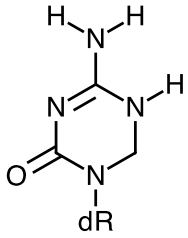
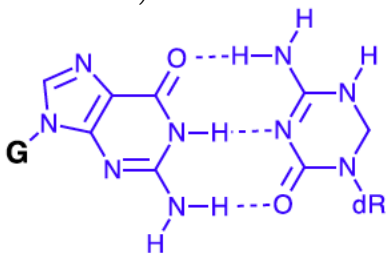
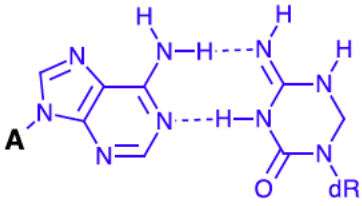
Problem Set #5 – Solutions

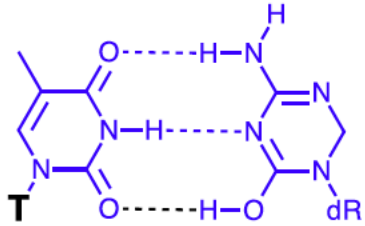
1. KP1212 is an antiviral drug candidate that interferes with replication. It adopts many different tautomers [see Li et al. (2014) paper on web page].

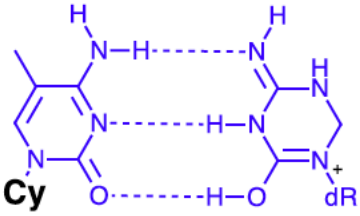


a. Produce tautomers of KP1212 that are capable of Watson-Crick base pairing with each of the four naturally occurring DNA bases, and show the hydrogen-bonding scheme for each (actually, C is pretty tricky – that's an extra credit effort).









b. In a study of mutation rates in DNA replication, it was found that KP1212 introduces guanine to adenine mutations 10% of the time, but no measurable mutation to cytosine or thymine. Why not?

KP1212 will presumably be placed opposite a purine during DNA replication. In subsequent rounds, KP1212 will allow either base to pair with it in standard purine/pyrimidine geometry (C and T are too small) so those will be the ones scrambled. DNA does have some preferences in size.

2. van't Hoff analysis of DNA melting is a little different than protein unfolding due to the stoichiometry of one duplex melting into two single strands. To find thermodynamic parameters in DNA melting, one plots $1/T_m$ vs. $\ln[C_T]$, where C_T is the <u>total</u> concentration of ssDNA. The equilibrium may be written as:

dsDNA $\Leftrightarrow$ 2 ssDNA.

At T_m, 50% of the duplex has melted.

a. Express [dsDNA] and [ssDNA] at the T_m as functions of C_T. Show that $K = C_T$ at T_m

If $[\text{strands}]_{total} = C_T$, then the number of total duplexes possible is $1/2C_T$. If 50% of duplexes have melted at T_m, then [ds]=1/4 C_T and [ss] = ½ C_T.

$K = [ss]^2/[ds] = (^1/_2C_T)^2/(^1/_4C_T) = C_T$!

b. Given the above, note that $\Delta G^\circ = \Delta H^\circ - T_m\Delta S^\circ = -RT_m\ln C_T$. Show that a plot of $1/T_m$ vs. $\ln C_T$ will have a slope of $-R/\Delta H^\circ$ and an intercept of $\Delta S^\circ/\Delta H^\circ$.

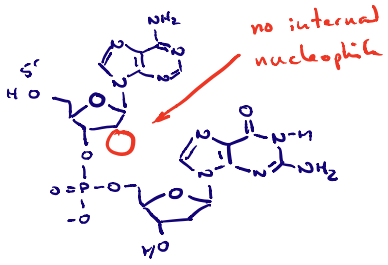
$\Delta H^\circ = -RT_m\ln C_T + T_m\Delta S^\circ$

$\Delta H^\circ/T_m = -R\ln C_T + \Delta S^\circ$

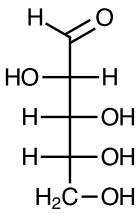
$1/T_m = -(R/\Delta H^\circ)\ln C_T + \Delta S^\circ/\Delta H^\circ$

