Exploring Bioinformatics: A Project-Based Approach

Key for Test Your Understanding Questions, Chapter 11

1. Find an α-helix in the sequence N-MDGPDFWEAMKRISTQTYSNGHKMPS-C using the Chou-Fasman rules.

   The Chou-Fasman rules direct that first a region of at least six contiguous residues with P(a)>103 be found. This occurs starting with amino acid #6, FWEAMK. Next, we try to extend the region until a set of four contiguous residues with P(a) < 100 is found. Going leftward, this takes us to the start of the peptide. Going rightward, we find the four at TYSN. This leaves MDGPDFWEAMKRISTQ as our potential α-helix. The average P(a) for this region is 107, which is greater than 103, and the sum of P(a), 1711, is greater than the sum of P(b), 1474. Thus, this region qualifies as an α-helix by the Chou-Fasman criteria.

2. Examine the Chou-Fasman rules carefully, and look at the P(a) and P(b) values for various amino acids in Table 11.1. What can you see that might reduce the ability of this algorithm to clearly distinguish between α-helices and β-sheets?

   Many amino acids that have P(a) above the initial cutoff of 103 also have P(b) above the initial cutoff of 105. Where a number of these amino acids occur, the averages and sums of P(a) and P(b) may be very close, reducing the ability of the algorithm to discriminate between the two.

3. How do we define a β-turn in a protein structure? Given this definition, can you think of a simple rule you could add to the algorithm for identification of β-turns that might increase its accuracy?

   A β-turn is a loop that occurs between two adjacent β-strands that will interact to form part of a β-sheet. We could therefore identify potential β-strands first and then require that a potential β-turn be located between two of them.

4. Would it improve the predictive ability of the algorithm to specify that a region should be identified as a β-strand only if it is either preceded or followed by a β-turn? Why or why not?

   This is essentially the converse of the answer to #3, but this will not work nearly as well. A may be composed of β-strands that are folded together from various regions of the protein; they do not have to follow one another in the sequence. So, an actual β-strand may not have a β-turn before or after it.

5. Proteins that are part of the cell membrane or an organelle membrane typically have one or several α-helical domains about 20 amino acids long that pass through the membrane. These membrane-spanning helices consist almost entirely of very hydrophobic amino acids such as L, I, V, F and W and are anchored in place by hydrophilic amino acids on their two ends. If you applied the Chou-Fasman algorithm

to a membrane protein, why would it likely fail to predict the membrane-spanning helices?

Although these hydrophobic amino acids do have high P(a) values, their P(b) values are even higher. Thus, a region of the protein where these amino acids are very common would most likely be identified as a β-strand rather than an α-helix according to the Chou-Fasman rules.