNA nucleotides have Adenine (A), Guanine (G), Thymine (T), or Cytosine (C).

RNA nucleotides have Adenine (A), Guanine (G), Uracil (U), or Cytosine (C).

Therefore:

tDNA nucleotides	A	G	T	C
pair with pre-mRNA nucleotides	U	C	A	G

Genes

Genes include three major **sequences of nucleotides**:

1) A promoter sequence for control of gene expression
2) Exon sequences coding for and orchestrating control of protein synthesis
3) Intron sequences with obscure functions that may include control of exon reassembly or control of expression of other genes., or to code for other non-protein coding RNAs

Cutting and Splicing of pre-mRNA to produce mRNA and other RNAs

The pre-mRNA is the complimentary copy of the nucleotide sequence of the tDNA, which contains most of the "Gene" except for the promoter sequence.

Introns are cut out of the pre-mRNA, and the exons along with the remainder of the pre-mRNA, are spliced together to form messenger RNA (mRNA). The introns may be used for control of exon reassembly or expression of other genes, or to code for other non-protein coding RNAs. See text chapter 3, p 8, Figure 3-7.

mRNA and Protein Synthesis (Translation)

The mRNA leaves the nucleus and attaches to ribosomes (composed of protein and ribosomal RNA (rRNA), to orchestrate the synthesis of proteins from amino acids. These amino acids are carried by another type of RNA called transfer RNA (tRNA).

mRNA with specific nucleotides triplets (Codons) will attach to tRNA with complementary nucleotide triplets (Anticodons).

Accordingly, specific nucleotides triplets of mRNA (Codons) will ultimately represent a particular amino acid, as shown in the following table.

The Genetic Code (mRNA)

1st position (5' end)	2nd position (middle)				3rd position (3' end)
	U	**C**	**A**	**G**	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP **Trp**	**U** **C** **A** **G**
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	**U** **C** **A** **G**
A	Ile Ile Ile **Met**	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	**U** **C** **A** **G**
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	**U** **C** **A** **G**

Given a strand of mRNA with the following nucleotide triplets, determine the amino acids that will be carried by the tRNAs.

mRNA Triplet:	G	A	U	C	C	U	G	A	C
Corresponding AA:									

Decoding the human "AVP" (8-arginine vasopressin) gene

"AVP" (8-arginine vasopressin) is a peptide hormone critical for control of renal water retention, blood vessel constriction, and thirst. This peptide contains only 9 amino acids (amino acid # 8 is arginine). The nucleotide sequence of the human "AVP" (8-arginine vasopressin) gene is obtained from the National Library of Medicine and reproduced on the following page. The human "AVP" gene contains 2500 nucleotide base pairs

position 1-168	A Promoter sequence that binds to RNA polymerase and determines which strand of DNA is transcribed
position 169-224	A code for controlling sequences
position 225-281	A Signal protein code that codes for a protein that directs the material protein to specific cellular organelles
position 282-308	The code for 8-Arginine Vasopressin (AKA antidiuretic hormone)
positions 318-344; 1718-1919; 2087-2136	The code for Neurophysin II that codes for a protein that transports vasopressin in blood
position 2138-2343	The code for a Glycoprotein
position 345-1717	The code for Intron 1
position 1920-2086	The code for Intron 2
position 2344-2500	The terminator code

Make a special note that DNA databases often publish DNA nucleotide sequences using the coding strand of DNA (cDNA). Using the cDNA data on the following page:

1) Locate each of the sequences above for the cDNA nucleotide sequence of the human "AVP" gene.
2) Locate the cDNA nucleotide sequence for 8-arginine vasopressin, determine the corresponding tDNA and mRNA nucleotide sequence, determine the corresponding amino acids (AA), and construct the amino acid sequence for this peptide.

position	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296
cDNA															
tDNA															
mRNA															
AA															

position	297	298	299	300	301	302	303	304	305	306	307	308
cDNA												
tDNA												
mRNA												
AA												

