ChE/BE 163
Problem Set #2
Due: 1 pm Thursday, November 3, 2022

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 www.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 °C for an on-target complex with target secondary structure:

```
((((((.....(((((((((+)))))))))).....((((((+)))))).((((((+)))))).))))))
```

and target concentration 1.0 μM.

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 Ψ 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 www.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 Ψ 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 (x, y, and z) to the output of an unstructured fluorophore-labeled strand.

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 Appendices A and B derived from the 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 °C. The HCR mechanism schematic is shown in Figure 2 (each DNA hairpin has a 12 nt toehold, 24 bp stem, and 12 nt loop) and the corresponding target test tubes are described in Figure 3. 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.

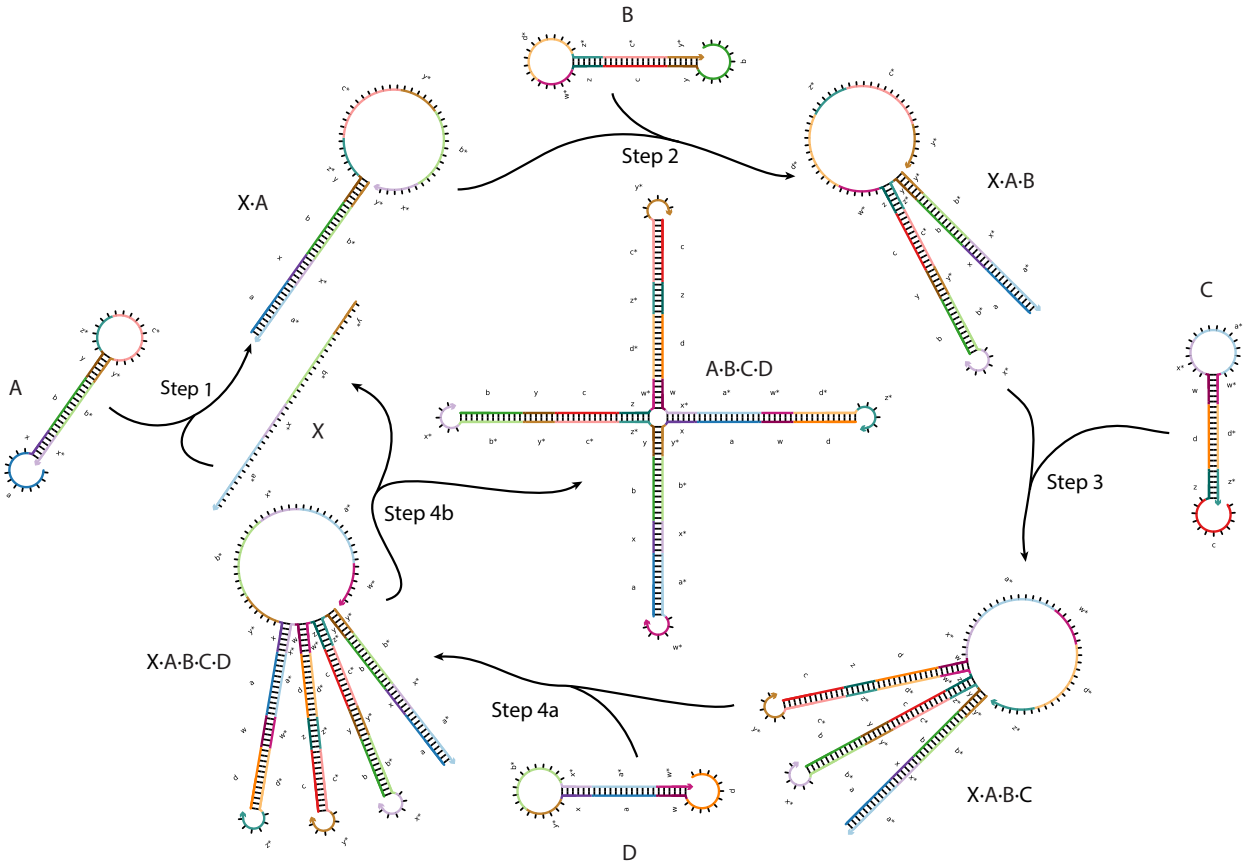
d) (5 pts) Plot the residual defects for your design ensemble and comment on any notable features (see Figure 4 for an example with 4 orthogonal HCR amplifiers). The TAs will provide Python commands to make plotting the residual defects straightforward.

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 Appendices A and C derived from the Supplementary Information. Consider sequence design for catalytic hairpin assembly (CHA) of a 4-arm junction with the elementary steps depicted in Figure 1.

a) (20 pts) Use the definitions and conventions of Appendices A and C to define the multistate test tube ensemble for CHA with $L_{\max} = 2$ for all tubes. Be sure to carefully read Appendix A 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 Appendix C to present your specification. Your specification should be written in markdown cells in your notebook. Note that the CHA example in Appendix C is closely related to your design problem. Be sure to include drawings of your target test tubes and a table itemizing the tube contents as in Appendix C.

b) (10 pts) Use executable `tube_design` of NUPACK 4 to design sequences over this multistate test tube design ensemble for DNA at 23 °C with a 2% stop condition. Include your final designed sequences outputted in your notebook containing your `tube_design` function calls.

c) (5 pts) Plot the residual defects for your design ensemble and comment on any notable features (see Figure 7 for an example with 3-arm junction CHA). The TAs will provide Python commands to make plotting the residual defects straightforward.



Step	Reaction	Function	Mechanism
1	X + A → X·A	assemble with catalyst X	toehold/toehold nucleation, 3-way branch migration
2	X·A + B → X·A·B	assemble	toehold/toehold nucleation, 3-way branch migration
3	X·A·B + C → X·A·B·C	assemble	toehold/toehold nucleation, 3-way branch migration
4a	X·A·B·C + D → X·A·B·C·D	assemble	toehold/toehold nucleation, 3-way branch migration
4b	X·A·B·C·D → X + A·B·C·D	disassemble from catalyst X and assemble 4-arm junction	intracomplex blunt-end strand invasion, 3-way branch migration

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: |a|=|b|=|c|=|d|=12 nt, |x|=|y|=|z|=|w|=6 nt. Top: Reaction pathway schematic. Bottom: Elementary step details.

