Biochemistry 674 **Your Name:** Key

Nucleic Acids **Prof. Jason Kahn**

Exam I (100 points total) **October 3, 2006**

You have 80 minutes for this exam.

Exams written in pencil or erasable ink will not be re-graded under any circumstances.

Explanations should be concise and clear. I have given you more space than you should need. There is a extra space on the last page if you need it.

You do not need a calculator for this exam, and no other study aids or materials are permitted.

Generous partial credit will be given, *i.e.*, if you don't know, guess.

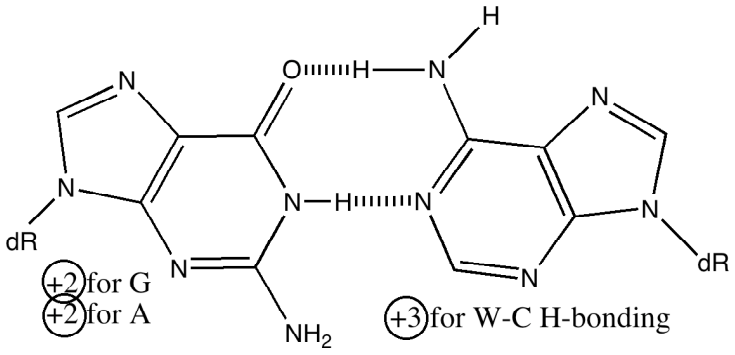
Honor Pledge: At the end of the examination time , please write out the following sentence and sign it, or talk to me about it:

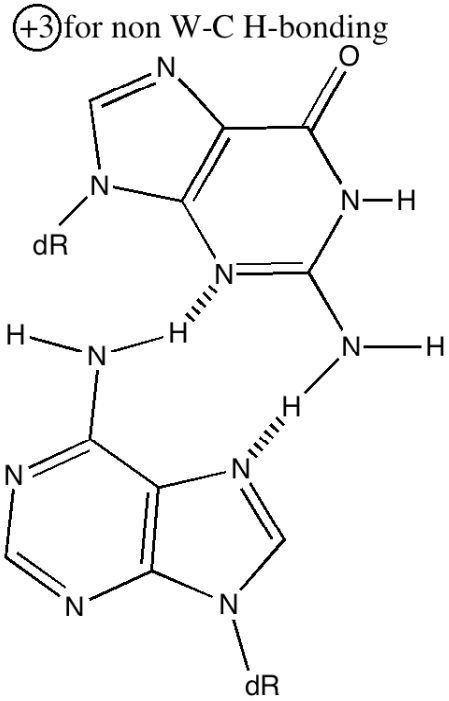
"I pledge on my honor that I have not given or received any unauthorized assistance on this examination."

1. DNA Structure and Stability (30 pts):

There are a lot of deviations from Watson Crick base pairing and the WC helix, both from nature and from synthesis. Eric Kool of Stanford has been particularly active in the latter field.

(a; 10 pts) Draw two G:A base pairs: (1) a structure in which the Watson-Crick H-bonding faces of G and A interact with each other. (2) Freeform H-bonding of your choice, with at least two H-bonds formed, different from (1). An example might be the G:A pair seen in the tetraloop. You do not need to draw out sugars.





Score for the page_____

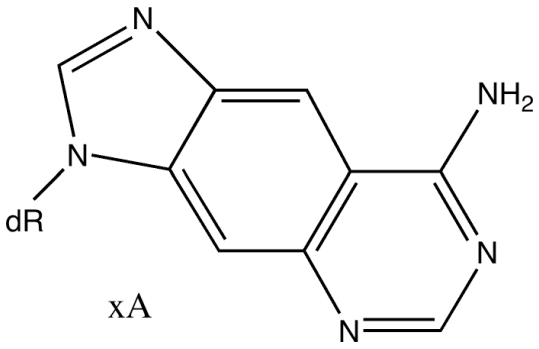
(b; 3 pts) We have emphasized over and over what it is that makes the WC base pairs special. What is it?

+3 The 4 WC base pairs all fit into the same regular helix geometry.

(c; 3 pts) G:A base pair # 1 from part (a) destabilizes B-form DNA. Why? It stacks and H-bonds well.

+3 The glycosidic bonds are further apart, therefore the G:A pair will not fit into the regular B-form helical geometry

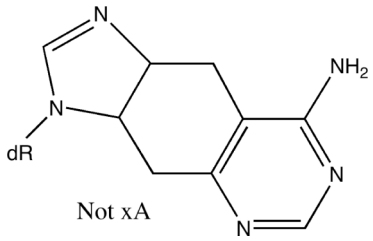
Kool's group has synthesized the extended adenine at the right (xA) and incorporated the xA:T base pair into DNA oligonucleotides.



xA

(d; 3 pts) Why did they choose to make the xA shown instead of "Not xA" shown below xA?

