Understanding Transcription, Translation, and the Effect of Mutations

Resource Type: Curriculum: Classroom

Publication Date: 7/2/2001

Authors

Linda Fisher
Associate Professor of Biology and Microbiology
University of Michigan-Dearborn
Dearborn, Michigan
USA
Email: Deceased

Abstract

In this activity, students will practice identifying a translation open reading frame in the context of a nucleotide sequence and recognizing the effects that mutations of various types have on the resulting polypeptide product. This can be done using short DNA "microsequences" written specifically for this activity, student- or instructor-written sequences, or actual published gene sequences (e.g., from GenBank).

Activity

Invitation for User Feedback. If you have used the activity and would like to provide feedback, please send an e-mail to MicrobeLibrary@asmusa.org. Feedback can include ideas which complement the activity and new approaches for implementing the activity. Your comments will be added to the activity under a separate section labeled "Feedback." Comments may be edited.

INTRODUCTION

Time Required.
Depending on how the exercise is used, it can take from 5 to 30 minutes or more.

Pedagogical Function.
Students often have difficulty understanding how a sequence of DNA is converted into a polypeptide and how changes in that DNA sequence affect the resulting protein. Although students generally can tell which sequences signal the start and stop for transcription and translation and know the definition of a codon, they do not recognize these elements in the context of a nucleotide sequence. This activity is designed to give students practice in identifying products of transcription and translation and determining the effect of gene sequence mutations on the resulting protein product.

p>**Background.**
For introductory students:
Students should know the basic mechanisms involved in transcription and translation, as well as how mutations in DNA affect the product transcribed from that DNA. This activity would be appropriate for introductory students or for a quick review of concepts for more advanced students.
For advanced students:
Students should understand the organization of genes and operons. They should know consensus sequences that define the position of the promoter and be familiar with the sequence shorthand used by GenBank.

PROCEDURE

Materials.

For introductory students or advanced student review:
Pen or pencil
Sheet(s) containing sequences to be studied

- Answer key for introductory students or advanced student review sequences (PDF)

A table of codon assignments could be provided

For advanced students:
Pencil and highlighter
Printout of selected GenBank nucleotide sequences

- Catalase gene sequence (GenBank #M21516) (PDF)
- Tryptophan synthetase gene sequence (GenBank #V00364) (PDF)

Codon assignment wheel

- Codon assignment wheel
  This wheel may be cut out, laminated, and mounted on cardboard, if desired. Both the three-letter and single-letter amino acid designations for the codon assignments are given.

Student Version.
For introductory students or advanced student review:
1. The following DNA sequences represent hypothetical (and very small) bacterial genes. Note that the sequences that signal start and stop for transcription *have not* been included here for simplicity's sake. Transcribe the DNA sequences into RNAs (label 5' and 3' ends of each transcript) and locate the complete open reading frame (ORF) for translating each message. How many amino acids are contained in the polypeptide that is translated from each message? Circle the translation start and stop codons.

A.

```
5'---CGCGCTGATGTGGCCCTGTGAAGACAAATGTTAATGAGGTCGACTC---3'
3'---GCGCGACTACACCGGGACACTTCTGTTTACAATTACTCCAGCTGAG---5'
```

B.

```
5'---ACAGTTCACTAAAGCGTGAAGCGCATCCGT---3'
3'---TGTCAAGTGATTTCGCACTTCGCGTAGGCA---5'
```

2. Each of the following "microsequences" represents a hypothetical bacterial cell gene. Note that each contains an ORF. Describe the effect of the mutation indicated by the asterisk (*) on the protein product of the gene sequence shown. You may use the codon assignment table to help in answering this question.

```
          *
5'---ATGCGTAGATAA---3'
3'---TACGCATCTATT---5'
          *
```

Substitution (point mutation) of TA base pair by a CG base pair.

```
            *
5'---ATGCGTAGATAA---3'
3'---TACGCATCTATT---5'
            *
```

Deletion of the GC base pair.

```
          **
5'---ATGCGCGACCCATAG---3'
3'---TACGCGCTGGGTATC---5'
          **
```

Insertion of a TA base pair between the CG and GC base pairs.

```
     *
5'---ATGCGTAGATAG---3'
3'---TACGCATCTATC---5'
     *
```

Substitution (point mutation) of a GC base pair for the AT base pair.

```
               *
5'---GTATGCCCGAGTAG---3'
3'---CATACGGGCTCATC---5'
               *
```

Substitution (point mutation) of an AT base pair for the CG base pair.

