I. You are trying to digest the protein of a South American bean in order to alter some of its functional properties. You choose the protease trypsin as the enzyme you would like to use to convert the bulk bean proteins to smaller peptides. Since you are interested in altering the molecular size of the proteins, you use size exclusion chromatography and SDS gel electrophoresis as your analytical tools. You run extracts of the raw beans on a Sephadex G-100 column and on an SDS gel. Then you incubate a good dose of trypsin with the original extract, then perform the chromatography and electrophoresis, again. You expect to see that new, smaller peptides appear after treatment with trypsin, which you ought to detect on the gels and by chromatography. To your amazement, essentially nothing happened during the treatment with trypsin. To try to determine why your experiment failed, you examine the gels and the elution profile more closely.

1) The SDS gels before and after treatment with trypsin look very similar except that there is a small peak where you expect trypsin to appear (so you didn't forget to add it, and you apparently added quite a large amount). The Sephadex G-100 column tells a slightly different story. This column packing is a size exclusion medium that is appropriate for globular proteins of molecular weight between 5 and \(\sim\)100 KDa. The column has a diameter of 1.5 cm and is packed to a height of 50 cm. When you run blue dextran over the column, its elution volume is 26.5 mL.

   a) Comparing the chromatograms of the original and treated samples you find a new peak in the treated sample with an elution volume of 61.6 mL. What is the Kav value for this molecule?

   b) You calibrate the column by running protein standards over the same column and obtain the following data:

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mr</th>
<th>Rs (Å)</th>
<th>Ve (mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New peak</td>
<td></td>
<td></td>
<td>61.6</td>
</tr>
<tr>
<td>Std A</td>
<td>6,670</td>
<td>11</td>
<td>78.7</td>
</tr>
<tr>
<td>Std B</td>
<td>12,500</td>
<td>14.4</td>
<td>74.9</td>
</tr>
<tr>
<td>Std C</td>
<td>13,600</td>
<td>14.5</td>
<td>74.8</td>
</tr>
<tr>
<td>Std D</td>
<td>14,320</td>
<td>17.4</td>
<td>71.5</td>
</tr>
<tr>
<td>Std E</td>
<td>33,600</td>
<td>24.7</td>
<td>63.25</td>
</tr>
<tr>
<td>Std F</td>
<td>35,000</td>
<td>21.6</td>
<td>66.7</td>
</tr>
<tr>
<td>Std G</td>
<td>45,800</td>
<td>24.7</td>
<td>63.3</td>
</tr>
<tr>
<td>Std H</td>
<td>67,900</td>
<td>30</td>
<td>57.8</td>
</tr>
</tbody>
</table>

   Using these data, estimate the molecular weights of the protein.

   c) Using these data, calculate the Stoke's radius of the protein.
2) You pool the fractions from the G-200 column that contain the new peak. You attempt to
purify the components more fully by using ion exchange chromatography. You quickly
find that most of the protein does not bind to DEAE cellulose at pH 8.2; most of the
protein runs through the column and appears in the void volume. Elution of the column
with a linear gradient of NaCl (from 0.01M to 0.8M, in 50 mM tris buffer, pH 8.2)
produces an ugly elution profile with many small peaks. The material from the void
volume does, however, adsorb to carboxymethyl cellulose at pH 6.2. Upon elution with a
linear gradient of NaCl (from 0.01M to 0.8M, 50 mM MES buffer, pH 6.2), you obtain a
single major peak. Interpret the results of the ion exchange chromatography experiments
(hint: it's very simple).

3) You want a second measurement of the molecular weight of this new protein, so you use
SDS gel electrophoresis. Since you want to calculate the molecular weight of the unknown,
you run an assortment of molecular weight standards (or markers) in lanes adjacent to the
lanes containing the unknown. To your surprise, you observe two protein bands in the
lanes containing the unknowns. You measure relative mobilities and obtain the following
data:

<table>
<thead>
<tr>
<th>Protein</th>
<th>Mr</th>
<th>Rm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown 1</td>
<td></td>
<td>0.52</td>
</tr>
<tr>
<td>Unknown 2</td>
<td></td>
<td>0.90</td>
</tr>
<tr>
<td>phosphorylase b</td>
<td>97,000</td>
<td>0.10</td>
</tr>
<tr>
<td>BSA</td>
<td>67,000</td>
<td>0.21</td>
</tr>
<tr>
<td>ovalbumin</td>
<td>45,000</td>
<td>0.33</td>
</tr>
<tr>
<td>carbonic anhydrase</td>
<td>30,000</td>
<td>0.45</td>
</tr>
<tr>
<td>lysozyme</td>
<td>14,400</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Fill in the molecular weights of the unknowns (yes, one of them is outside of the
standards, but assume that it is correct, anyway - just for this problem).

