Biochemistry 674: Nucleic Acids **Your Name:** _____

Prof. Jason Kahn **Exam II, November 15, 2001**

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.

You may need a calculator for this exam. No other study aids or materials are permitted.

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

Reminder of how microarray expression profiling works:

mRNA from an experimentally perturbed cell is isolated, and fluorescent cDNA is synthesized and then hybridized to a slide or surface bearing a different but known target DNA sequence at each one of many locations. Fluorescence at each position is quantitated and compared to an array hybridized with a control sample, e.g. one in which the perturbation was omitted. For the purposes of this exam, hybridization to the chip is perfect in all cases.

For simple binding equilibria at [DNA] $<<K_d$, $\theta = P / (P + K_d)$

1. ChIP on a chip (25 pts):

(a; 8 pts) Briefly describe how the chromatin immunoprecipitation assay works, using antibodies to YFTF (your favorite transcription factor).

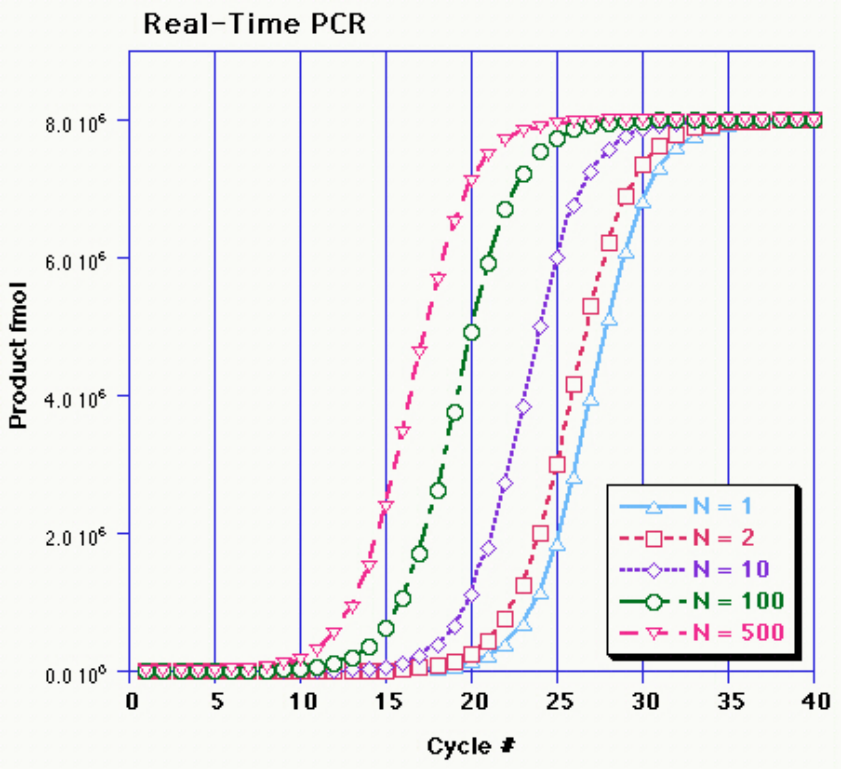
(b; 8 pts) The ChIP assay has recently been adapted to a microarray format, so that all the promoters in a genome can be examined at once for binding to YFTF. ChIP on a chip uses a method called ligation-mediated PCR (LM-PCR), in which known oligonucleotide duplexes are first ligated to the ends of target DNA, to amplify and fluorescently label DNA. Sketch the LM-PCR amplification method (you have to figure out what it is), and explain why it is used here instead of regular old PCR.

(c; 4 pts) How does ChIP on a chip provide different but complementary information to a simple mRNA expression profile done upon induction of YFTF activity?

(d; 5 pts) How would the sequences to put on the chip be determined, assuming that the promoters are always within 500 bp of the beginning of a gene? What other sequences might one throw in for good measure?

2. Methods (20 pts).

PCR can be used to quantitate the amount of template DNA in a sample. This is often done with "real-time PCR," in which the amount of PCR product is monitored in the reaction vessel after each cycle of denaturation/annealing/synthesis. The graph below shows reasonable simulated data for the amount of product vs. cycle number, for different starting amounts (N, in fmol) of template.