For advanced students:
1. Using the GenBank sequence given, locate the start signal for translation of the gene and highlight it. Remember that the gene sequence *may* contain the Pribnow and -35 sequences. Circle and label them, if they are present.
2. Locate the ORF. Mark off the codons.
3. Determine the protein sequence.
4. What stop codon(s) determine the end of the translation reading frame? Highlight the stop signal(s).

Safety Issues. None.

ASSESSMENT and OUTCOMES

Suggestions for Assessment.
These exercises can be used in a number of ways. They can be used directly as an assessment of understanding of the concepts of transcription, translation, and mutations. Short sequences can easily be used as exam questions. They are useful also as study problems to be done individually or in small groups either in or out of class with either written instructor feedback or class discussion following completion of the exercises.

Problems and Caveats.
Initially (particularly for introductory students), the sequences should be designed to include only a single ATG translation start codon (when read 5' to 3' on the DNA). More challenging sequences contain internal ATG codons in the ORF and/or ATGs in both strands of DNA. It is probably less confusing for students (at least initially), though, if only one of them will work as a start for a complete ORF (that is, one with a start codon, several codons, and a stop codon all in the same reading frame). Although students can actually write out the amino acid sequence of the translated peptides (using a codon assignment chart for reference), this is not absolutely necessary. For part 1 sequences, the students can simply report the number of amino acids that are in the ORF. The codon assignment table is more useful for part 2 sequences where mutations do not necessarily have any effect on the peptide that is translated. Even in more advanced courses (microbial genetics, molecular biology) these sequences can be useful as review exercises before having students read real gene sequences to identify the ORF encoded within them.

The GenBank sequences will often identify positions of Pribnow and -35 sequences as well as stop and start codons in the accompanying information. A single sequence (rather than a multigene operon) should be used initially as a test sequence. Examples are *E. coli* trpA gene (trp synthetase-GenBank #V00364) or *E. coli* katG gene (catalase-GenBank #M21516). More challenging gene operons include *E. coli* trp operon (GenBank #V00372) or *E. coli* lactose operon, including repressor gene (GenBank #J01636). The GenBank sequences can be obtained at http://www.ncbi.nlm.nih.gov, GenBank Search. This work is suited for individual homework assignments. The longer, more complex sequences could be worked on by small groups.

SUPPLEMENTARY MATERIALS

Possible Modifications.
This exercise can be modified using any sequences desired. These can be short ones designed by the instructor or actual published gene sequences for use by more advanced students. By manipulating the position of potential start signals for translation within the sequences, the exercise can be extended to a discussion on how the ribosome "knows" which to use.

Answer Key

1.

Sequence 1A

The start signal for translation begins at the 8th nucleotide from the left. The resulting mRNA would read:

5' AUG UGG CCC UGU GAA GAC AAA UGU UAAUGA 3'

There are 8 amino acids specified in this peptide. Stop codons UAA and UGA end the message.

Sequence 1B

The start signal for translation begins at the 5th nucleotide from the right. This illustrates that sometimes the sequence must be read from right to left to find the open reading frame. The resulting mRNA would read:

5' AUG CGC UUC ACG CUU UAGUGA 3'

There are 5 amino acids in this peptide with stop codons UAG and UGA ending the message.

2.

Mutation 2A

The original microsequence would be transcribed to read:

5' AUG CGU AGA UAA 3'

The point mutation changes the CGU codon to CGC. Mutations that change the nucleotide in the third position of a codon often result in a silent mutation that does not alter the amino acid specified. In this case the amino acid remains arginine.

Mutation 2B

The original microsequence would be transcribed as:

5' AUG CGU AAU AA 3'

Deletion of the designated GC base pair would alter the reading frame to 5' AUG CGA GAU AA 3' so that different amino acids are specified and so that there is no stop signal.

Mutation 2C

The original sequence would be transcribed as:

5' AUG CGC GAC CCA UAG 3'

Insertion of the TA base pair would result in the following message:

5' AUG CGC UGA CCC AUA G 3'

A premature stop codon (UGA) then occurs in the third position.

Mutation 2D

The original sequence would be transcribed:

5' AUG CGU AGA UAG 3'

The described point mutation produces a GUG beginning codon. There is no f-met start; however, GUG can serve as a start signal for bacteria in some contexts when there is not an AUG in an appropriate position.