+1 Not xA would not be planar due to the sp3 carbons
+2 Therefore it would not stack in a helix.



Not xA

(e; 3 pts) It is observed that the structure and backbone of the extended DNA (with an extendabase in each pair) is essentially a thick B-form, as shown below. This was somewhat surprising, at least to me. Why is it surprising? (Consider the limit of including more benzene "extenders.")

+3 The backbone has to stretch to encircle a thicker helix: apparently there's enough flexibility from all those torsion angles to allow this to happen.

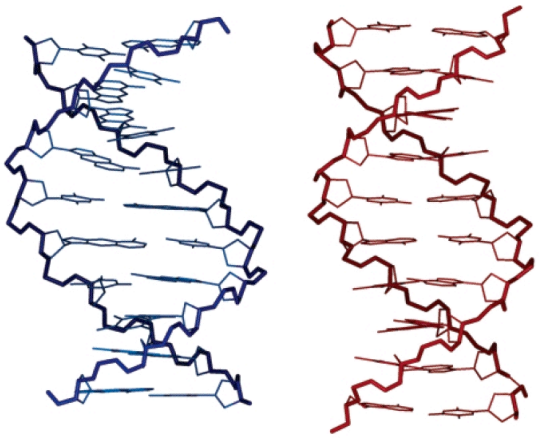


Figure 4. Mean structure of 24 randomly selected final structures of xDNA (at left) and a model structure of the control B-DNA (right).

Score for the page_____

(f; 8 pts) When xA:T is placed in an otherwise normal DNA it is observed to destabilize the helix (T_m goes down). However, the duplex in (e) that includes an extendabase in each base pair is observed to be more stable than normal DNA. For complicated reasons, the change in stability has not been parsed into ΔH° and ΔS° contributions yet. Sit on the fence: give a rationale for each possibility, i.e. explain how the extendabase could give a larger negative ΔH° for hybridization or else a smaller negative ΔS°.

+3 Larger negative ΔH° would be due to improved stacking from the larger ring structure. +1 The entropic penalty due to restraining conformational angles should change very little.

On the other hand...

+2 Smaller negative ΔS° (and smaller negative ΔH°) would arise if the ss xDNA is pre-organized in a helical geometry (remember LNA?), which would come about +2 if the stacking was already substantially present in the ss xDNA.

Remember, it's the change in going from ss to ds that is important, not the heat of formation of either.

2. Hybridization (25 pts):

(a; 10 pts) Briefly describe the DNA microarray and how it is used to study mRNA expression patterns.

+3 The microarray is prepared by spotting or synthesizing many different known DNA sequences on a glass slide. Usually Affymetrix or other commercial chip is used.

+2 Cells (often yeast) are grown under two different conditions and mRNA is isolated, or mRNA is isolated from two different tissues/patients/locations, etc.

+2 The mRNA is copied, or amplified and then copied, to make fluorescently labeled cRNA or cDNA. Each sample of the two above is worked up independently and labeled with a different dye.

+3 The mRNA samples are mixed and hybridized to the same chip, and the signal intensities for the two dyes are compared to determine increases/decreases in the level of mRNA.

Score for the page_____

Microarrays can also be applied in other ways. The "DNA tiling array" can be used to sequence variants of known genomes or look for things like polymorphisms or splice site changes. The idea is that 25mer probes that span a genome are synthesized on the array, and then labeled DNA is hybridized to the array. The picture below (Bertone *et al.*, *Genome Res.* 2006 16: 271-281) should give you the idea, with some extraneous detail.

