HUMHBB: Human Hemoglobin Beta Activity*

Background: This activity allows you to look at the actual nucleotide sequence of the beta-globin protein. Beta-globin is one of the protein subunits that makes up the protein hemoglobin. Hemoglobin is composed of two alpha subunits linked with two beta subunits. Each subunit binds a heme group, which in turn binds one iron atom (Fe). Oxygen can bind reversibly to this Fe atom. As a red blood cell passes through the lungs, oxygen diffuses into the cell and binds to the Fe of the heme group on all four subunits of the hemoglobin tetramer. Sickle Cell Anemia is caused by a single point mutation in the nucleotide sequence of beta-globin. The mutation is located in the sixth (seventh) codon. The sixth codon should read GAG (glutamic acid), but the middle nucleotide has changed to a thymine, which changes the codon to GTG which codes for valine. Beta-globin's three dimensional stucture places this sixth codon on the outside of the protein. Replacement of the normally charged glutamic acid with the hydrophobic valine alters the solubility of the protein so that at lower concentration of oxygen the protein changes to a sickled shape and it is unable to carry oxygen. This causes the symptoms of Sickle Cell Anemia.

This sequence was obtained from a internet source called **GenBank** under the key word: HUMHBB. There are actually 73,308 nucleotides for hemoglobin and it is located at the tip of chromosome 11. This includes the adult beta-globin gene as well as several other related globin genes (fetal, alpha..). If we cut out the entire sequence, which is 20 pages long and taped it together in the correct order, it would only represent 0.002% of the entire human genome.
The actual portion of this 20 page sequence that contains the beta-globin information is located between nucleotides 62,187 to 63,610. You are going to trace the flow of information from deoxyribonucleic acid (DNA) to the sequence of ribonucleic acid (messenger RNA) and finally to a sequence of amino acids that comprise the final beta-globin protein.

KEYWORDS:

hemoglobin, beta-globin, heme, Fe, nucleotide, codon-triplet, DNA, exon, intron, gene, spliceosome, messenger RNA, transcription, translation, protein synthesis, protein, mutation

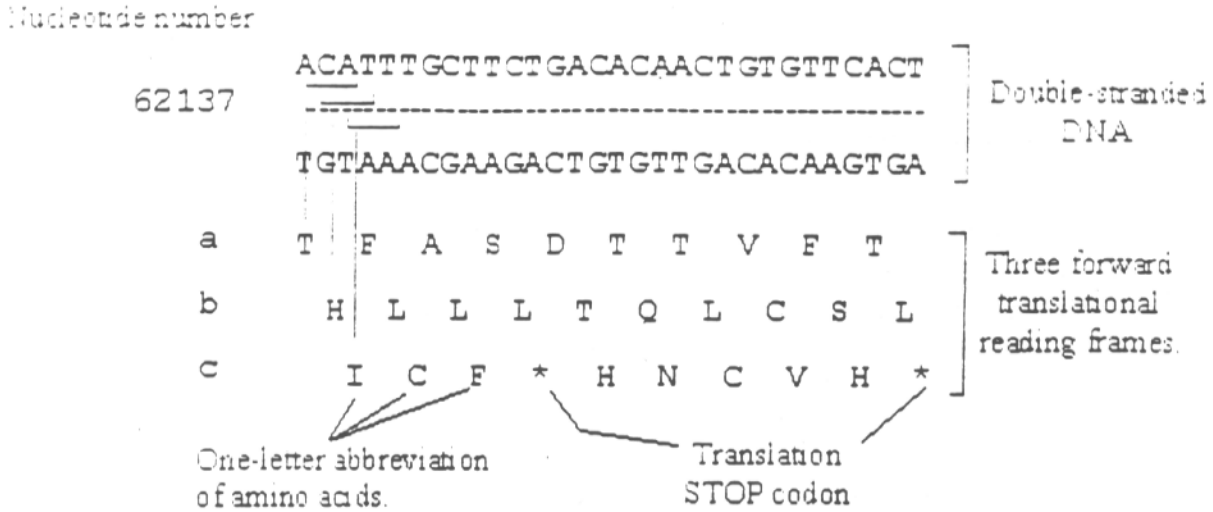
PROCEDURE:

1. Read the following page "How to interpret the map of the Beta-Globin Gene".
2. You will be given three pages of the actual nucleotide sequence in the format discussed in procedure step 1. Examine the pages carefully and familiarize yourself with amino acid codes. Below is the actual sequence of amino acids that comprise the beta-globin protein.

