

Conceptual Translation

Atg/M = start codon (beginning of translated protein region)

Taa/* = stop codon (end of translated protein region)

-- = exon/exon boundary

aataaa = polyadenylation signal

aggagg = shine-dalgarno sequence (ribosomal binding site)⁸

aaa = known upstream in-frame stop codon

agctt = promoter region

> - 8123 nucleotides

DNA: agctttttgacttgtaaaagtgggaaggtttcaaagggacttgtcagtactgctctgtg
+3fr: L F D L * K W * R F Q R D L S V L L C G

DNA: gaatggttggacttattttgcatcccccttttcttgtaaagaatcgagattcaccgagctg
+3fr: M V G L I L H P L F L * R I E I H R A E
60

DNA: agacttttctctttaagacagctagccatccacttttctccactgctaattgaccagtgg
+3fr: T F L F K T A S H P L S S T A N * P V V
120

DNA: tcgagcatccataggatttctccccagtgtgaaaggctttattttaaggcaagtgagac
+3fr: E H P * D F S P V * K A L F * R Q V R L
180

DNA: tgcttcaaataaaacaacttcaagcttccaagaaacagttaagaggaggcagagaagagc
+3fr: L Q I K Q L Q A S K K Q L R G G R E E Q
240

DNA: agaaacacttctttgctgacacttacactgttgccatggacctacataagcagtgaggaga
+3fr: K H F F A D T Y T V A * D L H K Q W E N
300

DNA: acacagagactaactggcataaggaaaagatggaattactggaccagtttgacaatgaaa
+3fr: T E T N W H K E K M E L L D Q F D N E R
360

DNA: gaaaggaatgggaaagtcaatggaagattatgcagaagaaaatagaagagctttgcccggg
+3fr: K E W E S Q W K I M Q K K I E E L C R E
420

DNA: aagtaaagctttggaggaaaatcaatatcaatgaaagtgctaagatcattgatctttacc
+3fr: V K L W R K I N I N E S A K I I D L Y H
480

DNA: atgagaagaccattccagagaaagtgatagaatcttcccaaattaccccgatttaggac
+3fr: E K T I P E K V I E S S P N Y P D L G Q

540

DNA: aaagtgaatttataaggacgaatcacaaagatggtctgagaaaagaaataaaagagagc
+3fr: S E F I R T N H K D G L R K E N K R E Q
600

DNA: agagcttagtcagtgaggaaatcaaagtgtgaaggaacaaaaagcaacaaaaaatcaa
+3fr: S L V S G G N Q M C K E Q K A T K K S K
660

DNA: aagtaggggttttggatcctttggctacagacaacccaaaaggaatgtgaggcctggcctg
+3fr: V G F L D P L A T D N Q K E C E A W P D
720

DNA: acctgaggacttctgaggaagacagcaagagctgttctggcgccctcagtacagctcttg
+3fr: L R T S E E D S K S C S G A L S T A L E
780

DNA: aagaacttgcaaggtgagtgagaattatgcagctttcaagaggaaattcgaaagcggg
+3fr: E L A K V S E E L C S F Q E E I R K R S
840

DNA: ctaaccatagaaggatgaagtcagattcttttctccaggaaatgccaaatgtaactaata
+3fr: N H R R M K S D S F L Q E M P N V T N I
900

DNA: tacctcatggggaccccatgatcaacaatgaccagtgacattcttccaatcagtttagaaa
+3fr: P H G D P M I N N D Q C I L P I S L E K
960

DNA: aagaaaaacagaaaaacaggaagaatctgagctgtaccaatgtgctccagagcaattcta
+3fr: E K Q K N R K N L S C T N V L Q S N S T
1020

DNA: cgaaaaaatgtggaattgatacaatcgatttaaaaagaaatgaaactccaccagttcctc
+3fr: K K C G I D T I D L K R N E T P P V P P
1080

DNA: ctccaagaagcacctctcgaaattttcccagctcggattctgaacaagcctatgaaagat
+3fr: P R S T S R N F P S S D S E Q A Y E R W
1140