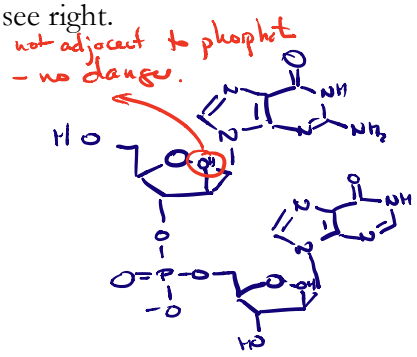
3. Draw each of the following assuming 3', 5' linkages unless otherwise indicated.

a. The 2-deoxyribodinucleotide with the sequence dAdG.

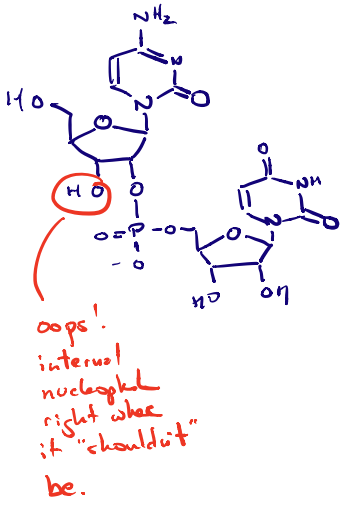


b. An arabinodinucleotide, aGaI. Note that β-D-arabinonucleotides uses the furanose form of D-arabinose instead of D-ribose – see right.





c. A **2', 5'** dinucleotide of RNA, whose sequence is rCrU.



d. Discuss whether the structures you have drawn in a-c are less susceptible to base hydrolysis than normal RNA.

As indicated in the diagrams, only the 2', 5' RNA has a hydroxyl group positioned to attack the adjacent phosphate with correct geometry, so it is the only one I would argue is as susceptible. The others will be more resistant.

4. The Kool lab has also worked on an unnatural base pair that forms between methylbenzimidazole (Z) and difluorotoluene (F). No hydrogen bonding takes place between the two, yet stable DNA duplexes form. You can see the complex in the file ZFbp.pse (on web-page).

duplex	$\Delta H°_{25}$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°_{25}$ (kcal/mol)
5'-C T T T T C **A** T T C T T 3'-G A A A A G **T** A A G A A	99.0	291	12.4
5'-C T T T T C **Z** T T C T T 3'-G A A A A G **F** A A G A A	47.4	129	8.9

a. Based on the melting data above, which duplex is more stable at 25°C? Which duplex has a higher melting temperature (show calculations to support your conclusion. T_m is where $\Delta G = 0$.).

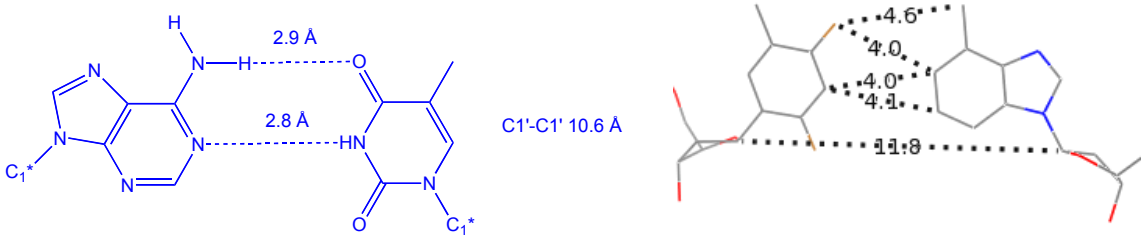
$T_m = \Delta H°/\Delta S°$: T_m = (99000cal/mol / 291 cal/molK) = 340 K for AT
T_m = (47400cal/mol / 129 cal/molK) = 367 K for AT

b. Explain why $\Delta S°$ of melting is likely less positive for the Z-F pair than for the A-T base pair.