The cDNA sequence for the human "AVP" (8-arginine vasopressin) gene

```
   1 gatcccctgc acagacaggc ccacgtgtgt ccccagatgc ctgaatcact gctgaccgct
  61 ggggacctgg cggccgtggg ctcctgggga gccactgggg agggggtggc ggccgcgtct
 121 cgcctccacg ggaacacctg cggacataaa taggcagcca gcaggaggcag cagcacagag
 181 ccaccaagca gtgctgcata cggggtccac ctgtgtgcac caggatgcct gacaccatgc
 241 tgcccgcctg cttccctggc ctactggcct tctcctccgc gtgctacttc cagaactgcc
 301 cgaggggcgg caagagggcc atgtccgacc tggagctgag acaggtactt cccactgtgg
 361 gccatctcag ggcagccata gcgggcagtg ctgacaccct gggtcagggg ctaggaaaga
 421 gggaagtcat gggtggtggt agcctttagg ggaagttcgg gggaggaaga gggaggcatg
 481 gcatggctgg gcagaggagc caatggggtg ggccagaggg gaccaggctt tggaggaggc
 541 tgggagaggc tgaaggcgct cctggtcact gtcgccatcc agacagggat gcaggaaaat
 601 gagggatgct tccccggtga ctgggcttgg ggctggatag ggagaacggg gcatcatggc
 661 ctcccctgtg cccatggcgt tcttgcatct ggactggctg gggcagcaga ggctccatcc
 721 tacctagcat tggaggcttt cctcatccag ccccagcctc ccagccacag gcgcccaggc
 781 ccccacacag aagatggcca ctggtctgag cgcgcttgag tggggcatcc tgtgggaagt
 841 tctgctggga acctggccta attctatagt gctggacgtt tcctccattt ccagcagagc
 901 tgaaggaaat ccaatcacga tgtgcatgca attctgtcca gctcaatgat gagcccttga
 961 gcaaattaga ccacaccagg ctcacgtaaa agtctaatcg cgtatccatt gcgccagaga
1021 accggctgtt gagcagatga gagtgcgcgc tcggcaaccc ccgcagcctc tcttcctcct
1081 gctaggctcc tttagggtcc tgagggcacc tgggtgtccg tgctcgcctc taggtctcag
1141 cccctgccac ccacctgata ggtcataggt ggctgagcag gggtcagggc tccagctgag
1201 gccgacaagc ttggcggggc cagggcgaag gcaagagagg agacaggaaa tgggaagggc
1261 cggggttctg gatgggtagg gcctctccgc atggtgtagt ggggaagggg gtgggcccgg
1321 gctcaagccg cagcagggcg aggaggaagg aggaagggtc tggagtggtg gagggtgggg
1381 cagctgcaac agtggcgccc accagcgatg accccgaggc tcgaggaagg gctccccacg
1441 ctgtagtcca cgggagaccc gaccctagct gagggtgagg acgctgaggg ctgtcaccga
1501 gaggtcatcc aagaaaccaa ggtgccgagc agatctggac gccccgcccg tgaccgcggt
1561 cgaggcccag tggcgcccga gcgtgcctgc agccgcagcc ccggtgtccc gcccgcactc
1621 cgagccctgg accccagcat ccccgcctcg ctgcgttccc ctccaacccc tcgactcccg
1681 gctcccctcc tcccgctcac cccgcccgtc cccgcagtgc ctcccctgcg gccccggggg
1741 caaaggccgc tgcttcgggc ccagcatctg ctgcgcggac gagctgggct gcttcgtggg
1801 cacggctgag gcgctgcgct gccaggagga gaactacctg ccgtcgccct gccagtccgg
1861 ccagaaggcg tgcgggagcg ggggccgctg cgccgccttc ggcgtttgct gcaacgacgg
1921 tgcgcggcgg gggcgggcct gggggggggc gcagacgctt gggtgggggg gacgcgggcc
1981 tgcggcgggg tgggggctcg tgcgggcccg gcagggaggg tgtgggcccc ccgcaccccg
2041 agctgcgccc gccccaggcg cccgtgctca cacgtcctcc cggcagagag ctgcgtgacc
2101 gagcccgagt gccgcgaggg ctttcaccgc cgcgcccgcg ccagcgaccg gagcaacgcc
2161 acgcagctgg acgggccggc cggggccttg ctgctgcggc tggtgcagct ggccgggggcg
2221 cccgagccct tcgagcccgc ccagcccgac gcctactgag cccgcgctgc ccccaccgcg
2281 ctcttccgcc cgcccctgca gcacggacaa taaacctccg ccaatgcacg gcctcgcgtc
2341 tgtctcagtc tctggcggga agagggaagg ggagagaggt gggagcgcgg acccccgcca
2401 ccacgcccac cggccagtcc ccggacctga ggtcgtgggc agatccaccc cagagaagca
2461 acaggtcccg tagaggaagc gatctgggac ccgcagaggt
```

DNA double helix for the human AVP gene

You can use JMol to view positions 282-308 of the DNA double helix for the human AVP gene. Click on "File", "Open", "Look In: Desktop", select "updated_Jmol_biomolecules"," PDB", and "vasopressin_DNA_web.spt".

Discussion:

1. Compare and contrast the structure and composition of tDNA, cDNA, and pre-mRNA,
2. Explain the role of nucleotide sequences in the genetic coding of proteins and non-coding RNA.
3. Explain the role of transcription factors and the promoter region in gene expression.
4. Speculate on the significance of the introns in gene expression.