A General formulation for target test tube specification

Consider specification of the multistate test tube ensemble, Ω, for the design of N orthogonal systems for a reaction pathway of M elementary steps, each corresponding to a self-assembly or disassembly operation in which complexes form or break. One elementary step tube is specified for each step $m = 0, 1, \ldots, M$ for each system $n = 1, \ldots, N$ (treating formation of the initial reactants as a precursor "Step 0"). Additionally, a single global crosstalk tube is specified to minimize off-pathway interactions between the reactive species generated during all elementary steps of all systems. The total number of target test tubes is then $|\Omega| = (M+1) \times N + 1$.

Elementary Step Tubes. Consider elementary step m for orthogonal system n with on-pathway products $\Psi_{m_n}^{\text{products}}$ that are intended to form at non-zero concentrations at equilibrium, and reactants $\Psi_{m_n}^{\text{reactants}}$ that are intended to fully convert into the on-pathway products at equilibrium. Furthermore, consider the set of off-pathway products, $\Psi_{m_n}^{\text{crosstalk}}$, corresponding to unintended interactions between these same reactants.

The *elementary step tube* for step m of system n is then:

$$\boxed{\text{Step } m_n \text{ tube: } \Psi_h^{\text{on}} \equiv \Psi_{m_n}^{\text{products}}, \qquad \Psi_h^{\text{off}} \equiv \Psi_{m_n}^{\text{reactants}} \cup \Psi_{m_n}^{\text{crosstalk}}}$$

where the on-targets are the on-pathway products, and the off-targets are the reactants and off-pathway crosstalk products. For step m of system n, this tube designs for full conversion of cognate reactants into cognate products and against local crosstalk between these same reactants. One elementary step tube is specified for each elementary step $m = 0, 1, \ldots, M$ for each system $n = 1, \ldots, N$.

The off-pathway crosstalk products for step m of system n are defined as:

$$\boxed{\Psi_{m_n}^{\text{crosstalk}} = \Psi_{m_n}^{L \leq L_{\max}} - \Psi_{m_n}^{\text{exclude}}}$$

where the set $\Psi_{m_n}^{L \leq L_{\max}}$ denotes the set of all complexes of up to $L_{\max}$ strands (that are not already on-targets in the Step m_n tube). The set $\Psi_{m_n}^{\text{exclude}}$ contains energetically favorable complexes that we wish to exclude from the ensemble for the current elementary step (e.g., downstream on-pathway products, or off-pathway products that are inhibited kinetically rather than thermodynamically, and hence are not suitable for inclusion in the equilibrium optimization ensemble).

Global Crosstalk Tube. To actively design against global crosstalk, we additionally specify a single *global crosstalk tube*:

$$\boxed{\text{Global crosstalk tube: } \Psi_h^{\text{on}} \equiv \Psi_{\text{global}}^{\text{reactive}}, \qquad \Psi_h^{\text{off}} \equiv \Psi_{\text{global}}^{\text{crosstalk}}}$$

where $\Psi_{\text{global}}^{\text{reactive}}$ denotes the set of all reactive species generated during all elementary steps for all systems and $\Psi_{\text{global}}^{\text{crosstalk}}$ denotes the set of undesired crosstalk products resulting from interactions between these species.

For the global crosstalk tube, we exploit *motif simplification* to enable specification of the on-target and off-target complexes using using only monomers and dimers. The presumption is that motif complexity will typically decrease rather than increase crosstalk between reactive species, so that for the global crosstalk tube, motif simplification is justified in the service of efficiency and simplicity. By contrast, for the elementary step tubes, reactant and product complexes are treated without motif simplification, ensuring that

any energetic effects associated with the full complexes (either unfavorable [e.g., 3-arm junction for CHA product] or favorable [e.g., nick stack for HCR]) are taken into consideration.

To define various forms of motif simplification, it is helpful to define input and output domains that participate in the elementary steps. Each scRNA or scDNA motif (monomer, dimer, trimer, etc) has one or more *input domains* that control the state of one or more *output domains*. An inactive output domain is toggled to the active state when sequestering input domains hybridize to active output domains generated by earlier elementary steps in the reaction pathway. Nucleation with an input domain occurs via hybridization to an accessible loop or toehold. Targets that serve as inputs to a reaction pathway may be viewed as unconditionally active output domains that are available to hybridize to complementary input domains at any step in a reaction pathway.

Using motif simplification, we specify the *reactive species* and *cognate products* for system n as follows:

- $\lambda_n^{\text{simple}}$: scRNA and scDNA motifs with multiple input or output domains are simplified so that only the input and output domains for a single elementary step are present in each simplified motif.
- $\lambda_n^{\text{ss-out}}$: single-stranded output domains are specified for each elementary step, removing other concatenated or hybridized domains that represent the history or future of the reaction (participating in previous or future elementary steps).
- $\lambda_n^{\text{ss-in}}$: single-stranded nucleation sites within input domains (toeholds or loops) are specified isolated from the surrounding domains representing the history or future of the reaction.[1]
- $\lambda_n^{\text{reactive}} \equiv \{\lambda^{\text{simple}} \cup_{\text{simp}} \lambda^{\text{ss-out}} \cup_{\text{simp}} \lambda^{\text{ss-in}}\}_n$: the set of reactive species for system n is specified using a union operator $\cup_{\text{simp}}$ that eliminates redundancies when one monomer species is an accessible subsequence of another monomer species.
- $\lambda_n^{\text{cognate}}$: cognate products expected to form from reactive species in $\lambda_n^{\text{reactive}}$ based on sequence complementarity imposed by the reaction pathway (e.g., an input domain within a motif in $\lambda_n^{\text{simple}}$ is expected to hybridize to a complementary output domain in $\lambda_n^{\text{ss-out}}$).