DNA: ggaaggaaaggttagaccacaacagctgggtgccccatgaggggtcgaagtaaaaggaatt
+3fr: K E R L D H N S W V P H E G R S K R N Y
1200

DNA: acaaccctcacttccctttgagacaacaagagatgtctatgttgtatccaaatgaagggga
+3fr: N P H F P L R Q Q E M S M L Y P N E G K
1260

DNA: aaacttcgaaagatgggtatcatcttttctcttttgggtaccagaagtcaaatagatagca
+3fr: T S K D G I I F S S L V P E V K I D S K
1320

DNA: agcctccaagtaatgaagatggttgacttagcatgtggatcatgtgacattgggataggtg
+3fr: P P S N E D V G L S M W S C D I G I G A
1380

DNA: caaaaaggagcccctctacttctggtggtttcagaaaacctgctctacccccagtaatccaa
+3fr: K R S P S T S W F Q K T C S T P S N P K
1440

DNA: aatatgaaatggatgatcccagatcacccctgctaaatctcatcctgatcttcatgtaagta
+3fr: Y E M V I P D H P A K S H P D L H V S N
1500

DNA: atgactgtagctcctcagtagcagagagcagtagcccacttagaaaatttcagttgtggct
+3fr: D C S S S V A E S S S P L R N F S C G F
1560

DNA: ttgaaaggactacaaggaatgagaagctggcagcaaagactgatgaatttaacagaactg
+3fr: E R T T R N E K L A A K T D E F N R T V
1620

DNA: tatttagaacagatagaaattgtcaggcaatacagcaaaatcacagctgctcaaaatcat
+3fr: F R T D R N C Q A I Q Q N H S C S K S S
1680

DNA: cggaggatctcaagccctgtgatacctcatctactcacacaggtagcatatcaciaaagta
+3fr: E D L K P C D T S S T H T G S I S Q S N
1740

DNA: acgatgtgtccggatatttggaaccacatgccacatgcctgtgcccattggaaaatgtgc
+3fr: D V S G I W K T N A H M P V P M E N V P
1800

DNA: ctgataatcccaccaagaaatccacaacaggcctagtaagacaaatgcagggacacctaa
+3fr: D N P T K K S T T G L V R Q M Q G H L S
1860

DNA: gtctctgcagttatcgaaatgatgctccacgagcatgactggagaccgagtaatttgtctg
+3fr: P R S Y R N M L H E H D W R P S N L S G
1920

DNA: gccgtccgaggtcagctgatcccaggtcaaattatggtggttggtggaaaagctgctgaaaa
+3fr: R P R S A D P R S N Y G V V E K L L K T
1980

DNA: cctatgagacagcaacagagtctgcattgcaaaattctaagtgcttccaggataattgga
+3fr: Y E T A T E S A L Q N S K C F Q D N W T
2040

DNA: ccaaagtgaattctgatgtcagtggtggtgcccacattaagtgcagcatttagaaatgctcc
+3fr: K C N S D V S G G A T L S Q H L E M L Q
2100

DNA: aaatggaacaacagtttcagcaaaagacagctgtgtgggggggacaggaagtgaagcaag
+3fr: M E Q Q F Q Q K T A V W G G Q E V K Q G
2160

DNA: gaatagatccgaaaaagataaacagaggaaatccatgtcagtgaaacgcctcacatggaaaag
+3fr: I D P K K I T E E S M S V N A S H G K G
2220

DNA: gattttcccgacctgctagaccagcaaatacgtcgtctcccctccagatgggcatccagat
+3fr: F S R P A R P A N R R L P S R W A S R S
2280

DNA: ctccatctgcacccccctgccttgcgaggagaactaccacaactataaccatttctctgcat
+3fr: P S A P P A L R R T T H N Y T I S L R S
2340

DNA: ccgaagcattgatggtttaagtctttggcctggattgctatattacagaagttctagtc
+3fr: E A L M V * V F G L D C Y I T E V L V P
2400 (stop @ 2420)