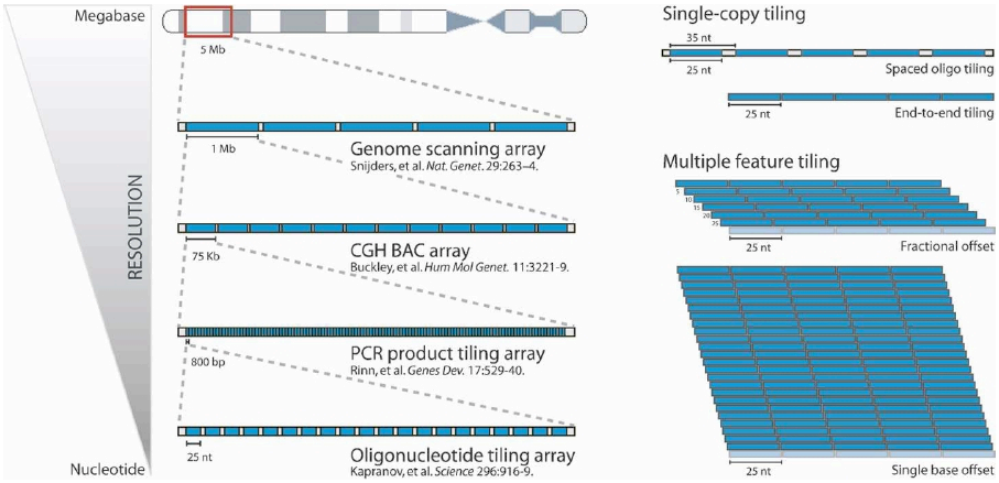


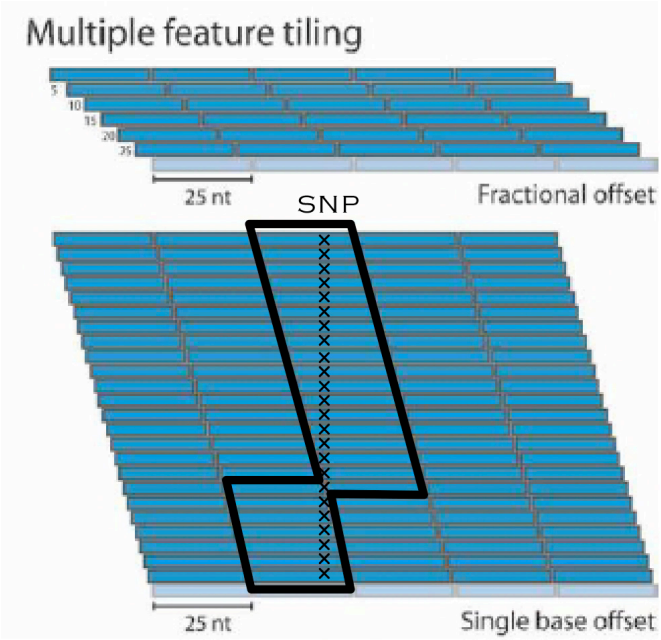
Figure 1. (*Left*) Evolution of genomic tiling arrays. Representing large spans of genomic DNA with bacterial artificial chromosome (BAC) clones facilitates global experimentation using relatively few array features, at the expense of low-tiling resolution. Higher-resolution designs using PCR products or oligonucleotides allow precise mapping of transcripts and regulatory elements, but require labor-intensive or technologically sophisticated approaches to implement. (*Upper right*) Linear feature tiling with gapped and end-to-end oligonucleotide placement. (*Lower right*) Overlapping tiles using fractional offset (e.g., one 25-mer probe placed every 5 nt) and single-base offset placement. The latter strategy provides a finer-resolution tiling of the genomic sequence, and can give a more precise indication of where hybridizing sequences are located on the chromosome.

(b; 6 pts) Experimentally, how could you identify a single nucleotide polymorphism or other slight change in a genome using a tiling array like the one at the bottom right above? In other words, what would be the signal that e.g. a new isolate of the bacterium had a sequence difference?

+3 The probes on either side that did not overlap the SNP would hybridize equivalently to the old and new isolates. This controls for e.g. amount of DNA loaded or labeling efficiency, and it confirms that the gene is present in both bugs.

+3 The probes that cover the SNP would hybridize less well due to the mismatch. If the new isolate is known, the array could include oligos that should pair perfectly with the new allele.

The diagram shows that only the boxed probes would be affected.



Score for the page_____

(c; 9 pts) Why 25mers? Why do much longer probes not work well for identifying SNPs? On the other hand, why not make the probes much shorter, so that synthesis would be easier? In general, what two somewhat conflicting goals do we have in any hybridization experiment?