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFE
SFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLG
NVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH

3. Can you find the gene? There are actually 1,423 nucleotides, enough to code for 474 triplet codons. But beta-globlin protein is only known to have 146 amino acids. What is all the excess DNA doing? It is serving as "introns" (intervening sequences) that separate the

How to Interpret the Map of the Beta-Globin Gene.



The beginning of the globin.map file is shown in the figure above. Using a printed version of this file, the following information can be explored with students:

The nucleotide sequence of both strands of DNA is shown, beginning with nucleotide 62,137. Remember that in the Watson-Crick double-helix, the two strands of DNA are anti-parallel. The top strand represents the "non-coding" strand and the bottom strand represents the "coding" strand. When this gene is expressed, RNA polymerase uses the bottom strand as a template to make an RNA copy that corresponds to the sequence shown in the top strand - with the T's replaced with U's.

The amino acid sequence encoded by the DNA is shown in the three lines below the nucleotide sequence. The one-letter abbreviation of the amino acids is used. Thus, T represents threonine, H is histidine, I is isoleucine and so forth. A * refers to one of the three translation STOP codons.

Why are there **three** lines of amino acid sequence?..... Because the triplet genetic code can be read in three possible "reading frames". These are designated "a, b, and c". In the "a" reading frame, the ribosome reads the triplet ACA to mean T (threonine). In reading frame "b", the ribosome shifts one nucleotide and reads the triplet CAT to mean H (histidine).

✳ Which reading frame is used to translate the beta-globin protein?.....and where does that translation begin? From the documentation found in the HUMHBB file, you know that the coding sequence starts at nucleotide 62,187. That establishes the "c" reading frame as the one that is used to translate the beta-globin mRNA. Therefore, the amino acid sequence of the beta-globin protein begins M V H L T P E E K..... and continues for a total of 146 amino acid residues. This coding region constitutes an "open reading frame". Examine the other 2 possible reading frames over this same region of the gene.....and note all of the translation STOP codons.

Finally, examine the codon that encodes the 7th amino acid - an E (glutamic acid). In sickle cell anemia, this GAG is changed to GTG as a result of a mutation in the second base of the codon. The new codon, GTG, encodes the hydrophobic amino acid valine. The replacement of a negatively-charged glutamic acid residue with a hydrophobic valine residue at this position of the beta globin protein is responsible for the aggregation of hemoglobin in red blood cells that results in sickle cell anemia.

Amino Acid Symbols

IUB Symbol 3-letter Meaning Codons Depiction

A Ala Alanine GCT,GCC,GCA,GCG

B Asp,Asn Aspartic, Asparagine GAT,GAC,AAT,AAC

C Cys Cysteine TGT,TGC

D Asp Aspartic GAT,GAC

E Glu Glutamic GAA,GAG

F Phe Phenylalanine TTT,TTC

G Gly Glycine GGT,GGC,GGA,GGG

H His Histidine CAT,CAC

I Ile Isoleucine ATT,ATC,ATA

K Lys Lysine AAA,AAG

L Leu Leucine TTG,TTA,CTT,CTC,CTA,CTG

M Met Methionine ATG

N Asn Asparagine AAT,AAC

P Pro Proline CCT,CCC,CCA,CCG

Q Gln Glutamine CAA,CAG

R Arg Arginine CGT,CGC,CGA,CGG,AGA,AGG

S Ser Serine TCT,TCC,TCA,TCG,AGT,AGC

T Thr Threonine ACT,ACC,ACA,ACG

V Val Valine GTT,GTC,GTA,GTG

W Trp Tryptophan TGG

X Xxx Unknown !XXX

Y Tyr Tyrosine TAT, TAC

Z Glu,Gln Glutamic, Glutamine GAA,GAG,CAA,CAG

* End Terminator TAA, TAG, TGA

protein-encoding "exons" (expressed sequences). The beta-globin gene consists of three exons and two introns.

Exon #1	62187-62278
Intron #1	62279-62408
Exon #2	62409-62631
Intron #2	62632-63481
Exon #3	63482-63610

4. In your group, locate the introns and highlight them with a transparent maker. **Check your work** with the teacher before going on to step 5.