These definitions for $\lambda_n^{\text{reactive}}$ and $\lambda_n^{\text{cognate}}$ are then used to define the on-targets for the global crosstalk tube:

$$\boxed{\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}}$$

and the off-targets for the global crosstalk tube:

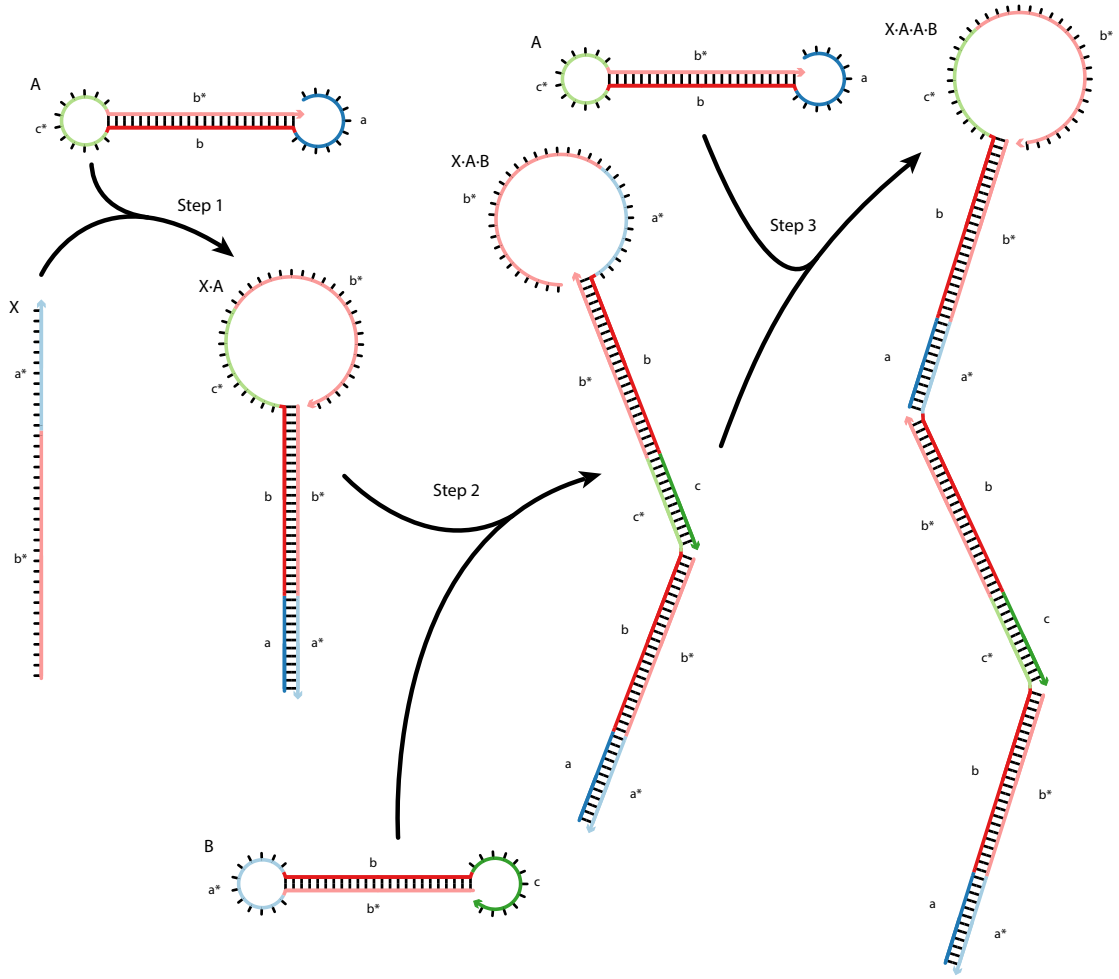
$$\boxed{\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L\leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}}$$

Here, $\Psi_{\text{global}}^{L\leq L_{\max}}$ denotes the set of all complexes of up to $L_{\max}$ strands (that are not already on-targets in the global crosstalk tube). The set $\cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$ contains all the cognate products that the reactive species in the N orthogonal systems are expected to form based on sequence complementarity. Crucially, by excluding these cognate products from $\Psi_{\text{global}}^{\text{crosstalk}}$, they do not appear in the global crosstalk tube as either on-targets or off-targets. Hence, all reactive species in the global crosstalk tube are forced to perform either no reaction (remaining as desired on-targets) or to undergo a crosstalk reaction (forming undesired off-targets), providing the basis for minimization of global crosstalk during sequence optimization.

[1] The role of $\lambda_n^{\text{ss-in}}$ is to enable $\Psi_{\text{global}}^{\text{crosstalk}}$ (the off-targets for the global crosstalk tube) to be specified without requiring any complexes larger than dimers. For example, consider a dimer motif with an exposed toehold. Crosstalk via kissing of this toehold with that of another dimer motif would yield a tetramer off-target; including these toeholds as isolated monomers in $\lambda_n^{\text{ss-in}}$ allows this crosstalk interaction to be described by an off-target dimer, which is automatically included in $\Psi_{\text{global}}^{\text{crosstalk}}$ by considering all off-targets of up to $L_{\max} = 2$ strands. Further, inclusion of loop nucleation sites in $\lambda_n^{\text{ss-in}}$ enables designing against pseudoknotted toehold/loop and loop/loop crosstalk interactions without needing to explicitly include pseudoknots in the structural ensemble of any complex.

B Conditional Self-Assembly via HCR

B.1 Reaction pathway for HCR



Step	Reaction	Function	Mechanism
1	$\mathrm{X} + \mathrm{A} \rightarrow \mathrm{X{\cdot}A}$	detect target, first A polymerization step	toehold/toehold nucleation, 3-way branch migration
2	$\mathrm{X{\cdot}A} + \mathrm{B} \rightarrow \mathrm{X{\cdot}A{\cdot}B}$	first B polymerization step, regenerate target sequence	toehold/toehold nucleation, 3-way branch migration
$2k{+}1$	$\mathrm{X{\cdot}(A)}_k{\cdot}\mathrm{(B)}_k + \mathrm{A} \rightarrow \mathrm{X{\cdot}(A)}_{k+1}{\cdot}\mathrm{(B)}_k$	generic A polymerization step, $k = 1, 2, \ldots$	toehold/toehold nucleation, 3-way branch migration
$2k{+}2$	$\mathrm{X{\cdot}(A)}_{k+1}{\cdot}\mathrm{(B)}_k + \mathrm{B} \rightarrow \mathrm{X{\cdot}(A)}_{k+1}{\cdot}\mathrm{(B)}_{k+1}$	generic B polymerization step, $k = 1, 2, \ldots$	toehold/toehold nucleation, 3-way branch migration

Figure 2: Reaction pathway for conditional self-assembly via hybridization chain reaction (HCR) (Dirks and Pierce, *Proc Natl Acad Sci USA*, **101**:15275-15278, 2004). Target X triggers self-assembly of metastable hairpins A and B into a long nicked dsDNA polymer via a chain reaction of alternating A and B polymerization steps. Top: Reaction pathway schematic. Bottom: Elementary step details.