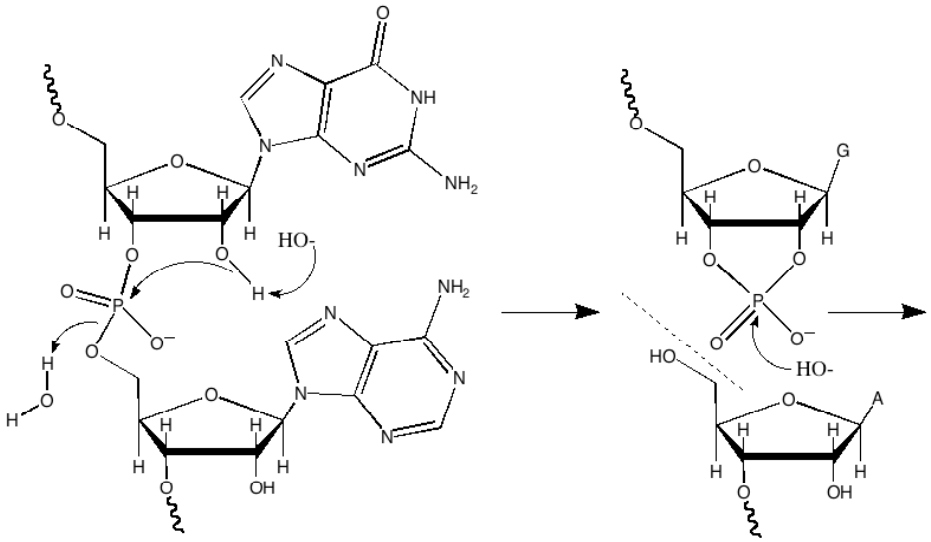
+3 If the probes are much longer, their hybridization will be so stable that it will be essentially unaffected by the mismatch at the SNP: Lose discrimination.

+3 If the probes are much shorter, they will not hybridize stably, or else they will match too many random sequences in the genome. Lose affinity and specificity.

+3 In general, we always want high stability and high specificity, and they are inherently in opposition.

3. RNA Structure (20 pts):

(a; 10 pts) Sketch base-catalyzed hydrolysis of the RNA phosphodiester backbone. Why are there no large RNA genomes?

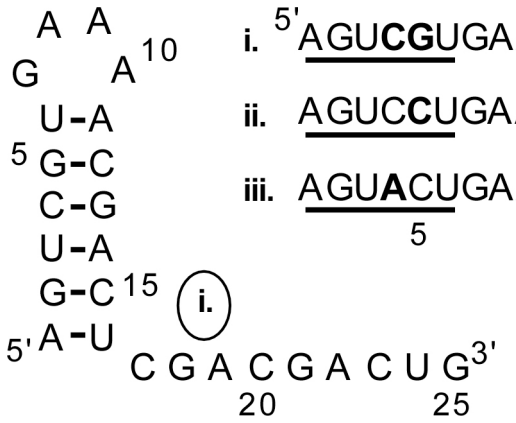


+3 for correct RNA backbone, +3 for correct attack mechanism and products, +1 for base-catalyzed

+3 Large RNA genomes could not survive because they could not survive long enough in water to be replicated successfully

Score for the page_____

The essential RNA (i) below was proposed to form the structure shown based on computer modeling. Then homologous sequences (ii) and (iii) were discovered. The bases that differ from RNA (i) are indicated in bold. The underlines are hints.



i. 5' AGU**CG**UGAAAA**C**G**A**CUCGA**CG**ACUG 3'

ii. AGUC**C**UGAAAA**C**A**A**CUCGA**G**GACUG

iii. AGU**A**CUGAAAA**C**C**A**CUCGAG**U**ACUG

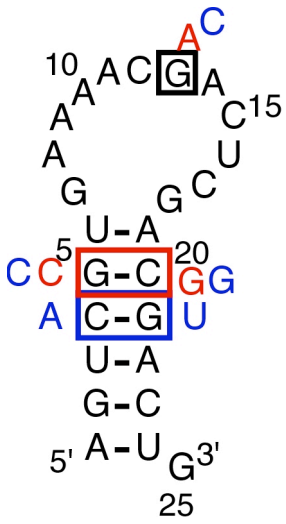
5 10 15 20 25

(b; 4 pts) Explain the notion of correlated invariants in phylogenetic studies of RNA structure.

+2 Structure is more conserved than sequence, so we expect consistent 2° and 3° structure in at least stable RNA's

+2 So we look for changes that are consistent with fitting into the same structure, so that if G->C then C-> G etc., whereas changed are uncorrelated if the nucleotides are not in contact.

(c; 6 pts) The sequences (ii) and (iii) support a structure different from the one shown. Draw the structure, for sequence (i). Why is this structure unexpected?

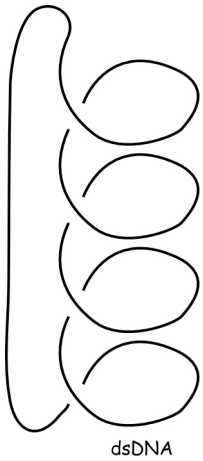


+2 for idea of using correlated invariants
+2 for correct base pairing and reasonable drawing
+2 for the original structure looks more stable because of the shorter loop and the tetraloop GAAA sequence.