(a; 4 pts) Why does the amount of product plateau at the end, even though I assert there is primer remaining? What explains the long lag before there is measurable signal?

(b; 6 pts) Based on the graph, why is real-time monitoring necessary for quantitation of starting template amounts that are even moderately different?

(c; 4 pts) In comparative microarray expression profiling experiments, it is often necessary to PCR-amplify RNA from tissue before doing a hybridization to the chip. Given what you have learned in this problem, what problems might be lurking in such microarray experiments?

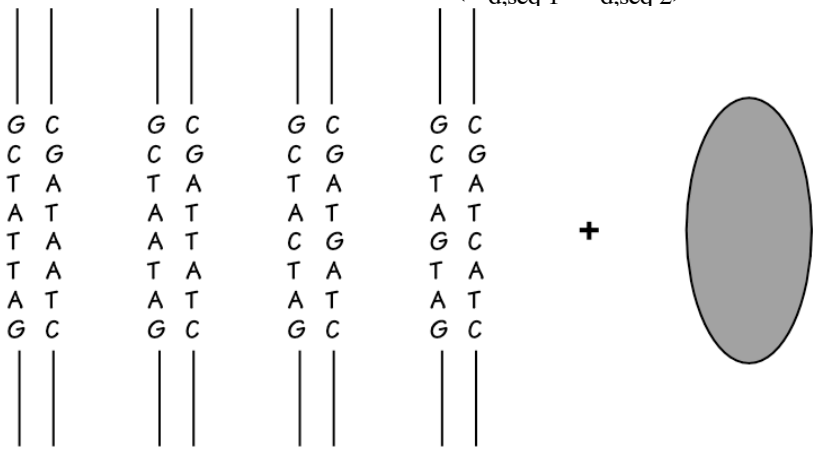
(d; 6 pts) (Unrelated.) What does it mean to "CIP/kinase" an RNA? Sketch the reactions, starting with 5′-phosphorylated cold RNA. Why is it important to remove the CIP aggressively between steps?

3. Sequencing and Footprinting (30 pts).

(a; 8 pts) Briefly describe dye-terminator automated fluorescent sequencing. What advantages do the "Big-Dye" terminators confer?

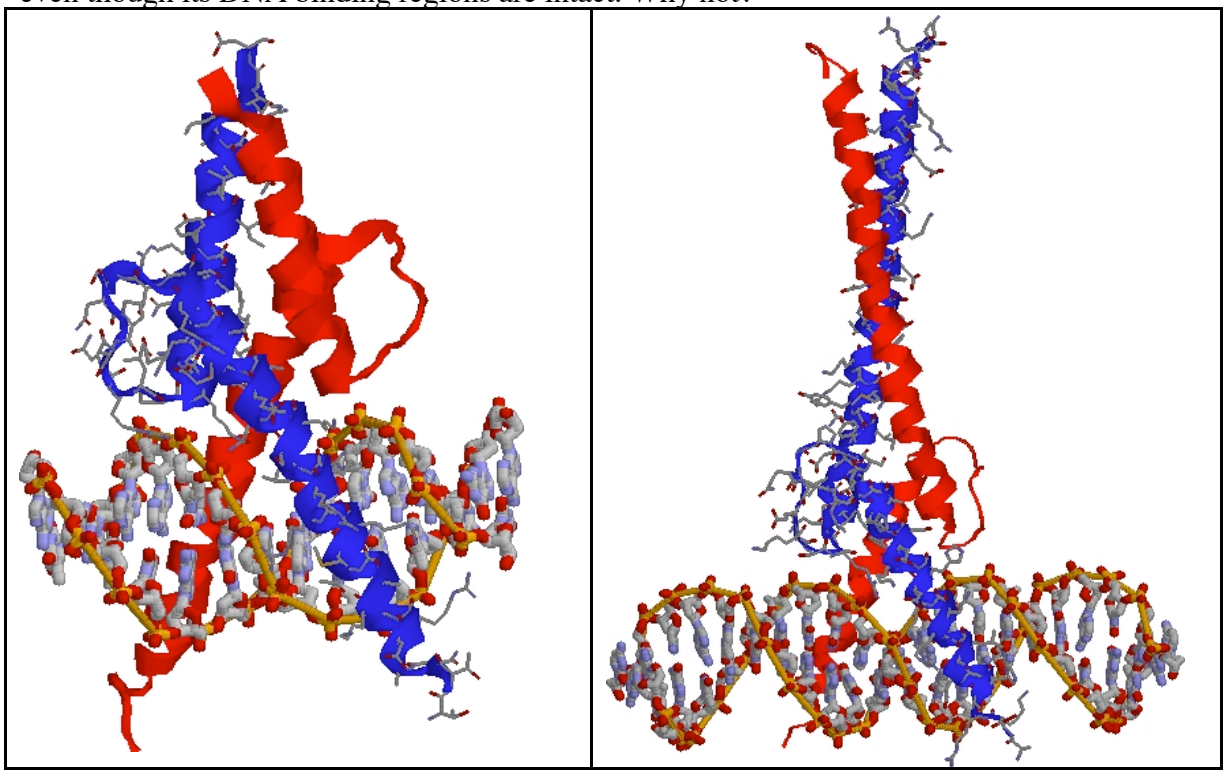
(b; 10 pts) It is possible to treat a transcription factor •DNA complex with DNAse I, then run a gel mobility shift, isolate free and bound, and finally analyze the two pools on a denaturing gel. What advantage might this have over a straightforward footprint? Why might it not work for footprinting a single-stranded RNA•protein complex with a ribonuclease? Why might it not work for footprinting with the extremely non-specific and reactive hydroxyl radical?