5. Translate your entire sequence into messenger RNA (mRNA). Which side is the coding side and is transcripted into RNA? If you aren't sure, see the sheet entitled "How to interpret the map of the Beta-Globin Gene". Record this in your journal.

6. Highlight the sections that will be removed by the spliceosome. How many nucleotides will be removed per intron section? Record this in your journal.

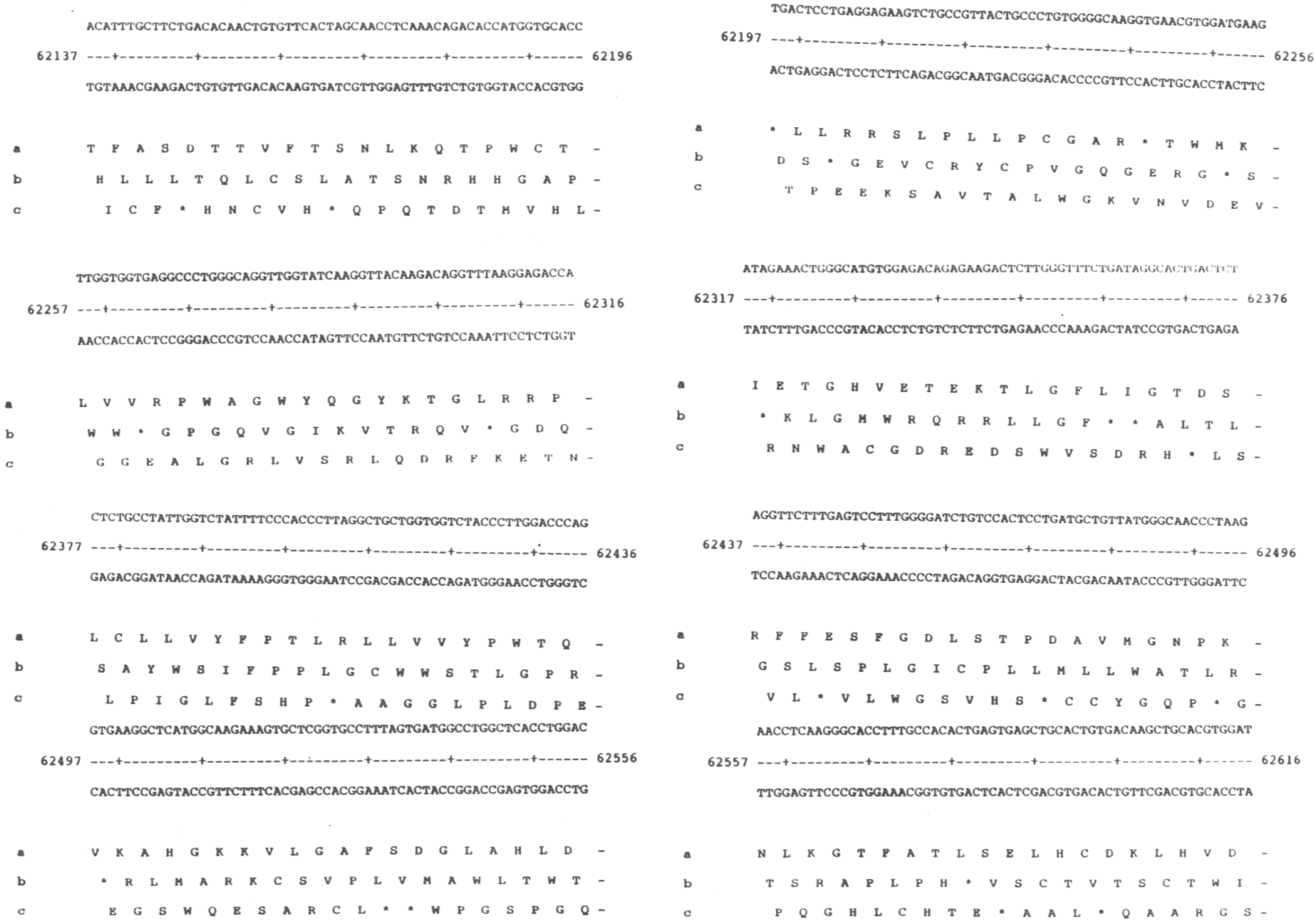
7. Recopy the new shorter messege. Place a slash mark between each codon. Does the amino acid sequence in procedure number 2, match your version of message? Record this in your journal.