B.2 Target test tube specification for HCR

Target test tubes are defined using the specification of Appendix A with the following definitions. The total number of target test tubes is $|\Omega| = \sum_{n=1,\ldots,N}\{\text{Step 0, Step 1, Step 2, Step 3}\}_n + \text{Crosstalk} = 4N+1$; the target test tubes in the multistate test tube ensemble, Ω, are indexed by $h = 1, \ldots, 4N+1$. $L_{\max} = 2$ for all tubes.

Reactants for system n

- Target: X_n
- Hairpins: $\{\mathrm{A}, \mathrm{B}\}_n$

Elementary step tubes for system n

- Step 0_n tube: $\Psi_{0_n}^{\text{products}} \equiv \{\mathrm{X}, \mathrm{A}, \mathrm{B}\}_n$; $\Psi_{0_n}^{\text{reactants}} \equiv \{\mathrm{A{\cdot}B}\}_n$ (dimer nucleus that inhibits leakage); $\Psi_{0_n}^{\text{exclude}} \equiv \{\mathrm{X{\cdot}A}\}_n$ (downstream on-pathway product)
- Step 1_n tube: $\Psi_{1_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A}\}_n$; $\Psi_{1_n}^{\text{reactants}} \equiv \{\mathrm{X}, \mathrm{A}\}_n$; $\Psi_{1_n}^{\text{exclude}} \equiv \emptyset$
- Step 2_n tube: $\Psi_{2_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A{\cdot}B}\}_n$; $\Psi_{2_n}^{\text{reactants}} \equiv \{\mathrm{X{\cdot}A}, \mathrm{B}\}_n$; $\Psi_{2_n}^{\text{exclude}} \equiv \emptyset$
- Step 3_n tube: $\Psi_{3_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A{\cdot}A{\cdot}B}\}_n$; $\Psi_{3_n}^{\text{reactants}} \equiv \{\mathrm{X{\cdot}A{\cdot}B}, \mathrm{A}\}_n$; $\Psi_{3_n}^{\text{exclude}} \equiv \emptyset$

Global crosstalk tube

- Crosstalk tube: $\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$; $\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L \le L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

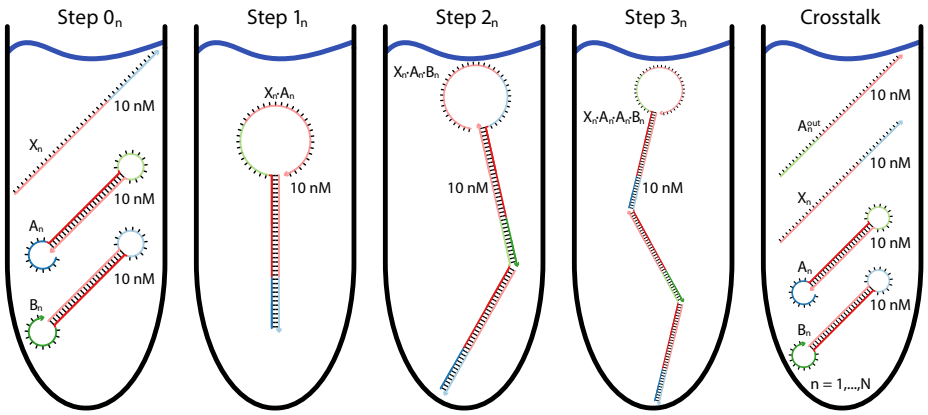
The reactive species and cognate products for system n are:

- $\lambda_n^{\text{simple}} \equiv \{\mathrm{A}, \mathrm{B}\}_n$
- $\lambda_n^{\text{ss-out}} \equiv \{\mathrm{X}, \mathrm{A}^{\text{out}}, \mathrm{B}^{\text{out}}\}_n$
- $\lambda_n^{\text{ss-in}} \equiv \{\mathrm{A}^{\text{toe}}, \mathrm{B}^{\text{toe}}\}_n$
- $\lambda_n^{\text{reactive}} \equiv \{\mathrm{A}, \mathrm{B}, \mathrm{A}^{\text{out}}, \mathrm{B}^{\text{out}}\}_n$
- $\lambda_n^{\text{cognate}} \equiv \{\mathrm{A}^{\text{out}}{\cdot}\mathrm{B}, \mathrm{B}^{\text{out}}{\cdot}\mathrm{A}\}_n$

based on the definitions (listed 5′ to 3′ using the sequence domain notation of Figure 2):

- $\mathrm{A} \equiv \mathrm{A}^{\text{in}}\text{-}\mathrm{A}^{\text{out}}$
- $\mathrm{A}^{\text{toe}} \equiv \mathrm{a}$
- $\mathrm{A}^{\text{in}} \equiv \mathrm{a\text{-}b}$
- $\mathrm{A}^{\text{out}} \equiv \mathrm{c^*\text{-}b^*}$
- $\mathrm{B} \equiv \mathrm{B}^{\text{out}}\text{-}\mathrm{B}^{\text{in}}$
- $\mathrm{B}^{\text{toe}} \equiv \mathrm{c}$
- $\mathrm{B}^{\text{in}} \equiv \mathrm{b\text{-}c}$
- $\mathrm{B}^{\text{out}} \equiv \mathrm{b^*\text{-}a^*}$
- $\mathrm{X} \equiv \mathrm{b^*\text{-}a^*}$

Note: X_n is identical to $\mathrm{B}_n^{\text{out}}$, so it is implicitly included in the definition of $\lambda_n^{\text{reactive}}$. To avoid redundancy, the toeholds of $\lambda_n^{\text{ss-in}}$ are not included in the definition of $\lambda_n^{\text{reactive}}$; these toeholds are already available to form dimer crosstalk products in the hairpin monomers of $\lambda_n^{\text{simple}}$.



