1. A) Covariation is based on the idea that there are many degenerate structures that retain function. For example, if a particular helix in an RNA requires only basepairing, then the identity of the bases that make up the helix is not important as long as Watson-Crick basepairing is conserved. Thus, if a C -> A change occurred at one position, the Watson-Crick partner would result in a concomitant compensatory G -> U change; however, this means that if a sequence were 100% conserved over the data set, there would be no covariation and this region would not be predicted to be base-paired by phylogenetic analysis. In order to perform a phylogenetic analysis, homologous RNAs are needed.

B) In terms of preservation of basepairing, the thermodynamic and phylogenetic structures agree fairly well since only the last structure does not retain the pairing in the P6 region. However, in the loop regions, the structures are different as the J5/6, J6/5, and L6 regions are open and unpaired in the phylogenetic structure while they are paired in the thermodynamic structure. This makes sense since it is entropically unfavorable for an RNA to have large unpaired regions, i.e. the loop sizes of these regions are too large in the phylogenetically determined structure and will probably pair or fold into some structure which would reduce the overall loop size. The phylogenetic structure is probably more biologically relevant and is the most likely structure in terms of the paired regions. However, the large loops in the phylogenetic structure are not very likely to exist, thus one of the thermodynamic structures which preserves the phylogenetic pairing but pairs these loops is probably the most likely RNA structure.

C) The protein will probably not bind to helical regions since the major groove of RNA is too narrow and deep. This would sterically prohibit the protein from interacting with the major groove. Areas of local structural deformations would be more likely especially hairpins, internal loops, and possibly junctions. These non-helical regions would result in a wider major groove allowing the protein access to potential interactions. Also, the protein could possibly bind to the minor groove of the helices (alternative answer). The protein would also probably bind to regions of high sequence conservation.

2. The different substituted bases are shown below:
—G,C,A,A— is the reference, Wildtype hairpin. In the context of this discussion, the loop nucleotides will be referred to as 1-4, i.e., G1, C2, A3, A4.

—ICA— Inosine is substituted for G1. This removes the exocyclic amino group which is involved in hydrogen bonds to the N7 of A4 as well as hydrogen bonds to the 5’ phosphate of A4. This should result in a lower melting temperature.

—dGCAA— Deoxyguanosine is substituted for G1. This removes the 2’ hydroxyl that hydrogen bonds to the exocyclic amino group of A4. This substitution should destabilize the loop, lowering the melting temperature.

—GCAP— Purine is substituted for A4. This removes the exocyclic amino group of A4 that hydrogen bonds to the 2’OH of G1 and N3 of G1. This should destabilize the hairpin, lowering the Tm.

—2CAA— 2-aminopurine is substituted for G1. This removes the carbonyl oxygen of G1. To a first approximation, this substitution does not remove any direct hydrogen in the loop and consequently should have no effect on the stability nor on the melting temperature. However, the substitution removes the G imino proton that is involved in a water-mediated hydrogen bond, lowering the Tm.
B. A hydrogen bond is energetically worth in the neighborhood of 1 ± 0.5 kcal/mol. The ΔΔG (at 70°C) of the substitutions are shown below:

<table>
<thead>
<tr>
<th>Hairpin (RNA)</th>
<th>ΔΔG° 70°C, kcal/mol</th>
<th>ΔΔG° 70°C, kcal/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCCGCAAGCC</td>
<td>-0.09 ± 0.05</td>
<td>-</td>
</tr>
<tr>
<td>GCC1CAAGCC</td>
<td>0.56 ± 0.05</td>
<td>+0.65</td>
</tr>
<tr>
<td>GCC(dG)CAAGCC</td>
<td>0.19 ± 0.03</td>
<td>+0.28</td>
</tr>
<tr>
<td>GCCGAPGCC</td>
<td>0.30 ± 0.04</td>
<td>+0.39</td>
</tr>
<tr>
<td>GCC2CAAGCC</td>
<td>0.39 ± 0.11</td>
<td>+0.48</td>
</tr>
</tbody>
</table>

All the hairpins in this data set are less stable relative to the Wildtype hairpin. However, the ΔΔG’s are less than the expected change in free energy for removal of a hydrogen bond, or are at best, on the lower range of the expected change. Thus, to a first approximation, the data do not strongly support the model that hypothesizes that the unusual stability of this tetraloop is due to these additional hydrogen bonds.