Two possible arguments – the ZF duplex may be more disordered than the AT duplex and may gain less entropy on melting, or the non-polar nature of the Z & F "bases" may lead to clathrate formation on melting, reducing solvent entropy. The usual suspects.
Clearly the AT base pair creates a more enthalpically stable duplex. One argument might be H-bonding stabilizes the AT base pair (not likely). Another argument might be strain destabilizes the Z•F base pair. A crystal structure exists for the Z•F base pair in a DNA duplex (download **ZFbp.pse.** I have colored major groove magenta and minor groove cyan so you can see these first hand. You may want to recolor by element. Objects are available for an AT and ZF base pair).

c. On the templates below, **draw in** the A•T and Z•F base pairs. Identify major groove and minor groove atoms in the ZF base pair and and indicate the distances between atoms indicated with dotted lines, including between C1' atoms.

A•T Z•F



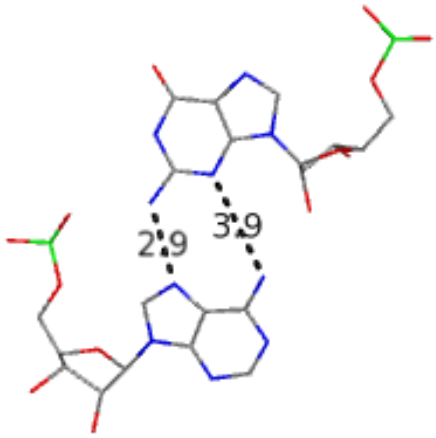
d. Aside from distances, indicate if there are any other differences in the geometry of the base pairs, either on the diagrams or in a brief description below.
Note that ZF has a displacement of F to the minor groove and Z to the major groove so that the closest atoms between "bases" are not as drawn for AT. Note that the methyl group of Z is 4.6 Å from the fluorine but the ring carbon of Z (eq. to position 1) is "only" 4.0 Å away.

e. Ignoring differences in hydrogen bonding, is there evidence that the A•T base pair duplex have a higher ΔH for melting than the Z•F base pair due to strain?
Sure – the greater separation of strands, the displacement of bases from standard geometry suggests that ZF doesn't fit and *may* be exerting strain on the duplex to accommodate their unusual properties.

6. RNA structure reveals that nucleic acids are not limited to simple double-stranded helices. The structure of a 158-base, folded RNA structure can be found in **1HR2.pse**. The initial view will show the entire structure with Mg^{2+} ions as spheres. Of interest is an tertiary interaction that leads to the overall structure of the RNA.

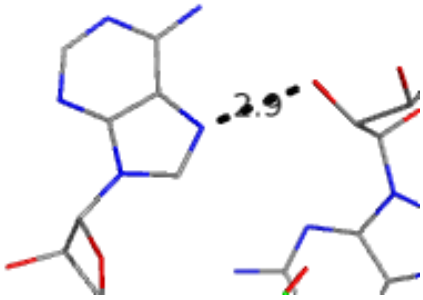
Zoom in on the object **GAAA** and hide the cartoon (and anything else that confuses your view). This is a "tetraloop", a four-base turn that comes at the end of a stem containing regular base pairs. The structure is "GNRA" type loop, where the first and last bases are always a G and A, and the third base is always a purine (R). The second base can be anything (N).

a. Sketch the interaction that clearly shows why the loop must be begin with an G and end with an A.



3.9 Å too long for H-bond, but will still be a dipole-dipole interaction.

b. Sketch the interaction that clearly shows why the third base (GNRA) must be a purine.

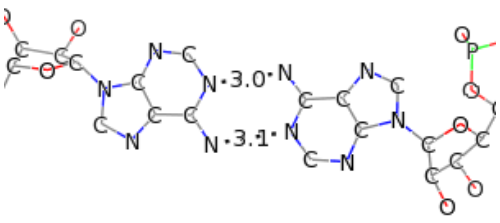


H-bond to N7 can't be formed with a pyrimidine.

c. Looking at the loop, explain why the second base can be A,C,G or U, the "N" in GNRA.

The base in the second position is not engaged in any H-bonding within the loop, though stacking might favor a purine.

d. In this structure, the tetraloop (GAAA) makes interactions with a "receptor region" to stabilize the tertiary fold. Sketch the interaction between the second base of the loop (GAAA), base 151, and adenine 248 (light green) that helps enable the loop-receptor interaction.