4) You are concerned about the purity of the protein because you saw two peaks on SDS
PAGE. You decide to determine the N-terminal amino acid as an additional check on
purity. You allow the protein to react with fluorodinitrobenzene. Luckily, the N-terminus
was not blocked and the reaction proceeded as planned. You hydrolyze the protein in 6N
HCl, dry the sample down, spot it on a TLC plate and do a two dimensional TLC. You
find three yellow spots, but one corresponds to \(\epsilon\)-labeled lysine, which you can ignore. The
other spots correspond to \(\alpha\)-DNP Arginine and \(\alpha\)-DNP-phenylalanine and are of equal
intensity. Phe is the N-terminal amino acid of the particular kind of trypsin you used.

   a) Discuss the possible interpretations concerning the purity of the protein.
b) How can you interpret the data so far?

5) In light of your interpretation of the results you have, you re-examine the G-100 chromatogram. You decide that at least one of the very low molecular weight proteins that appears in the chromatogram before treatment with trypsin was of lower intensity after addition of trypsin. You make more extract and run it over a Sephadex G-50 column, which gives better separation than G-100 among very small proteins (with larger proteins appearing in the void volume). You find that one of the protein bands that elute from the column is able to inhibit trypsin, so you can use inhibition as an assay for the presence of the protein. You continue to purify this protein, using ion exchange chromatography and your inhibition assay. When you are finished, you establish that the protein has an N-terminal Arg, has the same electrophoretic mobility on SDS gels as one of the unknown components you identified before, and is about 85% pure.

You submit the sample to sedimentation velocity ultracentrifugation. Corrected to 20° and water, and extrapolated to zero concentration, the sedimentation coefficient is $s_{20,w}^0 = 2.9 \times 10^{-13}$ s.

a) Assuming the viscosity of water to be 1.011 cp (convert to poise for the calculation), the density of water to be 0.998 g/cm$^3$, the partial specific volume of the protein to be 0.7320, calculate the molecular weight of the protein.

b) Using all the data you have so far, calculate the frictional ratio $f/f_0$ for the protein.

c) Using all the data you have so far, estimate the axial ratio $a/b$ for the protein, assuming it to be a prolate ellipsoid.
II. The amino acid sequences of three proteins are given below. Refer to them to answer the following questions.

6) Pick out regions of the sequence of protein A that might be expected to exist as an \( \alpha \)-helix.

7) Pick out regions of the sequence of protein A that might be expected to exist as a \( \beta \)-strand.

8) Assuming protein B exists as an \( \alpha \)-helix, does the helix have a strong hydrophobic moment? What does this suggest about the way it might pack in a folded protein? Show your work or supporting calculations.

9) Protein C is part of an integral membrane protein. Suggest in a general way what its structure might be. Explain your answer.

Sequences:

Protein A
NENWVVSAAH CYKSRVEVRL GEHNIKVTEG SEQFISSSRV IRHPNYSSYN

Protein B
EALAK IFARL AELQT HIAMA

Protein C
MVPFG GEQKK FYAIT TLVPA IAFTM YLSML LGYGL GYDAD SKAES

III. The following questions refer to five Brookhaven datasets that are listed under your name on the web page.

1) Is there a protein in this group that contains a \( \beta \)-barrel domain?
   Is it parallel, antiparallel or mixed? (If there is more than one, answer for both).

2) Which protein is a twisted sheet protein?
   Is it an \( \alpha/\beta \) sheet, or another kind?
   Is the sheet parallel, antiparallel or mixed?

3) Which protein is largely a helical protein?

4) Classify the proteins that you have not yet discussed into one of the following groups:
   \( \alpha \)-helix
   \( \beta/\alpha/\beta \) structure
   other \( \beta \)-sheet
   other \( \beta \) barrel
   mixed \( \alpha + \beta \)
IV. You will be given a nucleotide sequence as though you have just sequenced a gene (the sequence will be in the same directory as your unknown proteins). Answer the following questions about your sequence. We will not worry about molecular biological details such as ribosomal binding sites, etc.

1) What is the most likely amino acid sequence of the protein encoded by this gene?

2) Generate a hydropathy profile of the deduced protein sequence and discuss it. Is it likely to be a transmembrane protein?

3) a. Can you find any close relatives of your protein by looking for similar sequences?

   b. Comment on the significance of homologies or homologous proteins. (Some of these are actually in the database, so you will find its sequence. This question asks about the relatives of the protein, not its identity.)

   c. Can you suggest a probable function for this protein?