(c; 12 pts) A variant of modification-interference is the analysis of the effect of sequence changes at a single position. In the example below, the four DNAs indicated are mixed together, protein is added, free and bound are isolated, and then the two pools are analyzed to learn which DNA sequence is preferred at the variant position. How might they be analyzed using fluorescent sequencing methods, and what would the data look like? Given that you can determine the fraction bound (θ) for each of two sequences (instead of all four, just for simplicity), derive an expression for the ratio of dissociation constants ($K_{d,seq\ 1}/ K_{d,seq\ 2}$)for the two sequences.



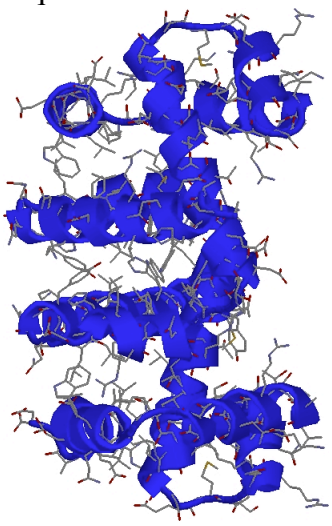
4. Protein-DNA interaction (25 pts)

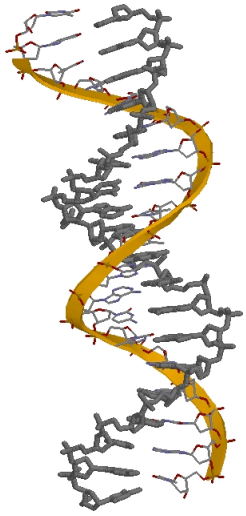
(a; 6 pts) The pictures below show the MyoD homodimer (left) and the Max homodimer (right) bound to DNA. Identify the leucine zipper, the helix-loop-helix region, and the basic region for each protein. When the leucine zipper of Max is removed, the protein no longer binds DNA even though its DNA binding regions are intact. Why not?



(b; 6 pts) Max/Max, Myc/Max, and Mad/Max bHLH dimers bind to the same DNA sequences but have every different downstream effects. How can expression of Mad control the activity of Myc even though the two proteins do not directly interact?

(c; 8 pts) The Trp repressor and its DNA partner are shown below at the same scale. Identify the helix-turn-helix motifs on the protein, and state how you identified them. Use the DNA at the right to demonstrate why dimeric DNA binding proteins generally bind to inverted repeat sequences.





(d; 5 pts) The main difference between aporepressor (-Trp, inactive) and holorepressor(+Trp, active) is that the helix-turn-helix motifs are separated by different distances in the two forms. Why is this so important?

Question	Score
1	/25
2	/20
3	/30
4	/25
Total	/100