Tube	On-targets (Ψ_h^{on})	Off-targets (Ψ_h^{off})
Step 0_n	$\{\text{X, A, B}\}_n$	$\{\text{A·B}\}_n \cup \Psi_{0_n}^{L\leq L_{\max}} - \{\text{X·A}\}_n$
Step 1_n	$\{\text{X·A}\}_n$	$\{\text{X, A}\}_n \cup \Psi_{1_n}^{L\leq L_{\max}}$
Step 2_n	$\{\text{X·A·B}\}_n$	$\{\text{X·A, B}\}_n \cup \Psi_{2_n}^{L\leq L_{\max}}$
Step 3_n	$\{\text{X·A·A·B}\}_n$	$\{\text{X·A·B, A}\}_n \cup \Psi_{3_n}^{L\leq L_{\max}}$
Crosstalk	$\cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$	$\Psi_{\text{global}}^{L\leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

Figure 3: Target test tubes for conditional self-assembly via HCR (reaction pathway of Figure 2). Top: Target test tube schematics. Bottom: Target test tube details. Each target test tube contains the depicted on-target complexes (each with the depicted target structure and a target concentration of 10 nM) and the off-target complexes listed in the table (each with vanishing target concentration). To simultaneously design N orthogonal systems, the total number of target test tubes is $|\Omega| = 4N + 1$. $L_{\max} = 2$ for all tubes. Design conditions: DNA in 1 M Na^+ at 25 °C.

B.3 Residual defects for design of conditional self-assembly via HCR

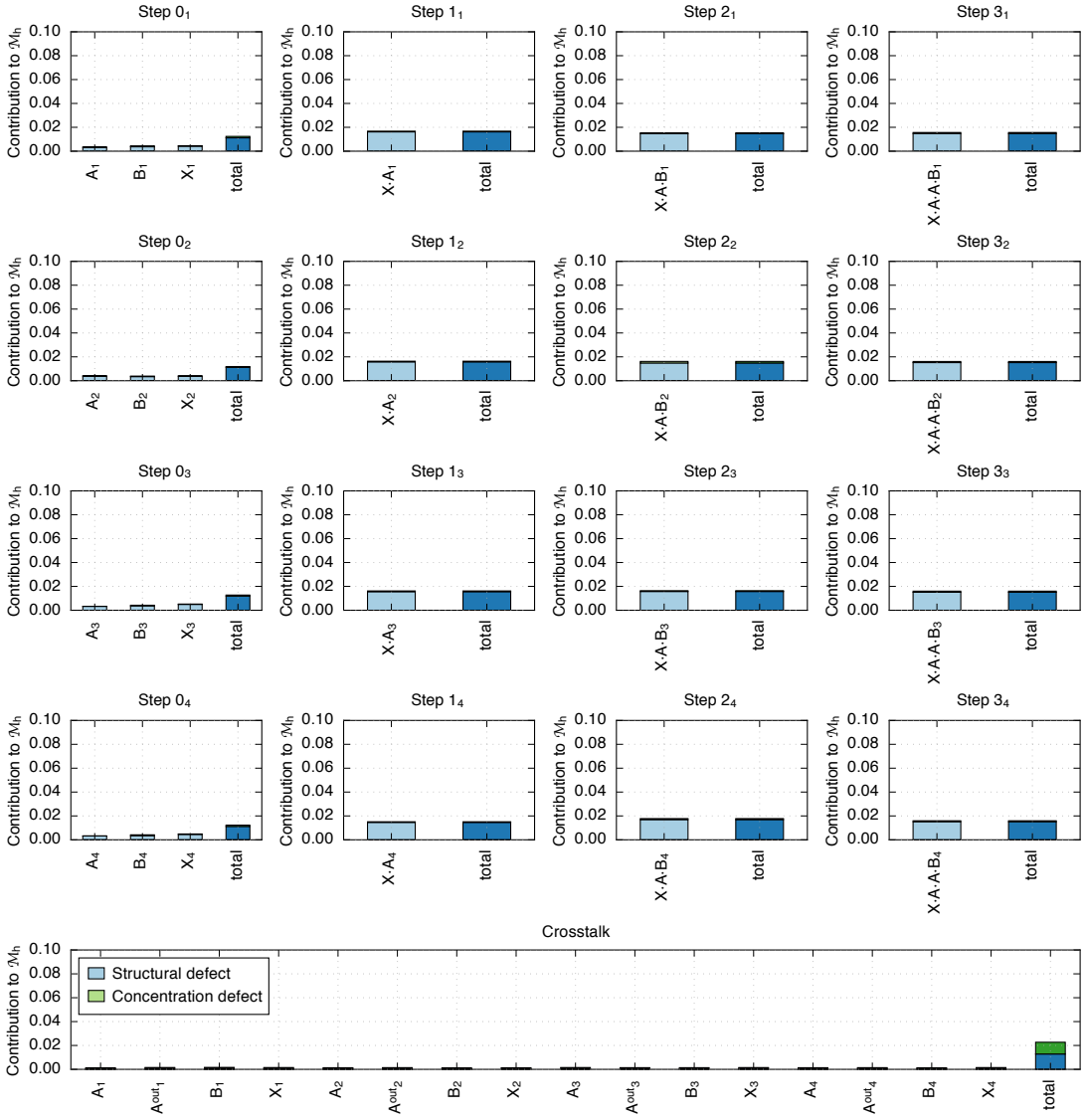
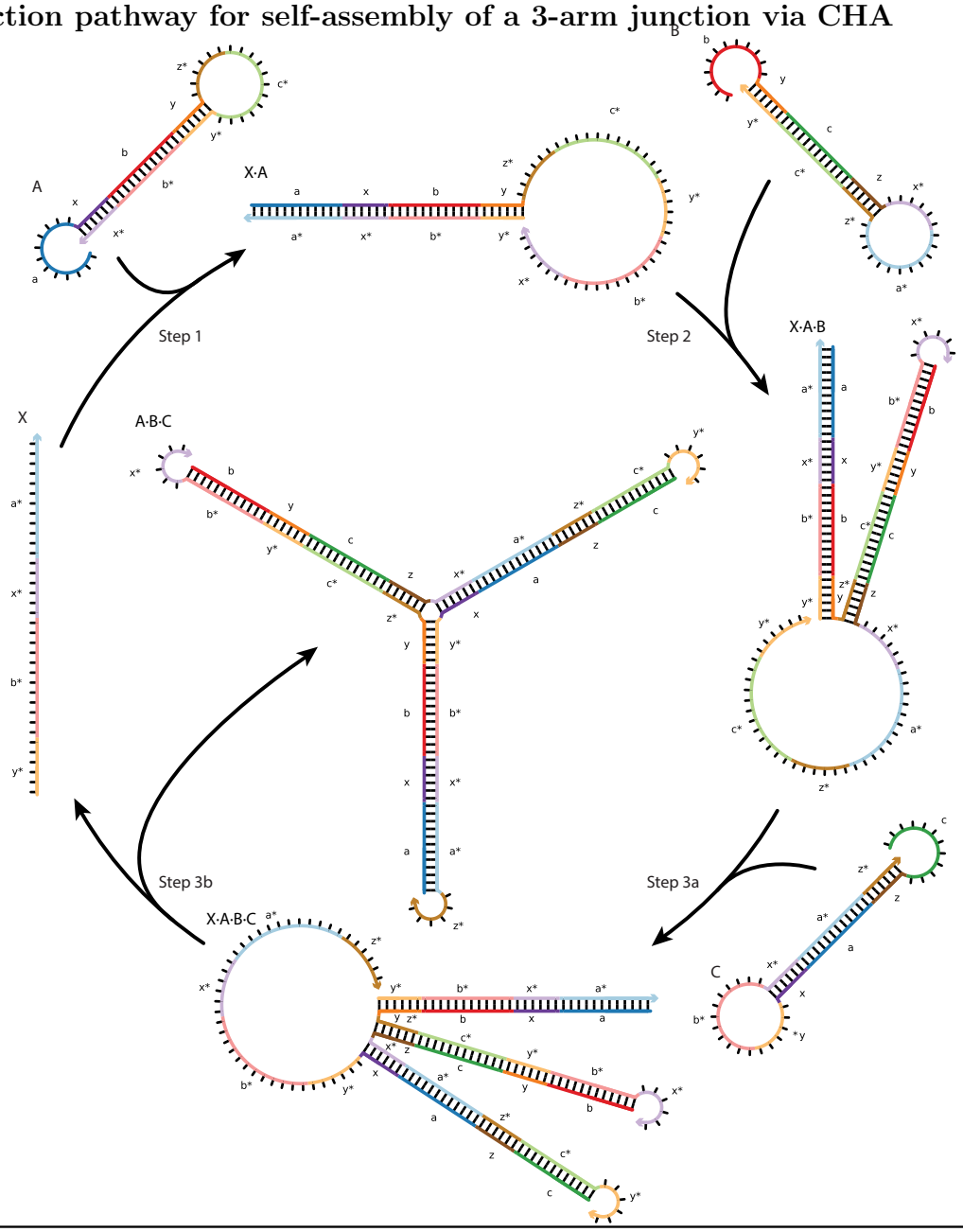