C. Limitations common to this type of mutational analysis does not only apply to RNA. Since a mutation often changes a number of factors (beside the one of interest) one cannot assume that changes in ΔG’s are solely due to the factor of interest. One assumption is that the mutation does not grossly perturb the overall structure and although these substitutions are relatively conservative, one has no guarantee that the overall structure is not affected by the mutation. Second order effects, such as conformational preferences as well as factors such as packing interactions, electrostatics, etc. can also affect the structure.

3. A) 1) Two-state approximation – i.e. the species are in either one of two states and perturbing the temperature affects the population of the two states (all-or-none transition) 2) The lower baseline corresponds to the change of ΔG of the folded (double stranded) DNA 3) The upper baseline corresponds to the change of ΔG of the unfolded (single stranded) DNA 4) The molecule is completely folded (ΔG = 1.0) at the beginning of the experiment 5) The molecule is folded “correctly,” for example a molecule that is designed to be a duplex is actually a duplex, and not a hairpin, etc. 6) The contribution of buffer absorbance over temperature is negligible. 7) ΔCp = 0, i.e. ΔH does not change with temperature
C) Calorimetry (Differential Scanning (DSC) and Isothermal Titration (ITC)), van’t Hoff Analysis. Calorimetry is arguably better since it is a direct measure of $\Delta H$; it is model independent. However, technically the DSC may not be a good measure since interpretation of the data (e.g. establishing good baselines) may be difficult. ITC is more accurate, however you need two binding partners for the experiment; it may not be useful to measure the $\Delta H$ for a hairpin.
The (noisy) melt B is an example of a molecule with the same Tm, but different\ DH.\ DH is smaller for molecule B, relative to molecule A. This causes a broader transition such that at temperatures below the Tm, the fraction folded is greater for molecule A while at temperatures above the Tm, the fraction folded is greater for molecule B.

E) The Tm for this data set is near 70 °C, thus this data set would be best at predicting parameters near 70 °C, possibly from 60-80 °C. This is because $K_{eq} = 1$ at Tm where [folded] = [unfolded] and the van’t Hoff equation is used to obtain $K_{eq}$ at other temperatures, with the assumption that $\Delta H$ will not change with temperature ($\Delta C_p = 0$). However, $\Delta H$ does change with temperature and extrapolation to temperatures far away from Tm will introduce greater error. It would probably not be good at predicting at 20 °C since you need to extrapolate. You could change the salt conditions (although this would also change other things too), strand concentrations, cosolvent, etc., however this will change other parameters also.

F) $M_2$(folded) $\propto$ M(unfolded)

$$K_{eq} = \frac{[M]}{[M_2]}$$

$\propto$ = Fraction folded and $c_T$ = total strand concentration.

$[M_2] = \propto c_T$

$[M] = (1-\propto)c_T$

Thus

$$K_{eq} = \frac{(1-\propto)c_T}{\propto c_T} = \frac{(1-\propto)}{\propto}$$

For a double stranded, self-complementary duplex, let M = single strand and $M_2$= double strand.

$$M + M \propto M_2$$

$$K_{eq} = \frac{[M_2]}{[M]^2}$$

$[M_2] = 1/2\propto c_T$

$[M] = (1-\propto)c_T$

$$K_{eq} = \frac{[M_2]}{[M]^2} = \frac{1/2\propto c_T}{(1-\propto)^2 c_T} = \frac{\propto}{(1-\propto)^2 c_T}$$
G) 5’ GGGCCC 3’
    3’ CCCGGG 5’

= (4) 5’GG 3’ + (1) 5’GC 3’ + init + sym
    3’CC  5’            3’CG  5’

\[ \Delta G = 4(-2.9) + 1(-3.4) + 1(+3.4) + 1(+0.4) = -11.2 \text{ kcal/mol} \]
\[ \Delta H = 4(-12.2) + 1(-14.2) + 1(0) + 1(0) = -63.0 \text{ kcal/mol} \]
\[ \Delta S = 4(-29.7) + 1(-34.9) + 1(-10.8) + 1(-1.4) = +165.9 \text{ cal/K-mol} \]

5’ GCGUACA 3’
3’ CGUGUG 3’

= (1) 5’GC 3’ + (1) 5’CG 3’ + (1) 5’GU 3’ + (1) 5’UA 3’
    3’CG  5’            3’GC  5’              3’UG  5’           3’AU  5’

+ (1) 5’AC 3’ + (1) 5’CA 3’ + init + sym
    3’UG  5’            3’G   5’

\[ \Delta G = 1(-3.4) + 1(-2.0) + 1(-0.5) + 1(-1.1) + 1(-2.1) + 1(-1.7) + 1(+3.4) + 0 \]

= -7.4 \text{ kcal/mol}