

Key for Test Your Understanding Questions, Chapter 9

1. DNA is double-stranded, and one strand may serve as the template (copied) strand for one gene (in one region) but the non-template (mRNA-like) strand for another (in another region). The algorithm above could find an ATG start codon in one of three reading frames by reading a sequence entered in the 5' to 3' direction, but really we should consider all six possible reading frames: three from the DNA as it was entered and three more on the complementary strand. What would we need to do to find ORFs in all six possible reading frames?

One solution would be to run the algorithm once on the DNA strand as entered, then get the reverse complement of the DNA and run the algorithm again. Or, you could reverse-complement the pattern (so CAT rather than ATG) and do another search with the same DNA string. The former method is probably better in the long run, as the user will be more comfortable with sequences that look like mRNA as output.

2. As noted above, the pattern-matching algorithm might find an ORF within another ORF, because within a gene there could be multiple ATG codons. How could your algorithm filter out these undesirable matches?

Several solutions are possible, depending on whether you want to allow for overlapping ORFs (two ORFs that overlap but are not in the same reading frame; short overlaps are not all that uncommon in bacteria and viruses may have long overlaps).

To disallow any overlaps, once an ORF is found, you could simply move the start point of the search to the nucleotide following the stop codon of the ORF. To allow overlaps in a different frame, you could take stock of the ATGs within an ORF and keep a list of those in-frame that you should not be considered valid start points for another ORF.

3. Identify parameters that could be used in the pattern-matching algorithm to search for a Shine-Dalgarno sequence once an ATG is found. Assume an exact match to the consensus sequence.

pattern = AGGAGG

start = 13 nt before the ATG (allowing for 6-nt pattern + max 7 nt spacing)

stop = 9 nt before the ATG (allowing for min 3 nt spacing)

increment = 1

threshold = 100%

4. Identify parameters that could be used in the pattern-matching algorithm to search for a promoter once an ATG is found. Assume that 5 of the 6 bases in the -10 and -35 sequences must match their consensus.

First, look for the -10:

pattern = TATAAT

start = 500 nt before ATG

stop = 50 nt before ATG

increment = 1

threshold = 83%

Then, look for -35 relative to -10:

pattern = TTGACA

start = 25 nt before -10

stop = 21 nt before -10

increment = 1

threshold = 83%