Paper of the week: Rothemund, "Folding DNA to create nanoscale shapes and patterns", *Nature* **440**, 297-302.

1. This paper describes the promise and limitations of AFM in atomic manipulation and then the use of AFM to evaluate the results of this study. Briefly describe how AFM works, preferably using a diagram/illustration to illuminate the process.

Nice answers will have words like "cantilever" optical reflection and contact mode. Basically you drag a tip across a surface and the deflection of the cantilever is monitored by reflected laser light.

2. Dimensions are important here. We have to beat Feynman's 8 nm. Explain the origin of the 6 nm resolution mentioned by the author. Let's pay particular attention to the helix gap.
a. What restrains the x-resolution for a "pixel" in this technique? Connect the measurement to a specific structural description of DNA.

The width of a pixel is the length of one turn of DNA (since contacts can't be randomly inserted mid-turn). 3.4 Å * 10.5 bp/turn (roughly) gives 35.7 Å or 3.6 nm.

b. Now y-resolution. Explain why the inter-helix gap may be 1-1.5 nm for a crossover. Draw a dinucleotide and estimate the greatest separation one might expect between adjacent C1' atoms.
I did a quick hunt and found an 8.3 Å (0.83 nm) gap in a tRNA molecule. I suspect 1 nm would be easy. 1.5? I think that must involve deformation of the helices that are being spanned.

He says 2 helix widths. That's 4-5 nm if the BDNA model is used (see my handout). So if the two helices has a 1.5 nm gap and each is 2.5 nm in diameter, I'd think we'd be looking at 5.5. You can see why in Figure 1d. Each staple spans two helices and staple "repeat" is a two helix repeat.

c. So given "b" how do we end up with a y-resolution of 6-7 nm?

3. Now the cross-overs. What is meant by "rastering progressively", how does that lead to a "scaffold cross-over" and what constrains the scaffold cross over to an odd number of half turns? Why are an even numbers of half-turns needed to return vertically? A diagram will probably be most useful.
Rastering is his word for tracing the DNA horizontally (like an old-style TV that shoots a cathode ray across the screen). When you reach the end of a "line" you need to turn the DNA around and head the other direction. I can draw a picture for odd/even.

4. In class, I said that one turn of B DNA includes about 10 bp. The author uses staples that come in 16 nucleotide chunks and generally uses 32 nucleotide staples. What assumptions are made in base pairs per turn and in nucleotides spanning the helices to guide that choice? What accommodations are made due to imperfections in those assumptions?

16 bp constitutes 1.5 turns (1.5 turns*10.5 bp/turn = 15.8 bp)

It's tricky to see the point in Figure 1d, but he designs cross-overs with +/- positions across the minor groove so that imperfections in one direction will be compensated by imperfections in the other. Also added base pairs as needed (not specified in my reading)

5. Onto the scaffold material.

a. What is M13mp18 viral DNA and why is it particularly suitable as a scaffold?

Single stranded! So you don't need to get rid of the native second strand and can use staples easily.

b. The author notes a single hairpin with a stem of 20 base pairs. That can be resolved by "digestion with BsrBI restriction enzyme". What is a restriction enzyme, what specifically does BsrBI digest, and what would happen to the M13 DNA after digestion? Why would BsrBI then be a hazard to the shapes shown in Figure 2? (Helpful info in pp 74-75 of supplemental).

c. What is the purpose of remainder strands?

Basicall to tie up all the ssDNA that didn't get stapled. Don't let it interfere with the pattern.

6. Explain how the patterning on the constructs shown in Figure 3b, d and f are produced.

Drop a dumbbell hairpin into a staple which will create a "bump" on the pattern. That can be registered by AFM as an alternate "pixel" and colored alternately to provide contrast based on height.

7. The author cites strand invasion as one key to success of this technique. What is strand invasion and how does it assist in creating the correct structures (see p. 79-80 supplemental)?

I would phrase this as equilibrium vs. kinetics. It's not the fastest strand who wins, but the one that ultimately provides best complex. Any imperfect duplex is possible prey for a better staple to displace it.