Score for the page_____

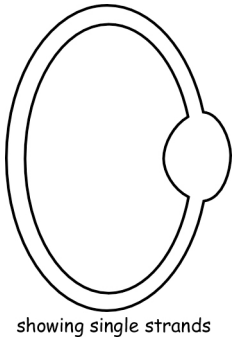
4. DNA Topology (20 pts):

(a; 15 pts) Sketch the three main ways in which negative supercoiling is manifested (one has $\Delta Tw < 0$ and the other two are writhed). Identify the thermodynamically stable form for typical supercoiled B-DNA in the absence of proteins. Give one biological function for each of the two writhed forms.



Solenoidal superhelix, +3 for sketch

+2 Functions in DNA packaging in chromatin



Untwisted form, +3

Plectonemic superhelix, +3 for sketch

+2 Thermodynamically stable

+3 Functions in bringing distant sites near each other.

Score for the page_____

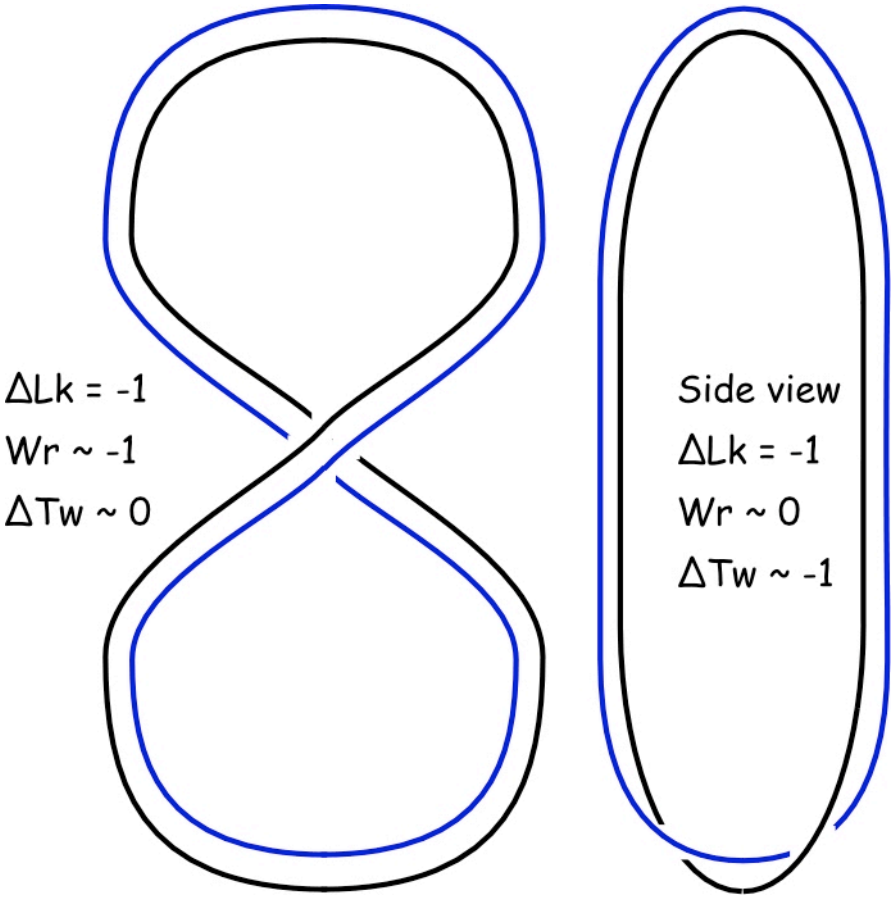
(b; 5 pts) The equation for determining linking number from a flat sketch of single strands winding around each other is Lk = (sum of nodes)/2. We determine writhe from a projected image of a double strand DNA similarly. What is the equation for that (and why is it different)? We know that Lk is always an integer, but writhe is not. How can this be given that the equations are so similar?

+1 Writhe = sum of nodes

+2 It's apparently a different equation because each superhelical writhe node actually comprises two crossovers of single strands on each other, so the two definitions are consistent.

+2 Writhe is an average over the number of crossings seen from all possible angles, and from some points of view there is no apparent writhe.

Extra explanation:
From those other points of view, one would always see twist crossovers. In the sketch below, the tighter the node, the more frequently the left side view would be seen and the more rarely the right side, the more closely the writher would approach to 1.



Page	Score
1	
2	
3	
4	
5	
6	
7	
8	
Total	

Score for the page_____