Mutation E

The original message reads:

5' AUG CCC GAG UAG 3'

The mutation changes the third codon to GAU. In this case change of the third nucleotide in the codon results in specifying a different amino acid (aspartate instead of glutamate).

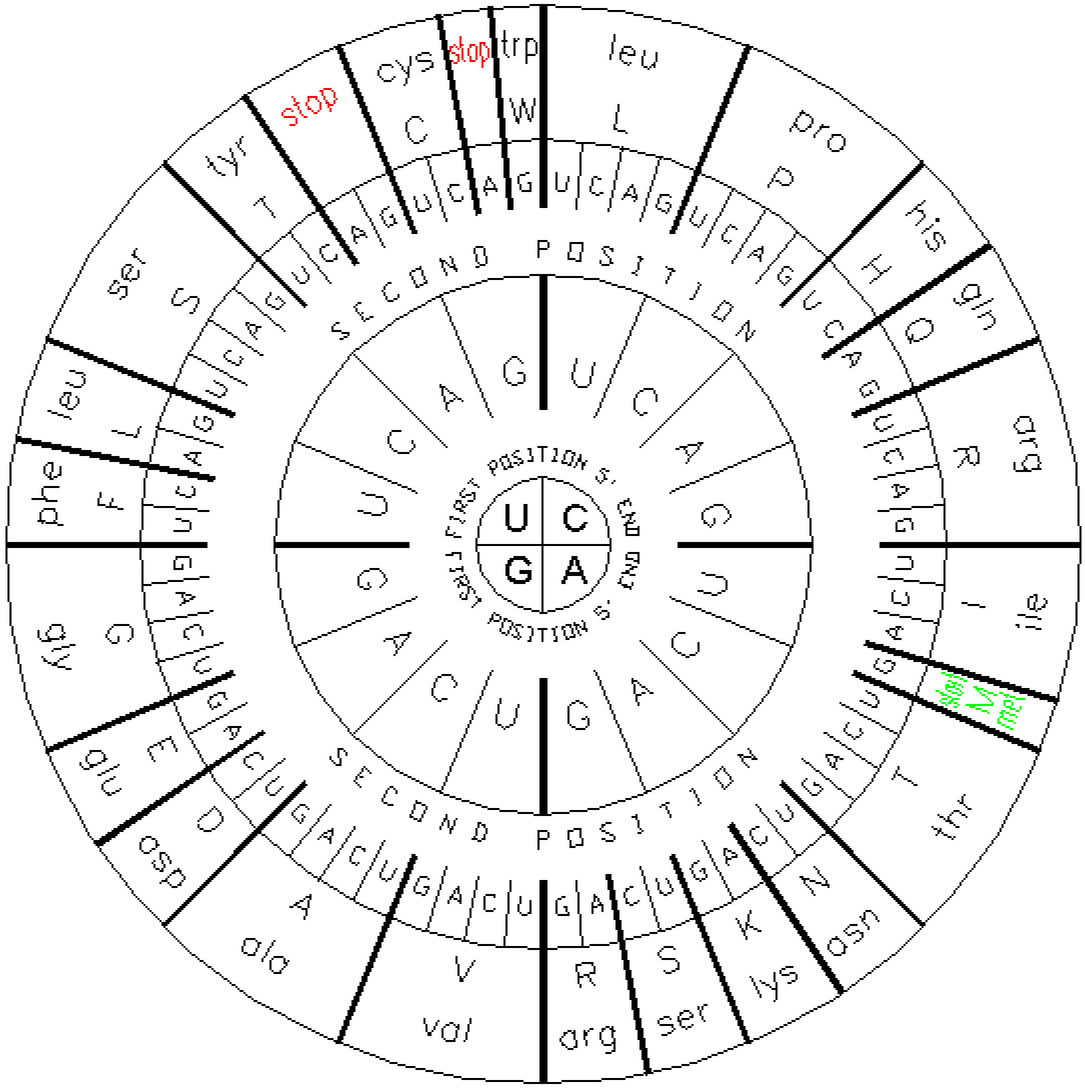
Catalase gene sequence
DEFINITION *E.coli* katG gene encoding catalase HP1, complete cds.
ACCESSION M21516

```
   1 aagcttaatt aagatcaatt tgatctacat ctctttaacc aacaatatgt aagatctcaa
  61 ctatcgcatc cgtggattaa ttcaattata acttctctct aacgctgtgt atcgtaacgg
 121 taacactgta gaggggagca cattgatgag cacgtcagac gatatccata acaccacagc
 181 cactggcaaa tgcccgttcc atcagggcgg tcacgaccag agtgcggggg cgggcacaac
 241 cactcgcgac tggtggccaa atcaacttcg tgttgacctg ttaaaccaac attctaatcg
 301 ttctaaccca ctgggtgagg actttgacta ccgcaaagaa ttcagcaaat tagattacta
 361 cggcctgaaa aaagatctga aagccctgtt gacagaatct caaccgtggt ggccagccga
 421 ctggggcagt tacgccggtc tgtttattcg tatggcctgg cacggcgcgg ggacttaccg
 481 ttcaatcgat ggacgcggtg gcgcgggtcg tggtcagcaa cgttttgcac cgctgaactc
 541 ctggccggat aacgtaagcc tcgataaagc gcgtcgcctg ttgtggccaa tcaaacagaa
 601 atatggtcag aaaatctcct gggccgacct gtttatcctc gcgggtaacg tggcgctaga
 661 aaactccggc ttccgtacct tcggttttgg tgccggtcgt gaagacgtct gggaaccgga
 721 tctggatgtt aactggggtg atgaaaaagc ctggctgact caccgtcatc cggaagcgct
 781 ggcgaaagca ccgctgggtg caaccgagat gggtctgatt tacgttaacc cggaaggccc
 841 ggatcacagc ggcgaaccgc tttctgcggc agcagctatc cgcgcgacct tcggcaacat
 901 gggcatgaac gacgaagaaa ccgtggcgct gattgcgggt ggtcatacgc tgggtaaaac
 961 ccacggtgcc ggtccgacat caaatgtagg tcctgatcca gaagctgcac cgattgaaga
1021 acaaggttta ggttgggcga gcacttacgg cagcggcgtt ggcgcagatg ccattacctc
1081 tggtctggaa gtagtctgga cccagacgcc gacccagtgg agcaactatt tcttcgagaa
1141 cctgttcaag tatgagtggg tacagacccg cagcccggct ggcgcaatcc agttcgaagc
1201 ggtagacgca ccggaaatta tcccggatcc gtttgatccg tcgaagaaac gtaaaccgac