8. What is the sixth codon? Record this in your journal. Confirm that if you replace the middle nucleotide, the amino acid sequence is changed to valine. Record this in your journal.

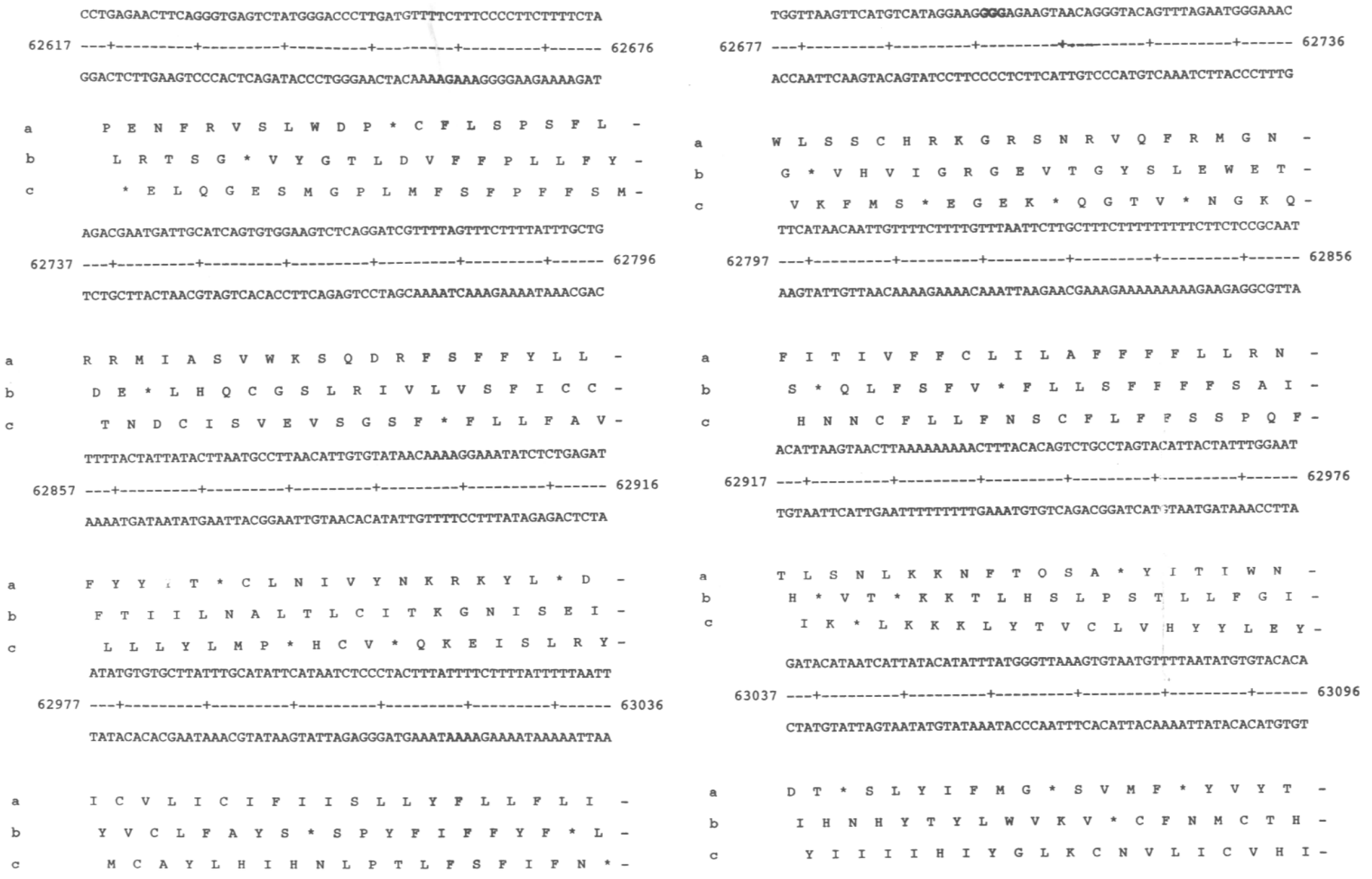
POST LAB QUESTIONS: (Typed for homework.)