Figure 4: Residual defects for conditional self-assembly via HCR ($N = 4$ orthogonal systems for target test tubes of Figure 3). Each panel corresponds to a different tube $h \in \{1, \ldots, 4N + 1\}$. For each tube h, the structural defect and concentration defect contributions to the tube ensemble defect, $\mathcal{M}_h$, are depicted for each complex (pale shaded bars). The total structural defect and total concentration defect contributions to the multistate test tube ensemble defect, $\mathcal{M}$, are also depicted for each tube (dark shaded bars). Each bar represents the mean over 30 independent design trials with stop condition $\mathcal{M} \leq 0.02$. All nucleotide, complex, and tube weights are left at the default value of 1 except for the global crosstalk tube which is assigned a weight of N to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases.

C Self-assembly of a 3-arm junction via catalytic hairpin assembly (CHA)

C.1 Reaction pathway for self-assembly of a 3-arm junction via CHA



Step	Reaction	Function	Mechanism
1	X + A → X·A	assemble with catalyst X	toehold/toehold nucleation, 3-way branch migration
2	X·A + B → X·A·B	assemble	toehold/toehold nucleation, 3-way branch migration
3a	X·A·B + C → X·A·B·C	assemble	toehold/toehold nucleation, 3-way branch migration
3b	X·A·B·C → X + A·B·C	disassemble from catalyst X and assemble 3-arm junction	intracomplex blunt-end strand invasion, 3-way branch migration

Figure 5: Reaction pathway for self-assembly of a 3-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, and C into 3-arm junction A·B·C. Top: Reaction pathway schematic. Bottom: Elementary step details.

C.2 Target test tube specification for self-assembly of a 3-arm junction via CHA

Target test tubes are defined using the specification of Appendix A with the following definitions. The total number of target test tubes is $|\Omega| = \sum_{n=1,\ldots,N} \{\text{Step 0, Step 1, Step 2, Step 3a, Step 3b}\}_n + \text{Crosstalk} = 5N+1$; the target test tubes in the multistate test tube ensemble, Ω, are indexed by $h = 1, \ldots, 5N+1$. $L_{\max} = 2$ for all tubes.

