## ChE/BE 163

## Problem Set \#2

Due: 1 pm Thursday, November 4, 2021

## Notes:

- Problems 3 and 4 are more computationally intensive; use one Jupyter notebook for Problems 1 and 2 combined, one notebook for Problem 3, and one notebook for Problem 4.
- Use physical model:

```
Model(material='dna04-nupack3', ensemble='some-nupack3', celsius=23)
```

so that the model you use for sequence design will match the one used for sequence analysis online at nupack.org.

Problem 1 (Test tube design, 20 pts).
Your primary reference is the paper Wolfe and Pierce, ACS Synth Biol, 4, 1086-1100, 2015. Consider DNA sequence design at $23{ }^{\circ} \mathrm{C}$ for an on-target complex with target secondary structure:
$((((((\ldots \ldots)(((((((+)))))))) \ldots . .(((((+)))))) .((((((+))))))))))$.
and target concentration $1.0 \mu \mathrm{M}$.
a) (3 pts) Write down the mathematical expression for the test tube ensemble defect $C$. Which term would become large if the on-target dimer is not dominated by the target secondary structure? Which term would become large if off-target monomers or tetramers form at appreciable concentration in the test tube?
b) ( 7 pts ) Use executable tube_design of NUPACK 4 (described in the NUPACK User Guide) to design sequences over a test tube ensemble $\Psi$ containing all off-target complexes of up to 4 strands. Your notebook should display your calls to the design executables, your final sequences, and the final value of the test tube ensemble defect. You may find it useful to view the Tube Design Example in the NUPACK User Guide under Getting Started/Example Jupyter notebooks).
c) ( 5 pts ) Use the Analysis page at nupack.org to document the salient properties of the test tube and complex ensembles for your final design.
d) ( 5 pts ) Now use complex design instead of test tube design to design for the target secondary structure without considering off-target complexes (Hint: use executable tube_design with max_size = 0 ). How does the new design compare to your previous design when you evaluate its properties over the ensemble $\Psi$ containing all complexes of up to 4 strands?

Problem 2 (Mechanism design for molecular logic, 15 pts ).
Design the mechanism for DNA logic gates to evaluate the conditional ( $x$ AND y) OR $z$. Here, $x, y$, and $z$ are all short unstructured DNA strands with unrelated sequences. Your primary reference is the paper Seelig et al., Science, 314, 1585-1588, 2006. Making use of sequence translator and logic gate motifs from the paper, sketch the secondary structure of the gates for your problem and annotate the structures to make toehold and domain complementarity clear. Be sure to clearly mark the steps taken from the inputs ( $\mathrm{x}, \mathrm{y}$, and z ) to the output of a single-stranded fluorophore.

Problem 3 (Sequence design for reaction pathway engineering: orthogonal HCR amplifiers, 30 pts ). Your primary reference is the paper Wolfe et al., J Am Chem Soc, 139, 3134-3144, 2017 and the corresponding Supplementary Information.
a) (5 pts) What is the conceptual advantage of multistate test tube design compared to multistate complex design in terms of design paradigms?
b) ( 5 pts ) Consider sequence design for a library of 2 orthogonal DNA HCR amplifiers intended to detect different initiator sequences and operate independently in the same test tube at $23{ }^{\circ} \mathrm{C}$. The HCR mechanism schematic is shown in Figure S1 (each DNA hairpin has a 12 nt toehold, 24 bp stem, and 12 nt loop) and the corresponding target test tubes are described in Sections S2.2.1 and S2.2.2 of the Supplementary Information. Sketch the target test tubes for your design (including target secondary structure and target concentration for each depicted on-target complex). How many tubes are there in your design ensemble?
c) ( 15 pts ) Use executable tube_design of NUPACK 4 (described in the NUPACK User Guide) to design 2 orthogonal HCR amplifiers over this multistate test tube design ensemble using a $2 \%$ stop condition (You may find it useful to view the Multi-Tube Design Examples for design of N orthogonal reaction pathways in the NUPACK User Guide under Getting Started/Example Jupyter notebooks). Use hard sequence constraints to ensure GC content in the range $45 \%-55 \%$ and to ensure sequence diversity with at least 2 nucleotide types in each word of length 4 . Leave all nucleotide, complex, and tube defect weights at the default value of 1 except for the global crosstalk tube which is assigned a weight of 2 (because you are designing $N=2$ orthogonal systems) to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases. Include your final designed sequences outputted in your notebook containing your tube_design function calls.
d) ( 5 pts ) Plot the residual defects for your design ensemble and comment on any notable features (see Supplementary Information Section S2.4 and Figure S13 for an example with 4 orthogonal HCR amplifiers). The TAs will provide Python commands to make plotting the residual defects straightforward.

Problem 4 (Sequence design for reaction pathway engineering: catalytic hairpin assembly of a 4 -arm junction, 35 pts .
Your primary reference is the paper Wolfe et al., J Am Chem Soc, 139: 3134-3144, 2017 and the corresponding Supplementary Information. Consider sequence design for catalytic hairpin assembly (CHA) with the elementary steps depicted in Figure 1.
a) (15 pts) Use the definitions and conventions of Section S2.2.1 and S2.2.4 to define the multistate test tube ensemble for CHA with $L_{\max }=2$ for all tubes. Be sure to carefully read the two pages of Section S2.2.1 including descriptions of motif simplification, input domains, output domains, reactive species, and cognate products that are used in defining the global crosstalk tube. For clarity, use the same set names, domain names, and presentation format as Section S2.2.4 to present your specification. Your specification should be written in markdown cells in your notebook. Note that the CHA example in Section S2.2.4 is closely related to your design problem and that there is an error in that example: X•B should be included in the sets $\Psi_{0_{n}}^{\text {exclude }}$ and $\lambda_{n}^{\text {cognate }}$. Be sure to include the test tube drawings and table on page 2 of the specification.
b) ( 10 pts ) Use executable tube_design of NUPACK 4 to design sequences over this multistate test tube
design ensemble for DNA at $23{ }^{\circ} \mathrm{C}$ with a $2 \%$ stop condition.
Include your final designed sequences outputted in your notebook containing your tube_design function calls.
c) ( 5 pts ) Update your script to stipulate that X is a subsequence of the $t p m 3 \mathrm{mRNA}$ sequence (available in the NUPACK User Guide under Getting Started/Example Jupyter notebooks: Multi-tube design (advanced): design a multi-step reaction pathway).
d) ( 5 pts ) Plot the residual defects for your design ensemble with and without the mRNA sequence constraint and comment on any notable features (see Supplementary Information Section S2.4 and Figure S15 for an example with 3 -arm junction CHA). The TAs will provide Python commands to make plotting the residual defects straightforward.


Figure 1: Reaction pathway for self-assembly of a 4 -arm junction via catalytic hairpin assembly (CHA) (Yin et al., Nature, 451:318-322, 2008 [see Figure 2]). Target X catalyzes self-assembly of metastable hairpins A, B, C, and D into 4 -arm junction A•B•C•D. Sequence domain lengths: $|\mathrm{a}|=|\mathrm{b}|=|\mathrm{c}|=|\mathrm{d}|=12 \mathrm{nt},|\mathrm{x}|=|\mathrm{y}|=|\mathrm{z}|=|\mathrm{w}|=6 \mathrm{nt}$. Top: Reaction pathway schematic. Bottom: Elementary step details.