1261 aatgctggtg accgacctga cgctgcgttt tgatcctgag ttcgagaaga tctctcgtcg
1321 tttcctcaac gatccgcagg cgttcaacga agcctttgcc cgtgcctggt tcaaactgac
1381 gcacagggat atggggccga aatctcgcta catcgggccg gaagtgccga aagaagatct
1441 gatctggcaa gatccgctgc cgcagccgat ctacaacccg accgagcagg acattatcga
1501 tctgaaattc gcgattgcgg attctggtct gtctgttagt gagctggtat cggtggcctg
1561 ggcatctgct tctaccttcc gtggtggcga caaacgcggt ggtgccaacg gtgcgcgtct
1621 ggcattaatg ccgcagcgcg actgggatgt gaacgccgca gccgttcgtg ctctgcctgt
1681 tctggagaaa atccagaaag agtctggtaa agcctcgctg gcggatatca tagtgctggc
1741 tggtgtggtt ggtgttgaga aagccgcaag cgccgcaggt ttgagcattc atgtaccgtt
1801 tgcgccgggt cgcgttgatg cgcgtcagga tcagactgac attgagatgt ttgagctgct
1861 ggagccaatt gctgacggtt tccgtaacta tcgcgctcgt ctggacgttt ccaccaccga
1921 gtcactgctg atcgacaaag cacagcaact gacgctgacc gcgccggaaa tgactgcgct
1981 ggtgggcggc atgcgtgtac tgggtggcaa cttcgatggc agcaaaaacg gcgtcttcac
2041 tgaccgcgtt ggcgtattga gcaatgactt cttcgtgaac ttgctggata tgcgttacga
2101 gtggaaagcg accgacgaat cgaaagagct gttcgaaggc cgtgaccgtg aaaccggcga
2161 agtgaaattt acggccagcc gtgcggatct ggtgtttggt tctaactccg tcctgcgtgc
2221 ggtggcggaa gtttacgcca gtagcgatgc ccacgagaag tttgttaaag acttcgtggc
2281 ggcatgggtg aaagtgatga acctcgaccg tttcgacctg ctgtaatctg accccgttca
2341 gcggctgctt gctggcagtc gctgaacgtt ctttaccagc gtatagtggg cgaacgaaaa
2401 ctacacactg gatctctcat gtctgccgca ggaaagagca acccactggc aatcagtggc
2461 ctggttgtgc tcacacttat ctggagttat agctggattt catgaagcaa gtcaccagtt
2521 acatcggtgc cttcgacttt accgccttac gctgcatttt cggcgctctc gttttattca
2581 tcgtcctttt attacgtggt cgcggaatgc gcccgacacc gtttaaatac accttagcca
2641 ttgccctgtt acaaacctgc gggatggttg gtctggcgca gtgggcgttg gtcagcggag
2701 gtgcggggaa ggtggcgatc ctgagctata ccatgccgtt ctgggtggtg attttcgccg
2761 cgttgtttct cggtgaacgc ctgcgacctg ggcaatattt cgcga
//
```

