1. A) RNA secondary structures are often determined by a technique called “phylogenetic analysis.” Explain how this technique works and what is needed (See Jaeger et al., Annu. Rev. Biochem. (1993), 62, 255-87).

B) Michael Zuker created a secondary structure prediction algorithm, called mfold, based in part on the thermodynamic parameters determined in Douglas Turner’s lab. The appropriate reference is:


The mfold server is located at: [http://www.bioinfo.rpi.edu/~zukerm/](http://www.bioinfo.rpi.edu/~zukerm/)

Telomerase is responsible for elongation of telomeres. It is composed of RNA and protein components. The following figure is a phylogenetically determined secondary structure model of a domain of telomerase RNA (Chen et. al, Cell, 100, 503-514 (2000)). Highly conserved bases are shown in red.
CR4-CR5 Domain Primary sequence:

5’
CCCGCCUGGAGGCCGCGGUCGGCCCGGGCUUCUCCGGAGGCACCCACUGCCACCG
GCGAAGAGAUUGGGCUCUGUCAGCCGCGGG 3’

Input the sequence into the mfold server and view the predicted structures as jpg files. How well do the thermodynamically predicted structures agree with the phylogenetically determined structure? Based on both the phylogenetically and thermodynamically determined structures, what do you think would be the most likely structure and why? Considering the data regarding the relationship between stability of a loop and its length, do you believe that the loop regions in the phylogenetically determined structure are reasonable?

C) It has been proposed that the telomerase protein interacts with this region of the telomerase RNA. Where do you think the protein would most likely bind? Explain why, taking into account the differences between RNA helices and DNA helices.

D) You have cloned a small RNA from six different organisms. Using phylogenetic analysis, propose a putative secondary structure for this RNA. (Hint: It may help to use mfold as a starting point).

Human: 5’ GCGCGAUUCCCUGAGCUGUGGGACGUGCAC 3’
Shark: 5’ UGCGUCUCCCCUGAGCAUUGGACCCGCA 3’
TreeShrew: 5’ GUGCGCGUCCCUGAGCUGUUGGGACUUGCAC 3’
Chinchilla: 5’ GGGCGCUUCUCUGAGCUUGGGACCGUGCC 3’
Horse: 5’ GUGCGCUUCCCUGAGCUGGGAUUGCCGAC 3’
Ferret: 5’ GUGCGCUUUCCCUGAGCUGGGAUACUCAC 3’

E) Which RNA structure proposed by mfold is the most unlike the phylogenetic consensus structure? Propose another structure that is more like the consensus structure.

2. The GNRA tetraloop where N= any nucleotide, R = purine (A,G) is an unusually stable RNA tetraloop. The unusual stability is hypothesized to be due to a hydrogen bond between the N7 of the purine to the 2’ hydroxyl of the G as well as another hydrogen bond from the G amino to the A phosphate (see figure below).

SantaLucia et. al. (Science, 1992, 256, 217-219) synthesized a number of mutant hairpins and determined their effects upon the stability by looking at melting curves of these hairpins. Their results are tabulated below:
<table>
<thead>
<tr>
<th>Hairpin</th>
<th>ΔGo 70°C, kcal/mol</th>
<th>ΔGo 37 °C, kcal/mol</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGCGCAAGCC</td>
<td>-0.09 ± 0.05</td>
<td>-3.00 ± 0.23</td>
<td>71.0</td>
</tr>
<tr>
<td>GGCICAAGCC</td>
<td>0.56 ± 0.05</td>
<td>-2.46 ± 0.17</td>
<td>63.9</td>
</tr>
<tr>
<td>GGC(dG)CAAGCC</td>
<td>0.19 ± 0.03</td>
<td>-3.00 ± 0.06</td>
<td>68.1</td>
</tr>
<tr>
<td>GGCPCAGCC</td>
<td>0.30 ± 0.04</td>
<td>-2.72 ± 0.15</td>
<td>66.8</td>
</tr>
<tr>
<td>GGC2CAAGCC</td>
<td>0.39 ± 0.11</td>
<td>-2.33 ± 0.25</td>
<td>65.2</td>
</tr>
</tbody>
</table>

Where I is inosine, P is purine, 2 is 2-aminopurine, dG is deoxyguanosine.

A. What effect would each substitution have on the hydrogen bonding interactions (i.e. what interaction would each substitution remove) if the model were true? What would be the expected effect on the melting temperature?

B. Do the data agree with the hypothesis for stability? Why or why not?

C. What are the limitations of using this type of mutational analysis to determine if these hydrogen bonds are responsible for the hairpin’s unusual stability?

3.

You have obtained the following melting curve for a 15 nt duplex DNA (molecule A).

A) List three or more assumptions that are normally made when analyzing a melting curve. When would these assumptions not be valid (give examples)?

B) Label the regions where you would expect the DNA to be 1) single stranded 2) double stranded 3) Label the Tm.
C) What are several ways of determining \( \Delta H \)? Which of these ways are better? Why?

D) You perform a temperature melt on another DNA (molecule B) at the same total strand concentration. Surprisingly, molecule B has the same Tm as molecule A, and is more folded at temperatures above the Tm, relative to molecule A. Show graphically how this is possible. What does this say about the thermodynamic parameters?

E) You perform melts on additional DNA sequences, which all have similar Tm’s to the molecules above. As a whole, at what temperatures would a model constructed from this data set be best? Would it be useful for predicting parameters at room temperature (20 °C)? Without redesigning these DNA sequences, how could you modify the conditions of the experiment such that this data set would be useful at higher and/or lower temperatures?

F) Recall that for a hairpin:

\[
K_{eq} = \frac{(1-\theta)}{\theta}
\]

where \( \theta \) represents the fraction folded. Let \( c_T \) to be the total strand concentration.

1) write an equilibrium expression in terms of \( M \) (the single strand) and \( M_2 \) (the double strand) 2) what is \( K_{eq} \) in terms of \( M \) and \( M_2 \)? 3) What is \( [M] \) in terms of \( \theta \) and \( c_T \)? What is \( [M_2] \) in terms of \( \theta \) and \( c_T \)? Show the derivation for \( K_{eq} \) for a double stranded duplex that is self-complementary.

F) Using the data from Frier et al. (PNAS, 83 9373 (1986)) \( \Delta G^\circ(37^\circ C) \), \( \Delta S^\circ \), and \( \Delta H^\circ \) for the following RNA duplexes (remember to check for self-complementarity):

- 5’ GGGCCC 3’
- 3’ CCCGGG 5’
- 5’ GCGUACA 3’
- 3’ CGUGUG 5’

\[\text{5’ GGGCCC 3’} \quad \text{5’ GCGUACA 3’} \]
\[\text{3’ CCCGGG 5’} \quad \text{3’ CGUGUG 5’} \]