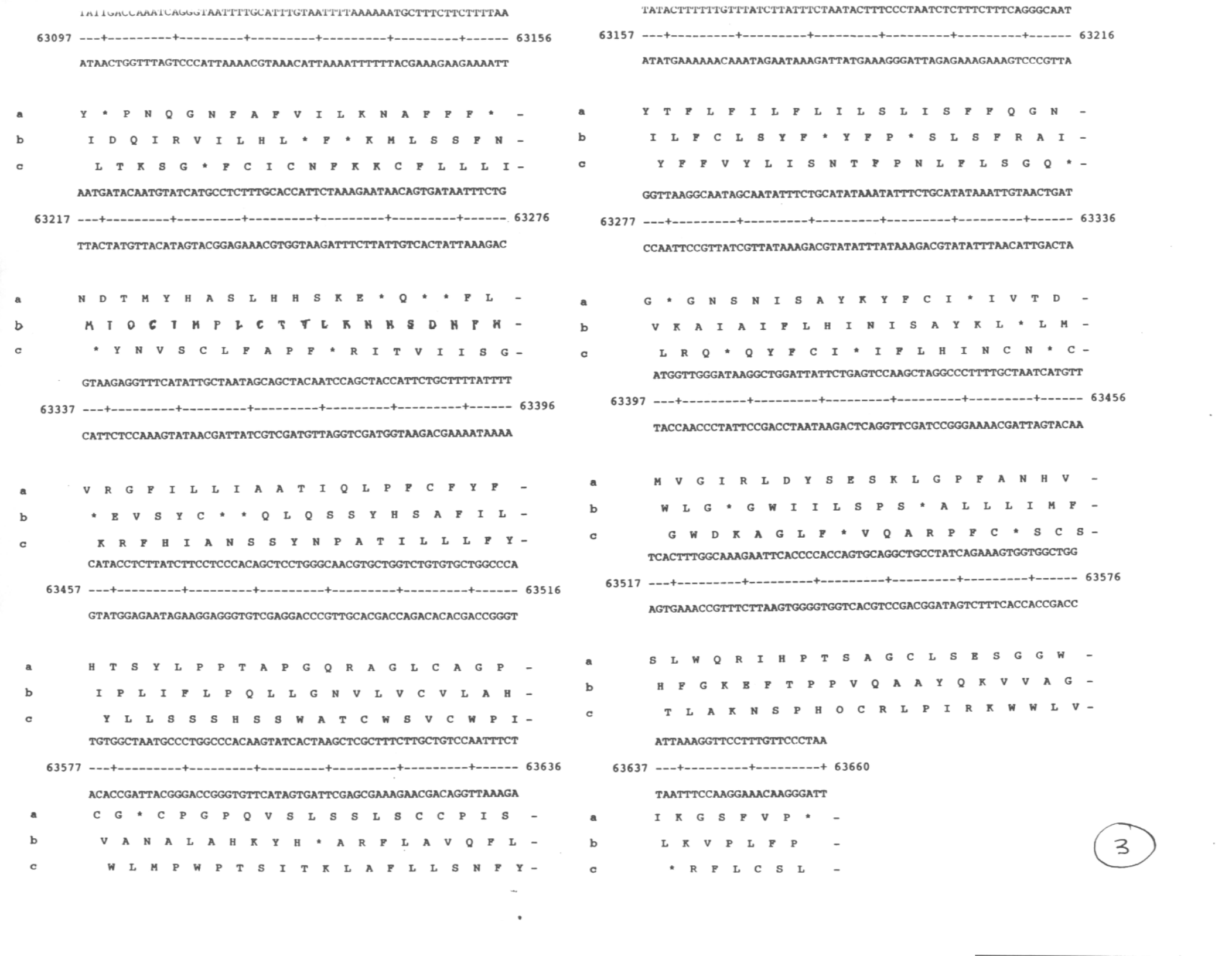
1. Define the key terms.
2. How is RNA different from DNA?
3. Besides trimming the introns out of mRNA, what else must be done to a mRNA before it can successfully leave the nucleus? What part do ribosomes and transfer RNA play in translation and protein synthesis?
4. There are other types of mutations besides point mutations (base substitutions), what are they? How do they destroy the integrity of the gene? How do mutations affect the final protein product?
5. Francis Crick's (of Watson and Crick) Central Dogma tells of the link between DNA, RNA and protein synthesis. How does this activity relate to the Central Dogma?
6. Look up sickle cell anemia. What are the symptoms? How does sickle cell anemia reduce the quality of life? What is the link to malaria? What is the genetic pattern of inheritance? (Include a bibliography of any sources you used.)