Tryptophan synthetase gene sequence

DEFINITION E. coli trpA gene (codes for tryptophan synthetase alpha-SU).
ACCESSION   V00364

```
   1 aacctttccg gtcgcggaga taaagacatc ttcaccgttc acgatatttt gaaagcacga
  61 ggggaaatct gatggaacgc tacgaatctc tgtttgccca gttgaaggag cgcaaagaag
 121 gcgcattcgt tcctttcgtc acgctcggtg atccgggcat tgagcagtca ttgaaaatta
 181 tcgatacgct aattgaagcc ggtgctgacg cgctggagtt aggtatcccc ttctccgacc
 241 cactggcgga tggcccgacg attcaaaacg ccactctgcg cgcctttgcg gcaggtgtga
 301 ctccggcaca atgttttgaa atgctggcac tgattcgcca gaaacacccg accattccca
 361 ttggcctgtt gatgtatgcc aatctggtgt ttaacaaagg cattgatgag ttttatgccc
 421 agtgcgaaaa agtcggcgtc gattcggtgc tggttgccga tgtgccagtt gaagagtccg
 481 cgcccttccg ccaggccgcg ttgcgtcaca acgtcgcacc tatcttcatc tgcccgccaa
 541 atgccgatga cgacctgctg cgccagatag cctcttacgg tcgtggttac acctatttgc
 601 tgtcacgagc aggcgtgacc ggcgcagaaa accgcgccgc gttacccctc aatcatctgg
 661 ttgcgaagct gaaagagtac aacgctgcac ctccattgca gggatttggt atttccgccc
 721 cggatcaggt aaaagcagcg attgatgcag gagctgcggg cgcgatttct ggttcggcca
 781 ttgttaaaat catcgagcaa catattaatg agccagagaa aatgctggcg gcactgaaag
 841 tttttgtaca accgatgaaa gcggcgacgc gcagttaatc ccacagccgc cagttccgct
 901 ggcggcattt taactttctt taatgaagcc ggaaaaatcc taaattcatt taatatttat
 961 ctttttaccg tttcgcttac cccggtcgat cgtyractta cgtcattttt ccgcccaaca
1021 gtaatataaa caaacaaatt aaacccgcaa cataacacca gtaaaatcaa taattttctc
1081 taagtcactt attcctcagg taattcttaa tatatccaga atgttcctca aaatatattt
1141 tccctctatc ttctcgttgc gcttaatttg actaattctc attagcgact aattttaatg
1201 agtgtcgaca cacaacactc atattaatga aacaatgcaa cgcaacggga gaaataacat
1261 ggccgaacat cgtggtggtt caggaaattt cgccgaagac cgtgagaagg catccgacgc
1321 agccgtaaag gcggtcagca tagcggcggt aattttaaaa atgatcgcaa cgcgcatctg
1381 aagcgggtaa aaaaggcggt yrac
```


FIRST POSITION 5' END
SECOND POSITION
U
C
A
G
leu
L
pro
P
his
H
gln
Q
arg
R
ile
I
start
M
met
thr
T
asn
N
lys
K
ser
S
arg
R
val
V
ala
A
asp
D
glu
E
gly
G
phe
F
leu
L
ser
S
tyr
T
stop
cys
C
stop
trp
W