Reactants for system n

- Target: X_n
- Hairpins: $\{\mathrm{A, B, C}\}_n$

Elementary step tubes for system n

- Step 0_n: $\Psi_{0_n}^{\text{products}} \equiv \{\mathrm{X, A, B, C}\}_n$; $\Psi_{0_n}^{\text{reactants}} \equiv \emptyset$; $\Psi_{0_n}^{\text{exclude}} \equiv \{\mathrm{X{\cdot}A, X{\cdot}B}\}_n$
- Step 1_n: $\Psi_{1_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A}\}_n$; $\Psi_{1_n}^{\text{reactants}} \equiv \{\mathrm{X, A}\}_n$; $\Psi_{1_n}^{\text{exclude}} \equiv \emptyset$
- Step 2_n: $\Psi_{2_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A{\cdot}B}\}_n$; $\Psi_{2_n}^{\text{reactants}} \equiv \{\mathrm{X{\cdot}A, B}\}_n$; $\Psi_{2_n}^{\text{exclude}} \equiv \emptyset$
- Step $3\mathrm{a}_n$: $\Psi_{3\mathrm{a}_n}^{\text{products}} \equiv \{\mathrm{X{\cdot}A{\cdot}B{\cdot}C}\}_n$; $\Psi_{3\mathrm{a}_n}^{\text{reactants}} \equiv \{\mathrm{X{\cdot}A{\cdot}B, C}\}_n$; $\Psi_{3\mathrm{a}_n}^{\text{exclude}} \equiv \emptyset$
- Step $3\mathrm{b}_n$: $\Psi_{3\mathrm{b}_n}^{\text{products}} \equiv \{\mathrm{X, A{\cdot}B{\cdot}C}\}_n$; $\Psi_{3\mathrm{b}_n}^{\text{reactants}} \equiv \{\mathrm{X{\cdot}A{\cdot}B{\cdot}C}\}_n$; $\Psi_{3\mathrm{b}_n}^{\text{exclude}} \equiv \emptyset$

Note: Step 3 combining an assembly operation (Step 3a; addition of C) with a disassembly operation (Step 3b; removal of X) is described using two target test tubes; the Step 3a tube prevents completion of the full operation by excluding the final product A·B·C from the ensemble ($L_{\max} = 2$ includes all off-targets up to dimers).

Crosstalk tube

- Crosstalk tube: $\Psi_{\text{global}}^{\text{reactive}} \equiv \cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$; $\Psi_{\text{global}}^{\text{crosstalk}} \equiv \Psi_{\text{global}}^{L \le L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

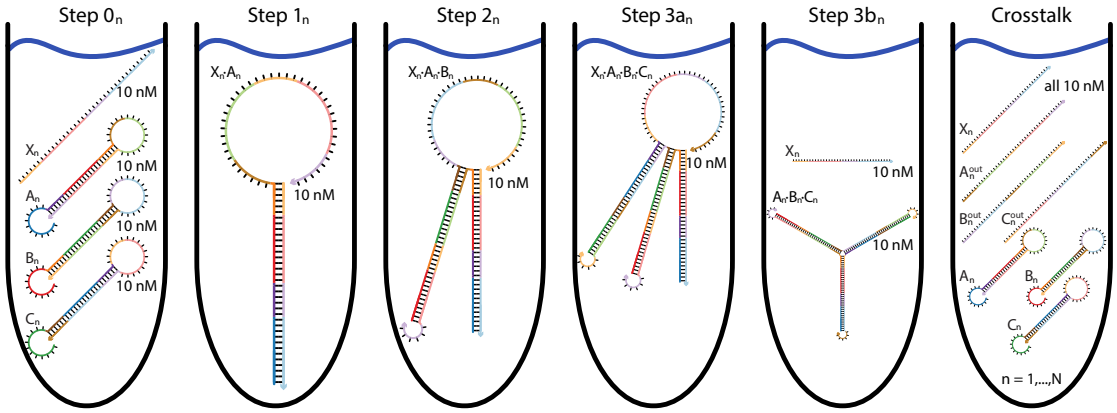
The reactive species and cognate products for system n are:

- $\lambda_n^{\text{simple}} \equiv \{\mathrm{A, B, C}\}_n$
- $\lambda_n^{\text{ss-out}} \equiv \{\mathrm{X, A^{out}, B^{out}, C^{out}}\}_n$
- $\lambda_n^{\text{ss-in}} \equiv \{\mathrm{A^{toe}, B^{toe}, C^{toe}}\}_n$
- $\lambda_n^{\text{reactive}} \equiv \{\mathrm{A, B, C, X, A^{out}, B^{out}, C^{out}}\}_n$
- $\lambda_n^{\text{cognate}} \equiv \{\mathrm{X{\cdot}A, X{\cdot}B, A^{out}{\cdot}B, B^{out}{\cdot}C, C^{out}{\cdot}A, C^{out}{\cdot}B, B^{out}{\cdot}A, A^{out}{\cdot}C}\}_n$

based on the definitions (listed 5′ to 3′ using the sequence domain notation of Figure 5):

- $\mathrm{A} \equiv \mathrm{A^{in}\text{-}A^{out}}$
- $\mathrm{A^{toe}} \equiv \mathrm{a}$
- $\mathrm{A^{in}} \equiv \mathrm{a\text{-}x\text{-}b\text{-}y}$
- $\mathrm{A^{out}} \equiv \mathrm{z^*\text{-}c^*\text{-}y^*\text{-}b^*\text{-}x^*}$
- $\mathrm{B} \equiv \mathrm{B^{in}\text{-}B^{out}}$
- $\mathrm{B^{toe}} \equiv \mathrm{b}$
- $\mathrm{B^{in}} \equiv \mathrm{b\text{-}y\text{-}c\text{-}z}$

- $B^{out} \equiv$ x*-a*-z*-c*-y*
- $C \equiv C^{in}$-C^{out}
- $C^{toe} \equiv$ c
- $C^{in} \equiv$ c-z-a-x
- $C^{out} \equiv$ y*-b*-x*-a*-z*
- $X \equiv$ y*-b*-x*-a*