*This activity was adapted from:
http://www.biochem.mcw.edu/~herman/science_ed/Pages/hemoglobin/real1.html



http://www.biochem.mcw.edu/~herman/science_ed/Pages/hemoglobin/files/bglobin.map


CCTGAGAACTTCAGGGTGAGTCTATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTA
62617 ---+---------+---------+---------+---------+---------+------ 62676
GGACTCTTGAAGTCCCACTCAGATACCCTGGGAACTACAAAAGAAAGGGGAAGAAAAGAT
a P E N F R V S L W D P * C F L S P S F L -
b L R T S G * V Y G T L D V F F P L L F Y -
c * E L Q G E S M G P L M F S F P F F S M -
AGACGAATGATTGCATCAGTGTGGAAGTCTCAGGATCGTTTTAGTTTCTTTTATTTGCTG
62737 ---+---------+---------+---------+---------+---------+------ 62796
TCTGCTTACTAACGTAGTCACACCTTCAGAGTCCTAGCAAAATCAAAGAAAATAAACGAC
a R R M I A S V W K S Q D R F S F F Y L L -
b D E * L H Q C G S L R I V L V S F I C C -
c T N D C I S V E V S G S F * F L L F A V -
TTTTACTATTATACTTAATGCCTTAACATTGTGTATAACAAAAGGAAATATCTCTGAGAT
62857 ---+---------+---------+---------+---------+---------+------ 62916
AAAATGATAATATGAATTACGGAATTGTAACACATATTGTTTTCCTTTATAGAGACTCTA
a F Y Y L T * C L N I V Y N K R K Y L * D -
b F T I I L N A L T L C I T K G N I S E I -
c L L L Y L M P * H C V * Q K E I S L R Y -
ATATGTGTGCTTATTTGCATATTCATAATCTCCCTACTTTATTTTCTTTTATTTTTAATT
62977 ---+---------+---------+---------+---------+---------+------ 63036
TATACACACGAATAAACGTATAAGTATTAGAGGGATGAAATAAAAGAAAATAAAAATTAA
a I C V L I C I F I I S L L Y F L L F L I -
b Y V C L F A Y S * S P Y F I F F Y F * L -
c M C A Y L H I H N L P T L F S F I F N * -
TGGTTAAGTTCATGTCATAGGAAGGGGAGAAGTAACAGGGTACAGTTTAGAATGGGAAAC
62677 ---+---------+---------+---------+---------+---------+------ 62736
ACCAATTCAAGTACAGTATCCTTCCCCTCTTCATTGTCCCATGTCAAATCTTACCCTTTG
a W L S S C H R K G R S N R V Q F R M G N -
b G * V H V I G R G E V T G Y S L E W E T -
c V K F M S * E G E K * Q G T V * N G K Q -
TTCATAACAATTGTTTTCTTTTGTTTAATTCTTGCTTTCTTTTTTTTTCTTCTCCGCAAT
62797 ---+---------+---------+---------+---------+---------+------ 62856
AAGTATTGTTAACAAAAGAAAACAAATTAAGAACGAAAGAAAAAAAAAAGAAGAGGCGTTA
a F I T I V F F C L I L A F F F F L L R N -
b S * Q L F S F V * F L L S F F F F S A I -
c H N N C F L L F N S C F L F F S S P Q F -
ACATTAAGTAACTTAAAAAAAAAACTTTACACAGTCTGCCTAGTACATTACTATTTGGAAT
62917 ---+---------+---------+---------+---------+---------+------ 62976
TGTAATTCATTGAATTTTTTTTTTGAAATGTGTCAGACGGATCATGTAATGATAAACCTTA
a T L S N L K K N F T Q S A * Y I T I W N -
b H * V T * K K T L H S L P S T L L F G I -
c I K * L K K K L Y T V C L V H Y Y L E Y -
GATACATAATCATTATACATATTTATGGGTTAAAGTGTAATGTTTTAATATGTGTACACA
63037 ---+---------+---------+---------+---------+---------+------ 63096
CTATGTATTAGTAATATGTATAAATACCCAATTTCACATTACAAAATTATACACATGTGT
a D T * S L Y I F M G * S V M F * Y V Y T -
b I H N H Y T Y L W V K V * C F N M C T H -
c Y I I I I H I Y G L K C N V L I C V H I -


TATTGACCAAATCAGGGTAATTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAA
63097 ---+---------+---------+---------+---------+---------+------ 63156
ATAACTGGTTTAGTCCCATTAAAACGTAAACATTAAAATTTTTTACGAAAGAAGAAAATT
a Y * P N Q G N F A F V I L K N A F F F * -
b I D Q I R V I L H L * F * K M L S S F N -
c L T K S G * F C I C N F K K C F L L L I -
AATGATACAATGTATCATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTG
63217 ---+---------+---------+---------+---------+---------+------ 63276
TTACTATGTTACATAGTACGGAGAAACGTGGTAAGATTTCTTATTGTCACTATTAAAGAC
a N D T M Y H A S L H H S K E * Q * * F L -
b M I Q C I M P L C T T L K N N S D N F N -
c * Y N V S C L F A P F * R I T V I I S G -
GTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTT
63337 ---+---------+---------+---------+---------+---------+------ 63396
CATTCTCCAAAGTATAACGATTATCGTCGATGTTAGGTCGATGGTAAGACGAAAATAAAA
a V R G F I L L I A A T I Q L P F C F Y F -
b * E V S Y C * * Q L Q S S Y H S A F I L -
c K R F H I A N S S Y N P A T I L L L F Y -
CATACCTCTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCA
63457 ---+---------+---------+---------+---------+---------+------ 63516
GTATGGAGAATAGAAGGAGGGTGTCGAGGACCCGTTGCACGACCAGACACACGACCGGGT
a H T S Y L P P T A P G Q R A G L C A G P -
b I P L I F L P Q L L G N V L V C V L A H -
c Y L L S S S H S S W A T C W S V C W P I -
TGTGGCTAATGCCCTGGCCCACAAGTATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCT
63577 ---+---------+---------+---------+---------+---------+------ 63636
ACACCGATTACGGGACCGGGTGTTCATAGTGATTCGAGCGAAAGAACGACAGGTTAAAGA
a C G * C P G P Q V S L S S L S C C P I S -
b V A N A L A H K Y H * A R F L A V Q F L -
c W L M P W P T S I T K L A F L L S N F Y -
TATACTTTTTTGTTTATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAAT
63157 ---+---------+---------+---------+---------+---------+------ 63216
ATATGAAAAAACAAATAGAATAAAGATTATGAAAGGGATTAGAGAAAGAAAGTCCCGTTA
a Y T F L F I L F L I L S L I S F F Q G N -
b I L F C L S Y F * Y F P * S L S F R A I -
c Y F F V Y L I S N T F P N L F L S G Q * -
GGTTAAGGCAATAGCAATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACTGAT
63277 ---+---------+---------+---------+---------+---------+------ 63336
CCAATTCCGTTATCGTTATAAAGACGTATATTTATAAAGACGTATATTTAACATTGACTA
a G * G N S N I S A Y K Y F C I * I V T D -
b V K A I A I F L H I N I S A Y K L * L M -
c L R Q * Q Y F C I * I F L H I N C N * C -
ATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATGTT
63397 ---+---------+---------+---------+---------+---------+------ 63456
TACCAACCCTATTCCGACCTAATAAGACTCAGGTTCGATCCGGGAAAACGATTAGTACAA
a M V G I R L D Y S E S K L G P F A N H V -
b W L G * G W I I L S P S * A L L L I M F -
c G W D K A G L F * V Q A R P F C * S C S -
TCACTTTGGCAAAGAATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGG
63517 ---+---------+---------+---------+---------+---------+------ 63576
AGTGAAACCGTTTCTTAAGTGGGGTGGTCACGTCCGACGGATAGTCTTTCACCACCGACC
a S L W Q R I H P T S A G C L S E S G G W -
b H F G K E F T P P V Q A A Y Q K V V A G -
c T L A K N S P H O C R L P I R K W W L V -
ATTAAAGGTTCCTTTGTTCCCTAA
63637 ---+---------+---------+ 63660
TAATTTCCAAGGAAACAAGGGATT
a I K G S F V P * -
b L K V P L F P -
c * R F L C S L -
3