Tube	On-targets (Ψ_h^{on})	Off-targets (Ψ_h^{off})
Step 0_n	$\{\text{X, A, B, C}\}_n$	$\Psi_{0_n}^{L\leq L_{\max}} - \{\text{X·A, X·B}\}_n$
Step 1_n	$\{\text{X·A}\}_n$	$\{\text{X, A}\}_n \cup \Psi_{1_n}^{L\leq L_{\max}}$
Step 2_n	$\{\text{X·A·B}\}_n$	$\{\text{X·A, B}\}_n \cup \Psi_{2_n}^{L\leq L_{\max}}$
Step $3a_n$	$\{\text{X·A·B·C}\}_n$	$\{\text{X·A·B, C}\}_n \cup \Psi_{3a_n}^{L\leq L_{\max}}$
Step $3b_n$	$\{\text{X, A·B·C}\}_n$	$\{\text{X·A·B·C}\}_n \cup \Psi_{3b_n}^{L\leq L_{\max}}$
Crosstalk	$\cup_{n=1,\ldots,N}\{\lambda_n^{\text{reactive}}\}$	$\Psi_{\text{global}}^{L\leq L_{\max}} - \cup_{n=1,\ldots,N}\{\lambda_n^{\text{cognate}}\}$

Figure 6: Target test tubes for self-assembly of a 3-arm junction via CHA (reaction pathway of Figure 5). Top: Target test tube schematics. Bottom: Target test tube details. Each target test tube contains the depicted on-target complexes (each with the depicted target structure and a target concentration of 10 nM) and the off-target complexes listed in the table (each with vanishing target concentration). To simultaneously design N orthogonal systems, the total number of target test tubes is $|\Omega| = 5N + 1$. $L_{\max} = 2$ for all tubes. Design conditions: DNA in 1 M Na^+ at 25 °C.

C.2.1 Residual defects for design of self-assembly of a 3-arm junction via CHA

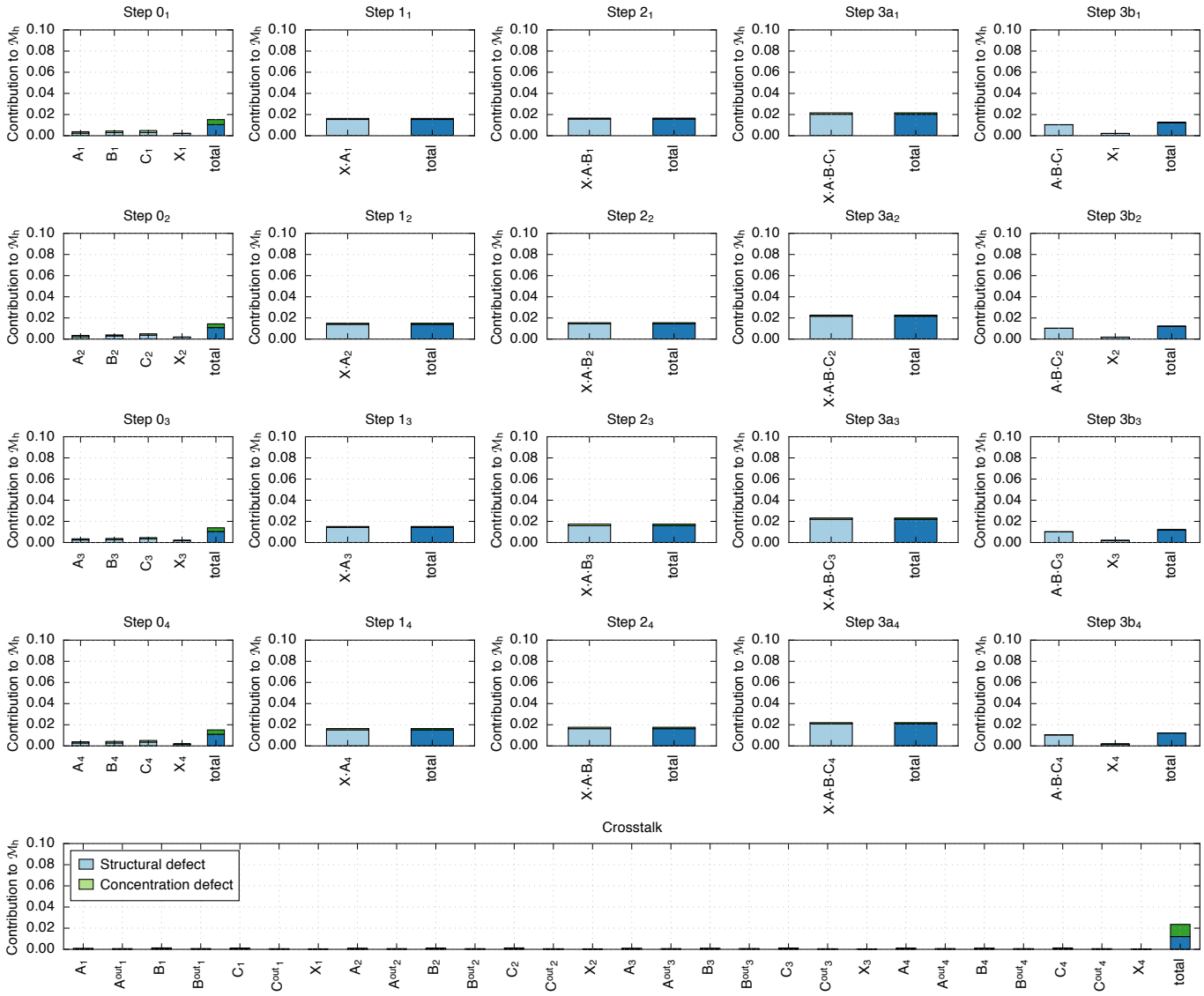


Figure 7: Residual defects for self-assembly of a 3-arm junction via CHA ($N = 4$ orthogonal systems for target test tubes of Figure 6). Each panel corresponds to a different tube $h \in \{1, \ldots, 5N+1\}$. For each tube h, the structural defect and concentration defect contributions to the tube ensemble defect, $\mathcal{M}_h$, are depicted for each complex (pale shaded bars). The total structural defect and total concentration defect contributions to the multistate test tube ensemble defect, $\mathcal{M}$, are also depicted for each tube (dark shaded bars). Each bar represents the mean over 30 independent design trials with stop condition $\mathcal{M} \leq 0.02$. All nucleotide, complex, and tube weights are left at the default value of 1 except for the global crosstalk tube which is assigned a weight of N to prevent the effect of crosstalk from being diluted in the design objective function as the number of orthogonal systems increases.